





Thèse de doctorat de l'Université Sorbonne Paris Cité
Préparée à l'Université Paris Diderot
Ecole doctorale Frontières du Vivant (ED 474)
Pasteur Paris-University International Doctoral Program
Unité Imagerie et Modélisation, Institut Pasteur

Deep Learning for Advanced Microscopy

Version 2.0 (incomplete)

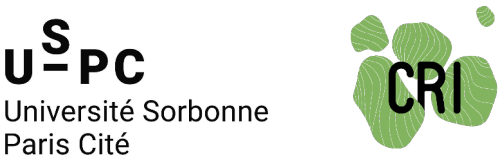
Wei OUYANG

Thèse de doctorat d'informatique

Dirigée par Christophe ZIMMER

Présentée et soutenue publiquement à Paris le 18/10/2018

Priscille BRODIN	Directeur de recherche, Institut Pasteur de Lille	Rapporteur
Maïté MOISAN COPPEY	Directeur de recherche, Institut Jacques Monod	Examinateur
Susan COX	Directeur de recherche, King's College London	Examinateur
Thierry DORVAL	Chef d'equipe, Institut de Recherches Servier	Examinateur
Sylvie HENON	Professeure, Université Paris Diderot	Président
Jean-Luc STARCK	Directeur de recherche, CEA	Rapporteur
Christophe ZIMMER	Directeur de recherche, Institut Pasteur	Directeur







Doctoral Thesis at Sorbonne Paris Cité University
Prepared at Paris Diderot University
Frontiers of Life Science PhD program (ED 474)
Pasteur Paris-University International Doctoral Program
Imaging and Model Unit, Pasteur Institute

Deep Learning for Advanced Microscopy

Version 2.0 (incomplete)

Wei OUYANG

Doctoral Thesis in Computer Science

Supervised by Christophe ZIMMER

Présentée et soutenue publiquement à Paris le 18/10/2018

Priscille BRODIN	Pasteur Institute of Lille	Reporter
Maïté MOISAN COPPEY	Jacques Monod Institute	Examiner
Susan COX	King's College London	Examiner
Thierry DORVAL	Servier Research Institute	Examiner
Sylvie HENON	Paris Diderot University	President
Jean-Luc STARCK	CEA	Reporter
Christophe ZIMMER	Pasteur Institute	Supervisor

Abstract

Background: Microscopy plays an important role in biology since several centuries, but its resolution has long been limited to $\sim 250nm$ due to diffraction, leaving many important biological structures (e.g. viruses, vesicles, nuclear pores, synapses) unresolved. Over the last decade, several super-resolution methods have been developed that break this limit. Among the most powerful and popular super-resolution techniques are those based on single molecular localization (single molecule localization microscopy, or SMLM) such as PALM and STORM. By precisely localizing positions of isolated fluorescent molecules in thousands or more sequentially acquired diffraction limited images, SMLM can achieve resolutions of 20-50 nm or better. However, SMLM is inherently slow due to the necessity to accumulate enough localizations to achieve high resolution sampling of the fluorescent structures. The drawback in acquisition speed (typically $1 \sim 30$ minutes per super-resolution image) makes it difficult to use SMLM in high-throughput and live cell imaging. Many methods have been proposed to address this issue, mostly by improving the localization algorithms to localize overlapping spots, but most of them compromise spatial resolution and cause artifacts.

Like other microscopy techniques, SMLM faces conflicting aims between improving spatial resolution, temporal resolution, signal-to-noise ratio and minimal phototoxicity for live cell imaging. However these trade-offs can be alleviated by introducing prior information about the object to be imaged. Suitable prior information can be learned from existing examples thanks to different computational techniques, including recently popularized techniques from *Deep Learning*.

Deep learning uses artificial neural networks to learn distributed feature representations and has been successfully applied to many computer vision tasks such as image classification with human-level or super-human level performance. For image restoration tasks, *Convolutional Neural Networks* such as U-net and Pix2Pix provide a generic image-to-image translation framework to transform noisy or blurry or otherwise corrupted images to uncorrupted versions.

Methods and results: In this work, we applied deep learning based image-to-image translation framework for improving imaging speed and quality by restoring information from rapidly acquired low quality SMLM images. By utilizing recent advances in deep learning including the U-net and *Generative Adversarial Networks*, we developed our method *Artificial Neural*

Network Accelerated PALM (ANNA-PALM) which is capable of learning structural information from training images and using the trained model to accelerate SMLM imaging by tens to hundreds folds. With experimentally acquired images of different cellular structures (microtubules, nuclear pores and mitochondria), we demonstrated that deep learning can efficiently capture the structural information from less than 10 training samples and reconstruct high quality super-resolution images from sparse, noisy SMLM images obtained with much shorter acquisitions than usual for SMLM. We also showed that ANNA-PALM is robust to possible variations between training and testing conditions, due either to changes in the biological structure or to changes in imaging parameters. Furthermore, we take advantage of the acceleration provided by ANNA-PALM to perform high throughput experiments, showing acquisition of 1000 cells at high resolution in $\sim$ 3 hours. Additionally, we designed a tool to estimate and reduce possible artifacts is designed by measuring the consistency between the reconstructed image and the experimental wide-field image. We demonstrated a novel way of accelerating SMLM, improving the image quality, and reducing artifacts. Our method enables faster and gentler imaging which can be applied to high-throughput, and provides a novel avenue towards live cell high resolution imaging. The results also give insights on how deep learning methods are enabling augmented microscopy view of the cellular environment with enhanced speed, resolution and colors.

Deep learning methods rely on training data and their performance can be improved even further with more training data. One cheap way to obtain more training data is through data sharing within the microscopy community. However, it often difficult to exchange or share localization microscopy data, because localization tables alone are typically several gigabytes in size, and there is no dedicated platform for localization microscopy data which provide features such as rendering, visualization and filtering. To address these issues, we developed a file format that can losslessly compress localization tables into smaller files, alongside with a web platform called *ShareLoc* (`https://shareloc.xyz`) that allows to easily visualize and share 2D or 3D SMLM data. We believe that this platform can greatly improve the performance of deep learning models, accelerate tool development, facilitate data re-analysis and further promote reproducible research and open science.

Keywords: SMLM, PALM, STORM, deep learning, neural networks, image reconstruction, super-resolution, data sharing

Résumé

Contexte: La microscopie joue un rôle important en biologie depuis plusieurs siècles, mais sa résolution a longtemps été limitée à environ 250 nm, de sorte que nombre de structures biologiques (virus, vésicules, pores nucléaires, synapses) ne pouvaient être résolues. Au cours de la dernière décennie, plusieurs méthodes de super-résolution ont été développées pour dépasser cette limite. Parmi ces techniques, les plus puissantes et les plus utilisées reposent sur la localisation de molécules uniques (microscopie à localisation de molécule unique, ou SMLM), comme PALM et STORM. En localisant précisément les positions de molécules fluorescentes isolées dans des milliers d'images de basse résolution acquises de manière séquentielle, la SMLM peut atteindre des résolutions de 20 à 50 nm voire mieux. Cependant, cette technique est intrinsèquement lente car elle nécessite l'accumulation d'un très grand nombre d'images et de localisations pour obtenir un échantillonnage super-résolutif des structures fluorescentes. Cette lenteur (typiquement 30 minutes par image super-résolutive) rend difficile l'utilisation de la SMLM pour l'imagerie cellulaire à haut débit ou en cellules vivantes. De nombreuses méthodes ont été proposées pour pallier à ce problème, principalement en améliorant les algorithmes de localisation pour localiser des molécules proches, mais la plupart de ces méthodes compromettent la résolution spatiale et entraînent l'apparition d'artefacts.

Méthodes et résultats: Nous avons adopté une stratégie de transformation d'image en image basée sur l'apprentissage profond dans le but de restaurer des images SMLM parcimonieuses et par là d'améliorer la vitesse d'acquisition et la qualité des images super-résolutives. Notre méthode, ANNA-PALM, s'appuie sur des développements récents en apprentissage profond, notamment l'architecture U-net et les modèles génératifs antagonistes (GANs). Nous montrons des validations de la méthode sur des images simulées et des images expérimentales de différentes structures cellulaires (microtubules, pores nucléaires et mitochondries). Ces résultats montrent qu'après un apprentissage sur moins de 10 images de haute qualité, ANNA-PALM permet de réduire le temps d'acquisition d'images SMLM, à qualité comparable, d'un facteur 10 à 100. Nous avons également montré que ANNA-PALM est robuste à des altérations de la structure biologique, ainsi qu'à des changements de paramètres de microscopie. Nous démontrons le potentiel applicatif d'ANNA-PALM pour la microscopie à haut débit en imageant 1000 cellules à haute résolution en environ 3 heures. Enfin, nous avons conçu un outil pour estimer et réduire les artefacts de reconstruction en mesurant la cohérence entre l'image reconstruite et l'image

en épi-fluorescence. Notre méthode permet une microscopie super-résolutive plus rapide et plus douce, compatible avec l'imagerie haut débit, et ouvre une nouvelle voie vers l'imagerie super-résolutive des cellules vivantes.

La performance des méthodes d'apprentissage profond augmente avec la quantité des données d'entraînement. Le partage d'images au sein de la communauté de microscopie offre en principe un moyen peu coûteux d'augmenter ces données. Cependant, il est souvent difficile d'échanger ou de partager des données de SMLM, car les tables de localisation seules ont souvent une taille de plusieurs gigaoctets et il n'existe pas de plate-forme de visualisation dédiée aux données SMLM. Nous avons développé un format de fichier pour compresser sans perte des tables de localisation, ainsi qu'une plateforme web (*ShareLoc*, `https://shareloc.xyz`) qui permet de visualiser et de partager facilement des données SMLM 2D ou 3D. A l'avenir, cette plate-forme pourrait grandement améliorer les performances des modèles d'apprentissage en profondeur, accélérer le développement des outils, faciliter la réanalyse des données et promouvoir la recherche reproductible et la science ouverte.

Mots clefs: SMLM, PALM, STORM, apprentissage profond, réseaux de neurones, reconstruction d'image, super-résolution, partage de données

List of Publication

1. **Wei Ouyang**, Andrey Aristov, Mickaël Lelek, Xian Hao, Christophe Zimmer. *Deep learning massively accelerates super-resolution localization microscopy.*

   Nature Biotechnology, 2018, doi: 10.1038/nbt.4106

2. **Wei Ouyang**, Christophe Zimmer, *The imaging tsunami: computational opportunities and challenges.*

   Current Opinion in Systems Biology, 2017, doi: 10.1016/j.coisb.2017.07.011

3. Aubin Samacoits, Racha Chouaib, Adham Safieddine, Abdel-Meneem Traboulsi, **Wei Ouyang**, Christophe Zimmer, Marion Peter, Edouard Bertrand, Thomas Walter, Florian Mueller. *A computational framework to study sub-cellular RNA localization.*

   Nature Communications, 2018, doi: 10.1038/s41467-018-0686-w

Acknowledgements

I would like to first thank my supervisor Christophe ZIMMER, for the patient guidance, encouragement, trust and support throughout my time as his student. I have been learning a lot from his knowledge, experience, openness and integrity. I have been very lucky to have a supervisor who cared so much about me, always thinking about what is good for me, not only for my studies but also for my career and life in general. He always actively answers my questions, responds to my requests, introduced me to people who may help me and coordinated the collaboration. He has been very helpful during the writing of this manuscript and I really appreciated his efforts for helping me to improve my writing skills.

I would like to thank all my colleagues in the Imaging and Modeling Unit, including Andrey Aristov, Aubin Samacoits, Benjamin Robert King, Benoît Lelandais, Christian Weber, Elena Rensen, Florian Mueller, Jean-Michel Arbona, Jyotsana Parmar, Mickael Lelek, Maxime Woringer and Xian Hao. It has been a pleasure to work with them, being friends with them and they have been very helpful for my studies and my life in general. I would like to thank Andrey, Mickael, and Xian for the great collaboration for my PhD project. I am particularly grateful to Christian who has been training me on the experimental skills and kindly helped me with many experiments. I really appreciate the assistance provided by Mickael and Xian during my work. I would like to thank Elena who helped me with my experiments, my thesis manuscript and lots of other stuff. I greatly appreciate her willingness for offering me help. I thank Jyotsana for helping with the thesis manuscript, and also for being supportive and encouraging. I would also like to thank Benoît who is always kind and generous for offering me help with many practical life problems. Florian and Maxime have been my great advisors, and I received many useful comments, advices and critiques which are crucial for my study, and also sometimes for life in general, thank you. I enjoyed the friendship with Andrey, Benjamin, Elena, Florian, Jean-Michel, Jyotsana and Xian, there have been many touching moments when I really feel that I was cared, encouraged and supported.

My special thanks goes to Cécile Leduc who trained me for immunolabeling, gave advices for my paper, and inspired me to introduce widefield images to ANNA-PALM which have been a very useful ingredient in my paper.

I would like to thank the following colleagues for providing data and having inspiring discussions:

Orestis Faklaris, Julien Sellès, Sandrine Lévêque-Fort, Nicolas Bourg, Pakorn Kanchanawong (NUS, Singapore).

I would like to thank the IT service of Institut Pasteur, particularly Stéphane Fournier, Youssef Ghorbal, Jean-Baptiste Denis, Sophie Créno, Nicolas Joly, Véronique Legrand, Eric Deveaud, for helping with software installation, GPU computing, and setting up the web servers.

I would also like to thank Quang Tru Huynh for helping me with the GPU machines, being responsive and supportive, and also for setting up the Inception GPU lab which was very helpful for my project.

My PhD is funded by the Pasteur-Paris University International Doctoral program (PPU) in Institut Pasteur. I would like to thank all the people from PPU office and the steering committee for accepting me to the program. I would like to thank Susanna CELLI and Monica SALA who have been very supportive and helpful for me. I thank Molly INGERSOLL who gave me many useful suggestions in improving my presentation.

I would also like to thank my doctoral school Frontières du vivant and many people from Center for Research and Interdisciplinarity. I would like to thank François Taddei for those inspiring discussions, and David Tareste who always promptly replied to my question and request, Elodie Kaslikowski who helped me a lot with the administrative steps.

Moreover, the resources and equipments used in PhD project is also funded by Agence Nationale de la Recherche grant (ANR 14 CE10 0018 02), Fondation pour la Recherche Médicale (Equipe FRM, DEQ 20150331762), and the Région Ile de France (DIM Malinf). We also acknowledge Investissement d'Avenir grant ANR-16-CONV-0005 for funding a GPU farm used in this work.

Finally, my deep and sincere gratitude to my family for their continuous and unparalleled love, help and support. Especially, I would like to thank my mom, my dad and my wife for their continuous support, I could not done it without them.

List of Abbreviations

AI	Artificial Intelligence
AGI	Artificial General Intelligence
ANN	Artificial Neural Network
ANNA-PALM	Artificial Neural Network Accelerated PALM
CARE	Content-Aware Image Restoration
cGAN	Conditional Generative Adversarial Networks
CryoEM	Cryogenic Electron Microscopy
CCD	Charge-Coupled Device
CNN	Convolutional Neural Networks
CPU	Central Processing Units
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CSV	Comma Separated Values
DNN	Deep Neural Network
DNA-PAINT	DNA-based Point Accumulation for Imaging in Nanoscale Topography
EM	Electron Microscopy
FCN	Fully Convolutional Networks
FWHM	Full Width Half Maximum
FoV	Field of View
GAN	Generative Adversarial Networks
GB	Gigabytes
GPU	Graphical Processing Units
HAWK	Haar Wavelet Kernel Analysis
LBP	Local Binary Patterns
MAML	Model-agnostic Meta-learning
MLP	Multi-layer Perceptrons
MS-SSIM	Multi-scale Structural Similarity Index
MSE	Mean Squared Error
MAE	Mean Absolute Error
NA	Numerical Aperture
NSOM	Near-field Scaning Optical Microscopy
PSF	Point Spread Function

PALM	Photoactivated Localization Microscopy
ReLU	Rectified Linear Unit
RESOLFT	Reversible Saturable Optical Linear Fluorescence Transitions
ROI	Region of Interest
RMSE	Root Mean Squared Error
sCMOS	Scientific Complementary Metal-oxide-semiconductor
SGD	stochastic gradient descent
SIFNE	SMLM Image Filament Network Extractor
SIM	Structured Illumination Microscopy
SMLM	Single Molecule Localization Microscopy
SQUIRREL	Super-resolution Quantitative Image Rating and Reporting of Error Locations
SVM	Support Vector Machines
SRRF	Super-resolution Radial Fluctuations
SSIM	Saturated Structured Illumination Microscopy
SSIM	Structural Similarity Index
STORM	Stochastic Optical Reconstruction Microscopy
STED	Stimulated Emission Depletion
SOFI	Super-resolution Optical Fluctuation Imaging
S3	Simple Storage Server
TIRF	Total Internal Reflection Fluorescence Microscopy
TSF	Tagged Spot File
WGA	Wheat Germ Agglutinin
3B	Bayesian analysis of the Blinking and Bleaching

Thesis objectives and manuscript overview

This work aims to bring the recent advances in the field of artificial intelligence, specifically deep learning, to bear on the field of advanced microscopy, with the specific goal of improving the speed and quality of high resolution imaging techniques.

This thesis aims at the following objectives:

- use *Single Molecule Localization Microscopy* (SMLM) as an example to show how deep learning can be used for improving the imaging speed of a particularly popular type of advanced microscopy.
- show the advantages of deep learning methods and discuss their limitations.
- develop software tools and applications to facilitate the adoption of deep learning methods and enable data sharing.

This manuscript is organized as follows:

- Chapter 1 is a general introduction to microscopy and deep learning, covering fluorescence microscopy and advanced super-resolution microscopy techniques including SMLM, neural networks, convolutional neural networks, generative neural networks and network architectures for image restoration.
- Chapter 2 includes two parts: the first part is our manuscript describing a method we proposed for accelerating SMLM, called ANNA-PALM; the second part describes Shareloc, a data sharing platform for localization microscopy, including a discussion about how data size can improve deep learning, as well as a new file format designed for storing and sharing SMLM data.
- Chapter 3 discusses the power and limitations of deep learning approaches in general, some practical considerations about using deep learning methods, perspectives on how to improve the method and extend to more applications, and a general conclusion.
- Appendix includes two papers: A perspective paper which discusses the computational opportunities and challenges caused by the rising amount of biological imaging data; another paper about a computational framework to study sub-cellular RNA localization.

Contents

Abstract **5**

Publications **9**

Acknowledgements **11**

List of Abbreviations **13**

Thesis objectives and manuscript overview **15**

1 Introduction **1**

1.1 Advanced microscopy for biology 1

1.1.1 Brief history 2

1.1.2 Fluorescence microscopy 3

1.1.3 Fluorescent labeling techniques 5

1.1.4 Resolution and diffraction limit 8

1.2 Super-resolution microscopy 11

1.2.1 Single Molecule Localization Microscopy 13

1.2.2 Limitations of Single Molecule Localization Microscopy 14

1.2.3 Methods for accelerating Single Molecule Localization Microscopy 18

1.3 Introduction to deep learning . . . 23

1.3.1 Deep Neural Networks . . . 26

1.3.2 Deep learning for image reconstruction . . . 35

2 Methods and Results **41**

2.1 Deep learning massively accelerates super-resolution localization microscopy . . 41

2.1.1 A deep-learning approach to super-resolution image reconstruction . . . 43

2.1.2 Neural net architecture and learning strategy . . . 45

2.1.3 Validating ANNA-PALM on simulated images . . . 46

2.1.4 ANNA-PALM reconstructions of immunostained microtubules . . . 47

2.1.5 ANNA-PALM enables high-throughput super-resolution imaging . . . 48

2.1.6 ANNA-PALM is robust to experimental perturbations . . . 48

2.1.7 ANNA-PALM adapts to different biological structures . . . 49

2.1.8 Discussion . . . 49

2.1.9 Online methods . . . 52

2.1.10 Supplementary Information . . . 57

2.2 ShareLoc: A platform for sharing localization data . . . 81

2.2.1 Deep learning performance scales with data size . . . 81

2.2.2 Maximize the usage of existing data by sharing . . . 82

2.2.3 Challenges in storing and sharing localization data . . . 83

2.2.4 The ‘SMLM’ file format . . . 83

2.2.5 Experimental validation on file size and loading time . . . 85

2.2.6 The flexibility of the ‘SMLM’ file format . . . 86

2.2.7 Saving raw data in the 'SMLM' file . . . 87

2.2.8 Data sharing with the ShareLoc platform . . . 88

3 Discussion 95

3.1 Summary . . . 95

3.2 Limitations and possible solutions . . . 98

3.2.1 Generalization and memorization . . . 98

3.2.2 Systematic extrapolation . . . 100

3.2.3 The complexity of the structure . . . 102

3.2.4 Reliability and uncertainty . . . 104

3.3 Practical considerations . . . 108

3.3.1 The impact of training datasets . . . 108

3.3.2 Data augmentation strategies . . . 110

3.3.3 Training objective functions . . . 111

3.3.4 Better reconstructions than the training data . . . 112

3.4 Perspectives . . . 117

3.4.1 Improving the method . . . 117

3.4.2 Extended applications . . . 119

3.4.3 Live cell imaging . . . 120

3.5 Conclusion . . . 125

A The imaging tsunami: Computational opportunities and challenges 127

B A computational framework to study sub-cellular RNA localization 137

Bibliography **147**

Chapter 1

Introduction



1.1 Advanced microscopy for biology

Microscopy has been widely used in life science. It plays an essential role in revealing the biological processes beyond the reach of our eye, and has been recognized as an essential source of major discoveries in biology. Among different types of microscopy techniques, light microscopy, especially fluorescence microscopy, is widely used because it is less invasive and is suitable for living samples unlike other imaging techniques such as *Electron Microscopy* (EM). When combined with fluorescent labeling techniques such as immunostaining, microscopy has the ability to perform molecule-specific imaging. For example, by labeling several proteins of interest simultaneously, this multi-color imaging technique allows to disentangle the roles or functions of different molecules in complex cellular environment. Since the first microscope was invented, light microscopy was and still is continuously improved in terms of, for example, sensitivity, resolution and speed. Over the last decade, advances in chemistry, biology, physics, computer science and engineering have furthered the entire microscopy field into a new era of advanced microscopy.

1.1.1 Brief history

The first microscope can be dated back to the late 17th century and was invented by Antonj van Leeuwenhoek (1632–1723). He used a simple hand-held microscope with a single lens to study biological samples [Baker and Leeuwenhoek, 1739]. This important invention has led to the discovery of microorganisms. Later, in 1842, the Swiss botanist Karl Wilhelm von Nägeli discovered chromosomes in plant cells with the help of a microscope. Between 1911 and 1913, Otto Heimstaedt and Heinrich Lehmann developed the first ultraviolet fluorescence microscope, which was being used to investigate e.g. autofluorescence of bacteria, protozoa, and animal tissues. Shortly, in 1914, this microscope was then instrumental to Stanislav Von Prowazek who used it to study fluorescence binding to living cells. In the early 1940s, Albert Coons developed the first immunofluorescence technique using antibodies with fluorescent dyes to label biological structure. Since then, more and more variations of fluorescence microscopy including *Confocal Microscopy* [Minsky, 1988], *Phase Contrast Microscopy* [Zernike, 1955], *Total Internal Refection Fluorescence Microscopy* (TIRF) [Axelrod, 1981] and *Multi-photon Excitation Microscopy* [Zipfel et al., 2003] have been developed. The invention of the *Charge-Coupled Device* (CCD), a sensor digitalizing optical images, in 1970s, and of digital cameras brought light microscopy into the digital era. Nowadays, almost every microscope is equipped with a digital camera and connected to a computer or micro-controller. Images and movies are acquired in digital formats, stored on hard-drives or storage servers allowing to be further visualized and analyzed. This digitalization of images thus allows for digital image processing using modern computers. Microscopy images can be processed to e.g. enhance the contrast or to improve *Signal-to-Noise Ratio* (SNR) for better visualization. Furthermore, the automated analysis of microscopy images has become a standard method for biological objects such as cells, their organelles or even viruses. The structure of interest is labeled with fluorescent dyes, then images and/or movies are recorded, which are then being used to extract information such as number, size, shape, or moving trajectories of the structure of interest. These analysis are facilitated by open software tools such as *ImageJ* [Schneider et al., 2012] or *Cellprofiler* [Carpenter et al., 2006]. Recently, advanced microscopy techniques are even more closely integrated into digital techniques. Lenses, mirrors and other optical devices are controlled precisely with micro

controllers and computers; images are processed in real-time and used as a feedback signal to direct the imaging process. Very recently, more advanced signal processing techniques such as machine learning are combined with powerful computers to further enhance and augment the capability of microscopy. Such augmented microscopy techniques can resolve smaller details, capture faster motions, image deeper into the specimen and are less harsh for the imaged sample.

1.1.2 Fluorescence microscopy

Fluorescence is a physical property of certain molecules or proteins (also called *fluorophores*) which can emit light. This light emission is caused by the excitation of electrons, typically after absorbing photons of a specific wavelength. As shown in Figure 1.1 (a), (b), when a fluorophore absorbs an excitation photon, it enters an excited state. A fluorescence signal is emitted when the electronic state of the fluorophore falls from an excited state back to the ground state. In most of the case, the emit light has a longer wavelength and less energy than the excitation light. The difference in energy is also called *Stokes Shift*. Fluorophores are typically characterized by their brightness and their spectra of excitation wavelength and emission wavelength. Depending on the type of fluorophores, many of them have limited lifetime. After some cycles of excitation and emission, photochemical alteration can occur and the fluorophore will be permanently unable to fluorescent, this is also called *photobleaching*.

The most commonly used fluorophores are organic small molecules of 20-100 atoms and include e.g. DAPI, Cy3, Cy5, Alexa Fluor® 647. While some dyes such as DAPI and Phalloidin can directly label a specific target in the cell (DNA and F-actin respectively), many of them are typically coupled to antibodies which bind to a specific protein target in the cell thus linking the fluorophore to the structure of interest. Nowadays, it is also common to combine fluorophores with protein tagging systems (e.g. HaloTag® and SNAP-tag®), or Click Chemistry [Kolb et al., 2001] which provides specific binding to the target protein. More details about labeling techniques is described in Section 1.1.3.

Moreover, proteins emitting light themselves, such as the *Green Fluorescent Protein* (GFP)

[Tsien, 1998] have been widely used in cell biology since it enables fluorescence imaging in live cells. GFP was first isolated from a jelly fish and its gene was inserted into *E.coli* bacteria in 1994. The discovery of GFP has opened many doors in molecular biology. It is commonly used as a reporter of gene expression– the GFP gene is typically inserted adjacent to the target protein and shares the same regulatory sequence (e.g. promoters and enhancers) such that GFP is expressed and fused with the target protein. For instance, GFP has been used as a marker for gene expression in living adult fruit flies, *Drosophila melanogaster*, plant and mammalian cells [Plautz et al., 1996].

Although GFPs can be used alone, it is often combined with fluorescence microscopy which can give detailed spatial localization and distribution of the target protein. As an example, by combining two fluorophores which bind to different protein targets or different parts of the same protein.

When acquiring fluorescence images, the information about the imaging sample is transmitted by the emitted fluorescence signal. This signal is often relatively weak and can only be detected with highly sensitive image sensors. Besides that, it is important to separate it from the excitation light signal. A typical design in a fluorescence microscope (Figure 1.1 (c)) is to use an optical filter to filter out the excitation light before it reaches the image sensor, i.e. the camera. This excitation light source can be a white light lamp with a band-pass filter or a laser which can generate narrow bandwidth light around a specific wavelength. In addition, fluorophores which have a good separation between the excitation wavelength and emission wavelength are preferred because they result in less background in the acquired fluorescence images.

Preferred specifications of a good fluorophore are high brightness, photo-stability to avoid photo-bleaching and thus loss of signal and little non-specific binding. Meanwhile, photo-bleaching of the dye is caused by reacting with molecular oxygen, which further leads to photo-toxicity or even irreversible photo-damage because it produces free radicals that can damage sub-cellular components. Therefore, when imaging living samples, photo-toxicity of the dye should be minimal since it may affect the physiological state or alter the behavior of the imaged

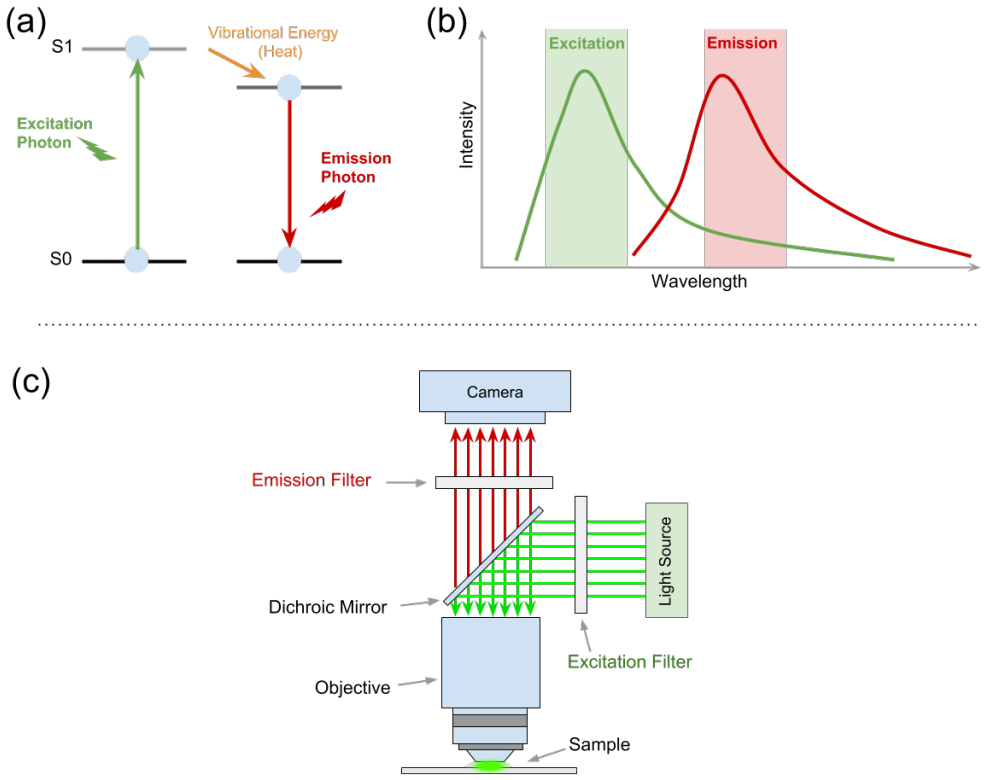


Figure 1.1: The excitation and emission of fluorescence. (a) When the excitation photon from an external light source hits the fluorophore, it will force the fluorophore to go to an unstable excited electronic state (S1) from the ground state (S0). Shortly after (typically in a few nanoseconds) the excited molecule generally relaxes towards the vibrational energy level within the excited state and emits a small amount of heat. When it falls from the excited state to the ground state again, all the rest of the energy is transformed into emission of a single photon at a wavelength determined by the energy gap (energy $E = hf$ where h designates the *Planck* constant, and f designates the frequency); (b) A fluorophore typically responds to a spectrum of excitation wavelength and the emission wavelength lays in a shifted spectrum; (c) A basic configuration for a fluorescence microscopy consists of a light source which emits the excitation light which passes through an excitation filter, is then reflected by the dichroic mirror and passes through the objective and hits the sample labeled with fluorophores. Once excited, the fluorophore emits a fluorescent signal which goes back to the objective, passes through the dichroic mirror and is registered by the camera. Importantly, an emission filter is inserted between the dichroic mirror and the camera to filter out excitation light.

sample.

1.1.3 Fluorescent labeling techniques

Among those various ways to perform fluorescence imaging, *immunofluorescence* (IF) imaging, which uses antibodies conjugated with fluorescent dyes to label protein targets, is one the most

commonly used techniques. The first immunofluorescence experiment was done by Albert H. Coons and his colleagues in 1941, they described the identification of tissue antigens using an IF protocol in [Coons et al., 1941].

IF can be divided into direct IF and indirect IF depending on where the fluorophore is attached. In indirect IF, the protein of interest in the biological sample is first labeled with a primary antibody which specifically detects an epitope of the protein of interest, and this primary antibody is then detected by a second antibody conjuncted to a fluorescence dye. Compared to direct IF which does not require the first step, indirect IF is more flexible, less expensive, and the signal can be amplified by using multiple fluorophores bind with the same second antibody, or multiple secondary antibody bind to the same primary antibody.

However, they can not be used for living samples mainly because the penetration of the antibody into the cell is difficult. Thus, immunostaining is applied to cells which have been fixed and permeabilized. A common protocol is to use protein crosslinking reagents such as *Paraformaldehyde* (PFA) for fixation and Triton, a detergent, for permeabilization. Also, sometimes Methanol is used for both fixation and permeabilization. Whilst allowing the antibody to penetrate the sample after fixation, one major drawback is that its ultrastructure is not always preserved and causing artifacts in the resulting image. Furthermore, depending on the fixation method, the antibody may not recognize its target epitope and causing low SNR in the image.

For living cells, fluorescent proteins such as GFP are introduced to the cell and are expressed and fused together with the target protein. Specifically, a piece of fluorescent protein gene which consists of 238 amino acids is inserted to the genome of the target cell at the location of the target protein (also know as knock-in), for example using a technique called CRISPR/Cas9 (where CRISPR stands for *Clustered Regularly Interspaced Short Palindromic Repeats* and Cas9 is the enzyme which the two strands of DNA) [Cong et al., 2013, Hsu et al., 2014], a gene editing tool which allows to cut the host cell DNA at a specific location and replace the target sequence by a new DNA sequence, for example that of GFP. While genome editing may take a considerable amount of efforts to implement, there are also alternative approaches such as transfection. In transfection, the fluorescent protein sequence is introduced with the

target protein sequence by using a plasmid DNA or a viral DNA. For instance, a standard plasmid for transfection is a circular DNA consisting of the DNA sequence of the target protein, the fluorescent protein, and other utility sequences which facilitate transfection, regulatory sequences to regulate the expression, or provide drug resistance for selecting transfected cells.

Although fluorescent proteins are widely used in live cell imaging, their photophysical properties can be less optimal than organic dyes, due to their comparatively low brightness, which often leads to images with low signal to noise ratios. Therefore, many protein tagging systems such as HaloTag® and SNAP-tag® have been developed, which provide a way to use bright organic dyes in live cell by introducing a specific tag sequence into the cell using CRISPR/Cas or standard transfection methods. When the tagged protein is expressed, it can be recognized by a ligand caring the organic fluorophore. Since the ligand binds to the protein specifically, the labeling is specific. Protein tagging systems are also used a better alternative to IF for fixed samples due to the advantage of allowing more direct labeling since the distance between the dye and the protein is much smaller.

Once the target protein is labeled with the fluorophore, the emission of the fluorophore will reflect the presence of the target protein in live or fixed cells. The quality of the produced image can be evaluated with qualitative or quantitative measurements including the signal-to-noise ratio, contrast, background intensity, noisiness, resolution, the amount of artifacts etc. The imaging experiment can be seen as a production pipeline, where every process occurring in the pipeline will affect the quality of the final product – the fluorescence image. Problems leading to the degradation of the final image can be caused by any process or step involved in the experiment. For example, the sample preparation (e.g. fixation and IF) may introduce artifacts, the physical properties of the fluorescent protein or dye which could lead to fast photo bleaching, auto fluorescence which causes high background intensity, background noise caused by the camera, the mismatch between the fluorophore and the filters attached to the microscope, all contribute to the quality of the final product, the image.

1.1.4 Resolution and diffraction limit

The purpose of using a microscope is to visually resolve small objects such as cells, bacteria, viruses and molecules. The resolution of a fluorescence image is one of the most important features when interpretations are made based on the image. However, even if all parts of an imaging system can be considered to be perfect, most microscopes suffer from optical resolution limit which is called *diffraction limit*. This limit restricts limits the optical resolution to around $200 \sim 300nm$; objects smaller than these length scales will be unresolvable with a standard light microscope.

The limit of diffraction was first described by Ernst Abbe in 1873 and can be characterized by the *Rayleigh criterion*, which is defined as :

$$r = \frac{0.61\lambda}{NA} = \frac{0.61\lambda}{nsin(\theta)}$$

Where r is the distance between the two objects which we try to observe (e.g.: two fluorophores), λ is the wavelength of the light, n is the refractive index of the light medium. NA is the *numerical aperture* of the objective lens. With the Rayleigh criterion we can calculate a minimum resolvable distance between objects under a microscope. For example, using an objective lens with $NA = 1.4$ and a fluorescent dye which emits light at 647 nm wavelength, we can get r =282 nm which means the microscope can only resolve details larger than 282 nm. Figure 1.2 shows three pairs of fluorophores with different distances. The corresponding images captured by the image sensor and the intensity profiles show three different cases when the fluorophores are unresolved, just resolved (at the Rayleigh criterion distance) and fully resolved.

In terms of the spatial frequency in a image, the Rayleigh criterion defines the cut-off frequency of the signal acquired with a microscope, meaning that the imaging system can not resolve objects which are smaller than the distance corresponding to the cut-off frequency. For fluorescence microscopy, this means that a microscope that is constrained by the diffraction limit can be used to observe biological samples and resolve structures at the level of organs,

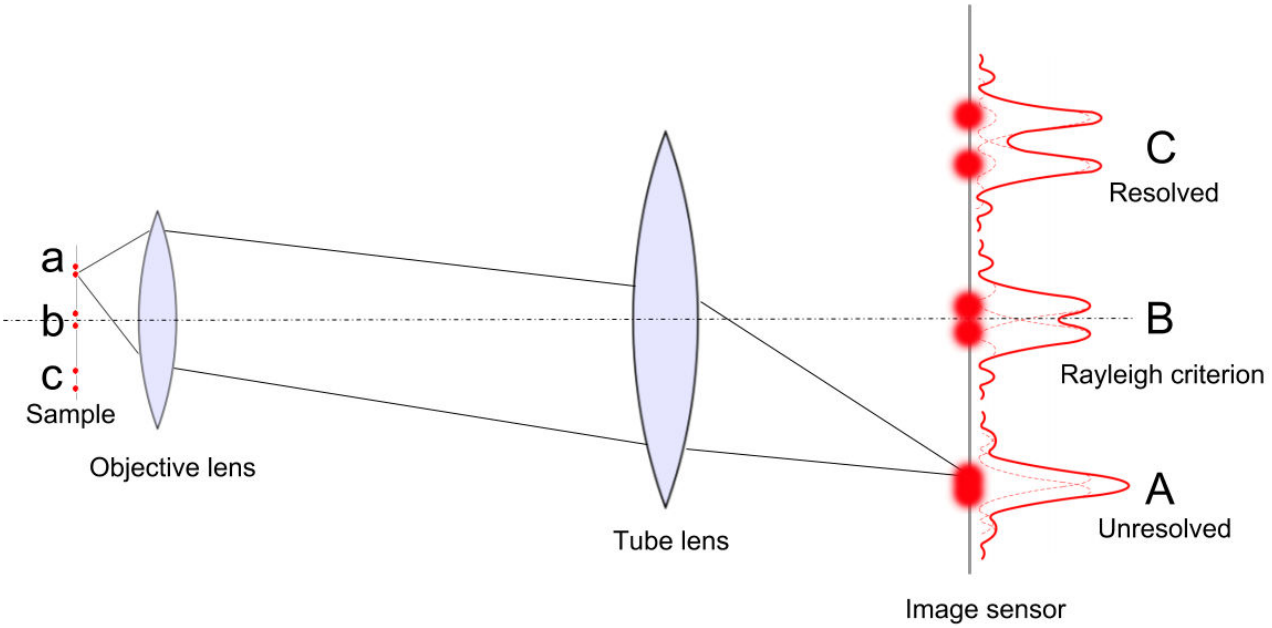


Figure 1.2: Diffraction limit and Rayleigh criterion. As shown in the figure from left to right, 3 pairs of fluorophores with different distances between them are placed at three different positions (**a**,**b** and **c**). The distance between two fluorophores at position **b** is equal to the Rayleigh criterion, fluorophores at the other two positions have distances smaller (**a**) or bigger (**c**) than the Rayleigh criterion. When an illumination beam excites these fluorophores, they will emit light. The emission light passes through the objective lens and the tube lens, and is then captured by the image sensor (a camera). Due to the diffraction of light, the image of each points is a blurry spot. Depending on their physical distance, the image corresponding to fluorophores at position **a** is unresolved (**A**), **b** is just resolved (**B**), or **c** is fully resolved (**C**).

tissue, or cells as small as bacteria (typically 1-2 micrometers in length), but it gives blurry images when observing sub-cellular structures, bacteria or viruses. In order to see smaller details beyond the Rayleigh criterion, one needs to choose an objective with a bigger NA, which implies using large refractive index immersion medium, closer distance to the sample, and also smaller wavelength λ. However, in practice, it is difficult to get an objective lens with large NA and most of the objectives have NAs smaller than 1.5. Meanwhile the wavelength of light is limited by the fluorophore itself. For live cell imaging, smaller wavelength means higher energy and is more toxic to cells. Therefore, the diffraction limit becomes the resolution barrier for most of the light microscopy. A commonly used term to describe the response of an imaging system to a point source or point object is the *Point Spread Function* (PSF). Intuitively the image of PSF represents an object whose *full-width-half-maximum* (FWHM) is about the size defined by Rayleigh criterion.

In cell biology, most structures of interest such as most proteins, DNA, RNA, and organelles are smaller than the diffraction limit. Therefore, going beyond the diffraction limit is essential for the understanding of many biological phenomena and thus answering important scientific questions.

1.2 Super-resolution microscopy

Many techniques were proposed for improving the resolution of optical microscopy beyond the diffraction limit, and they were termed *super-resolution microscopy*.

Depending on the working distance between the objective lens and the sample, they can be divided into near-field and far-field techniques. For example, [Pohl et al., 1984] proposed the *Near-field Scaning Optical Microscopy* (NSOM) which can achieve a resolution of $20 \sim 50nm$. However, NSOM works at a distance of tens of nm from the sample, and therefore has limited applications in biology. Farfield techniques which have larger working distance have more practical meaning for real-life applications in biology. There are several widely used far-field techniques that can moderately improve the resolution including confocal microscopy and multi-photon microscopy (also known as *Non-linear microscopy*) [Zipfel et al., 2003]. Confocal microscopy uses focused laser beams for excitation and a pinhole for detection, and improves the theoretical resolution by a factor of $\sqrt{2}$. However, in practice it offers little resolution improvement mainly due to the fact that the pinhole reduces the amount of photons being detected and lower the SNR [Cox and Sheppard, 2004]. Multi-photon microscopy utilizes nonlinear absorption processes to reduce the excitation PSF size, however it uses a longer wavelength and can only achieve a resolution similar to confocal microscopy (around 250 nm) [Cox and Sheppard, 2004].

Bigger effective NA have already been explored by various techniques including the 4Pi microscopy and I^5M [Gustafsson et al., 1999]. They use multiple objective lens to maximize the coverage of photon detection, and have shown to improve the spatial resolution to around 100 nm [BEWERSDORF et al., 2006].

While in most of the regular microscopy techniques a homogeneous light field or spot light source is used to illuminate the sample, patterned illumination has been explored for increasing the resolution. Among them, *Structure-illumination Microscopy*(SIM) has been proposed based on the Moiré effect – a phenomenon which allows to obtain high frequency information from lower frequency observations. As shown in Figure 1.3 (A), by acquiring multiple images illuminated

by light patterns with different phases and orientations, for example periodic light stripes generated with interference of multiple light sources, an image with higher resolution can be reconstructed.

However, the illumination pattern used in SIM is also diffraction limited, it improves the resolution only by a factor of two, i.e. at most to $100nm$. One way to improve this resolution is to incorporate sub-diffraction sized features within the illumination pattern. Several approaches have been developed along this direction including *Stimulated Emission Depletion* (STED) [Westphal and Hell, 2005], and *REversible Saturable Optical Linear Fluorescence Transitions* (RESOLFT) [Hell, 2007], *Saturated Structured Illumination Microscopy* (SSIM) [Gustafsson, 2005].

In STED microscopy, a laser beam is used to excite the fluorophore and the sample is scanned in a similar way as in regular confocal microscopy. To achieve higher resolution, STED uses a second donut-shaped beam (STED laser) to deplete the fluorescence emission from the first excitation beam, which results in an effective PSF with sub-diffraction size (Figure 1.3 B). As a result, the resolution of STED can reach down to 30 nm [Westphal and Hell, 2005]. Saturable processes can be also applied to improve SIM and the technique is called *Saturated Structured Illumination Microscopy*(SSIM). It uses sinusoidal pattern with strong excitation power, and as a result fluorophores receiving higher doses of power will be saturated and flattened in the peak position of the sinusoidal, while the fluorophore located at the valley are still in the dark state and appear as black stripes with sub-diffraction sizes. This introduces high frequency features and improves the resolution in addition to the one obtained by the Moiré effect. It has been demonstrated that SSIM can achieve 50 nm resolution [Gustafsson, 2005].

However, since the saturated depletion only happens when using a high laser power($>10^9 W/cm^2$) and lower depletion laser power will decrease the resolution, it is challenging to use STED for live cell imaging. To address that, an extended method called *REversible Saturable Optical Linear Fluorescence Transitions* (RESOLFT) [Hell et al., 2003] was proposed. It is similar to STED, but employs photo-switchable fluorescent probes which greatly reduce the power of the depletion power down to 600 W/cm^2. The term RESOLFT is also being used

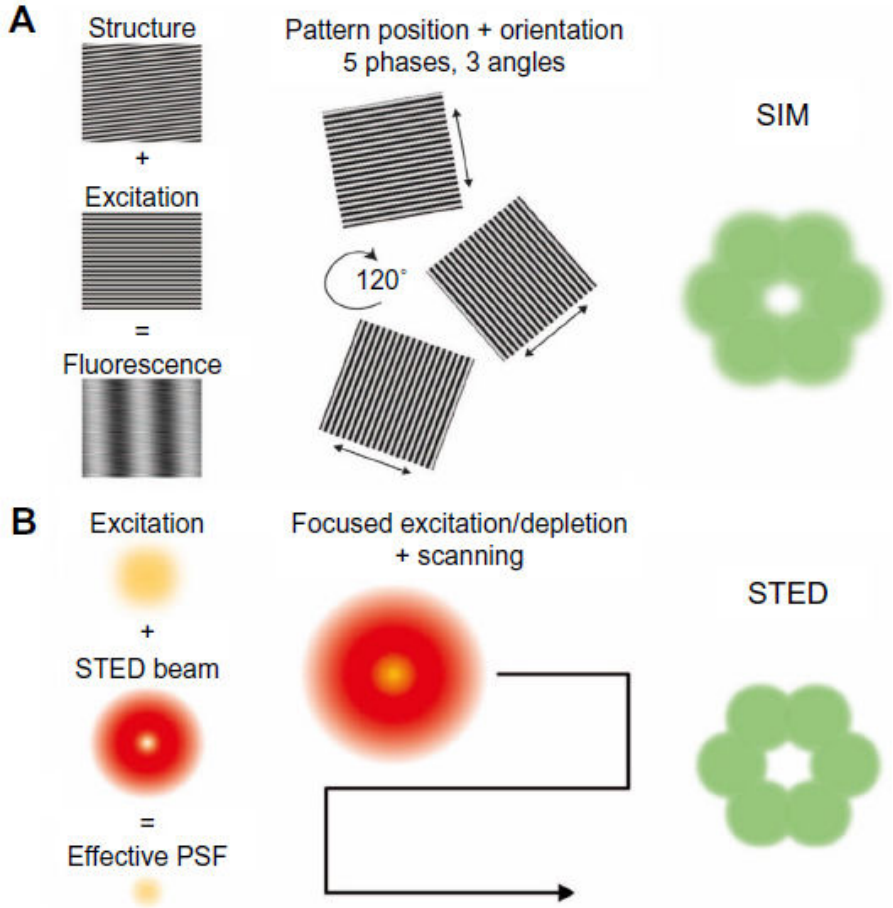


Figure 1.3: Three super-resolution microscopy techniques: SIM and STED [Nollmann and Georgieva, 2015]. (A) In SIM, an excitation beam applied on a structure with higher spatial frequency results in fluorescence emission with smaller spatial frequency. By shifting and rotating the excitation pattern and acquiring multiple images, the underlying structure with details at higher spatial frequency can be reconstructed (shown on the right column). (B) In STED, an excitation beam is applied together with a donut shaped STED beam which results in an effective PSF with sub-diffraction size. The two beams are used to scan the sample, by collecting all the fluorescence signal and assembling into an image as in confocal microscopy.

as more generalized concept for saturable depletion based super-resolution microscopy [Hell, 2007] such as STED, SSIM and RESOLFT with different switchable dyes.

1.2.1 Single Molecule Localization Microscopy

A fluorescence image reflects the spatial distribution of the underlying fluorophores, which is further linked to the labeled biological structure. Determining the spatial localization of these fluorophores is therefore a suitable way to assemble an image. Following this strategy, a different type of approach for achieving super-resolution was proposed in 2006, which was initially called

Photoactivated Localization Microscopy (PALM) [Betzig et al., 2006] or *Stochastic Optical Reconstruction Microscopy* (STORM) [Rust et al., 2006]. These light microscopy techniques (and many other variants) are now commonly referred as *Single Molecule Localization Microscopy* (SMLM).

Instead of exciting all the fluorophores at the same time, in SMLM, specific types of fluorophores are used to make sure that only a subset of fluorescent molecules are excited and emit fluorescent signal at any given time. As shown in Figure 1.4, for a labeled sample under the microscope, a movie of these blinking molecules is recorded which typically consists of several thousands or more of frames. For each frame of the movie, the fluorescent signal consists of isolated bright spots with diffraction limited sizes. Since the intensity profile of these spots approximately follows a Gaussian distribution [Zhang et al., 2007], the precise peak position can be localized using a Gaussian fitting algorithm. The resulting localizations of the fluorophores are stored in a table which can be visualized as a 2D histogram of the x, y coordinates. Since the localization precision of the Gaussian fitting can be very high (a few nanometers for images with good SNR), SMLM pushes the resolution limit down to 10 $30nm$ or less. Figure 1.5 shows an example of STORM image compared with conventional total *internal reflection fluorescence* (TIRF) imaging which has a resolution similar to widefield microscopy. More recently, by combining STED with SMLM, a method called *MINimal emission FLUXes* (MINFLUX) [Balzarotti et al., 2017] has been proposed for pushing the resolution of light microscopy even further down to $\sim 1nm$ – a size close to or smaller than the physical size of a fluorescent molecule (e.g.: GFP is about $\sim 5nm$ long).

1.2.2 Limitations of Single Molecule Localization Microscopy

In order to achieve nano-scale resolution in SMLM, at least two conditions need to be fulfilled:

1. For each frame of the diffraction limited images, the fluorescence spots need to be spatially separated from each other in order to be localized with a high localization precision. Even though overlapping spots can be potentially identified by algorithms that simulta-

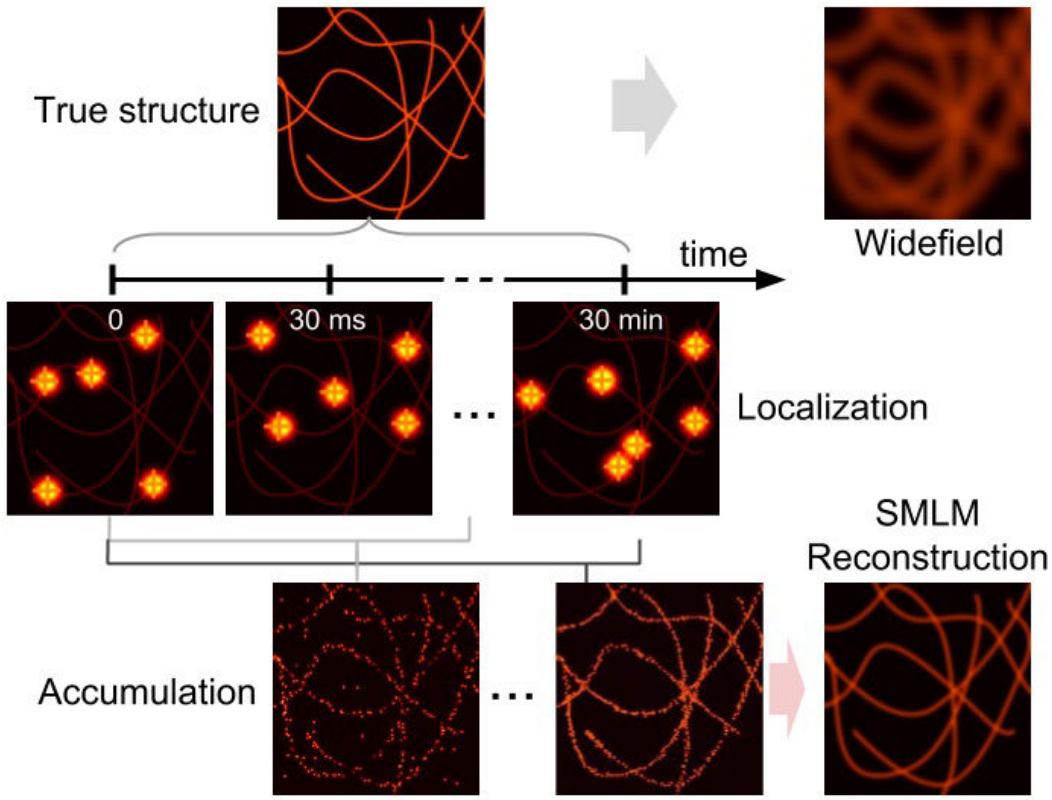


Figure 1.4: Image reconstruction with SMLM. One single image in SMLM imaging contains isolated fluorescence spots. As an example, 60,000 such images (30 ms exposure time, 30 minutes in total) are recorded sequentially and each isolated fluorescent spot in each image is then localized with a appropriate algorithm (e.g. Gaussian fitting) resulting in a series of localized coordinates saved in a localization table. X and Y coordinates can be used to render a 2D histogram image. By adding more and more localizations, the rendered image becomes closer and closer to the underlying biological structure. This super-resolved image has a much higher spatial resolution than the wide-field image.

neously fit the positions of multiple molecules, a high density of fluorescent spots in a small area will necessarily increase the uncertainty of the localization [Ram et al., 2006]. This then leads to poor localization accuracy, which further lowers the resolution of the reconstructed image.

2. A sufficient total number of localizations needs to be acquired in order to correctly represent the underlying true structure at a high resolution. When rendering the reconstructed image with a 2D histogram, the acquisition can be seen as a Poisson sampling process over time dimension, and the reconstructed image is a Poisson realization of the true underlying structure. A sufficient number of localizations need to be accumulated in order to better approximate the true image. The minimal requirement of localizations for a

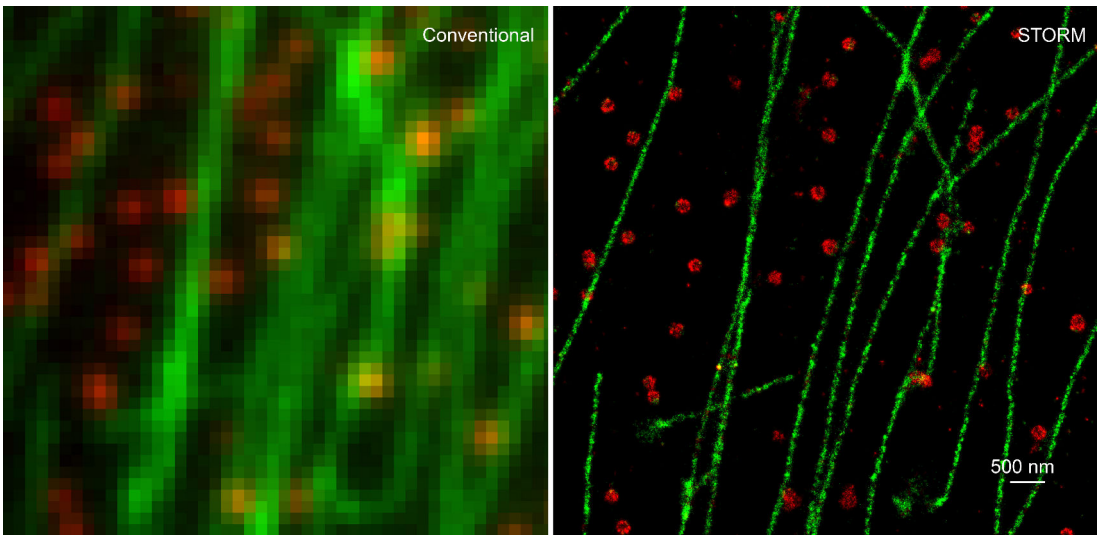


Figure 1.5: Comparison between STORM and conventional TIRF microscopy on two-color imaging [Bates et al., 2007]. The image shown on the left is produced with TIRF and the right is from STROM. In both images, microtubules are shown as green and clathrin-coated pits are red.

giving underlying image is often linked with Nyquist criterion by calculating the density of the localizations and matched with $(2/r)^2$, where r is the expected spatial resolution one would like to achieve [Shannon, 1949, Shroff et al., 2008]. For example, to achieve 10 nm resolution, the density of localization must be higher than $0.04/nm^2$. However, this does not mean that the theoretical resolution of 10 nm is achieved once the density is reached, since the resolution is also bounded by the localization precision. More recently, [Legant et al., 2016] suggest to use fivefold or more molecules compared to the number defined by Nyquist criterion in order to generate images qualitatively comparable to *electron microscopy* (EM).

These two conditions lead to one drawback of SMLM, which is low imaging speed, since high spatial localization can only be achieved by acquiring a large number of diffraction-limited images. SMLM requires the control of the density of excited fluorophores to obtain partially separated localizations in one given image which makes SMLM inherently slow. Typically $10^3 - 10^5$ diffraction-limited frames are required for the reconstruction, and it typically takes minutes to hours to acquire one super-resolution image. Obviously, this limits its application in high-throughput imaging where imaging speed is important or live cell imaging where temporal resolution is crucial.

An additional constraint for SMLM is that the sample should stay fixed relative to the objective lens, otherwise the quality of the image will decrease dramatically because of motion blur. Due to the long acquisition times needed for SMLM imaging, the relative motion occurring between the sample and the objective is captured and can result in blurring of several microns in both dimensions. Even though this motion also occurs during low resolution imaging, it is usually negligible due to the short exposure time of a single frame. This movement of the sample, often called drift, is mostly due to gravity, fluctuations of temperature or vibrations and it cannot be avoid completely, even with a fixed sample stage. However, drift can be corrected by first using a software or device to track the sample drift, and then applying it to the localization table and resulting super-resolution image.

There are two types of most commonly used drift correction algorithms: 1) adding fiducial markers such as fluorescent beads into the sample, and then track their trajectory ; 2) dividing the localization table into subsets with each of them containing several subsequent frames and reconstructing super-resolution images with these groups, then using cross correlation analysis to extract the relative motion between the localizations of these groups which is then being interpolated to the entire movie. Once the drift trajectory is recovered by these methods, opposite offsets are applied to correct the localization table.

While drift correction algorithms can efficiently reduce the motion blur in the reconstructed high resolution image, they have limitations. For example, inserting beads as fiducial markers will affect the surrounding regions of the beads by e.g. blinding the camera due to its brightness when imaging structures with low intensity; or the beads are not fixed on the sample and thus move; or the beads can photo-bleach before the end of the acquisition; using cross-correlation may fail or decrease its performance when the noise level is too high.

1.2.3 Methods for accelerating Single Molecule Localization Microscopy

Multiple approaches have been proposed to address the limitation on imaging speed by improving the localization algorithm such that a higher density of fluorescent signal, possibly highly overlapping spots can be localized. With higher laser power, smaller camera exposure time or dyes with fast switching kinetics, a shorter movie with higher localization number per frame can be acquired, which further results in faster SMLM imaging with a few seconds or even video-rate (>24 frame per second) temporal resolution [Huang et al., 2013, Jones et al., 2011]. While most of the localization algorithms can only fit a single fluorescent emitter, algorithms dealing with high density data often fit the acquired image with a multi-emitter model using maximum likelihood estimators [Huang et al., 2011, Simonson et al., 2011]. Since the search space is much larger than single-emitter fitting and solving the optimization is a iterative process, multi-emitter fitting algorithms are typically much slower than single-emitter fitting algorithms. For example, depending on different molecular density, a commonly used algorithm called DAOSTORM [Holden et al., 2011] process images at the speed of $1 \sim 10$ second per frame [Takeshima et al., 2018], while recent single emitter fitting algorithm [Przybylski et al., 2017, Li et al., 2018] can achieve real-time speed (usually more than 30 frames per second).

For an even higher density of excitation, where no single molecule can be found in each frame of the movie, some other approaches were proposed by analyzing or modeling the blinking and bleaching process. For example *Bleaching/blinking assisted Localization Microscopy* (BaLM) uses a theoretical Gaussian [Burnette et al., 2011] to detect and localize fluorophores from their bleaching and blinking events. Another example is *Bayesian analysis of the Blinking and Bleaching* (3B analysis) [Cox et al., 2012], which uses factorial hidden Markov model [Ghahramani and Jordan, 1997] to model high density frames. It can use information from bleaching and blinking events for high density localization. While 3B analysis produces a localization table, *Super-resolution Optical Fluctuation Imaging* (SOFI) [Dertinger et al., 2009] uses higher-order statistical analysis of temporal fluorophore intensity fluctuations to generate the super-resolution image directly. Most of these algorithms are computationally intensive and processing

large datasets can be difficult due to iterative computation. Recently, *Super-resolution Radial Fluctuations* (SRRF) was proposed as a fast alternative. It combines radial symmetry which can be calculated analytically and temporal analysis which can merge information across multiple frames, and has demonstrated its efficiency for live cell imaging [Gustafsson et al., 2016]. These algorithms can reconstruct super-resolution images from very high density fluorescence movies and enable live cell imaging with high resolution.

Even though all the above mentioned approaches substantially increase the speed of SMLM, they all degrade the spatial resolution when compared to low density data as a result of lower localization precision of overlapping fluorophores, and in most cases they generate artifacts. In general, any algorithm working with high density data will have to deal with the inherent uncertainty in the observation, thus compromise its performance. Figure 1.6 (a) shows comparisons on reconstruction speed and quality between different high density localization methods(SRRF, 3B, SOFI, DeconSTORM, DAOSTORM, FALCON and Integrated Gaussian implementation in ThunderSTORM). When processing images with higher and higher density, different algorithm performances differently, but all of them face common problems such as low spatial resolution, missing filaments, incorrect merging of filaments, artificial sharpen, etc. However, a recent work by [Marsh et al., 2018] introduces a preprocessing method to generate low density frames from high density images with a set of Haar wavelet filters, such that single emitter fitting algorithms can be used for processing high density data. As shown in Figure 1.6, it greatly reduced the amount of artifacts compared to other algorithms such as multi-emitter fitting (ME), SOFI and SRRF.

Beside the drawbacks of decreasing spatial resolution and generating artifacts with high density methods, high power excitation light used to increase the density can cause photobleaching of the fluorophores, which further reduces the number of fluorophores actively emitting light. When used with live samples, high excitation power increases phototoxicity leading to changes in physiological state of the cells. Although recent techniques such as *Lattice Light Sheet Microscopy* [Chen et al., 2014] can dramatically reduce the light dose by using a thin slice (usually a few hundred nanometers to a few micrometers) of light to illuminate the sample. However, the implementation is less affordable and still limited to a small number of laboratories.

Altogether, these techniques are limited by the amount of information flowing from the sample to the final reconstructed image. Since the bandwidth of the information flow is limited, they are inherently struggling between the trade-off of resolution, sensitivity and speed (and also sample health for live cell imaging) which is also known as the *eternal triangle* [Resch-Genger and Barker, 2008]. To overcome this problem, a fundamentally different approach is needed, such as the one we put forward in this work. We make use of recent advances in the field of artificial intelligence, called *deep learning*.

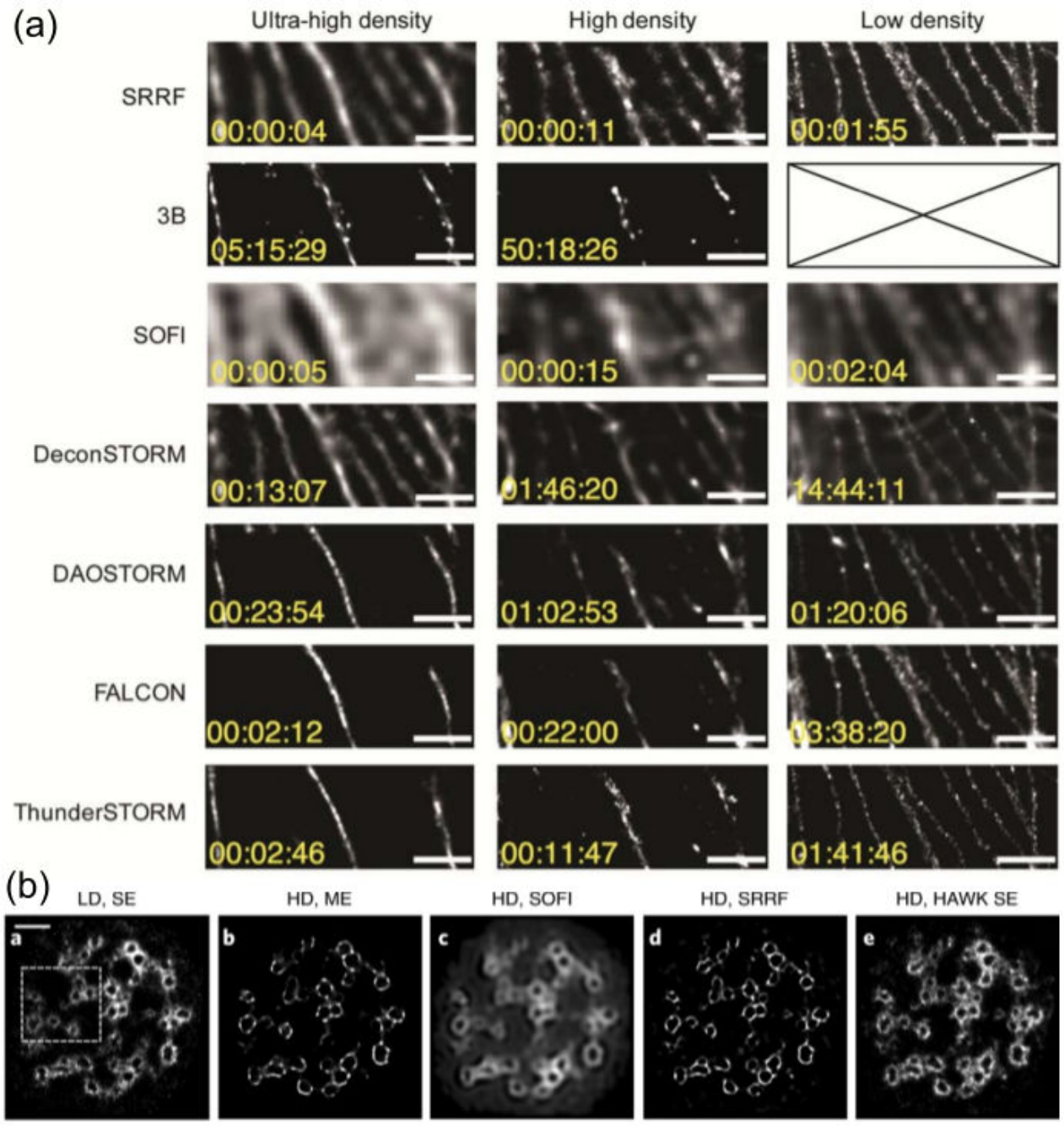


Figure 1.6: Comparisons between different high-density localization algorithms [Gustafsson et al., 2016, Marsh et al., 2018]. (a) shows comparisons between 7 different high density localization algorithms in terms of processing ultra-high density, high density and low density microtubule images. All these algorithms decrease their performance when dealing with higher density images, artifacts can be easily identified in the left and middle column when compared with the right column. The numbers in yellow shows the reconstruction time for each experiment. A detailed description can be found in Supplementary Note 1 of [Gustafsson et al., 2016]. (b) is an example from [Marsh et al., 2018] which shows how the HAWK preprocessing followed by single emitter fitting (SE) leads to "artifact-free" results compared to low density data processed with a single emitter fitting algorithm (labeled as "LD, SE"), high density data processed with multi-emitter fitting algorithm(labeled as "HD, ME"), SOFI (labeled as "HD, SOFI") and SRRF (labeled as "HD, SRRF").

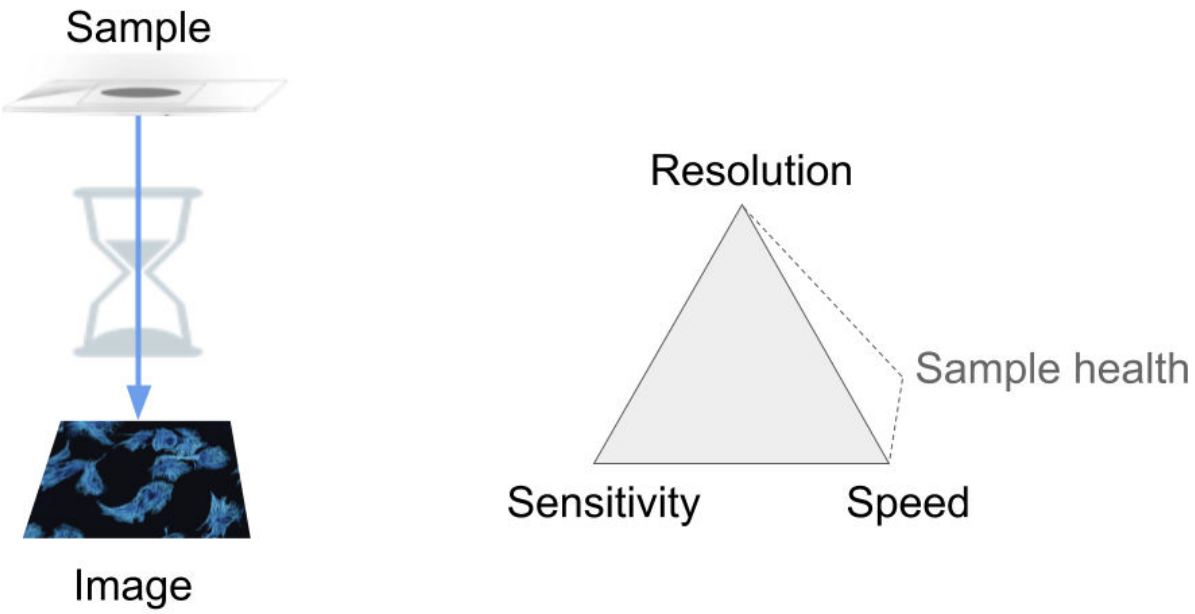


Figure 1.7: The information bottleneck and eternal triangle in imaging. When imaging with a microscopy, the information flow from the sample to the image has a limited flow rate or bandwidth, this further lead to the *eternal triangle*(or pyramid) which describes the trade offs between resolution, sensitivity and speed (also sample health for living samples).

1.3 Introduction to deep learning

Artificial intelligence (AI) is a thriving field which aims to make machines and algorithms capable of solving tasks like humans. Machine learning is one of the most commonly used techniques in AI with the goal is to design learning algorithms which can adapt to data and process new data. *Deep learning* is a particular type of machine learning technique which is inspired by the human brain. It uses artificial neural networks to perform learning tasks. Thanks to the ever growing amount of data and computing power we can afford nowadays, training large deep learning models with large datasets has become possible. In the past few years, deep learning was continuously making breakthroughs in many fields including image recognition, speech recognition, text-to-speech synthesis, machine translation, just to name a few [LeCun et al., 2015]. Many of the deep learning algorithms have been designed with human level or super-human level performance. For example, as shown in Figure 1.8, in the ImageNet [Deng et al., 2009] , a visual recognition challenge for object detection and image classification, the error rate has dropped from 20% (in 2010) to 2.25% (in 2017), while for an average human, the error rate is around 5%. Since 2012, all the top winners of the challenge are deep learning algorithms. Other examples of successes empowered by deep learning include human-level gaming control [Mnih et al., 2015], super-human level performance in the Go game [Silver et al., 2016, Silver et al., 2017], super-human level visual reasoning [Santoro et al., 2017], dermatologist-level skin cancer classification [Esteva et al., 2017] and super-human level traffic sign classification [Cireşan et al., 2012].

Traditional machine learning algorithms often require hand-engineered features as input such as support vector machines, random forests etc. For example, a commonly used approach for face recognition, before the use of deep learning, is to describe the face images with hand-crafted descriptors such as *Local Binary Patterns* (LBP) [Ahonen et al., 2006]. A typical workflow of such an approach is shown in Figure 1.9(a).

By contrast, deep learning algorithms perform end-to-end processing, which means that they start from the raw inputs without any feature extraction step and directly provide the result. For example, in computer vision tasks, deep learning algorithms process the raw images in the

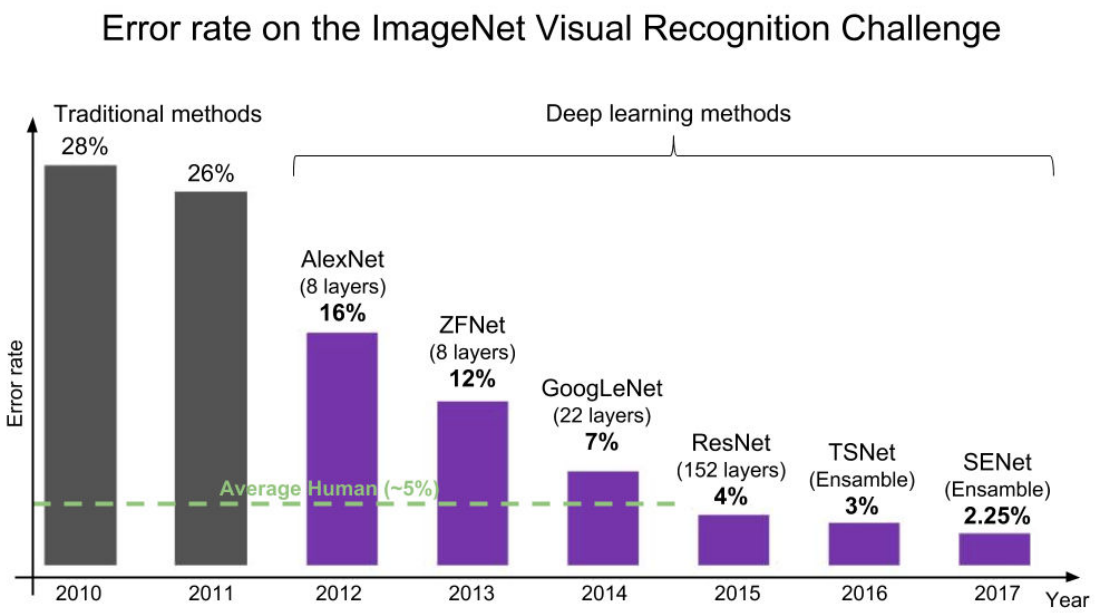
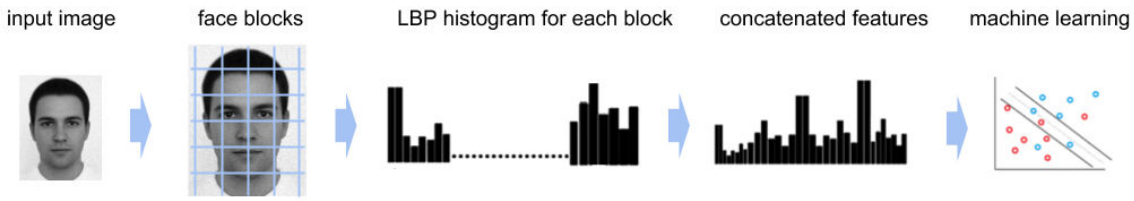


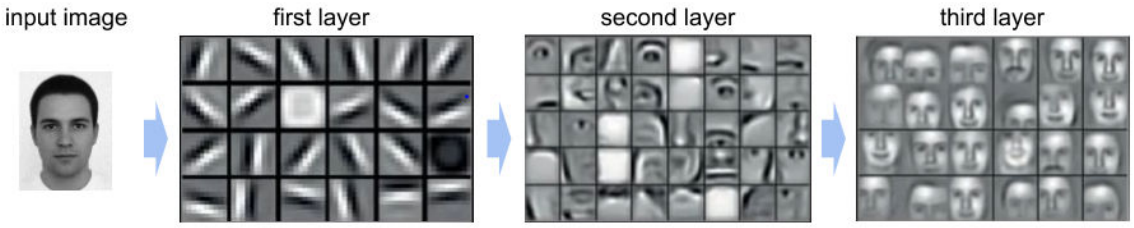
Figure 1.8: Error rate of the ImageNet visual recognition challenge drops dramatically since the introduction of deep learning. Before 2012, the winner of the challenge are traditional methods, which produce error rates of more than 20%. After 2012, all the winning algorithms use deep learning models (AlexNet, ZFNet, GoogLeNet, ResNet, TSNet and SENet), and the error rate drops dramatically to 2.25% in 2017. In comparison, the average error rate for human is 5%.

form of 2D pixel matrices. Deep learning algorithms consist of neural networks with multiple layers and work by learning a hierarchical representation of the input data. Details about neural networks will be described in Section 1.3.1.

As an example, Figure1.9 (b) shows the learned representations of faces in a deep neural network. It is a hierarchical representation of the input image which comprises edges on the first layer, and parts of the face such as e.g. nose, eye, mouth on the second layer, which is further assembled into different types of faces in the third layer. Avoiding the manually designed feature extraction step makes deep learning a pure data driven approach. As a result, the algorithm has the chance to better utilize all the possible detailed information in the data, which may otherwise be filtered out in the feature extraction step. Therefore, in general, deep learning exhibits better scalability than conventional learning algorithms when a large amount of data can be provided for training. Figure 1.10 illustrates how deep learning can improve its performance with the increasing amount of training data, while other types of learning algorithms can saturate quickly.



(a) face recognition in traditional machine learning



(b) face recognition in deep learning

Figure 1.9: Conventional machine learning vs deep learning for face recognition. Hierarchical representations of face images [Lee et al., 2009]. (a) In a conventional face recognition approach which uses *Local Binary Patterns* (LBP) for face description: the input face image is first divided into small blocks, LBP histograms are generated for the blocks, then the concatenated histogram features are used as inputs for conventional machine learning algorithms such as *support vector machines* (SVMs). (b) In a deep neural network trained with faces, the input face image is first represented as combinations of edges, and these edges are assembled into parts on the face such as eye, noise and ear in the second layer. Going deeper to the third layer, a higher level representation of faces are made with the representation from the second layer.

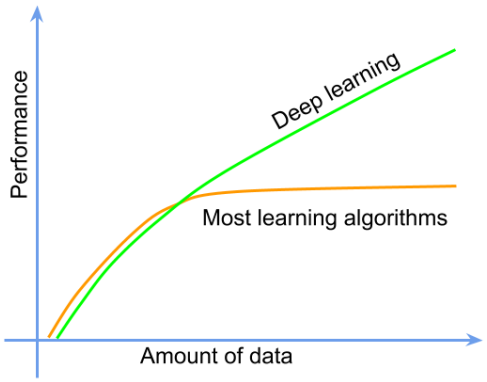


Figure 1.10: Deep learning models scale better with the amount of data used for training than most learning algorithms. By providing more data for training, deep learning methods can constantly increase their performance, while most other learning algorithms are saturated quickly in terms of performance.

The building blocks in deep learning, the first *Artificial Neural Network* (ANN) model was created by a neurophysiologist Warren McCulloch and a mathematician Walter Pitts in 1943. However, the success of neural networks only occurred in the last ten years, which is mainly due to the increasing amount of data and computing power, alongside with breakthroughs

in the training of the deep neural networks. One of the large-scale dataset, the ImageNet dataset [Deng et al., 2009] consists of more than 3.2 million images labeled by human, in which categories such as cats or dogs were manually assigning to each image. This challenge played a crucial role in the development of the deep learning field. In addition to the data, advances in computing power, especially the use of *graphical processing units* (GPUs), enabled the training of large scale neural networks in a parallel fashion. Practically, the advances in open-source softwares such as Theano [Bergstra et al., 2010], Tensorflow [Abadi et al., 2016], PyTorch [Paszke et al., 2017] made it much easier to develop deep learning models with GPU computing, often resulting in 10 to 100 times faster training compared to implementations on *central processing units* (CPUs).

1.3.1 Deep Neural Networks

Under the hood, deep learning models are artificial neural networks. Artificial neurons are the basic units of a neural network and are connected to each other in computational graphs.

In most neural networks, neurons are arranged in layers stacked on top of each other as shown in Figure 1.11 (b). Neural networks consisting of 2-3 or more layers are called *deep*, hence the name *deep learning*.

Multi-layer Perceptrons

Multilayer perceptrons (MLP) are a type of neural network which consist of 3 layers of neurons that was widely used in the early days before the advent of deep learning. The basic unit of a MLP is the perceptron, which is a simplified mathematical model of a single neuron.

The following equation defines the output of the j-th perceptron o_j in the hidden layer of a MLP:

$$o_j = \varphi(\sum_{i=1}^{n} w_{ij} x_i + b_j)$$

Where x_i are the inputs, w_{ij} designates the weights applied on the i-th input, b_j is the bias. φ

is a function which provide non-linearity called *activation function*. For example, the *sigmoid function* (defined as $y = \frac{1}{1+e^{-x}}$), or the *rectified linear unit* (ReLU, defined as $y = max(0, x)$), are commonly used activation functions in deep learning. The weights w_{ij} are initialized with a small random number which follows normal distribution, and the biases b_j are set to 0. Therefore MLPs produces random outputs before training. They can only be used to perform tasks after training. Compared to MLPs, deep learning have at least several differences: 1)

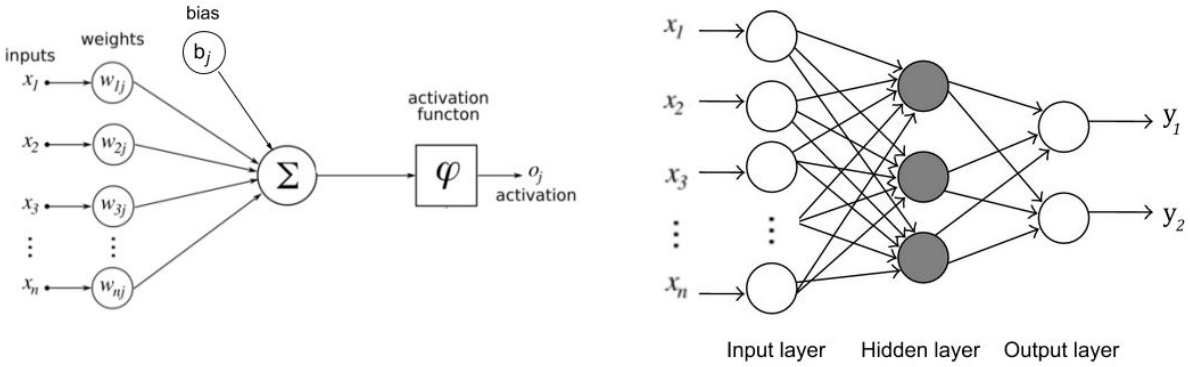


Figure 1.11: A diagram of the perceptron (a single artificial neuron) and a Multilayer Perceptron (MLP) with an input layer, a hidden layer and an output layer.

deep neural networks typically contain many more layers, sometimes hundreds or more; 2) they also contain other units such as pooling units or dropout units [Srivastava et al., 2014]; 3) which use different weights initialization schemes [He et al., 2015, Mishkin and Matas, 2015]; 4) Weight-sharing is common to reduce the number of parameters; 5) some deep learning models contain recurrent units such as long short-term memory unit [Hochreiter and Schmidhuber, 1997] or Gated recurrent unit [Cho et al., 2014] which contains feedback loops in the network, while MLPs or others are called "feed-forward neural networks".

However, most deep neural networks use neurons which are the same as perceptrons. In fact, MLP layers are commonly used in deep neural networks, and they are typically referred as *fully connected layer* or *dense layer* because each node has a connection to all other nodes in the previous layer. Sometimes, a neural network only contains *fully connected layer* is also called *fully connected network*.

Most importantly, the training algorithm used by MLPs, which is called "back-propagation", is used by most of the deep learning algorithms. The Back-propagation algorithm was originally

introduced in the 1970s and was used for training neural networks in [Rumelhart et al., 1986].

Training with Back-propagation

In deep learning, the standard training algorithm uses back-propagation, which is an iterative approach for optimizing parameters according to a criterion called *objective function*, *cost function* or *loss function*.

Take the training of a neural network $f(W, x)$, for example, W designates the parameters of the neural network, consisting of all the weights and biases. The training is an iterative process which updates W in each iteration. With back-propagation and *stochastic gradient descent* (SGD), W is randomly initialized before training, and at each training iteration, the algorithm performs the following steps:

1. Randomly take a sample from the training dataset which consists of input x and target y.

2. Feed input x to the neural network, run through all the layers and produce an output $\hat{y} = f(W, x)$.

3. Use the loss function g to get the loss value $L(\hat{y}, y)$, which expresses a measure of the error between $\hat{y}$ and y. For example, the commonly used object function for regression tasks is the *mean-squared-error* (MSE, also called L_2 loss) where, $L(\hat{y}, y) = ||\hat{y} - y||^2$.

4. Calculate the gradient ∇L which consists of partial derivatives $\frac{\partial L}{\partial W}$ of the loss function L with respect to any parameter of the neural network W (weights and bias).

5. Update all the parameters $W \rightarrow W - l_r * \frac{\partial L}{\partial W}$ for a small step with the gradient ∇L, where the step size is controlled by l_r (also known as *learning rate*). Intuitively, this will make the network move towards the direction that reduces the loss value. In practice, the learning rate is an important hyper-parameter of the algorithm. It is also common to use optimizers such as Adam [Kingma and Ba, 2014], RMSprop [Hinton et al.,],

Adagrad [Duchi JDUCHI and Singer, 2011] etc., to update the parameters in a more adaptive and efficient manner.

6. Repeat the above steps.

In SGD, the gradient is computed from only one sample, and it often results in fluctuations and noise. On one hand, the fluctuation is important for escaping local minima [Ge et al., 2015], on the other hand, it makes the training harder to converge. Therefore, back-propagation is often combined with mini-batch gradient descent, which first computes the gradient with a subset of samples (called *mini-batch*, e.g. with 32 samples) and then compute an averaged gradient in step 4. Computing the averaged gradient can lead to more stable convergence. In practice, the mini-batch size is another hyper-parameter which needs to be selected.

Regularization and data augmentation

When working with limited amounts of data, a common problem of training a neural network is overfitting – meaning that the model gives good predictions on the training dataset but fails to make correct predictions for unseen data (Figure 1.12).

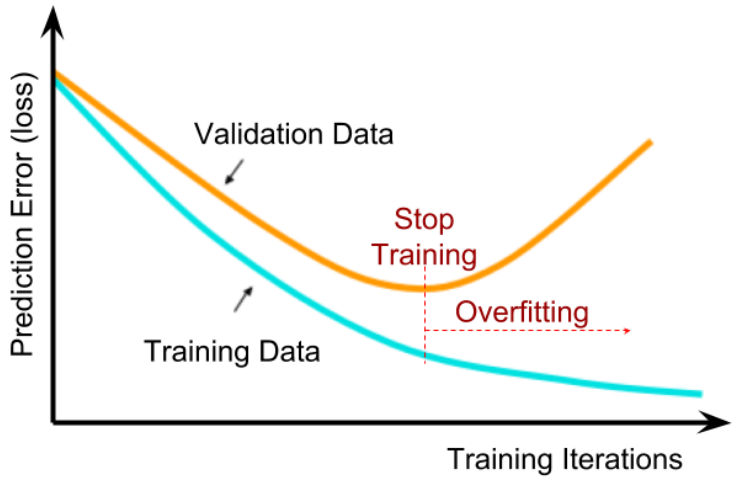


Figure 1.12: Overfitting and early-stopping. During training of neural networks, prediction error on both training data and validation data decrease continuously. After some point, the network starts to overfit the training data, meaning the prediction error on training data may still decline but increases on validation data. A commonly used strategy to avoid this phenomenon is called *early-stopping*, which stops the training when overfitting is detected.

Therefore, it is important to prevent overfitting. A commonly used strategy is called *early-stopping* [Caruana et al., 2000] – the algorithm keeps monitoring overfitting and stops the training when it starts to overfit. As shown in Figure 1.12, by monitoring the prediction error (loss) with samples from the validation dataset, which has never been used to update the network parameters, one can stop the training when the prediction error on validation dataset starts to increase. Besides using early-stopping and adding more data for training, there are various methods which efficiently regularize the training and improve the generalization. For example, dropout [Srivastava et al., 2014], which injects noise into the training by randomly dropping the connection at each iteration. This is a widely used method to prevent overfitting. Other methods include Batch Normalization [Ioffe and Szegedy, 2015] which reduce the so-called *internal covariate shift* and speed up the convergence; L1 regularization and L2 Regularization which add penalty terms ($||W||_1$ for L1 and $||W||_2$ for L2) to the loss function in order to encourage sparsity or reduce model complexity.

Since acquiring data can be very expensive (for example taking one super-resolution image can take tens of minutes to hours in SMLM), it is important to increase the efficiency of using existing data. Data augmentation is a cheap way to increase the amount of data solely by transforming existing data computationally. It can also reduce overfitting by preventing the neural network from memorizing the training data. For image data, random rotations, elastic transformations and injection of random noise are commonly used as effective ways increase the size of a small dataset. While potentially there are many choices of transformation that could be used for data augmentation, it is important to keep the same statistics of important features before and after the transformation. Figure 1.13 shows an example of data augmentation which has been used in [Chen and Tao, 2017] for visual tracking tasks. For microscopy images, random rotations (with an angle ranging from 0 to 360 degrees) could be used to generate new training samples. However, applying a random rotation angle for other types of images will not be appropriate, because it may generate images with objects such as houses or trees upside down.

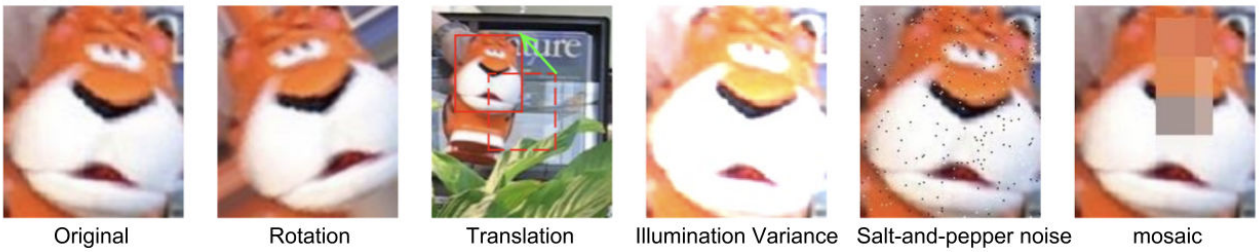


Figure 1.13: Data augmentation examples [Chen and Tao, 2017]. For the same original image patch, different data augmentation strategies (rotation, translation, illumination variance, salt-and-pepper noise, mosaic) are used for increasing the amount of data and so preventing the neural network from memorizing the original data.

Convolutional Neural Networks

Besides MLP or fully connected networks, the most commonly used neural networks in computer vision are *Convolutional Neural Networks* (CNNs). CNNs are heavily inspired by how the mammalian vision system works, for example, the primary visual cortex. In order to achieve shift-invariance, CNN uses a mathematical operation called "convolution" which is similar to cross-correlation. For 2D images, the convolution operation works as shown in Figure 1.14 (a). The input image is convolved with a kernel (the weights) and results in another image which is called the *feature map*. By adjusting the convolution kernel (weights) during training, CNNs can extract different features. For example, Figure 1.14 (b) shows a convolution kernel which can detect edges in the image. By using convolutions as a means of achieving trainable feature extraction, CNNs with multiple layers, can learn hierarchical representations of the input images as illustrated in Figure 1.9 (b).

There are two important properties which are imposed by the design of CNN: sparse connection and parameter sharing. Unlike fully connected layers, where each node has a connection to all other nodes in the previous layer, CNNs use sparse connections, meaning that every node is only connected to a small number of nodes (e.g, a 3x3 squared region) in the previous layer. In addition, the weights and biases to the previous layers are shared between adjacent nodes, which impose the transnational invariance constraint. There properties make CNNs computationally more effective and less sensitive to the absolute position of the object in an image. CNNs are a successful example of embedding domain-specific knowledge (in this case of vision tasks) to

the neural network architecture design.

A typical CNN contains layers that perform the following three steps: 1) convolve the image with the convolution kernel, which is a linear operation; 2) apply a non-linear activation function to the output feature map; and, 3) apply sub-sampling using a type of layer called pooling. The most commonly used sub-sampling is max-pooling, which simply computes a single (1x1) value as the maximum of 2x2 pixel regions in the input feature map. This pattern of calculations can be commonly seen in CNNs for computer vision tasks. For example, one of the earlier work on CNNs done by [Lecun et al., 1998] applies these calculations in the first few layers, and then feeds the result into a fully connected layer. The full network is named *LeNet* and is shown in Figure 1.14 (c).

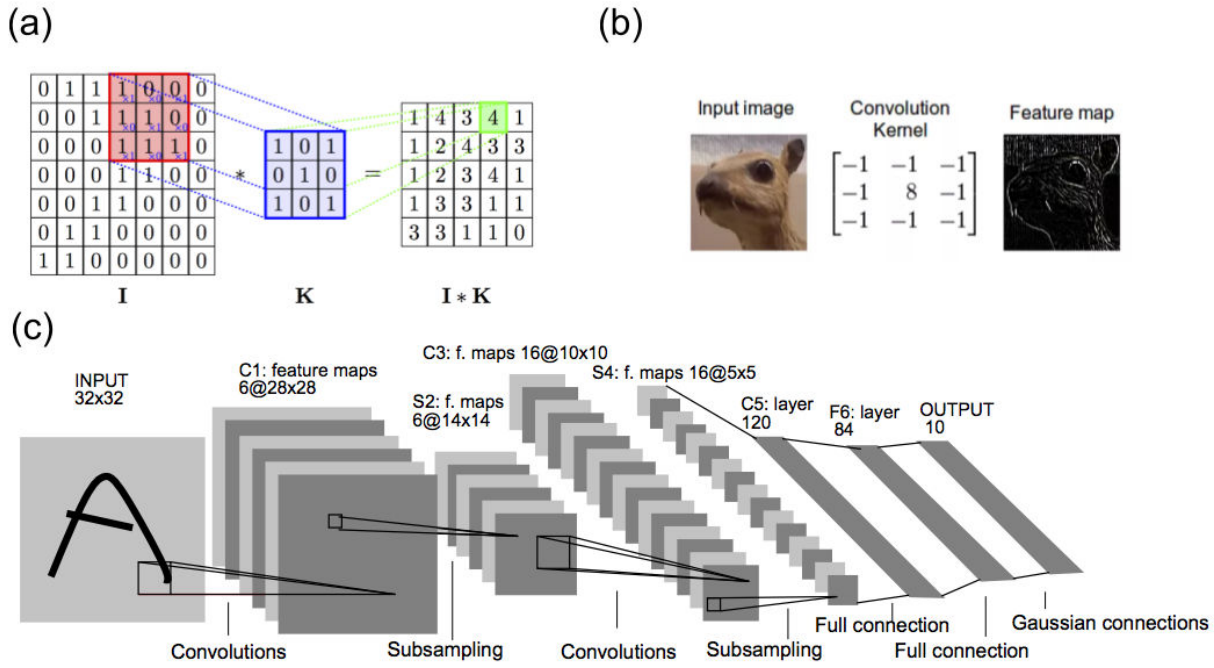


Figure 1.14: Feature extraction with convolution and an example CNN for image recognition. (a) illustrate a basic step of convolution when an image I is convolved with a kernel K, the result image $I * K$ is called *feature map*. (b) shows a feature map produced by a convolution with a 3x3 convolution kernel. Such a convolution kernel can be used to detect edges in the image. (c) A convolutional neural network network architecture designed by [Lecun et al., 1998] called *LeNet-5*. It consists of 2 convolution layers, 2 sub-sampling layers, 2 full connection layers and 1 Gaussian connection.

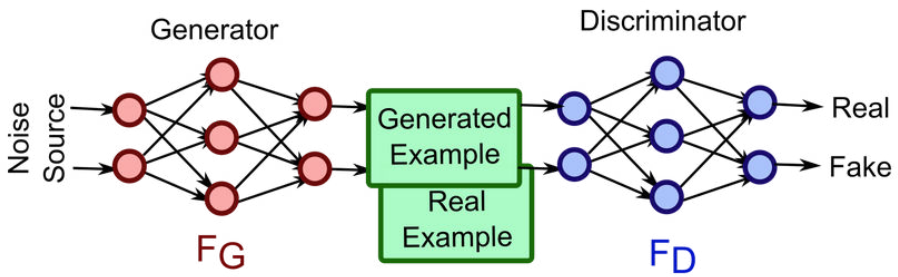
.

Generative Adversarial Networks

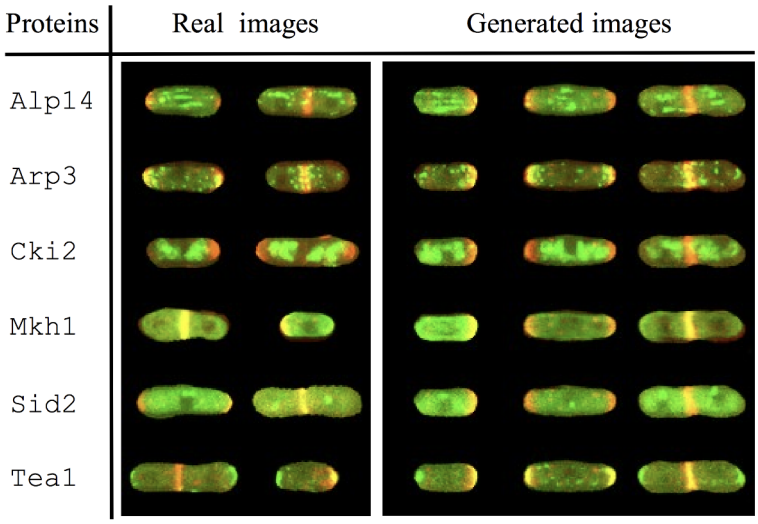
Despite the success of CNNs and fully connected neural networks, they are mostly used for extracting information of interest from the data, such as classification. Generative models such as *variational autoencoders* (VAEs) [Kingma and Welling, 2013] , contrarily, are designed to model the distribution of data and are capable of generating new samples as well.

A *Generative Adversarial Network* (GANs) is a type of generative model that is now commonly used in deep learning. GANs were first proposed by [Goodfellow et al., 2014], and work by introducing adversarial training between two neural networks. As shown in Figure 1.14 (a), a GAN consists of a generator network F_G and a discriminator network F_D which compete against each other. While the generators transform random numbers into meaningful data (e.g. a face image), the discriminator is basically a classifier which tries to determine if the input data is real or fake (synthetic). During training, the generator is trained to fool the discriminator such that the generated image will be classified as real, while the discriminator attempts to recognize all samples from the generator as being synthetic, and all samples from the training set as being real. The goal of the training is to reach an equilibrium state where the generator can generate indistinguishable results which cannot be identified by the discriminator even with further training. GANs are widely used for generating new data, for example, designing cloths [Zhu et al., 2017] for the fashion industry, or drug molecules [De Cao and Kipf, 2018] for the pharma industry, images conditioned on text [Zhang et al., 2016b] etc. Figure 1.14 (b) shows one application of GAN for generating microscopy images of fission yeast cells [Osokin et al., 2017].

In addition, GANs are also used as a way to regularize the training of other deep learning models using a mechanism called *adversarial training*. In that case, discriminators are used as part of the loss in combination with other loss terms such as MSE, and the corresponding loss is called *adversarial loss*. In comparison with MSE (L2), the effect of using an adversarial term in a network is to push the network to choose one possible solution, while L2 tends to produce "the average of all possible solutions". Figure 1.15 (c) shows an example of how an adversarial loss term can improve the sharpness of the face image. However, the details which are generated



(a) Composition of a GAN



(b) GANs generate fission yeast cells

élément sous droit, diffusion non autorisée

(c) MSE loss vs adversarial loss

Figure 1.15: Generative Adversarial Networks are trained to generate images from noise. (a)The network is trained with a Discriminator which discriminates the output of the generator during training. (b) After performing adversarial training on real fission yeast cell images labeled with different proteins, the generator alone can produce realistic-looking images of fission yeast cells [Osokin et al., 2017]. (c) An example shown in [Lotter et al., 2015] which illustrates the effect of using the adversarial loss term compared to MSE. The task is predicting the next frame in a video sequence. A model trained with MSE loss produces a blurred image, while another model trained with a weighted MSE and adversarial loss outputs a sharp image with more details. Notice that the small details produced with the adversarial loss do not necessary reflect the truth when compared with the Ground Truth, for example the eye and ears are slightly different.

.

using the adversarial loss are not necessary correct. The risk of using adversarial losses for image prediction are potential artifacts, especially in ambiguous regions. To benefit from the regularization power of adversarial training and reducing the risk of generating artifacts, it is important to enforce data fidelity loss terms such as L1 and L2 by choosing a small weight for the adversarial loss term, and also using a conditional version of GAN (also called *cGAN*, to be described in the following section) which further conditions the generation with additional input data.

1.3.2 Deep learning for image reconstruction

For image data, image reconstruction methods are commonly used for solving image restoration problems such as image denoising and deconvolution. A typical process is to a) transform the input image into a different domain (e.g. Fourier domain, Wavelet etc.), b) process the transformed data (e.g. remove high frequency components for denoising), and c) perform an inverse transform to obtain the reconstructed image. With deep learning, image reconstruction can be formulated as a dense prediction problem where both the input and output data are images. Deep Learning based reconstruction algorithms work by training on example image pairs. The model will learn the transformation between the input image(s) and output image(s) according to a criterion defined by the loss function.

Image super-resolution

Image super-resolution is a common image reconstruction technique in computer science for recovering details form corrupted observations. Specifically, the goal of image resolution is to produce a image with more details in higher resolution from a single or multiple images in low resolution.

Different from the super-resolution microscopy described previous section which aims to improve the image resolution with specific acquisition device or method, image super-resolution are purely relying on computational algorithms. Image super-resolution methods can work with

a single image as input or multiple of them. In order to recover details in higher resolution, the algorithm typically learn prior information from a set of training samples. For example, sparsity can be used as prior information for performing image super-resolution [Jianchao Yang et al., 2010], however, recent deep learning based image super-resolution shows better results than conventionally ones [Lai et al., 2017, Dong et al., 2016a, Kim et al., 2016, Kim et al., 2015].

Figure 1.16 shows different types of CNNs which have been used for image super-resolution. Most of them performing image transformation after a bicubic interpolation.

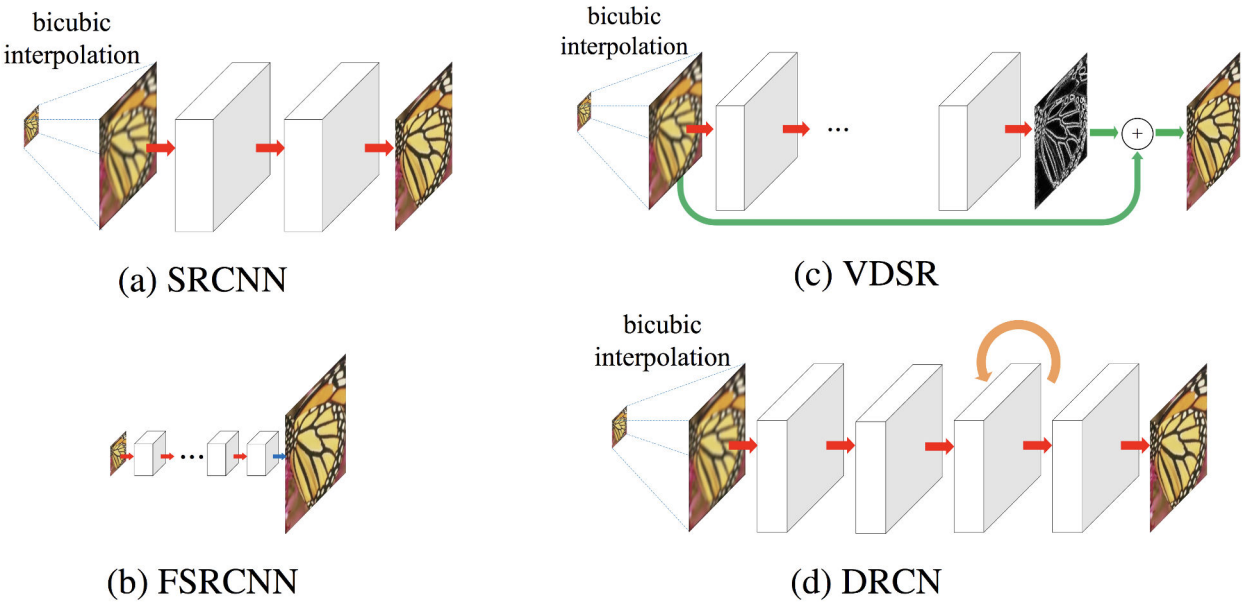


Figure 1.16: Deep learning based singe image super-resolution methods [Lai et al., 2017] using different types of deep neural networks: a) SRCNN from [Dong et al., 2016a] b) FSRCNN from [Dong et al., 2016b] c) VDSR from [Kim et al., 2016] d) DRCN from [Kim et al., 2015]. While a), b) and c) apply bicubic interpolation to enlarge the image before feeding into the neural network, b) takes the original image directly and directly outputs an image with bigger size assembled from different channels. .

Deep learning based image reconstruction can not only be used for solving image restoration such as denoising but also for more generic *image-to-image translation*, for solving tasks such as image segmentation [Ronneberger et al., 2015], for depth map prediction [Eigen et al., 2014] or for in-silico-labeling [Christiansen et al., 2018], to name a few examples.

U-net

A widely used type of CNN for image reconstruction is called *Fully Convolutional Network* (FCN) which mainly contains convolutional layers (including pooling layers) but no fully connected layers. Image segmentation for example, an earlier type of FCN, is proposed in [Long et al., 2014]. Later, another type of FCN called *U-net* was proposed by [Ronneberger et al., 2015]. It is more computationally efficient than alternatives such as [Ciressan et al., 2012], and it can work with small dataset. Figure 1.17 shows the architecture of the U-net, it consists of two paths: a contracting path which encodes the information into different scales and an expanding path which decodes the information and outputs the reconstructed image. A key ingredient of U-nets are the skip-connections between the contracting path and the expanding path, which facilitate the decoding process by providing more precise localizations information.

Pix2Pix Network

U-nets can be used with different types of loss functions during training, and the choice has a great impact on the image. Figure 1.18 shows a comparison which reflects the effectiveness of the skip-connections in U-net and also the effect of using *conditional GAN* (cGAN) [Isola et al., 2016]. The difference between a cGAN and a regular unconditional GAN is that both the input and output image of the generator (or the target image) are fed into the discriminator in a cGAN, while the discriminator in a regular GAN do not see the input of the generator. Because of the additional input to the discriminator, the generation of GAN is conditioned on the input image. This is a way to regularize the generation and reduce potential artifacts. [Isola et al., 2016] also show that the conditional version is better than an unconditional GAN in image transformation tasks. These comparison experiments led to the design of the *Pix2Pix* network, which uses a U-net as the generator and is trained with a loss function consisting of a weighted sum of the L1 loss and the conditional adversarial loss (labeled as cGAN in the figure).

Pix2Pix is a generic framework for *image-to-image translation*. It is suitable for image trans-

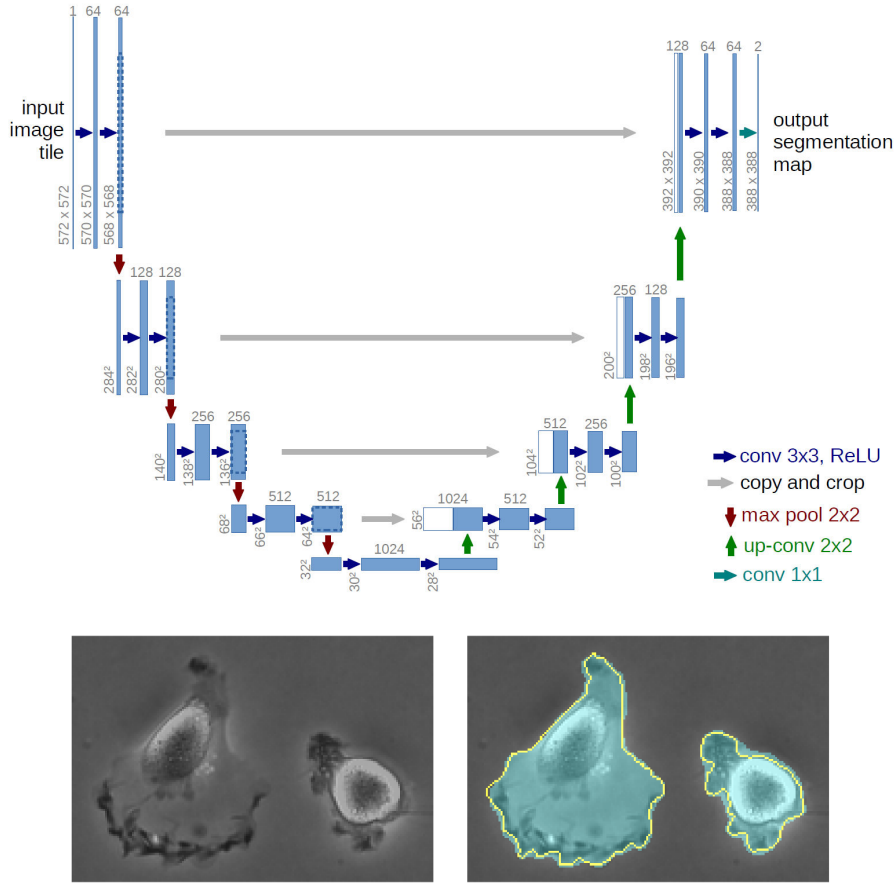


Figure 1.17: Architecture of the U-net and its application in image segmentation [Ronneberger et al., 2015]. U-net consists of two arms, the encoder (left) on the left and the decoder (right) on the right, which are bridged with 4 skip connections in the middle. The building block of the encoder consists of two 3x3 convolutional layer and is followed by a max-pooling layer. The decoder path is similar, except the max-pooling layer is replaced by a up-convolutonal layer. The output is produced by a 1x1 convolutional layer. The two images shown below are an example of using U-net to perform image segmentation on images with HeLa cells. The image on the right is the input image, and the image on the right is an image with segmented regions.

formations, mapping the image from one modality to another modality (Figure 1.18 (b)). For microscopy images, it can be used to solve inverse problems including denoising, deconvolution etc. when treating the corrupted image as input and the expected image with higher quality, higher resolution as output. Figure 1.19 shows an example from [Su et al., 2018], which demonstrates the use of Pix2Pix network for image denoising with synthetic *Cryogenic Electron Microscopy* (CryoEM) data.

In this work, we explored how deep neural networks can be used for reconstructing microscopy images from noisy and incomplete observations. Specifically, we adapted U-net and Pix2Pix to our method for recovering missing information from sparsely acquired SMLM and wide-field



(a) Comparisons on skip connections and cGAN loss



(b) Image-to-image translation examples

Figure 1.18: Comparisons on adversarial loss, U-net, and examples produced with Pix2Pix network [Isola et al., 2016] . Comparisons with or without skip connections, and L1 with or without cGAN loss (b) Image-to-image translation examples produced with the Pix2Pix network. Once trained, Pix2Pix networks can generate realistic facade from labels represented with colors in a segmentation map, or generate photo-realistic bags from images with only edges.

images, resulting in faster acquisition or better image quality.

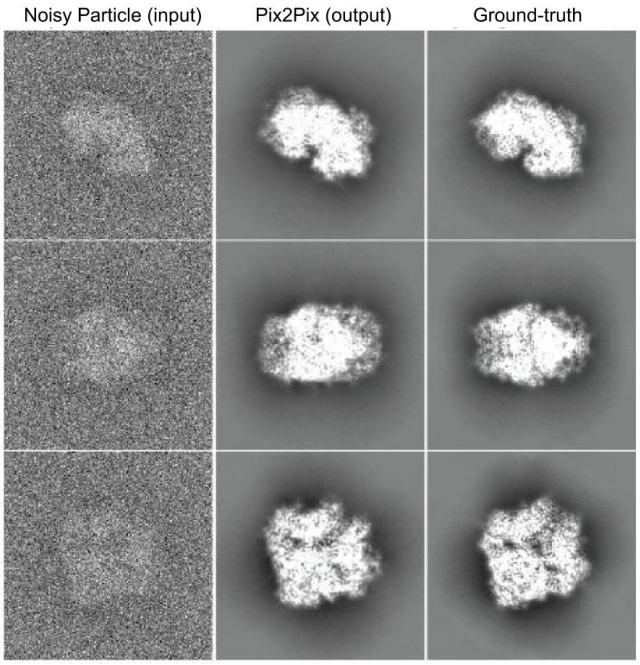


Figure 1.19: CryoEM image denoising with Pix2Pix network [Su et al., 2018]. It shows results of a Pix2Pix network on three different forward projections of β-galactosidase. After training with 2000 pairs of synthetic particle forward projection images with and without noise, the trained Pix2Pix network can produce output images from noisy inputs, which agree with the corresponding ground-truth images.

Chapter 2

Methods and Results

2.1 Deep learning massively accelerates super-resolution localization microscopy

By utilizing recent advances in deep learning, we developed a computational method called *Artificial Neural Network Accelerated PALM* (ANNA-PALM), which massively speeds up the imaging speed of SMLM by 10-100 folds and improves the image quality. With only a small fraction of diffraction limited images used in conventional SMLM, ANNA-PALM is able to reconstruct high quality super-resolution images without compromising the spatial resolution. The dramatic improvement in speed and quality afforded by ANNA-PALM can find use in biological projects which involve high-throughput imaging or live-cell imaging.

Here is an outline of the results:

1. We designed a deep neural network architecture called *A-net*, which is able to learn structural information about a biological samples and reconstruct high quality images from rapidly acquired SMLM images.

2. A training strategy with extensive data augmentation is proposed, which typically requires training images from less than 10 fields of view for a certain type of biological structure.

3. Validation of the method was carried out with various types of biological structures including microtubules, nuclear pore complexes, and mitochondria, and variations of the imaging method were tested, including STORM, DNA-PAINT.

4. Experimental perturbations of microtubules with drugs demonstrate the robustness of the method regarding the differences between training and testing data.

5. To demonstrate the advantages of the method in terms of imaging speed, a high-throughput imaging experiment is performed. With ANNA-PALM, we were able to scan a 2x2 mm region in roughly 3 hours and reconstruct an image measuring from 20 nm to 2 mm.

6. To reduce and evaluate the potential artifacts generated by the method, we implemented a tool that produces an error map by measuring the consistency between the output image and the widefield image.

7. A set of tools are developed to facilitate the adaptation of the method, including a web application, an ImageJ plugin and a python library.

ARTICLES

nature biotechnology

Deep learning massively accelerates super-resolution localization microscopy

Wei Ouyang[1–3], Andrey Aristov[1–3], Mickaël Lelek[1–3], Xian Hao[1–3] & Christophe Zimmer[1–3]

The speed of super-resolution microscopy methods based on single-molecule localization, for example, PALM and STORM, is limited by the need to record many thousands of frames with a small number of observed molecules in each. Here, we present ANNA-PALM, a computational strategy that uses artificial neural networks to reconstruct super-resolution views from sparse, rapidly acquired localization images and/or widefield images. Simulations and experimental imaging of microtubules, nuclear pores, and mitochondria show that high-quality, super-resolution images can be reconstructed from up to two orders of magnitude fewer frames than usually needed, without compromising spatial resolution. Super-resolution reconstructions are even possible from widefield images alone, though adding localization data improves image quality. We demonstrate super-resolution imaging of >1,000 fields of view containing >1,000 cells in ~3 h, yielding an image spanning spatial scales from ~20 nm to ~2 mm. The drastic reduction in acquisition time and sample irradiation afforded by ANNA-PALM enables faster and gentler high-throughput and live-cell super-resolution imaging.

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

Fluorescence microscopy methods that overcome the diffraction limit of resolution (~200–300 nm) allow imaging of biological structures with molecular specificity closer to the molecular scale. Among super-resolution microscopy approaches, those based on single-molecule localization, such as PALM[1] and STORM[2] (hereafter referred to collectively as PALM), are particularly attractive owing to their exquisite spatial resolution and ease of implementation. In these methods, random subsets of fluorophores are imaged in many consecutive diffraction-limited frames, computationally localized to high precision, and the combined localizations are used to generate a super-resolution view. In practice, typically 10^3–10^5 diffraction-limited frames are needed to assemble a single super-resolution image. This requirement follows from two conditions that must be simultaneously satisfied to ensure high spatial resolution: (i) a small number (~10–10^2) of active fluorophores per frame, to avoid overlaps between diffraction-limited spots and enable precise localization of individual molecules, and (ii) a large number of independent localizations to ensure a sufficiently dense sampling of the underlying biological structures[3,4]. The large number of required frames makes localization microscopy inherently slow, thereby limiting its potential for high-throughput imaging, where many fields of view (FoVs) are to be imaged, and for imaging live-cell dynamics. As a result, most localization microscopy studies are restricted to analyzing a small number of cells (typically less than ten).

Multiple approaches have been explored to accelerate localization microscopy. Using bright dyes with rapid switching kinetics, high-power lasers, and fast cameras allows minimization of exposure time without decreasing the signal-to-noise ratio[5,6], but reaching submillisecond exposure remains challenging, and intense irradiation exacerbates phototoxicity in live-cell imaging[7,8]. Increasing the number of active fluorophores per frame can reduce acquisition time, but despite algorithms designed to handle overlapping fluorescent spots[9–13], this approach necessarily degrades spatial resolution[4,14].

Here, we introduce a computational strategy that allows the total number of frames and independent localizations to be reduced without trading off spatial resolution. Unlike previous approaches, our method exploits the structural redundancy of most biological images to reconstruct high-quality images from vastly undersampled localization microscopy data. Our method leverages deep learning, which employs artificial neural networks (ANNs) to learn complex non-linear mappings between numerical inputs and outputs[15]. Accordingly, we call it 'artificial neural network accelerated PALM', or ANNA-PALM.

RESULTS

A deep-learning approach to super-resolution image reconstruction

We aim to reconstruct a super-resolution image of approximately similar information content as a standard PALM acquisition (with K frames and N localizations) from a much smaller number of raw frames ($k << K$) without changing the average density of localizations, ρ, that is, from a much smaller number of total localizations ($n = \rho k << N = \rho K$). If PALM images are defined as two-dimensional (2D) histograms of independent localizations, this task can be formulated as restoring an image corrupted by Poisson noise (and potentially additional forms of noise). Image restoration is an ill-posed problem that admits an infinity of solutions in the high-dimensional space of all possible images, unless additional constraints (priors) are imposed that restrict the solution to a lower dimensional manifold. Suitable manifolds exist because most natural images are highly redundant, and can be represented to very good approximation with a much smaller number of coefficients than pixels, via appropriate functions that map feature space to pixel

[1]Institut Pasteur, Unité Imagerie et Modélisation, Paris, France. [2]UMR 3691, CNRS, Paris, France. [3]C3BI, USR 3756, IP CNRS, Paris, France. Correspondence should be addressed to C.Z. (czimmer@pasteur.fr).

Received 6 July 2017; accepted 16 February 2018; published online 16 April 2018; doi:10.1038/nbt.4106

space[16,17]. In recent years, ANNs with multiple layers (deep nets) have proven very successful at learning meaningful features and nonlinear mappings for image classification, segmentation, restoration, and many other tasks[15,17,18]. Inspired by these developments, we designed ANNA-PALM, a deep-learning approach for restoring super-resolution views from undersampled (sparse) localization microscopy data.

ANNA-PALM comprises a training stage and an inference stage (**Fig. 1**). For training (**Fig. 1a**), a few super-resolution images

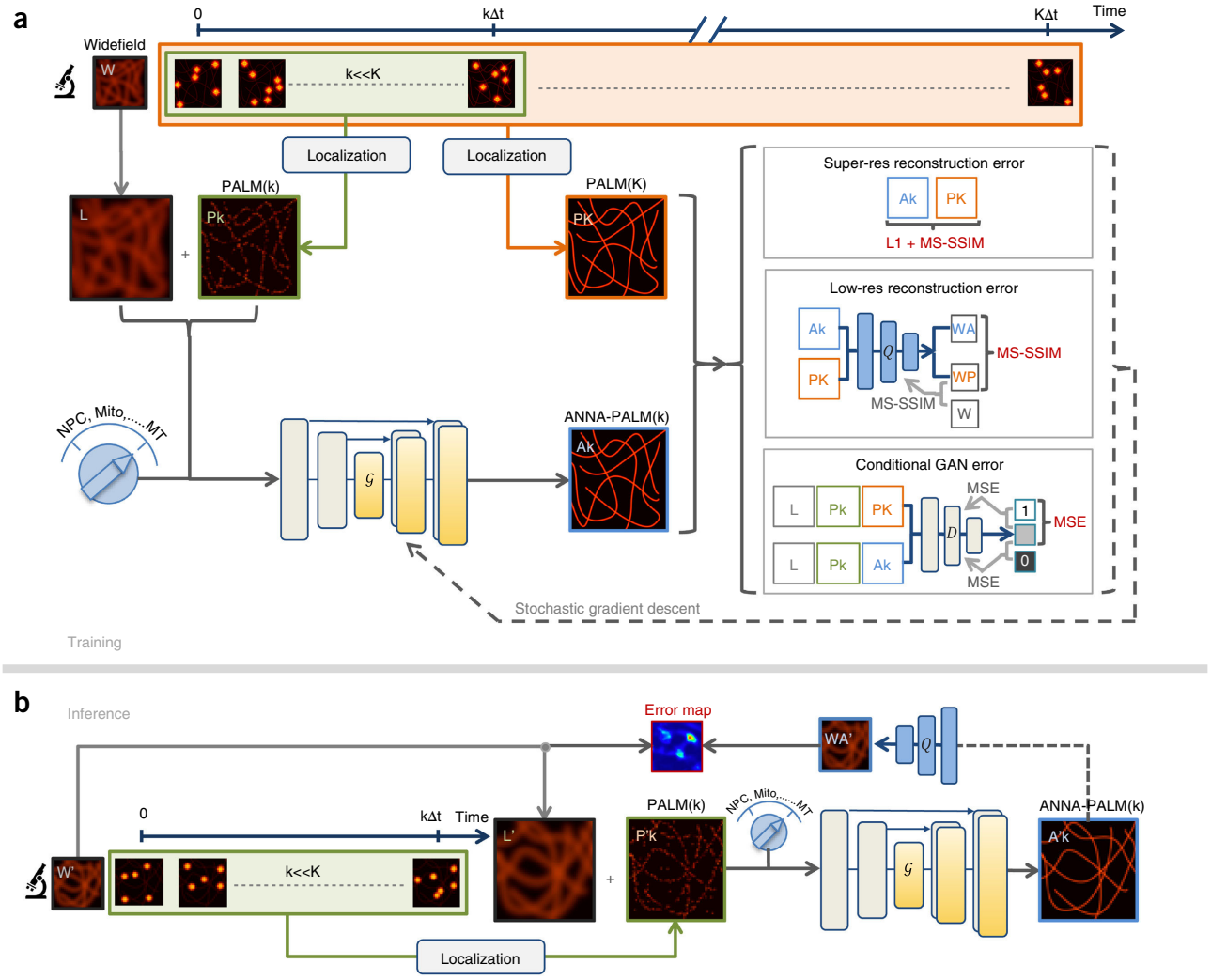


Figure 1 Overview of ANNA-PALM. (**a**,**b**) ANNA-PALM consists of two main stages: (i) acquisition of training images using standard localization microscopy (PALM) followed by artificial neural network (ANN) training (**a**), and (ii) reconstruction of super-resolution views and low-resolution error maps from new sparse PALM and/or widefield images (inference) (**b**). (**a**) Training images are obtained by acquiring one or a few long sequences, of $K \sim 10^3$–10^5 diffraction limited, single-molecule image frames, as in standard PALM experiments; optionally, a widefield image W can also be acquired (top). The acquisition time for each image sequence is $K\Delta t$, where Δt is the single-frame exposure time. Standard localization microscopy algorithms ('Localization' boxes) are used to generate super-resolution images. For each experiment, a highly sampled (dense) super-resolution image (PK) is generated using all (or in practice ≥95%) available K frames. Sparse PALM images (Pk) from the same experiment are obtained by using only $k << K$ frames. A switch (light blue) can be set to distinguish different types of structures, for example, nuclear pore complexes (NPC), mitochondria (Mito) or microtubules (MT). An ANN (labeled $\mathcal{G}$ for 'generator') is trained by using the sparse PALM images Pk (plus an upsampled version L of the widefield image W and the switch setting) as inputs and the corresponding dense PALM image PK as target output. During training, the output of the generator $\mathcal{G}$ (Ak) is compared to the target image PK and the widefield image L (if available) via three loss, or error functions (gray bordered boxes): (i) the 'super-resolution reconstruction error' measures the difference between the reconstructed image Ak and the target PK using a combination of the L1 norm and the MS-SSIM; (ii) the 'low-resolution reconstruction error' measures the MS-SSIM between the low-resolution image WA predicted from Ak and the low-resolution image WP predicted from PK. Images WA and WP are predicted using a second ANN, called low-resolution estimator (labeled Q, blue) that is trained to produce an approximation of the actual widefield image W based on the MS-SSIM metric; (iii) the 'conditional GAN error' uses a third ANN (labeled D for 'discriminator', red) that attempts to distinguish between real dense PALM images PK and the generator's output Ak. The combined loss functions are iteratively optimized using stochastic gradient descent. (**b**) A short sequence of diffraction-limited single-molecule images (with $k << K$ frames, i.e., acquisition time $k\Delta t$), and an optional widefield image (W') are acquired. Standard localization algorithms generate a sparse (undersampled) PALM image (P′k). This sparse image P′k, the upscaled widefield image L′ and switch setting are fed as inputs to the trained generator $\mathcal{G}$, which outputs a reconstructed ANNA-PALM image (A′k). In addition, the low-resolution estimator Q predicts a low-resolution image WA′, which can be compared to the input widefield image W′ via the MS-SSIM to produce a low-resolution error map (top).

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

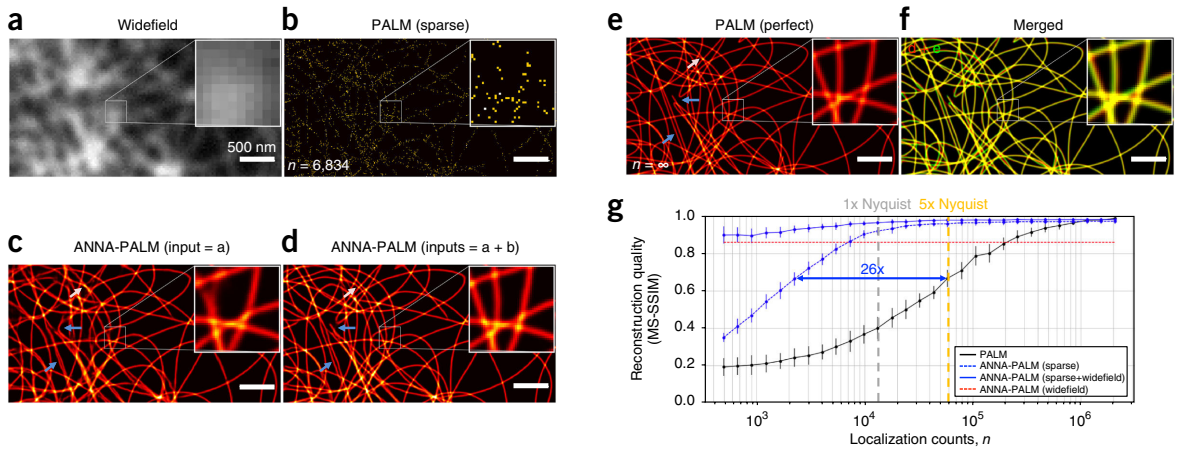


Figure 2 Validation of ANNA-PALM on simulated images. (**a**) Simulated widefield image of microtubules. (**b**) Simulated sparse PALM image of microtubules with $n = 6{,}834$ localizations. (**c**) ANNA-PALM reconstruction using only the widefield image **a** as input. (**d**) ANNA-PALM reconstruction using both the widefield image **a** and the sparse PALM image **b** as inputs. (**e**) Simulated "perfect" PALM image, equivalent to a PALM image with an infinite number of localizations ($n = \infty$) and a resolution of 23 nm. This image was used to generate **a** (by convolution with a Gaussian kernel approximating the microscope point spread function) and **b** (by application of Poisson noise). (**f**) Merged image showing the perfect PALM image **e** in green and the ANNA-PALM reconstruction **d** in red. Note that the ANNA-PALM images **c**,**d** provide many high-resolution details that are absent from the widefield image **a** and the sparse PALM image **b** and that are in good (**c**) or very good (**d**) agreement with the perfect PALM image **e**; some reconstruction errors are highlighted by arrows. Blue arrows in **c** point to errors of ANNA-PALM reconstruction from the widefield image only, the white arrow in **d** points to an error of ANNA-PALM reconstruction from both widefield and sparse PALM images combined. Reconstruction errors diminish for larger numbers of localizations, n (**Supplementary Fig. 4**). (**g**) Reconstruction quality of PALM and ANNA-PALM images, measured by the MS-SSIM with the perfect PALM image **e**, as function of localization number n. Black curve: reconstruction quality of the standard PALM images. Dashed blue curve: reconstruction quality of ANNA-PALM using the sparse PALM images as input. Solid blue curve: reconstruction quality of ANNA-PALM using both the sparse PALM and widefield images as inputs. Red dashed line: reconstruction quality of ANNA-PALM using the widefield images as input only. Dots are averages from ten simulations; error bars show s.d. The vertical dashed orange line indicates the minimum number of localizations needed to achieve a resolution of $R_{5\times Nyq} = 23$ nm according to the fivefold Nyquist criterion[3]. The dashed gray line indicates the minimum number of localizations needed to achieve a double mean nearest neighbor distance less than 23 nm. ANNA-PALM reconstructions from sparse PALM images only (i.e., without widefield images) achieve the same average MS-SSIM as standard PALM at the fivefold Nyquist sampling limit with 26 times fewer localizations (blue double arrow). ANNA-PALM reconstruction quality is highest when using both widefield and sparse PALM images as inputs.

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

representative of the structure of interest (e.g., microtubules, nuclear pores, or mitochondria) are obtained using standard PALM imaging, that is, by acquiring long diffraction-limited image sequences (e.g., $K \sim 10^4 - 10^5$, $N \sim 10^5 - 10^7$) and processing them with standard localization software[19], resulting in highly sampled (dense) PALM images. In addition, a low-resolution (widefield) image can also be acquired, as is commonly done before single-molecule imaging when bleaching out preactivated fluorophores. Next, the dense PALM images are undersampled by using a much smaller number of diffraction-limited frames, $k<<K$, thus yielding sparse PALM images from the same localization data. Then, an ANN is trained to recover approximations of the dense PALM images from these sparse PALM images (and the optional widefield image). Once trained, the ANN can be applied to new sparse PALM images (with or without a widefield image), obtained from new image sequences with small numbers of frames ($k<<K$)—and hence in much shorter time—in order to reconstruct high-quality, super-resolution images not previously seen (inference, **Fig. 1b**).

Neural net architecture and learning strategy

Our ANN, hereafter called A-net, contains a total of 25 convolutional layers, and roughly 42 million trainable parameters. A-net is adapted from the pix2pix network[20], which itself builds on U-nets[21] and generative adversarial networks (GANs)[22], two recent successful deep-learning techniques. U-nets are special types of convolutional neural networks (CNNs) that have proven effective at learning multiscale representations of images and accurate, pixel-wise mappings[21,23]. GANs can generate new samples from real image distributions using a generator network that outputs synthetic images, and a discriminator network that outputs the probability that an input image is real or synthetic, both networks being trained simultaneously to compete against each other[22]. Importantly, the generator can be conditioned on input data (conditional GAN, or cGAN)[20,22], for example on images, as in the pix2pix network. We modified the pix2pix architecture to accept a computational switch as additional input to handle multiple types of data, and introduced an additional network to evaluate the consistency between the reconstructed image and the widefield input image.

Training of our A-net proceeds as follows. Randomly undersampled (i.e., sparse) versions of PALM images are fed as input to the A-net, while the corresponding dense PALM images are defined as the A-net's targets, that is, desired outputs (**Fig. 1a**). Additional, optional inputs are widefield images, if available, and the switch, which indicates the image type when multiple types of images (e.g., microtubules and nuclear pores) are used during training. ANN training requires defining an objective function (also called loss), which measures how well the outputs match the targets. We implemented a loss function containing three terms. The first term measures the difference between the A-net output and the dense PALM image. Instead of the widely used mean squared error (MSE), which poorly reflects visual quality[24], we used a combination of the absolute difference (L1 norm) with a multiscale version of the structural similarity index, a perceptually motivated quality metric shown to improve image restoration with deep learning (MS-SSIM)[25]. The second term measures the consistency between the A-net output and the widefield image. Although in theory the latter should simply be a blurred version of

the PALM image, this is often not the case in practice[26]. Therefore, we introduced another CNN (with four layers), called 'low-resolution estimator' to predict the widefield image from the super-resolution image. The corresponding loss was defined as the MS-SSIM between this CNN's output and the observed widefield image. (In the absence of a widefield image, this loss is set to zero.) The third term contains a cGAN discriminator loss[20], where the discriminator is a five-layer CNN, whose inputs are the sparse PALM (and widefield) image(s) and either the dense PALM image or the output of the generator above; the discriminator's output is compared to 0s and 1s (for synthetic and real, respectively), via the MSE. We use dropout[27] and extensive data augmentation, including random rotations, translations and elastic deformations, with addition of noise in the input image to mimic false detections and unspecific labeling. As a result, only a few dense PALM images are required for successful training without overfitting. On graphical processing units (GPUs), training ANNA-PALM from scratch takes on the order of hours to days, but when starting from a previously trained A-net, retraining can be done in an hour or less.

Once trained, the A-net can take sparse localization data with an optional widefield image as input(s), and output a reconstructed super-resolution image in less than a second (**Fig. 1b**). In addition, the A-net produces an 'error map' that measures the consistency of this super-resolution image with the widefield image[26] (when available) and can be used to estimate the degree of reliability and highlight potential reconstruction artifacts. For more details, see Online Methods and **Supplementary Note 1**.

Validating ANNA-PALM on simulated images

We first sought to validate ANNA-PALM on synthetic data. For this, we used Brownian dynamics simulations[28] to generate 200 dense PALM images of semi-flexible filaments mimicking microtubules, with a resolution $R_{loc} \approx 23$ nm. These represent 'perfect' PALM images that would be obtained with an infinite number of localizations. We applied varying levels of Poisson noise to these perfect images to create sparse PALM images corresponding to finite numbers of localizations (**Supplementary Fig. 1**). We then trained our A-net using the perfect images as targets and the sparse images (and widefield image) as inputs, varying the number of localizations over a large range.

Next, we applied the trained A-net to a distinct set of PALM images generated by the same stochastic simulation (**Supplementary Fig. 2**). **Figure 2a** shows a widefield image and **Figure 2b** a corresponding sparse PALM image obtained from $n = 6{,}834$ localizations. Although curvilinear structures can be seen in this image despite its sparsity, small-scale features remain highly ambiguous (**Fig. 2b**, inset), and the resolution according to a recently proposed fivefold Nyquist criterion[3] is limited by sampling to $R_{5\times Nyq} \approx 85$ nm; according to this criterion, $N > N_{5xNyq} \approx 60{,}000$ localizations are needed to achieve 23-nm resolution (**Supplementary Fig. 3d**). **Figure 2c** shows the ANNA-PALM image reconstructed from the widefield image alone, which exhibits clear and continuous filaments that were not previously recognizable. Most of the relatively isolated filaments roughly agree with the perfect PALM image (**Fig. 2e**). In the denser regions, however, many small features are erroneous, for example, filaments are incorrectly joined, displaced, split or merged (**Fig. 2c**, blue arrows and **Supplementary Fig. 4**, top). By contrast, the ANNA-PALM image reconstructed from the sparse PALM image alone or in combination with the widefield image, exhibits continuous and sharp filaments in very good agreement with the perfect PALM image (**Supplementary Fig. 3b**,**c** and **Fig. 2d**–**f**). The spatial resolution of these reconstructed images is limited neither by diffraction nor sampling, but only by the localization precision, and is thus $R_{loc} \approx 23$ nm, as in the perfect images (**Supplementary Fig. 3e**,**f**). These results indicate that high-quality, super-resolution images can be obtained from only a small fraction of the number of localizations traditionally

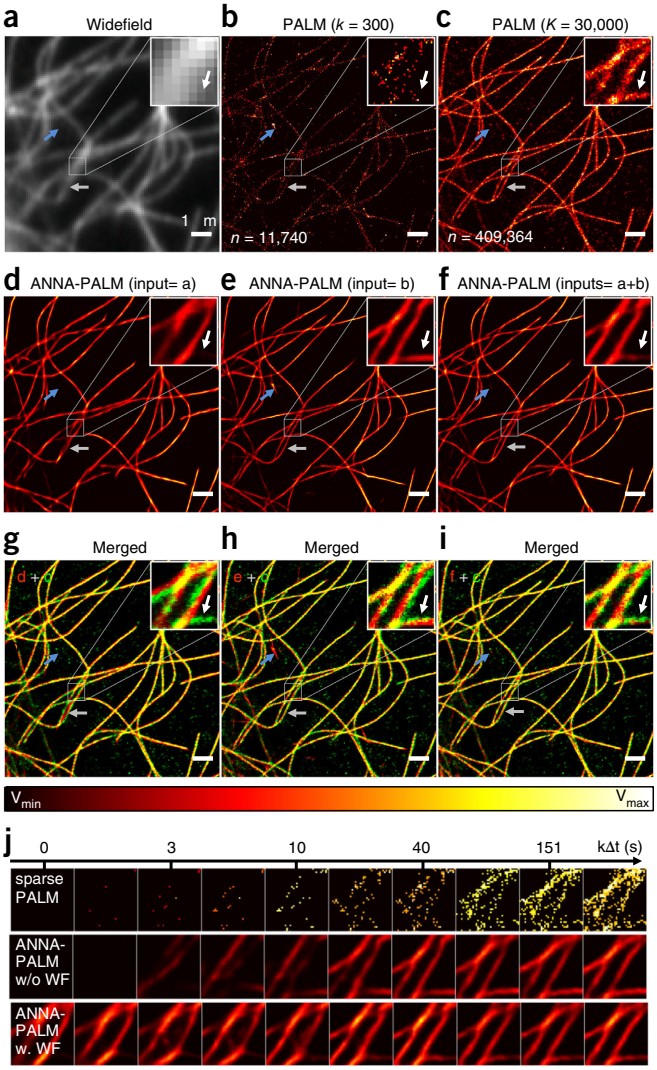


Figure 3 ANNA-PALM imaging of microtubules. ANNA-PALM reconstructions of a localization microscopy image of immunostained microtubules. (**a**) Widefield image. (**b**) Sparse PALM image obtained from the first 9 s of acquisition ($k = 300$ frames, $n = 11{,}740$ localizations). (**c**) Dense PALM image obtained from a 15-min-long acquisition ($K = 30{,}000$ frames, $N = 409{,}364$ localizations). (**d**) ANNA-PALM reconstruction from the widefield image **a** only. (**e**) ANNA-PALM reconstruction from the sparse PALM image **b** only. (**f**) ANNA-PALM reconstruction from the widefield image **a** and sparse PALM image **b** combined. In panels **b**–**f**, pixel values are linearly mapped to colors from the look-up table shown below. Black and white correspond to values V_{min} and V_{max}, respectively, with $V_{min} = 0$ for all panels, $V_{max} = 3$, 24, 102, 102 and 102 for panels **b**, **c**, **d**, **e** and **f**, respectively. (**g**–**i**) Merged images comparing ANNA-PALM reconstructions from panels **d**–**f** to the dense PALM image **c**. ANNA-PALM reconstructions are shown in red, the dense PALM image in green. (**j**) Gradual improvement of image quality for increasing acquisition time $k\Delta t$, shown for the area highlighted in the insets of **a**–**i**. Top row: sparse PALM images. Middle row: ANNA-PALM reconstructions from the sparse PALM images only (without widefield). Bottom row: ANNA-PALM reconstructions from the widefield and sparse PALM images combined. **Supplementary Video 1** shows the gradual increase in quality of PALM and ANNA-PALM images with increased acquisition time for the larger region of interest shown in **a**–**i**.

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

required (here, ~11% of N_{5xNyq} above; **Supplementary Fig. 3d**), hence enabling a strong reduction in acquisition time. Nevertheless, reconstruction errors can still occur in areas where the sparse localization data are most ambiguous, for example, where filament density is highest (**Fig. 2d**,**e**,**f**, white arrow). These errors can be reduced by increasing the localization number n, implying a trade-off between acquisition time and reconstruction quality (**Supplementary Fig. 4**).

To quantify this trade-off, we computed the MS-SSIM between reconstructed ANNA-PALM and perfect PALM images ($n = \infty$) as a function of localization number, from $n \sim 200$ to $n \sim 2 \times 10^6$, in comparison with the standard PALM images (**Fig. 2g**). The MS-SSIM lies between 0 and 1 and reaches 1 for perfect reconstructions. For standard PALM images, the MS-SSIM increases monotonically, as expected, from <0.2 to >0.95 for $n = 2 \times 10^6$ localizations (**Fig. 2g**, black curve). Using only the sparse image as input, ANNA-PALM reconstructions achieve MS-SSIM that are consistently higher and increase with localization number n much more rapidly than standard PALM, exceeding 0.9 for $n \approx 10{,}000$ localizations (**Fig. 2**, dashed blue curve). ANNA-PALM achieves the same MS-SSIM as standard PALM at the fivefold Nyquist sampling level (≈0.65) with only $n = 2{,}248$ localizations instead of $n = 58{,}588$, suggesting a ~26-fold speed-up. If the widefield image is used as additional input, the MS-SSIM further increases, and dramatically so for low localization numbers (**Fig. 2g**, solid blue curve). For example, with $n = 7{,}477$ localizations, ANNA-PALM achieves an MS-SSIM (≈0.95) similar to standard PALM with $n = 644{,}844$, implying a speed-up of roughly two orders of magnitude. (Note that, if the perfect PALM image was not available for these quantifications, it could be replaced by the ANNA-PALM reconstruction of a dense PALM image with a large number of localizations, for example, $n = 10^5$, with similar results; **Supplementary Fig. 5**).

As any image restoration method, ANNA-PALM can make errors. The low-resolution error map described above (**Fig. 1b**) provides a means to estimate where errors are most likely to occur. When applied to ANNA-PALM reconstructions of a sparse PALM image, this error map highlights regions containing the highest density of filament crossings, where reconstructions tend to be least accurate (**Supplementary Fig. 6i**,**k**). When we artificially displace a small piece of filament in this image to simulate a false positive and a false negative in the reconstruction (**Supplementary Fig. 6b**,**d**, white and blue arrows, respectively), the affected regions also light up in the error map (**Supplementary Fig. 6j**,**l**). Thus, the error map offers a useful tool to highlight regions most likely to contain reconstruction errors, and conversely, to outline regions where reconstructions are most trustworthy. Thus our, simulations suggest that ANNA-PALM can considerably reduce acquisition time in localization microscopy and also map reconstruction reliability.

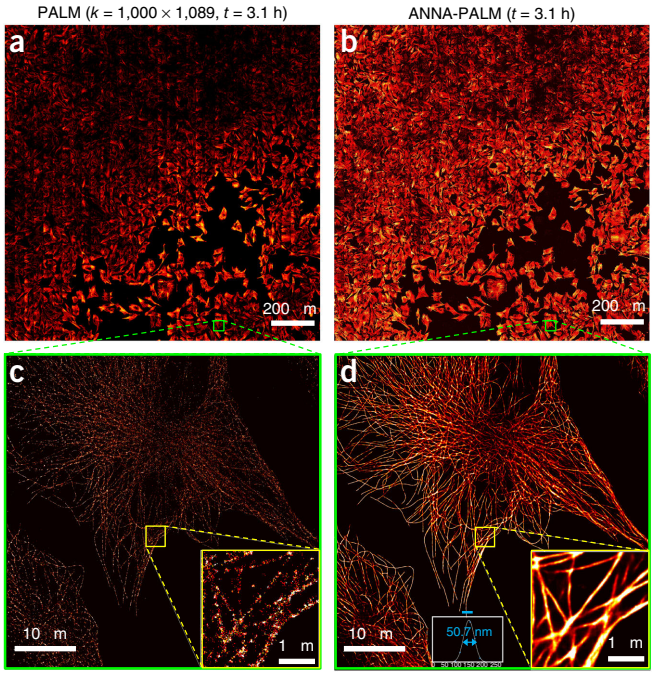


Figure 4 High-throughput imaging with ANNA-PALM. Application of ANNA-PALM to high-throughput imaging of a 1.8 mm × 1.8 mm area containing more than 1,000 cells. (**a**) Sparse PALM image of this area, constructed by assembling a mosaic of 33 × 33 = 1,089 sparse PALM images of individual fields of view, obtained from $k = 1{,}000$ raw frames each (with $\Delta t = 10$ ms exposure time per frame, i.e., in $k\Delta t = 10$ s). Total image-acquisition time was 1,089 × 10s, that is, ~3.1 h. The sparsity of the image is not apparent at this large scale. (**b**) ANNA-PALM reconstruction of the image in **a**, obtained by assembling a mosaic of 1,089 individual reconstructions (one per field of view). (**c**) Magnified view of the green boxed region in **a**. The inset shows a further magnified view of the yellow boxed region, highlighting the sparsity of the image. (**d**) Same as **c**, but for the ANNA-PALM reconstruction. A line profile across a microtubule is shown, with a FWHM ≈51 nm. Nonlinear contrast adjustment was applied manually for **a** and **b**, with black corresponding to values of zero in both panels. In **c** and **d**, pixel values were linearly mapped to colors from the look-up table in **Figure 3**; black and white correspond to values V_{min} and V_{max}, respectively, with $V_{min} = 0$ for all panels, and $V_{max} = 3$ and 51 for panels **c** and **d**, respectively. See also **Supplementary Video 2** for an animated 'zoom-in' highlighting the spatial scales covered by the assembled image. See also **Supplementary Figure 12** for ANNA-PALM reconstructions of the same area from the widefield images only.

ANNA-PALM reconstructions of immunostained microtubules

We next tested ANNA-PALM on real images of immunolabeled microtubules (**Fig. 3**). We trained our A-net on seven dense PALM images with corresponding widefield images obtained during 10-min-long acquisitions ($K = 60{,}000$; $\Delta t = 10$ ms exposure time) (data not shown). We then considered a sparse PALM image of microtubules in a distinct FoV obtained from only 9 s of acquisition ($k = 300$; $\Delta t = 30$ ms), together with a widefield image ($\Delta t = 2 \times 50$ ms) (**Fig. 3a**,**b**). Whereas microtubule filaments can already be seen in this sparse PALM image, structural details below the diffraction limit are hard to discern, making it difficult to follow the path of individual filaments in the denser regions and to identify features such as filament crossings (**Fig. 3b**). By contrast, the ANNA-PALM images, whether reconstructed from the widefield image alone, the sparse PALM image alone, or both, all display sharp and continuous filaments and clearly reveal many structural details (**Fig. 3d**–**f**). Their resolution is similar to or even better than the dense PALM image (**Supplementary Fig. 7a**). As for the simulations, in regions where microtubule filaments are isolated, the ANNA-PALM image reconstructed from the widefield image alone is in good agreement with the dense PALM image (**Fig. 3d**,**g**). However, it is often incorrect in areas of high microtubule density (e.g., **Fig. 3d**,**g**, white and gray arrows). Most of these reconstruction errors are corrected when applying ANNA-PALM to the sparse PALM image instead (**Fig. 3e**,**h**). For example, parallel sections of two microtubules unresolved in the widefield image and incorrectly merged in **Figure 3d** are now clearly separated and positioned correctly, and missed portions of other filaments are recovered (**Fig. 3h**, white and gray arrows). Counterintuitively, the sparse PALM image exhibits high signal in some locations where

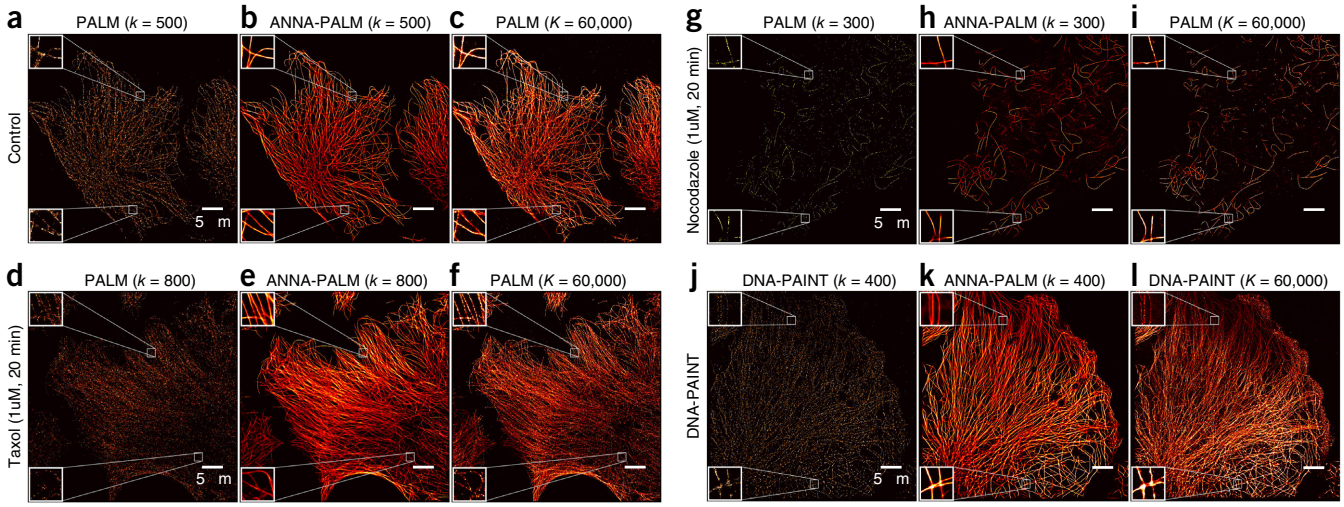


Figure 5 Robustness of ANNA-PALM to experimental perturbations. (**a**–**l**) Sparse PALM images and ANNA-PALM reconstructions using an ANN trained on PALM images of microtubules in untreated cells and applied without retraining to sparse localization images of microtubules in different experimental conditions: untreated control cells (**a**–**c**); cells treated with 1 μM of taxol (**d**–**f**); cells treated with 1 μM of nocodazole (**g**–**i**); untreated cells imaged with DNA-PAINT (**j**–**l**). (**a**,**d**,**g**,**j**) Sparse localization images obtained from the first *k* frames of the acquired image sequence, with *k* = 500, 800, 300, and 400 for **a**, **d**, **g**, and **j**, respectively. (**b**,**e**,**h**,**k**) ANNA-PALM reconstructions using the sparse localization images immediately to the left as input. (**c**,**f**,**i**,**l**) Dense localization images obtained from *K* = 60,000 frames. Pixel values are linearly mapped to colors from the look-up table in **Figure 3**. Black and white correspond to values V_{min} and V_{max}, respectively, with V_{min} = 0 for all panels, and V_{max} = 10, 120, 90, 25, 150, 40, 18, 150, 50, 18, 120, and 200 for panels **a**, **b**, **c**, **d**, **e**, **f**, **g**, **h**, **i**, **j**, **k**, and **l**, respectively.

the dense PALM image does not, presumably because of spurious localizations due, for example, to unspecific binding (**Fig. 3b**, blue arrow). Such signal can lead to incorrect features in the ANNA-PALM reconstruction from the sparse localization data alone (**Fig. 3e**,**h**, blue arrows). However, when the widefield and sparse PALM data are combined, these artifacts are largely removed and reconstructions agree very well with the dense PALM image (**Fig. 3f**,**i**). Reconstruction quality increases with the number of frames *k* (**Fig. 3j**, **Supplementary Figs. 8** and **9** and **Supplementary Video 1**). More quantitatively, an MS-SSIM analysis similar to that for the simulated data above (with the ANNA-PALM output of the dense PALM image defined as ground truth; **Supplementary Figs. 5** and **9f**) suggests that ANNA-PALM allows a 100-fold reduction of acquisition time compared to standard PALM (**Supplementary Fig. 9g**). **Supplementary Figure 10** shows other examples of sparse microtubule images reconstructed by ANNA-PALM.

As for simulations above, we used the widefield image to compute an error map (**Supplementary Figs. 6** and **11**). Bright areas in this error map highlight regions where the reconstruction indeed disagrees with the dense PALM image; conversely, reconstructions are of high quality in the majority of regions where the error map is dim (**Supplementary Fig. 11d**–**f**). These results demonstrate experimentally that ANNA-PALM can restore high-quality approximations of super-resolution images from much shorter acquisition times than typical for PALM imaging, and also predict where reconstruction errors are most likely.

ANNA-PALM enables high-throughput super-resolution imaging

The drastic improvement in imaging efficiency afforded by ANNA-PALM permits super-resolution imaging of orders of magnitude more cells and FoVs per unit time. To demonstrate this, we used an automated acquisition protocol to image >1,000 cells with immunolabeled microtubules in 1,089 (33 × 33), partly overlapping, FoVs of 55.3 μm × 55.3 μm each (**Fig. 4** and **Supplementary Fig. 12**). We first acquired widefield images at each of these positions, in a total of ~12 min, mostly consisting of stage stabilization delays (**Supplementary Fig. 12a**). Next, we obtained 1,089 sparse PALM images using only 10 s of imaging time per FoV (*k* = 1,000 frames, Δ*t* = 10 ms), in a total of only ~3.1 h (**Fig. 4a**). Neither the widefield nor the sparse PALM images provide much small-scale information (**Fig. 4c** and **Supplementary Fig. 12c**,**e**). However, ANNA-PALM reconstruction produces high-quality, super-resolution images, allowing the microtubule network to be visualized with clarity and microtubule filaments to be distinguished in dense areas that appeared as unstructured regions in the sparse PALM image (**Fig. 4b**,**d**). The full width at half maximum across filaments in the reconstructed image is ~51 nm (**Fig. 4d**), within the range measured for the training data (**Supplementary Fig. 7a**). Similar images are obtained by ANNA-PALM using the widefield images alone (**Supplementary Fig. 12**), although with lower quality, as discussed above. Stitching the reconstructed images together yields a single super-resolution image that contains approximately 7 billion 20 nm × 20 nm pixels and covers an area of 1.8 mm × 1.8 mm, thereby spanning almost five orders of magnitude in spatial scales (**Fig. 4b**,**d**, **Supplementary Fig. 12b**, and **Supplementary Video 2**).

ANNA-PALM is robust to experimental perturbations

ANNA-PALM can reconstruct accurate super-resolution images from sparse data because it was trained on similar images before. This raises the question of whether an ANN trained in one experimental condition can be successfully applied to another condition. To test this, we used the A-net as trained above to analyze microtubule images of cells exposed to drugs affecting the cytoskeletal network. We first treated U373 cells with 1 μM of taxol, an antimitotic agent that inhibits the depolymerization of microtubules and increases their bending rigidity[29,30]. In a sparse PALM image of these cells

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

(k = 800, $k\Delta t$ = 8 s), the microtubule network was barely recognizable (**Fig. 5d**). By contrast, the ANNA-PALM reconstructions clearly display a complex microtubule network and agree well with the dense PALM image obtained from K = 60,000 frames ($K\Delta t$ = 10 min) (**Fig. 5d–f**). These images show a greater density of straighter and more parallel filaments with less frequent crossings than in the untreated cells (**Fig. 5a–c**), consistent with microtubule stabilization and increased rigidity[30].

Next, we treated cells with 1 µM of nocodazole, a drug that, in contrast, promotes microtubule depolymerization and is expected to more dramatically alter the cytoskeletal network[31]. Again, whereas the sparse PALM image obtained from k = 300 frames ($k\Delta t$ = 3 s) contained little exploitable information (**Fig. 5g**), the ANNA-PALM reconstruction offer clear and detailed views of the disrupted microtubule network (**Fig. 5h**), exhibiting a much smaller number of filaments, with higher curvature, than in untreated cells. These reconstructions are in good (though not perfect) agreement with dense PALM images obtained from K = 60,000 frames ($K\Delta t$ = 10 min) (**Fig. 5i**). Thus, ANNA-PALM, when trained on microtubules in one experimental condition, may be successfully applied to new experimental conditions without retraining, thereby highlighting the method's robustness to biologically relevant structural alterations.

We further asked if ANNA-PALM is also robust to changes in technical imaging conditions. To address this, we performed localization microscopy on microtubules by simultaneously changing multiple important imaging parameters relative to the training data. Instead of PALM/STORM, we used DNA-PAINT, a technique where single-molecule detection relies on transient binding of fluorophore-conjugated short DNA strands to complementary, antibody-conjugated, DNA strands[32], rather than on fluorophore blinking. The continuously emitting, freely diffusing dyes lead to higher background noise in DNA-PAINT compared to PALM/STORM. Moreover, we used primary mouse antibodies instead of rat antibodies, Cy3 dyes instead of Alexa-647 dyes, and an EMCCD instead of a sCMOS camera. Despite all these differences, when ANNA-PALM is applied without retraining on a sparse microtubule image (k = 400 frames, $k\Delta t$ = 12 s), the reconstructed image still agrees very well with the dense DNA-PAINT image obtained from K = 60,000 frames ($K\Delta t$ = 30 min) (**Fig. 5j–l**). These data demonstrate the high robustness of ANNA-PALM to changes in experimental imaging conditions.

ANNA-PALM adapts to different biological structures

To demonstrate that ANNA-PALM is not restricted to filamentary structures, we turned to nuclear pores, a very different biological structure, and another popular target of super-resolution imaging studies[33–35]. We retrained A-net simultaneously on microtubule images and on a single PALM image of the nucleoporin gp210 in immunolabeled nuclear membranes of *Xenopus* frog eggs[33,35] (K = 30,000). With the switch (**Fig. 1**) set to microtubules ('MT'), this newly trained A-net can still reconstruct sparse images of microtubules as when trained exclusively on microtubule data (**Supplementary Fig. 13a–c**). We then applied the same A-net with the switch set to nuclear pores ('NPC') to reconstruct a new sparse PALM image of gp210 obtained from the first k = 3,000 frames (**Fig. 6a**). The sparsity of this image makes it difficult to clearly distinguish individual nuclear pores. ANNA-PALM, however, produces a much clearer image, containing many easily identifiable ring-like structures, as expected for nuclear pores[33] (**Fig. 6b**), and in good agreement with the dense PALM image obtained from K = 30,000 frames (even though the latter shows mostly incomplete, open rings, presumably due to suboptimal labeling) (**Fig. 6c**). An automated procedure based on cross-correlation with a ring template indeed identified ~2.7 times more putative nuclear pores from the ANNA-PALM image than the sparse PALM image (**Supplementary Fig. 14a–c**). Moreover, computed pore locations are in good agreement with a corresponding, second color PALM image of wheat germ agglutinin (WGA), a lectin that concentrates in the inner nuclear pore channel[35] (**Supplementary Fig. 14 d–f**). These results show that ANNA-PALM can successfully analyze non-filamentary structures, when properly retrained, and that a single ANN, with a simple computational switch, can reconstruct very different types of structures.

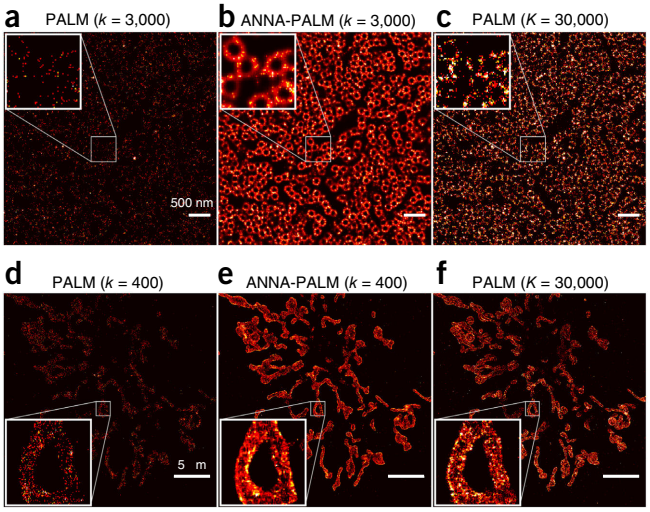


Figure 6 ANNA-PALM reconstructions of nuclear pores and mitochondria. (**a–f**) Sparse PALM and ANNA-PALM images of nuclear pores (**a–c**) and mitochondria (**d–f**). (**a**) Sparse PALM image of the immunolabeled *Xenopus* nucleoporin gp210 obtained from the first k = 3,000 frames. Note that individual nuclear pores are hard to identify. (**b**) ANNA-PALM reconstruction of image **a**. (**c**) Dense PALM image obtained from all K = 30,000 frames. (**d**) Sparse PALM image of the immunolabeled mitochondrial outer membrane protein TOM22, obtained from the first k = 400 frames. (**e**) ANNA-PALM reconstruction of image **d**. (**f**) Dense PALM image obtained from all K = 30,000 frames. Pixel values are linearly mapped to colors from the look-up table shown in **Figure 3**. Black and white correspond to values V_{min} and V_{max}, respectively, with V_{min} = 0 for all panels, and V_{max} = 3, 51, 3, 3, 128, and 18 for panels **a**, **b**, **c**, **d**, **e**, and **f**, respectively.

Finally, we imaged TOM22, a protein of the mitochondrial outer membrane[36]. Whereas, at the resolution of our experiments, microtubules and nucleoporins are essentially one-dimensional and zero-dimensional structures, respectively, mitochondrial membranes are two-dimensional surfaces. Furthermore, their complex three-dimensional (3D) morphology might seem less predictable than filaments or nuclear pores, potentially hampering ANNA-PALM reconstruction. Despite these differences, after being trained on nine PALM images of TOM22 (with frame numbers ranging from K = 24,000 to K = 40,000), ANNA-PALM reconstructions of distinct sparse PALM images (k = 400 frames) display mitochondrial morphologies in good agreement with the dense PALM images (**Fig. 6d–f**), although the protein's localization along the membrane is less well reproduced. Taken together, our results illustrate the versatility of ANNA-PALM and its applicability to images of very different structural content.

DISCUSSION

We introduce ANNA-PALM, a computational method based on deep learning, that reconstructs high-quality, super-resolution images from

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

sparse, rapidly acquired, single-molecule localization data and/or widefield images. Our method enables considerable gains in acquisition time compared to standard localization microscopy without increasing active fluorophore density, thereby preserving spatial resolution. In fact, ANNA-PALM even improves spatial resolution when applied to images of lower resolution than the training data (**Supplementary Figs. 7a** and **10**), and greatly diminishes the detrimental effect of drift. The improvement in imaging efficiency afforded by ANNA-PALM alleviates the incompatibility between high-resolution and high-throughput microscopy by enabling super-resolution imaging of thousands of cells within a few hours or even less (**Fig. 4** and **Supplementary Fig. 12**). This will facilitate super-resolution studies of rare events, cellular heterogeneity, and of partly stochastic structures such as cytoskeletal polymers or chromosomes, whose characterization requires statistics on many configurations[37,38]. ANNA-PALM may also be beneficial for high-throughput imaging screens, for example, of drug treatments or gene knockouts[39–41]. In addition, we envision applications to super-resolution imaging of large samples by stitching together multiple images of spatially adjacent fields. The ability to generate images spanning many orders of magnitude in scale could be well adapted to expansion microscopy, a super-resolution technique that physically increases sample size, but often requires tiling many fields of view to image even a single cell[42,43]. With correlative microscopy[44], it might also be possible to train ANNA-PALM to reconstruct electron microscopy (EM) images from fluorescence images, potentially extending the method to molecular resolutions currently out of reach of localization microscopy. Adaptation of ANNA-PALM to 3D[4,45] and multicolor[46,47] localization microscopy should be relatively straightforward. Localization microscopy of cellular dynamics remains very challenging[3,48]. By using far fewer frames (or even only widefield images), ANNA-PALM could dramatically improve the temporal resolution of live-cell localization microscopy without sacrificing spatial resolution or increasing phototoxicity and photobleaching. Thus, ANNA-PALM provides multiple novel avenues for multiscale imaging beyond conventional spatiotemporal resolution limits.

Nevertheless, important caveats should be stressed. First, although ANNA-PALM can be applied successfully to very different types of images (**Fig. 6**), the method fails in the absence of statistical redundancies between molecular localizations, for example, for entirely random distributions of molecules. Second, ANNA-PALM requires prior training on dense PALM images with structures similar to those in the images to be reconstructed. We showed that ANNA-PALM is robust, that is, does not require retraining, for some experimentally induced changes in structures and variations in imaging parameters (**Fig. 5**). However, indiscriminate application of ANNA-PALM to very different structures without retraining, or incorrect setting of the switch, may result in artifacts (**Supplementary Fig. 13h**). Third, even when applied to data similar to the training images, ANNA-PALM can produce errors, as can any reconstruction method in a context of information scarcity. The frequency of errors can be reduced by increasing the number of recorded frames, at the cost of reduced acceleration (**Figs. 2**, **3** and **Supplementary Figs. 4**,**8** and **9**). In addition, ANNA-PALM can use widefield images to estimate the reliability of reconstructions, thereby helping their interpretation, providing some protection against artifacts and indicating when retraining may be needed (**Supplementary Figs. 11** and **15**). Future work, for example, using Bayesian deep learning[49], may provide additional assessments of reconstruction uncertainty and shed more light on the scope and limitations of our approach.

As a purely computational technique, ANNA-PALM does not necessitate any changes to existing microscopy systems, but only requires one or a few standard PALM images for training. To facilitate its adoption and future development, we make our source code, an ImageJ plugin and a cloud-computing-based web application, available on https://annapalm.pasteur.fr/ along with sample data. Because the performance of deep-learning methods improves with the amount and variety of training data, we designed our web application to enable sharing of data and trained ANNs. As ANNA-PALM will learn from an increasingly large and diverse collection of images, it should reach even higher accuracy or efficiency and further expand its scope of application in the future.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank the following colleagues for useful discussions and suggestions and/or critical reading of the manuscript: C. Leduc, S. Etienne-Manneville, S. Lévêque-Fort, N. Bourg, A. Echard, J.-B. Masson, T. Rose, P. Hersen, F. Mueller, M. Cohen, Z. Zhang, and P. Kanchanawong. We also thank the four anonymous reviewers for their constructive criticism, which led to significant improvements of ANNA-PALM. We further thank O. Faklaris, J. Sellés and M. Penrad (Institut Jacques Monod), and F. Montel (Ecole Normale Supérieure de Lyon) for providing *Xenopus* nuclear pore data, J. Bai (Institut Pasteur) for TOM22 antibodies, and C. Leterrier for fixation protocols. We thank E. Rensen and C. Weber for help with experiments and suggestions, B. Lelandais for help with PALM image processing, J.-B. Arbona for polymer simulations and J. Parmar for suggestions that led to the name ANNA-PALM. We thank the IT service of Institut Pasteur, including J.-B. Denis, N. Joly, and S. Fournier, for access to the HPC cluster and relevant assistance, and T. Huynh for help with GPU computing. This work was funded by Institut Pasteur, Agence Nationale de la Recherche grant (ANR 14 CE10 0018 02), Fondation pour la Recherche Médicale (Equipe FRM, DEQ 20150331762), and the Région Ile de France (DIM Malinf). We also acknowledge Investissement d'Avenir grant ANR-16-CONV-0005 for funding a GPU farm used in this work. A.A. and X.H. are recipients of Pasteur-Roux fellowships from Institut Pasteur. W.O. is a scholar in the Pasteur–Paris University (PPU) International PhD program.

AUTHOR CONTRIBUTIONS

W.O. conceived the method, developed ANNA-PALM software and web application, and performed experiments and analyses. A.A., M.L., and X.H. performed experiments. C.Z. conceived the method, supervised the project, and wrote the manuscript.

COMPETING INTERESTS

W.O. and C.Z. are listed as inventors on European patent applications EP17306022 and EP18305225.7 filed by Institut Pasteur.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Betzig, E. *et al.* Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642–1645 (2006).
2. Rust, M.J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **3**, 793–796 (2006).
3. Legant, W.R. *et al.* High-density three-dimensional localization microscopy across large volumes. *Nat. Methods* **13**, 359–365 (2016).
4. Deschout, H. *et al.* Precisely and accurately localizing single emitters in fluorescence microscopy. *Nat. Methods* **11**, 253–266 (2014).
5. Jones, S.A., Shim, S.H., He, J. & Zhuang, X. Fast, three-dimensional super-resolution imaging of live cells. *Nat. Methods* **8**, 499–505 (2011).
6. Huang, F. *et al.* Video-rate nanoscopy using sCMOS camera-specific single-molecule localization algorithms. *Nat. Methods* **10**, 653–658 (2013).
7. Carlton, P.M. *et al.* Fast live simultaneous multiwavelength four-dimensional optical microscopy. *Proc. Natl. Acad. Sci. USA* **107**, 16016–16022 (2010).
8. Stelzer, E.H.K. Light-sheet fluorescence microscopy for quantitative biology. *Nat. Methods* **12**, 23–26 (2015).
9. Huang, F., Schwartz, S.L., Byars, J.M. & Lidke, K.A. Simultaneous multiple-emitter fitting for single molecule super-resolution imaging. *Biomed. Opt. Express* **2**, 1377–1393 (2011).

10. Burnette, D.T., Sengupta, P., Dai, Y., Lippincott-Schwartz, J. & Kachar, B. Bleaching/blinking assisted localization microscopy for superresolution imaging using standard fluorescent molecules. *Proc. Natl. Acad. Sci. USA* **108**, 21081–21086 (2011).
11. Simonson, P.D., Rothenberg, E. & Selvin, P.R. Single-molecule-based super-resolution images in the presence of multiple fluorophores. *Nano Lett.* **11**, 5090–5096 (2011).
12. Zhu, L., Zhang, W., Elnatan, D. & Huang, B. Faster STORM using compressed sensing. *Nat. Methods* **9**, 721–723 (2012).
13. Cox, S. *et al.* Bayesian localization microscopy reveals nanoscale podosome dynamics. *Nat. Methods* **9**, 195–200 (2012).
14. Ram, S., Ward, E.S. & Ober, R.J. Beyond Rayleigh's criterion: a resolution measure with application to single-molecule microscopy. *Proc. Natl. Acad. Sci. USA* **103**, 4457–4462 (2006).
15. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
16. Michael, E. *Sparse and Redundant Representations: from Theory to Applications in Signal and Image Processing* (Springer, 2010).
17. Hinton, G.E. & Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. *Science* **313**, 504–507 (2006).
18. Schmidhuber, J. Deep learning in neural networks: an overview. *Neural Netw.* **61**, 85–117 (2015).
19. Sage, D. *et al.* Quantitative evaluation of software packages for single-molecule localization microscopy. *Nat. Methods* **12**, 717–724 (2015).
20. Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A.A. Image-to-image translation with conditional adversarial networks. Preprint at http://arxiv.org/abs/1611.07004 (2016).
21. Ronneberger, O., Fischer, P. & Brox, T. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015.* (eds., Navab, N., Hornegger, J., Wells, W.M. & Frangi, A.F.), 234–241 (Springer; 2015).
22. Goodfellow, I. *et al.* Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **27**, 2672–2680 (2014).
23. Ciresan, D.C., Giusti, A., Gambardella, L.M. & Schmidhuber, J. Mitosis detection in breast cancer histology images with deep neural networks. *Med. Image Comput. Comput. Assist. Interv.* **2013**, 8150 (2013).
24. Wang, Z.W.Z. & Bovik, A.C.A.C. Mean squared error: love it or leave it? A new look at Signal Fidelity Measures. *IEEE Signal Process. Mag.* **26**, 98–117 (2009).
25. Zhao, H., Gallo, O., Frosio, I. & Kautz, J. Loss functions for image restoration with neural networks. *IEEE Trans. Comput. IMAGING* **3**, 47–57 (2017).
26. Culley, S. *et al.* Quantitative mapping and minimization of super-resolution optical imaging artifacts. *Nat. Methods* **15**, 263–266 (2018).
27. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
28. Arbona, J.-M., Herbert, S., Fabre, E. & Zimmer, C. Inferring the physical properties of yeast chromatin through Bayesian analysis of whole nucleus simulations. *Genome Biol.* **18**, 81 (2017).
29. Arnal, I. & Wade, R.H. How does taxol stabilize microtubules? *Curr. Biol.* **5**, 900–908 (1995).
30. Wu, S. *et al.* Microtubule motors regulate ISOC activation necessary to increase endothelial cell permeability. *J. Biol. Chem.* **282**, 34801–34808 (2007).
31. De Brabander, M., De May, J., Joniau, M. & Geuens, G. Ultrastructural immunocytochemical distribution of tubulin in cultured cells treated with microtubule inhibitors. *Cell Biol. Int. Rep.* **1**, 177–183 (1977).
32. Schnitzbauer, J., Strauss, M.T., Schlichthaerle, T., Schueder, F. & Jungmann, R. Super-resolution microscopy with DNA-PAINT. *Nat. Protoc.* **12**, 1198–1228 (2017).
33. Löschberger, A. *et al.* Super-resolution imaging visualizes the eightfold symmetry of gp210 proteins around the nuclear pore complex and resolves the central channel with nanometer resolution. *J. Cell Sci.* **125**, 570–575 (2012).
34. Szymborska, A. *et al.* Nuclear pore scaffold structure analyzed by super-resolution microscopy and particle averaging. *Science* **341**, 655–658 (2013).
35. Sellés, J. *et al.* Nuclear pore complex plasticity during developmental process as revealed by super-resolution microscopy. *Sci. Rep.* **7**, 14732 (2017).
36. Bellot, G. *et al.* TOM22, a core component of the mitochondria outer membrane protein translocation pore, is a mitochondrial receptor for the proapoptotic protein Bax. *Cell Death Differ.* **14**, 785–794 (2007).
37. Boettiger, A.N. *et al.* Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **529**, 418–422 (2016).
38. Zhang, Z., Nishimura, Y. & Kanchanawong, P. Extracting microtubule networks from superresolution single-molecule localization microscopy data. *Mol. Biol. Cell* **28**, 333–345 (2017).
39. Neumann, B. *et al.* Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature* **464**, 721–727 (2010).
40. Beghin, A. *et al.* Localization-based super-resolution imaging meets high-content screening. *Nat. Methods* **14**, 1184–1190 (2017).
41. Ouyang, W. & Zimmer, C. The imaging tsunami: computational opportunities and challenges. *Curr. Opin. Syst. Biol.* **4**, 105–113 (2017).
42. Chen, F., Tillberg, P.W. & Boyden, E.S. Optical imaging. Expansion microscopy. *Science* **347**, 543–548 (2015).
43. Chang, J.-B. *et al.* Iterative expansion microscopy. *Nat. Methods* **14**, 593–599 (2017).
44. de Boer, P., Hoogenboom, J.P. & Giepmans, B.N.G. Correlated light and electron microscopy: ultrastructure lights up!. *Nat. Methods* **12**, 503–513 (2015).
45. Huang, B., Wang, W., Bates, M. & Zhuang, X. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science* **319**, 810–813 (2008).
46. Bates, M., Huang, B., Dempsey, G.T. & Zhuang, X. Multicolor super-resolution imaging with photo-switchable fluorescent probes. *Science* **317**, 1749–1753 (2007).
47. Jungmann, R. *et al.* Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT. *Nat. Methods* **11**, 313–318 (2014).
48. Shroff, H., Galbraith, C.G., Galbraith, J.A. & Betzig, E. Live-cell photoactivated localization microscopy of nanoscale adhesion dynamics. *Nat. Methods* **5**, 417–423 (2008).
49. Kendall, A. & Gal, Y. What uncertainties do we need in Bayesian deep learning for computer vision? *Adv. Neural Inf. Process. Syst.* **30**, 5580–5590 (2017).

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

ONLINE METHODS

Artificial neural network. Our ANN, called 'A-net', is based on the pix2pix architecture[20], which is a special conditional generative adversarial network (cGAN)[22] for image-to-image 'translation', that is, mapping from one type of image to another. The A-net consists of three distinct neural networks: (i) a generator network $\mathcal{G}$ that produces the reconstructed super-resolution image, (ii) a network Q called 'low-resolution estimator' that produces the low-resolution error map, (iii) a cGAN discriminator network $\mathcal{D}$ that provides the adversarial loss (**Fig. 1a**). The generator network $\mathcal{G}$ builds on the U-net architecture, and consists of an encoder-decoder network with skip connections[21] and 16 convolutional layers. Its inputs and outputs are image patches containing $(256m) \times (256m)$ pixels, where m is an integer (we used $m = 1$ or 2, but this can be adjusted for different sizes of CPU/GPU memory or input images). The input of $\mathcal{G}$ is a sparse PALM image, a widefield image scaled up to the same size (see below), and a computational switch number that allows the network to switch between different types of image structures (e.g., nuclear pores or microtubules). The switch setting is encoded numerically and coupled by convolutional operations into the A-net encoder. The output of the generator $\mathcal{G}$ is a reconstructed image (called ANNA-PALM reconstruction or ANNA-PALM image elsewhere) of the same size as the input images. The low-resolution estimator Q has four convolutional layers. It takes the $(256m) \times (256m)$ dense PALM image patch or the ANNA-PALM image patch as input and outputs a low-resolution image with $(64m) \times (64m)$ pixels. The cGAN discriminator network D has five convolutional layers. Its inputs are three $(256m) \times (256m)$ pixel image patches (the sparse PALM image, the scaled-up widefield image, and either the ANNA-PALM reconstruction or the corresponding dense PALM image), and its output is a $(30m) \times (30m)$ image whose pixel values indicate whether the corresponding input patch is real (i.e., an experimentally obtained conventional PALM image) or produced by the generator $\mathcal{G}$. All convolutional layers are followed by batch normalization[50]. Dropout layers[27] (with dropout probability $P = 0.5$) are introduced in the central layers of the A-net generator and turned on during training, but switched off during inference. Activation functions are rectified linear units (ReLU) $x \to \sup(x, 0)$ or 'leaky' ReLUs $x \to \sup(x, 0) + \inf(\varepsilon x, 0)$ with $\varepsilon = 0.2$ (ref. 51), except for the last layer of $\mathcal{G}$, which uses the hyperbolic tangent $x \to \tanh(x)$ and the last layer of Q, which uses a sigmoid function $x \to (1+\exp(-x))^{-1}$. The A-net architecture is fully described in **Supplementary Note 1** and **Supplementary Tables 1–4**.

Training objectives and error map. Each of the three networks mentioned above ($\mathcal{G}$, Q and D) is associated to a distinct objective function—also called loss—and hereafter denoted $\mathcal{L}_{\mathcal{G}}$, $\mathcal{L}_{Q}$, and $\mathcal{L}_{\mathcal{D}}$, respectively. These loss functions are specified in detail below. In the following equations, for notational simplicity, we designate the sparse input image as $\boldsymbol{S}$, the low-resolution (widefield) input image as $\boldsymbol{W}$, the corresponding dense PALM image (i.e., the target) as $\boldsymbol{T}$, and the A-net's output as $\boldsymbol{A} = \mathcal{G}(\boldsymbol{S}, \boldsymbol{W})$ (in **Fig. 1a**, images $\boldsymbol{S}$, $\boldsymbol{T}$ and $\boldsymbol{A}$ are labeled Pk, PK, and Ak, respectively); the computational switch indicating the image type is denoted $\boldsymbol{M}$. Low-resolution images produced by the low-resolution estimator network Q from $\boldsymbol{A}$ and $\boldsymbol{T}$ are designated as $\boldsymbol{W}_A = Q(\boldsymbol{A})$ and $\boldsymbol{W}_T = Q(\boldsymbol{T})$, respectively.

The generator loss function $\mathcal{L}_{\mathcal{G}}$ is the sum of three terms. The first term of $\mathcal{L}_{\mathcal{G}}$ is the super-resolution reconstruction error, hereafter called $\mathcal{L}_{\text{SuperRes}}(\mathcal{G})$. This term penalizes the difference between the generator output $\boldsymbol{A}$ and the target image $\boldsymbol{T}$. Based on a recent analysis of loss functions for image restoration with neural networks[25], we defined this difference as a weighted average of two quantities: (i) the multiscale structural similarity index (MS-SSIM) between $\boldsymbol{A}$ and $\boldsymbol{T}$ and (ii) a modification of the L1 norm, where the absolute difference between $\boldsymbol{A}$ and $\boldsymbol{T}$ is smoothed by a Gaussian kernel:

$$\begin{aligned}\mathcal{L}_{\text{SuperRes}}(\mathcal{G}) &= \mathbb{E}_{(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})\sim p_{\text{data}}(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})}\Big[\rho\big(1-MS_SSIM(\boldsymbol{A},\boldsymbol{T})\big) \\ &\quad +(1-\rho)\big\langle G_\sigma * |\boldsymbol{A}-\boldsymbol{T}|\big\rangle\Big] \\ &= \mathbb{E}_{(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})\sim p_{\text{data}}(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})} \\ &\quad \Big[\rho\big(1-MS_SSIM(\mathcal{G}(\boldsymbol{S},\boldsymbol{W},\boldsymbol{M}),\boldsymbol{T})\big) \\ &\quad +(1-\rho)\big\langle G_\sigma * |\mathcal{G}(\boldsymbol{S},\boldsymbol{W},\boldsymbol{M})-\boldsymbol{T}|\big\rangle\Big]\end{aligned}$$

where $\mathbb{E}$ denotes expectation; $p_{\text{data}}(\boldsymbol{S}, \boldsymbol{T}, \boldsymbol{W}, \boldsymbol{M})$ is the joint probability density of the sparse PALM images $\boldsymbol{S}$, dense PALM images $\boldsymbol{T}$, widefield images $\boldsymbol{W}$ and switch settings $\boldsymbol{M}$ from the training data set; $MS_SSIM(\boldsymbol{A},\boldsymbol{T})$ is the multiscale structural similarity index between $\boldsymbol{A}$ and $\boldsymbol{T}$; G_σ is a Gaussian smoothing kernel; $*$ denotes convolution; $|\boldsymbol{A} - \boldsymbol{T}|$ is the absolute difference image (i.e., pixel (i,j) has value $|\boldsymbol{A}(i,j) - \boldsymbol{T}(i,j)|$) and $\rho \in [0,1]$ is a scalar weight that balances the relative contributions of MS-SSIM and the modified L1 norm and is set to $\rho = 0.84$ as in ref. 25.

The second term of $\mathcal{L}_{\mathcal{G}}$ is called $\mathcal{L}_{\text{LowRes}}(\mathcal{G},Q)$ and measures the consistency between the low-resolution images $\boldsymbol{W}_A$ and $\boldsymbol{W}_T$ predicted by the low-resolution estimator network Q:

$$\begin{aligned}\mathcal{L}_{\text{LowRes}}(\mathcal{G},Q) &= \mathbb{E}_{(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})\sim p_{\text{data}}(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})}\big[1-MS_SSIM(\boldsymbol{W}_A,\boldsymbol{W}_T)\big] \\ &= \mathbb{E}_{(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})\sim p_{\text{data}}(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W},\boldsymbol{M})}\big[1-MS_SSIM\big(Q(\mathcal{G}(\boldsymbol{S},\boldsymbol{W},\boldsymbol{M})),Q(\boldsymbol{T})\big)\big]\end{aligned}$$

Alternatively, in the above objective function, $\boldsymbol{W}_T$ can be replaced by the actually observed widefield image $\boldsymbol{W}$, although with our data this led to slightly lower reconstruction quality. The low-resolution estimator network Q is trained simultaneously with the generator $\mathcal{G}$ to produce a low-resolution image from the dense PALM image $\boldsymbol{T}$ that is consistent with the observed low-resolution image $\boldsymbol{W}$. This training is done based on the following objective function:

$$\mathcal{L}_Q(Q) = \mathbb{E}_{(\boldsymbol{T},\boldsymbol{W})\sim p_{\text{data}}(\boldsymbol{T},\boldsymbol{W})}\big[1-MS_SSIM\big(Q(\boldsymbol{T}),\boldsymbol{W}\big)\big]$$

Note that the reconstructed low-resolution image $Q(\boldsymbol{T})$ is four times smaller than the dense PALM image $\boldsymbol{T}$, as described in **Supplementary Note 1**. Because the input widefield image $\boldsymbol{W}$ can have a different size, we use bilinear interpolation to resize $\boldsymbol{W}$ to the same size as $Q(\boldsymbol{T})$ (If needed, a scaling factor different from 4 can be obtained by adding or removing downsampling layers in network Q). At inference time, the low-resolution estimator Q is also used to produce the error map, as shown in **Figure 1b**, and **Supplementary Figures 6**, **11** and **15**. This error map is defined as:

$$\boldsymbol{E}_Q(\boldsymbol{A},\boldsymbol{W}) = \big(1-MS_SSIM\big(Q(\boldsymbol{A}),\boldsymbol{W}\big)\big)\times\big(Q(\boldsymbol{A})+\boldsymbol{W}\big)$$

High (respectively low) values of the error map indicate large (respectively small) inconsistencies between the reconstructed super-resolution image $\boldsymbol{A}$ and the observed widefield image $\boldsymbol{W}$.

The third term of $\mathcal{L}_{\mathcal{G}}$ draws from recent work on generative adversarial networks (GAN)[20,22,52] and is noted $\mathcal{L}_{\text{cGAN}}(\mathcal{G},\mathcal{D})$. In a GAN, a generator network $\mathcal{G}$ learns to transform random input vectors $\boldsymbol{z}$ (drawn from a probability density $p_z(\boldsymbol{z})$) into new samples of a data probability density $p_{\text{data}}(\boldsymbol{x})$. In our case, the data samples $\boldsymbol{x}$ are the dense PALM images $\boldsymbol{T}$. The generator $\mathcal{G}$ learns by working against a discriminator network $\mathcal{D}$ that simultaneously learns to discriminate between original data samples and samples generated by $\mathcal{G}$. Adversarial training thus consists in playing a minmax game such that $\left(\mathcal{G}^*,\mathcal{D}^*\right) = \arg\min_{\mathcal{G}}\max_{\mathcal{D}}\mathcal{L}_{\text{GAN}}(\mathcal{G},\mathcal{D})$, with an objective function of the form[22]: $\mathcal{L}_{\text{GAN}}(\mathcal{G},\mathcal{D}) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}\big[\log\mathcal{D}(\boldsymbol{x})\big] + \mathbb{E}_{\boldsymbol{z}\sim p_z(\boldsymbol{z})}\log\big[1-\mathcal{D}(\mathcal{G}(\boldsymbol{z}))\big]$, or equivalently by simultaneous optimization of two coupled loss functions:

$$\begin{cases}\mathcal{D}^* = \arg\max_{\mathcal{D}}\Big(\mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}\big[\log\mathcal{D}(\boldsymbol{x})\big]+\mathbb{E}_{\boldsymbol{z}\sim p_z(\boldsymbol{z})}\log\big[1-\mathcal{D}(\mathcal{G}(\boldsymbol{z}))\big]\Big) \\ \mathcal{G}^* = \arg\min_{G}\Big(\mathbb{E}_{\boldsymbol{z}\sim p_z(\boldsymbol{z})}\log\big[1-\mathcal{D}(\mathcal{G}(\boldsymbol{z}))\big]\Big)\end{cases}$$

In a conditional GAN (cGAN), the generator and the discriminator have an extra input vector $\boldsymbol{c}$ and the first objective function above becomes: $\mathcal{L}_{\mathcal{D}}(\mathcal{G},\mathcal{D}) = \mathbb{E}_{(\boldsymbol{c},\boldsymbol{x})\sim p_{\text{data}}(\boldsymbol{c},\boldsymbol{x})}\big[\log\mathcal{D}(\boldsymbol{c},\boldsymbol{x})\big] + \mathbb{E}_{\boldsymbol{c}\sim p_{\text{data}}(\boldsymbol{c}),\boldsymbol{z}\sim p_z(\boldsymbol{z})}\log\big[1-\mathcal{D}(\boldsymbol{c},\mathcal{G}(\boldsymbol{c},\boldsymbol{z}))\big]$, such that the generator learns a conditional probability density $p_{\text{data}}(\boldsymbol{x}|\boldsymbol{c})$; the second objective function likewise becomes $\mathcal{L}_{\text{cGAN}}(\mathcal{G},\mathcal{D}) = \mathbb{E}_{\boldsymbol{c}\sim p_{\text{data}}(\boldsymbol{c}),\boldsymbol{z}\sim p_z(\boldsymbol{z})}\log\big[1-\mathcal{D}(\mathcal{G}(\boldsymbol{z}))\big]$. In our A-net, we replaced the logarithmic losses above by least square losses[52], as they empirically yielded better results. Thus, we used the objective functions:

$$\mathcal{L}_{\mathcal{D}}(\mathcal{G},\mathcal{D}) = \mathbb{E}_{(\boldsymbol{c},\boldsymbol{x})\sim p_{\text{data}}(\boldsymbol{c},\boldsymbol{x})}\big[\mathcal{D}(\boldsymbol{c},\boldsymbol{x})-1\big]^2 + \mathbb{E}_{\boldsymbol{c}\sim p_{\text{data}}(\boldsymbol{c}),\boldsymbol{z}\sim p_z(\boldsymbol{z})}\big[\mathcal{D}(\boldsymbol{c},\mathcal{G}(\boldsymbol{c},\boldsymbol{z}))\big]^2$$

and $\mathcal{L}_{\text{cGAN}}(\mathcal{G},\mathcal{D}) = \mathbb{E}_{\boldsymbol{c}\sim p_{\text{data}}(\boldsymbol{c}),\boldsymbol{z}\sim p_z(\boldsymbol{z})}\big[1-\mathcal{D}(\boldsymbol{c},\mathcal{G}(\boldsymbol{c},\boldsymbol{z}))\big]^2$. In our case, the input $\boldsymbol{c}$ is the sparse PALM image $\boldsymbol{S}$ combined with the upsampled version

$\boldsymbol{L} = \mathcal{B}(\boldsymbol{W})$ of the widefield image $\boldsymbol{W}$, where $\mathcal{B}$ denotes bilinear interpolation. Note that in practice the noise $\boldsymbol{z}$ in our ANN was introduced only through the use of dropout layers, as in the pix2pix implementation[20]. Thus, the objective functions are:

$$\mathcal{L}_{\mathcal{D}}(\mathcal{D}) = \mathbb{E}_{\boldsymbol{S},\boldsymbol{T},\boldsymbol{W}\sim p_{\text{data}}(\boldsymbol{S},\boldsymbol{T},\boldsymbol{W})}\left(\mathcal{D}(\mathcal{B}(\boldsymbol{W}),\boldsymbol{S},\boldsymbol{T})-1\right)^2 + \mathbb{E}_{z\sim p_z(z),\boldsymbol{S},\boldsymbol{W},\boldsymbol{M}\sim p_{\text{data}}(\boldsymbol{S},\boldsymbol{W},\boldsymbol{M})}\left(\mathcal{D}(\mathcal{B}(\boldsymbol{W}),\boldsymbol{S},\mathcal{G}(\boldsymbol{S},\boldsymbol{W},\boldsymbol{M}))\right)^2$$

and:

$$\mathcal{L}_{\text{cGAN}}(\mathcal{G},\mathcal{D}) = \mathbb{E}_{(\boldsymbol{S},\boldsymbol{W},\boldsymbol{M})\sim p_{\text{data}}(\boldsymbol{S},\boldsymbol{W},\boldsymbol{M})}\left(\mathcal{D}(\mathcal{B}(\boldsymbol{W}),\boldsymbol{S},\mathcal{G}(\boldsymbol{S},\boldsymbol{W},\boldsymbol{M}))-1\right)^2$$

In the end, combining the three loss terms described above, we implemented the following optimization problem:

$$\begin{cases} Q^* = \arg\min_Q \mathcal{L}_Q(Q) \\ \mathcal{D}^* = \arg\max_{\mathcal{D}} \mathcal{L}_{\mathcal{D}}(\mathcal{D}) \\ \mathcal{G}^* = \arg\min_{\mathcal{G}}\left[\alpha\mathcal{L}_{\text{SuperRes}}(\mathcal{G}) + \beta\mathcal{L}_{\text{LowRes}}(\mathcal{G},Q) + \gamma\mathcal{L}_{\text{cGAN}}(\mathcal{G},\mathcal{D})\right] \end{cases}$$

The weights α, β, and γ are hyperparameters, which we set manually to $\alpha = 50$, $\beta = 25$, and $\gamma = 1$ for most experiments. In absence of widefield images $\boldsymbol{W}$, β was simply set to zero. The reported results are not very sensitive to these parameters.

We trained the A-net end-to-end using stochastic gradient descent (SGD) with Adam[53] and a batch size of 1 with 200,000 or more iterations (back-propagation steps). Our implementation was adapted from affinelayer's TensorFlow[54] implementation, which is ported from the Torch implementation of pix2pix[20]. Both network training and inference were performed on Tesla P100, Tesla M40, Tesla K80, or GTX Titan XP graphical processing units (GPUs) from Nvidia. A-net training from scratch typically takes from hours to days on a single GPU. Once trained, the A-net takes only ~1 second or less to reconstruct a super-resolution image of 2,560 × 2,560 pixels (corresponding to an entire FoV). Training time could be further reduced by pretraining (or transfer learning), use of GPU clusters, or optimized data augmentation.

Experimental training images and data augmentation. Experimental training data are obtained from standard localization microscopy data (dense PALM images). To achieve good performance, ANNs generally necessitate large amounts of training data. However, ANNA-PALM typically requires PALM images from no more than 10 FoVs (of 55 µm × 55 µm each) and can even be trained with a single FoV. This is possible thanks to an extensive on-the-fly data augmentation strategy, as described below. Each of the dense PALM training images corresponds to a list of localizations $(f_i, x_i, y_i)_{i=1..n}$, where $f_i \in [1,K]$ is the index of the diffraction-limited frame from which localization (x_i, y_i) originates, and K is the total frame number. PALM images are obtained as plain 2D histograms of these localizations with typical pixel sizes of 10–20 nm. From each list of localizations (corresponding to a dense PALM image of a single FoV), we generate 10–30 pairs of input and target images $(\boldsymbol{S},\boldsymbol{T})$ for training. To define the target image $\boldsymbol{T}$, we take a random consecutive subset $[k_T, k_T + 0.95K]$ of 95% of all available K frames (k_T is chosen randomly between 0 and $0.05K$) and create the 2D histogram image based on localizations from those frames only, i.e. from all (x_j, y_j) such that $f_j \in \left[k_T, k_T + 0.95K\right]$. To define the sparse input image $\boldsymbol{S}$, we take random subsets of 300–500 consecutive frames from the first half of the image sequence and similarly create a 2D histogram of the localizations from those frames only. When a widefield image $\boldsymbol{W}$ is available, this image must first be aligned with the corresponding dense PALM image $\boldsymbol{T}$. This is done using an FFT-based phase correlation algorithm[55] after histogram equalization of image $\boldsymbol{T}$, smoothing by convolution with a Gaussian kernel of standard deviation 6 pixels, and resizing with bilinear interpolation to the same size as image $\boldsymbol{W}$. The registered widefield image $\boldsymbol{W}$ is then scaled up using bilinear interpolation to an image $\boldsymbol{L} = \mathcal{B}(\boldsymbol{W})$ with the same size as image $\boldsymbol{T}$.

During training, for each iteration of SGD, we crop the images $\boldsymbol{S}$, $\boldsymbol{L}$ and $\boldsymbol{T}$ with a randomly placed 712 × 712 pixel-sized region $R = [x_{\min}, x_{\min} + 712] \times [y_{\min}, y_{\min} + 712]$. We then use random geometric transformations and apply them identically to the three images. Specifically, we rotate the images by a random angle between 0 and 360 degrees, apply elastic transformations[56], and then crop the center region of size 512 × 512 pixels. In addition to geometric transformations, we also introduce realistic noise from experimental background images. This is done by manually outlining regions of background in selected PALM training images, splitting these regions into small patches of 40 × 40 pixels, grouping them according to their summed pixel values, then assembling them into a larger image with the same size as $\boldsymbol{T}$. During training, these semi-synthetic noise images are randomly selected and added to the input image $\boldsymbol{S}$ without altering $\boldsymbol{T}$ or $\boldsymbol{W}$. Finally, we normalize the input image $\boldsymbol{S}$ by subtracting its mean and dividing by the standard deviation. If a widefield image $\boldsymbol{W}$ is provided, its pixel values are scaled to a minimum of 0 and a maximum of 1. Otherwise, $\boldsymbol{W}$ is replaced by an image containing zeros only. The target image $\boldsymbol{T}$ is truncated at a maximum value of 255 and then scaled to have a minimum of 0 and a maximum of 1. For the switch $\boldsymbol{M}$, we used an integer number to define the type of training images (e.g., 0 for microtubules, 1 for nucleoporins, 2 for mitochondria). When training on different types of images (e.g., microtubules and nucleoporins; **Supplementary Fig. 13**), we assign the corresponding switch value to $\boldsymbol{M}$ and use it as additional input to the A-net together with images $\boldsymbol{S}$ and $\boldsymbol{W}$, as described in **Supplementary Note 1**.

Image simulations. Our procedure to simulate localization microscopy (PALM) images of microtubules is illustrated in **Supplementary Figure 1**. To simulate microtubule filaments, we used a Langevin dynamics simulation[28] that generates random configurations of semiflexible curves with a specified rigidity (persistence length), starting from a random initial configuration (**Supplementary Fig. 1a**,**b**). The initial configurations were generated with a Python library named cpolymer and the Langevin dynamics was implemented using the molecular dynamics code LAMMPS[57]. Although the simulation generates 3D polymer chains, we only considered their 2D projections, consisting of N_p connected positions $(x_k, y_k)_{k=1..N_p}$. To obtain smooth filaments, we further interpolated these connected segments using spline functions with the Scipy function scipy.interpolate.splev. Next, we turned these 2D curves into a gray scale image of 800 × 800 pixels, with an assumed pixel size of 7.8 nm, using the Python library Matplotlib. This image was further convolved with a Gaussian kernel of standard deviation 1.5 pixels, resulting in a smooth image $I(i,j)$ as shown in **Supplementary Figure 1c**, and normalized to a probability density $\left(\sum\sum_{i,j} I(i,j) = 1, \text{with all } I(i,j) \geq 0\right)$. This image was used to mimic a "perfect" PALM image of filaments corresponding to an infinite number of localizations ($n = \infty$). Such perfect images were used as targets during ANN training for simulated data and defined as ground truth for the quantification of reconstruction quality by MS-SSIM (**Fig. 2g**). During training, we applied the same rotations and elastic transformations described for experimental data in the previous section.

Localization microscopy images obtained from a finite number of localizations $n < \infty$ (sparse PALM images $\boldsymbol{S}$) can be considered as a sampling of the probability density $I(i,j)$ with n samples. These images can therefore be simulated by applying Poisson noise to a rescaled version of the perfect PALM image, that is: $\boldsymbol{S}(\lambda,I) = \mathcal{P}(\lambda I/I_{\max})$, where $I_{\max}$ is the maximum value of I, $\mathcal{P}(\mu)$ denotes the Poisson probability distribution of mean µ and where the peak parameter λ controls the level of sampling. In order to simulate sparse PALM images for various levels of sampling, we varied the peak value λ following a log-normal distribution where $\ln(\lambda)$ has mean −0.5 and standard deviation 0.001 and applied Poisson noise using the numpy library function random.poisson. An example of a simulated sparse PALM image is shown in **Supplementary Figure 1d**. Besides finite sampling, localization microscopy images are corrupted by additional noise sources such as false detections from background noise due to out-of-focus light or unspecific binding of antibodies. To mimic this, we first created a probability density $I_b = I * G_{\sigma b}$ for the background noise by convolving I with a Gaussian kernel $G_{\sigma b}$ of large standard deviation $\sigma_b = 25$ pixels, and applied Poisson noise with $\lambda = 0.06$. To create training images, we added this background noise image to the sparse PALM image $\boldsymbol{S}$ above (**Supplementary Fig. 2**). We did not add background noise to the test images used during inference (**Fig. 2**). To simulate the widefield images (**Fig. 2a**), we first blurred the perfect PALM image by convolution with a Gaussian kernel of standard deviation 8 pixels, then added Gaussian noise with zero mean and standard deviation chosen randomly between 0.5 and 1.5.

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

doi:10.1038/nbt.4106

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

For simulations of nuclear pore images (**Supplementary Fig. 15**), we applied a similar procedure, except that the perfect PALM images were obtained by randomly distributing circles of diameter 150 nm in the plane (avoiding overlaps) and placing eight Gaussian spots (of standard deviation 1.7 pixels) at equal distance from each other on each circle to mimic the octogonal shape of nuclear pores.

Sample preparation. For microtubule imaging experiments (**Figs. 3–5** and **Supplementary Figs. 8–10**), except those using DNA-PAINT (**Fig. 5j**,**l**), U-373 MG (Uppsala) cells were cultured in Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (DMEM/F12; Gibco), supplemented with 10% (v/v) fetal bovine serum (FBS; Gibco), 1% (v/v) penicillin-streptomycin (Gibco), in a 5% CO_2 environment at 37 °C on 18-mm cleaned coverslips in 12-well plates. 24 hours after plating, cells were pre-extracted for 10 s in 0.25% (v/v) Triton X-100 (Triton) in BRB80 (80 mM PIPES, 1 mM MgCl2, 1 mM EGTA, adjusted to pH 6.8 with KOH) supplemented with 4 mM EGTA, and immediately fixed for 10 min with 0.25% (v/v) Triton + 0.5% glutaraldehyde in BRB80, followed by reduction for 7 min with 0.1% $NaBH_4$ solution in PBS and another washing step in PBS. Cells were directly incubated for 1 h at room temperature in PBS with 1:500 rat alpha-tubulin antibodies (Bio-Rad MCA77G), followed by three washing steps with PBS, and then incubated for 45 min in PBS with 1:500 anti-rat Alexa-647 conjugated secondary antibodies from donkey (Jackson ImmunoResearch Laboratories, ref. 712-605-153).

For the DNA-PAINT experiment on microtubules (**Fig. 5k–m**), U-373 cells stuck on 18 mm diameter coverslips were fixed at 37 °C with 4% PFA and 0.25% glutaraldehyde in PHEM buffer and permeabilized in 0.3 % Triton X-100, followed by reduction in a 0.1% $NaBH_4$ solution as above. Next, cells were incubated for 1 h with 1:500 primary mouse antibodies against alpha-tubulin. The sample was washed three times in PBS, then incubated with 1:100 anti-mouse oligo-conjugated antibodies from Ultivue Kit 2 for DNA-PAINT imaging[32]. After washing the sample three times in PBS, and just before imaging, 2 nM of complementary oligos coupled to Cy3 fluorophores were added to the sample.

Nuclear pore imaging data of gp210 and WGA (**Fig. 6a–c** and **Supplementary Fig. 14**) were kindly provided by J. Sellés and O. Falklaris (Institut Jacques Monod, Paris) and obtained from nuclear membranes of *Xenopus* frog eggs prepared as described previously[35].

For mitochondria imaging experiments (**Fig. 6e**,**f**), COS7 cells were cultured under the same conditions as U-373 cells above using phenol-red free DMEM medium and fixed with 4% PFA in PBS for 10 min. The sample was blocked with 3% BSA in PBS for 20 min and immunostained with 1:500 mouse antibodies against TOM22 (Sigma, ref. T6319) in wash buffer (PBS with 0.5% BSA) for 1 h. After extensive washing with wash buffer, the sample was incubated with 1:500 anti-mouse secondary antibodies from donkey conjugated to Alexa-647 dyes (Jackson ImmunoResearch Laboratories, ref. 715-605-151) in wash buffer for 30 min. After washing five times with wash buffer and two times with PBS, samples were post-fixed with 2% PFA in PBS for 10 min and washed five times with PBS.

For all localization microscopy experiments except DNA-PAINT, we used a photoswitching buffer[58] composed of 50 mM Tris-HCl + 10 mM NaCl + 10% (w/v) glucose + 168 AU/mL glucose-oxidase + 1,404 AU/mL catalase + 1% 2-mercaptoethanol. For microtubule imaging experiments, we used this buffer to fill a square hole that was manually cut in a parafilm sheet, which was deposited on a rectangular coverglass. The round coverslips were sealed with nail polish.

Image acquisition in localization and high-throughput microscopy. We performed single-molecule localization microscopy experiments (PALM/STORM and DNA-PAINT) on custom built microscopy systems, as previously described[58–60]. The system used for PALM/STORM imaging of microtubules is based on an inverted microscope body (Nikon Ti Eclipse) equipped with either a 60× 1.49 NA oil immersion objective (Nikon) or a 60× 1.2 NA water immersion objective (Nikon) and with the Perfect Focus System active. A 642-nm wavelength laser with 500 mW power was used to excite Alexa-647 fluorophores and an AOTF (AA Opto Electronic) was used to modulate laser excitation. Sequences of diffraction-limited single-molecule image frames were acquired either on a sCMOS camera (Hamamatsu ORCA-Flash4.0), which can capture images of 2,048 × 2,048 pixels (for **Figs. 4** and **5a–i**), or on an EMCCD (Andor IXON ULTRA 897) with 512 × 512 pixels (for **Figs. 3** and **6d–f**). Both cameras were controlled by MicroManager software[61]. For experiments using the sCMOS camera, the effective pixel size was 108 nm and we used a 512 × 512 pixels region of interest, which resulted in an imaged FoV of 55.3 µm × 55.3 µm. For experiments using the EMCCD camera, we used a 2× telescope and the effective pixel size was 107 nm, resulting in a FoV of 54.8 µm × 54.8 µm. The exposure time was set to Δt = 10 ms or 30 ms per frame. The number of frames acquired ranged from k = 1,000 (**Fig. 4**) to K=60,000 (e.g., **Fig. 5c**) per FoV.

For the DNA-PAINT experiment (**Fig. 5j**,**l**), we used an inverted Nikon Ti-E Eclipse microscope equipped with a 100× 1.49 NA TIRF objective (Nikon) and with the Perfect Focus System active. A 561-nm wavelength laser with 500 mW power was used to excite Cy3 dyes. Highly inclined laser illumination was used to reduce out-of-focus background signal. Images were acquired on an EMCCD camera as above, with a 1.5× telescope, resulting in an effective pixel size of 106 nm and a FoV of 54 µm × 54 µm. The sample was mounted in a magnetic sample holder filled with the imaging buffer provided in the Ultivue kit. Exposure time was set to Δt = 30 ms and the EM gain of the EMCCD was set to 300. The laser power was increased until isolated fluorescent spots were observed. For the experiment shown in **Figure 5g**,**i**, K = 60,000 frames were acquired.

The *Xenopus* nuclear pore data (**Fig. 6a–c** and **Supplementary Fig. 14**), were acquired on a Zeiss Elyra P.S.1 microscope as described previously[35].

For high-throughput imaging of microtubules (**Fig. 4**), we used the Multi-Dimensional Acquisition tool in Micro-manager to define the positions of 1,089 FoVs of 55.3 µm × 55.3 µm on a 33 × 33 grid, with overlaps of 1 µm; the stage was automatically shifted to each of these 1,089 positions. We first acquired only widefield images, taking five frames at each of these positions (the first two were ignored because of motion blur), in a total acquisition time of 12 minutes. Then, the laser power was raised to bleach out preactivated molecules and k =1,000 frames of single-molecule images were acquired at each of the 1,089 positions, in a total acquisition time of 3 hours and 8 minutes. Raw image frames were written directly to a remote storage server via Samba networking protocol.

Localization microscopy image analysis. The input to ANNA-PALM reconstruction is a localization image, defined as a 2D histogram of n single-molecule positions $(x_i, y_i)_{i\,=\,1..n}$. The histogram bin, that is, the pixel size of the localization image, was set to 7.8 nm for the simulated data (**Fig. 2** and **Supplementary Figs. 1–6**,**15**) and 20 nm for the experimental data (**Figs. 3–6** and **Supplementary Figs. 7–14**). The positions $(x_i, y_i)_{i\,=\,1..n}$ were obtained by analyzing sequences of diffraction-limited frames using standard single-molecule localization algorithms. For experimental microtubule images, we used the ThunderSTORM[62] plugin of ImageJ, applying wavelet filters for detection and weighted least squares Gaussian fitting for precise estimation of subpixelic positions. We used the cross-correlation feature in ThunderSTORM for drift correction, and filtered out the least certain localizations based on the fitted Gaussian's standard deviation and the χ^2 of the residual. Localizations in consecutive frames separated by less than 20 nm were assumed to originate from the same molecule and merged into a single localization. The final number n of localizations was ~7 million for the full 55 µm x 55 µm FoV of the images shown in **Figure 3** and **Supplementary Figures 8** and **9** (obtained from K = 30,000 frames). For the high-throughput experiment (**Fig. 4** and **Supplementary Fig. 12**), the number of localizations per 55 µm × 55 µm FoV ranged from n = 2,949 to n = 1,442,048 with an average $\langle n \rangle$ = 610,983 and standard deviation $\sigma(n)$ = 273,606. The total number of localizations across all 1,089 FoVs was ≈665 million. ThunderSTORM analyses were performed either on high-end workstations or on Institut Pasteur's high-performance computer (HPC) cluster. For the high-throughput experiments, we used Python scripts to run ThunderSTORM in batch mode (without user intervention) on the HPC cluster and assembled mosaic images (**Fig. 4a**,**b** and **Supplementary Fig. 12a**,**b**) using a stitching plugin of ImageJ[63].

The nuclear pore images were analyzed using the ZEN software from Zeiss as previously described[35]. For the DNA-PAINT experiments, we used PALMTT, a modified version of the single-molecule tracking algorithm MTT[64], based on Matlab (Mathworks). This algorithm uses Gaussian

smoothing and thresholding for detection, and Gaussian fitting for precise estimation of subpixelic positions. Drift correction was performed computationally by tracking fluorescent beads used as fiducial markers.

Quality metrics and sampling resolution. In order to quantitatively assess the quality of PALM images and ANNA-PALM reconstructions, we calculated the multiscale structural similarity index (MS-SSIM) between either image and the ground truth (**Fig. 2g** and **Supplementary Figs. 5**,**9**). For the simulated data, the ground truth was simply defined as the "perfect" PALM image, corresponding to an infinite number of localizations (see 'Image simulations' above, **Fig. 2e** and **Supplementary Fig. 1c**). For the experimental data, the ground truth was defined as the ANNA-PALM reconstruction of a dense PALM image obtained from all available frames (e.g., **Supplementary Fig. 9c**). Before calculation of the MS-SSIM, all simulated images were linearly normalized without clipping to a maximum value of 255.

In order to evaluate the effect of sampling on the resolution of PALM images, we computed the double mean distance, $R_{\mathrm{Nyq}} = 2<d>$ between nearest neighbors in the underlying sets of localizations[48] as a function of localization number (**Supplementary Fig. 3d**). For the simulated sparse PALM data, sets of localizations were obtained by interpreting each image $\boldsymbol{S}$ as a 2D histogram of localizations, and creating a random subpixelic position $(x_k, y_k)_{k=1..m}$ within each pixel (i,j), as many times as given by the pixel value $m = S(i,j)$ (therefore resulting in a set of $n = \sum_{i,j} S(i,j)$ localizations). The quantity $R_{\mathrm{Nyq}}(n) = 2<d(n)>$ decreases towards zero with increasing number of localizations, n. A Nyquist criterion introduced in ref. 48 suggests that the resolution is limited by sampling to no less than R_{Nyq}, implying that at least $n \geq N_{\mathrm{Nyq}} = R_{\mathrm{Nyq}}^{-1}(R)$ localizations are needed to achieve a given resolution R. However, a more stringent and realistic criterion[3] prescribes a fivefold larger number of localizations to reach resolution R: $n \geq N_{5\times\mathrm{Nyq}} = 5 \times N_{\mathrm{Nyq}}$, implying that the sampling limit to resolution is $R_{5\times\mathrm{Nyq}} = 2<d(n/5)>$. Accordingly, if this condition is met, the resolution is no longer limited by sampling, but only by the localization precision, $R_{\mathrm{loc}} \approx 2.3\sigma_{\mathrm{loc}}$, where σ_{loc} is the standard deviation of localization errors along each coordinate. In general, the resolution of a PALM image, as limited by both sampling and localization precision, can be written: $R = \max\left(R_{5\times\mathrm{Nyq}}, R_{\mathrm{loc}}\right) = \max\left(2<d(n/5)>, 2.3\sigma_{\mathrm{loc}}\right)$.

Life Sciences Reporting Summary. Further information is available in the Life Sciences Reporting Summary.

Code availability. The source code of ANNA-PALM is available from https://annapalm.pasteur.fr/.

Data availability. The localization data used in this paper can be downloaded directly from https://annapalm.pasteur.fr/.

50. Xu, L., Choy, C.S. & Li, Y.W. in *2016* IEEE *International Workshop on Acoustic Signal Enhancement, IWAENC 2016* (IEEE, 2016) doi:10.1109/IWAENC.2016.7602891.
51. Maas, A.L., Hannun, A.Y. & Ng, A.Y. Rectifier nonlinearities improve neural network acoustic models. *Proc. 30th Int. Conf. Mach. Learn.* **30**, 3 (2013).
52. Mao, X. *et al.* Least squares generative adversarial networks. *ICCV* 2794–2802 (2017) doi:10.1109/ICCV.2017.304.
53. Kingma, D.P. & Ba, J. Adam: a method for stochastic optimization. *ICLR* 1–15. http://doi.acm.org.ezproxy.lib.ucf.edu/10.1145/1830483.1830503 (2015).
54. Rampasek, L. & Goldenberg, A. TensorFlow: biology's gateway to deep learning? *Cell Syst.* **2**, 12–14 (2016).
55. Reddy, B.S. & Chatterji, B.N. An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE Trans. Image Process.* **5**, 1266–1271 (1996).
56. Simard, P.Y., Steinkraus, D. & Platt, J.C. Best practices for convolutional neural networks applied to visual document analysis. *Seventh Int. Conf. Doc. Anal. Recognition, 2003. Proceedings* **1**, 958–963 (2003).
57. Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).
58. van de Linde, S. *et al.* Direct stochastic optical reconstruction microscopy with standard fluorescent probes. *Nat. Protoc.* **6**, 991–1009 (2011).
59. Henriques, R. *et al.* QuickPALM: 3D real-time photoactivation nanoscopy image processing in ImageJ. *Nat. Methods* **7**, 339–340 (2010).
60. Lelek, M. *et al.* Superresolution imaging of HIV in infected cells with FlAsH-PALM. *Proc. Natl. Acad. Sci. USA* **109**, 8564–8569 (2012).
61. Edelstein, A., Amodaj, N., Hoover, K., Vale, R. & Stuurman, N. Computer control of microscopes using μmanager. *Curr. Protoc. Mol. Biol.* **92**, 14.20.1–14.20.17 (2010).
62. Ovesný, M., Křížek, P., Borkovec, J., Svindrych, Z. & Hagen, G.M. ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. *Bioinformatics* **30**, 2389–2390 (2014).
63. Preibisch, S., Saalfeld, S. & Tomancak, P. Globally optimal stitching of tiled 3D microscopic image acquisitions. *Bioinformatics* **25**, 1463–1465 (2009).
64. Sergé, A., Bertaux, N., Rigneault, H. & Marguet, D. Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. *Nat. Methods* **5**, 687–694 (2008).

© 2018 Nature America, Inc., part of Springer Nature. All rights reserved.

doi:10.1038/nbt.4106

Supplementary Information
for
Deep learning massively accelerates super-resolution localization microscopy

Wei Ouyang[1,2,3], Andrey Aristov[1,2,3], Mickaël Lelek[1,2,3], Xian Hao[1,2,3], Christophe Zimmer[1,2,3]

[1] Institut Pasteur, Unité Imagerie et Modélisation, Paris, France
[2] UMR 3691, CNRS, Paris, France
[3] C3BI, USR 3756, IP CNRS, Paris, France

Supplementary Note 1 : Neural network architecture—A-net 2
Supplementary Table 1 : Architecture of generator network..... 4
Supplementary Table 2 : Architecture of low-resolution estimator network 5
Supplementary Table 3 : Architecture of cGAN discriminator network 5
Supplementary Table 4 : Neural network layers and operators 6
Supplementary References :..... 7
Supplementary Figure 1 : Simulating dense and sparse PALM images of microtubules 8
Supplementary Figure 2 : Simulated PALM training images of microtubules..... 9
Supplementary Figure 3 : Sampling and resolution in simulated PALM images ... 10
Supplementary Figure 4 : ANNA-PALM reconstructions for simulated images of varying sparsity..... 11
Supplementary Figure 5 : Reconstruction quality for different ground truths..... 12
Supplementary Figure 6 : ANNA-PALM error maps for simulated data..... 13
Supplementary Figure 7: Resolution of PALM and ANNA-PALM in experiments.. 14
Supplementary Figure 8: ANNA-PALM reconstructions of experimental images for varying frame numbers 15
Supplementary Figure 9: Quantifying ANNA-PALM reconstruction quality in experiments. 16
Supplementary Figure 10: ANNA-PALM reconstructions of immunolabeled microtubules. 17
Supplementary Figure 11: ANNA-PALM error maps for experimental images..... 20
Supplementary Figure 12: ANNA-PALM reconstruction of a large field from widefield images 21
Supplementary Figure 13: Reconstructing microtubule and nuclear pore images with a single ANN and a switch..... 22
Supplementary Figure 14: Identifying and counting nuclear pores in ANNA-PALM and PALM images 23
Supplementary Figure 15: ANNA-PALM reconstructions and error maps for mismatch between training and testing data..... 24

Nature Biotechnology: doi:10.1038/nbt.4106

Supplementary Note 1: Neural network architecture—A-net

As explained in the text and Online Methods, ANNA-PALM employs a convolutional neural network (CNN) based on the pix2pix[1] network, which is a special conditional generative adversarial neural network (cGAN)[2] that uses a U-net[3] as generator network. Our neural network architecture (which we refer to as "A-net", version 1, or A-net(V1)) contains three individual networks: (i) the generator network $\mathcal{G}$, (ii) the low-resolution estimator network Q and (iii) the cGAN discriminator network $\mathcal{D}$ (see **Figure 1a**). Below, and in **Supplementary Tables 1-3**, we specify the architecture of each network in more detail, with the different types of layers and operations listed in **Supplementary Table 4**.

The generator network $\mathcal{G}$ architecture is similar to the pix2pix[1] and U-net[3] architecture and consists of two parts (**Supplementary Table 1**): an encoder network (**Supplementary Table 1**, upper half) and a decoder network (**Supplementary Table 1**, lower part). The encoder network contains multiple convolutional layers (labeled 'Conv'; see **Supplementary Table 4** for a definition of each layer) stacked on top of each other, resulting in feature maps of halving size and doubling number. The encoder takes as input two images, namely the sparse super-resolution image $\boldsymbol{S}$ (labeled 'INPUT_SR') and the low-resolution image $\boldsymbol{W}$ (labeled 'INPUT_LR'), plus the switch $\boldsymbol{M}$ (labeled 'SWITCH'). The low-resolution image INPUT_LR is upsampled using bilinear interpolation (labeled 'RESIZE') and then concatenated ('Cat') with the sparse super-resolution image INPUT_SR as a second input channel. The switch value $\boldsymbol{M}$ is represented using 1-hot encoding (labeled '1-HOT') and then coupled to the encoder via convolution layers with kernel size 1x1 (labeled 'Conv1'). The decoder network contains multiple upsampling convolutional layers (labeled 'UpConv'), resulting in feature maps of doubling size and halving number. The last layer of the decoder network (**Supplementary Table 1**, bottom) produces an output image of the same size as the input image fed to the encoder network (**Supplementary Table 1**, top). Unlike the U-net[3], but as in the pix2pix network[1], we used convolution layers with a stride of 2 instead of max pooling layers, and also replaced rectified linear unit layers (labeled 'ReLU') in the encoder by "leaky ReLUs"[4] (labelled 'LReLU'). The last layer of the decoder uses a hyperbolic tangent activation function ('Tanh'). We introduce batch normalization[5] ('BN') after all convolutional layers and dropout layers[6] ('Dropout') (with dropout probability p=0.5) in the second, third and fourth layers of

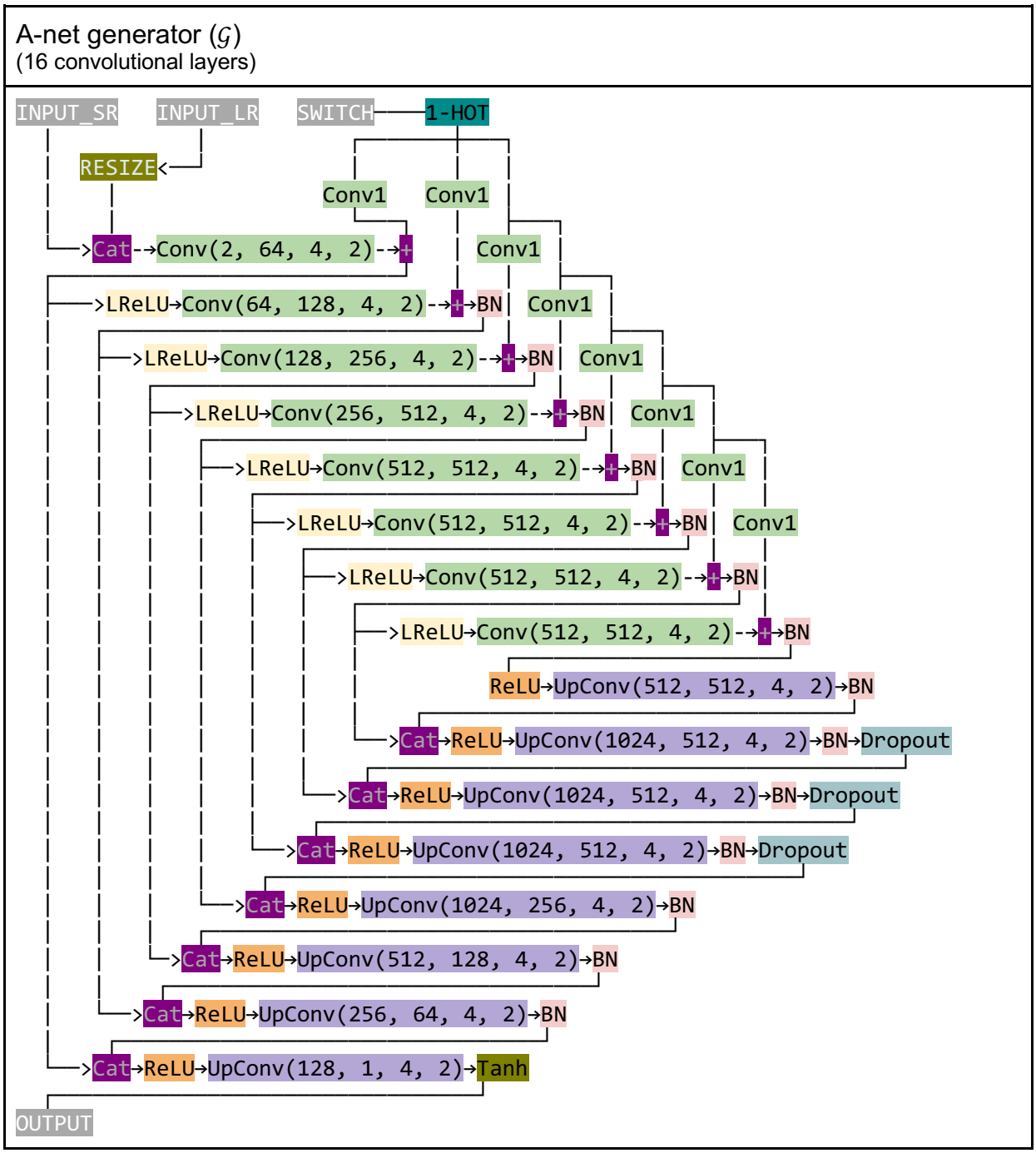
Nature Biotechnology: doi:10.1038/nbt.4106

decoder. As in the U-net[3], the generator contains direct ("skip-layer") connections between feature maps located symmetrically in the encoder and decoder. These feature maps, which have identical size, are concatenated in the corresponding layers (labelled 'Cat') of the decoder. The skip-layer connections allow the network to combine local, pixel-wise information, with the larger context information that is propagated through consecutive convolutional layers. We called our network 'A-net' because the encoder and decoder network with skip-connections evoke the letter 'A'.

The low-resolution estimator network Q (**Supplementary Table 2**) has an architecture consisting of 4 convolutional layers ('Conv') and 2 max-pooling layers ('MaxPool'). Dropout layers ('Dropout') and leaky ReLUs ('LReLU') are used after each convolutional layer except for the last layer, where we used a sigmoid activation function ('Sigmoid'). The low-resolution estimator Q takes a super-resolution image as input (INPUT_LD) and outputs the corresponding low resolution image (OUTPUT_LD), as also described in Online Methods.

Our cGAN discriminator network $\mathcal{D}$ (**Supplementary Table 3**) has an architecture identical to the discriminator network in the pix2pix[1] network, except for the last layer, where we removed the sigmoid function because we replaced the logarithmic losses in the objective function by least squares losses[7] (see Online Methods). The discriminator $\mathcal{D}$ takes as input three images ($\boldsymbol{S}$, $\mathcal{G}(\boldsymbol{S})$ or $\boldsymbol{T}$, $\boldsymbol{W}$) concatenated into a 3-channel image (labelled 'INPUT_GD'). The output of the discriminator is a small image (labelled 'OUTPUT_GD'), whose pixels correspond to image patches in $\boldsymbol{S}$ and $\mathcal{G}(\boldsymbol{S})$ or $\boldsymbol{T}$. These pixel values are large when the discriminator believes that the corresponding patch in $\mathcal{G}(\boldsymbol{S})$ or $\boldsymbol{T}$ is an experimentally obtained image, and small when it believes that it was synthesized by the generator network.

Nature Biotechnology: doi:10.1038/nbt.4106

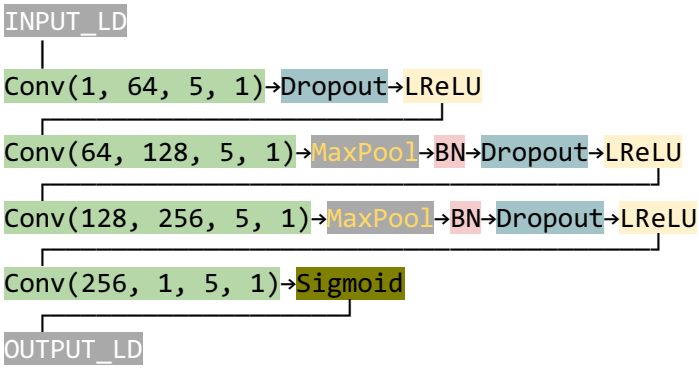


Supplementary Table 1: Architecture of generator network

This table specifies the architecture of the generator network $\mathcal{G}$ used in A-net(V1). Each line corresponds to a distinct layer in the neural network. Colors are used to distinguish different types of operations, as indicated in **Supplementary Table 4**. Solid and dashed black lines show connections between layers. The inputs of the network are: a sparse PALM image $\boldsymbol{S}$ (INPUT_SR), a low resolution image $\boldsymbol{W}$ (INPUT_LR), and a switch value $\boldsymbol{M}$ (SWITCH). The output $\mathcal{G}(\boldsymbol{S}, \boldsymbol{W}, \boldsymbol{M})$ is a reconstructed ANNA-PALM image (OUTPUT) of the same size as the sparse PALM image INPUT_SR.

Nature Biotechnology: doi:10.1038/nbt.4106

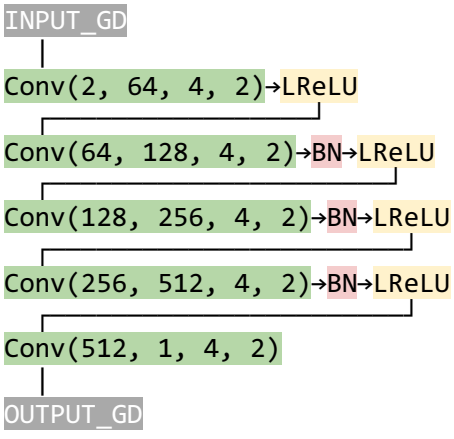
A-net low-resolution estimator (Q)
(4 convolutional layers)



Supplementary Table 2: Architecture of low-resolution estimator network

This table specifies the architecture of the low-resolution estimator network Q used in A-net(V1). This network takes as input (INPUT_LD) either an experimentally acquired dense PALM image $\boldsymbol{T}$ or the generator output $\mathcal{G}(\boldsymbol{S},\boldsymbol{W},\boldsymbol{M})$. The output of the network (OUTPUT_LD) is a low resolution image which is 4 times smaller than the input (INPUT_LD) and has the same size as the experimental widefield image $\boldsymbol{W}$. Notice that, if necessary, convolutional layers ('Conv') with max-pooling layer ('MaxPool') can be added or removed to account for different size ratios between the input and output images (i.e. between the PALM images and the widefield images).

A-net cGAN discriminator ($\mathcal{D}$)
(5 convolutional layers)



Supplementary Table 3: Architecture of cGAN discriminator network

This table specifies the architecture of the discriminator network $\mathcal{D}$ used in A-net(V1). This network takes a 3-channel image (INPUT_GD) as input: the first channel is an upsampled version of low-resolution image $\boldsymbol{W}$; the second channel is a sparse PALM image $\boldsymbol{S}$; the third channel is either the generator output $\mathcal{G}(\boldsymbol{S},\boldsymbol{W},\boldsymbol{M})$ or a corresponding dense PALM image $\boldsymbol{T}$. The output of the discriminator network is a smaller image, where each pixel corresponds to a patch of the input images. The value of each pixel indicates whether the discriminator believes that the corresponding patch in the third input image channel is an experimentally acquired real dense PALM image $\boldsymbol{T}$ (high value) or the generator output $\mathcal{G}(\boldsymbol{S},\boldsymbol{W},\boldsymbol{M})$ (low value).

Nature Biotechnology: doi:10.1038/nbt.4106

Layer name	Layer type	Arguments or comment	Reference
`Conv(a, b, c, d)`	2D Convolution	a: input feature map number b: output feature map number c: kernel size (k=4 means we use a 4x4 convolution kernel) d: stride (d=2 reduces output size by half)	8
`Conv1`	2D Convolution	a special case of 2D convolution with a 1x1 convolution kernel and stride=1 input size matches the one-hot vector encoding the switch output size matches the feature maps of the `Conv` layer output it is added to (see **Supplementary Table 1**)	8
`UpConv(a, b, c, d)`	Up-convolution	a: input feature map number b: output feature map number c: kernel size d: stride (d=2 doubles output size) Nearest neighbor up-sampling followed by a convolutional layer with kernel 1x1.	3
`MaxPool`	Max Pooling	2x2 max-pooling with stride=2	8
`RESIZE`	2D Up-sampling	rescales the input image using bilinear interpolation	
`1-HOT`	One-hot Encoding	encodes an integer number using one-hot representation	
`Cat`	Concatenation	concatenates two groups of input feature maps along channel or feature dimension	3
`+`	Addition	add two groups of input feature maps aligned by channel.	
`ReLU`	Rectified Linear Unit	applies the function $x \to \max(x, 0)$	9
`LReLU`	Leaky Rectified Linear Unit	applies the function $x \to \max(x, 0) + \min(\varepsilon x, 0)$ with ε =0.2	4
`Tanh`	Hyperbolic Tangent	applies the hyperbolic tangent function $x \to \tanh(x)$	
`Sigmoid`	Sigmoid	applies the sigmoid function $x \to (1 + \exp(-x))^{-1}$	
`BN`	Batch Normalization		10
`Dropout`	Dropout layer	with dropout probability p=0.5	6

Supplementary Table 4: Neural network layers and operators

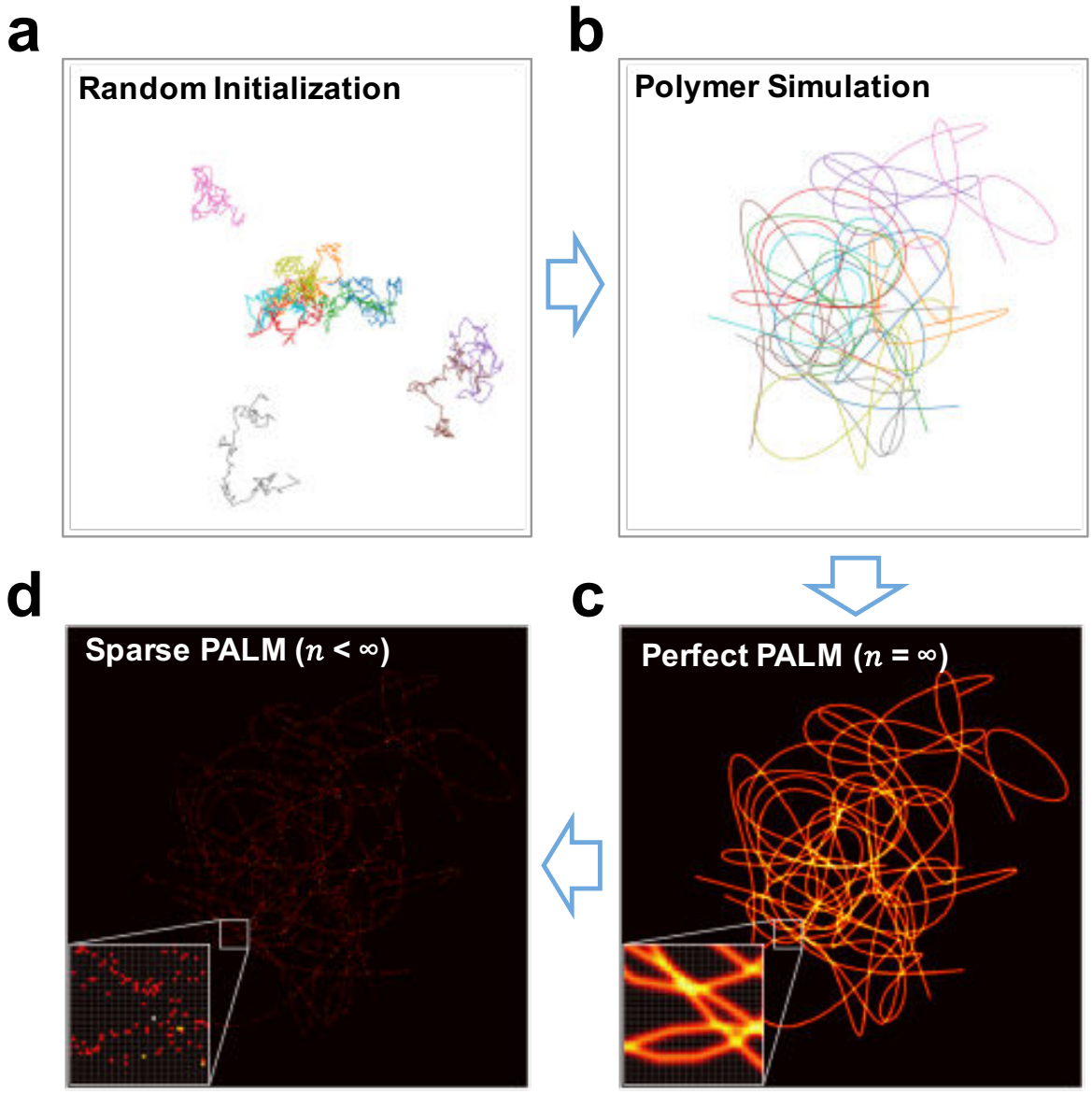
This table lists the 14 types of layers/operators and their parameters used in A-net(V1). The first column indicates the notation used in Supplementary Tables 1-3; the second column indicates the full name of the layer or operator; the third column describes the applied function and/or its arguments; the fourth column provides a reference.

Nature Biotechnology: doi:10.1038/nbt.4106

Supplementary references:

1. Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A. A. Image-to-Image Translation with Conditional Adversarial Networks. *http://arxiv.org/abs/1611.07004* (2016).
2. Goodfellow, I. *et al.* Generative Adversarial Nets. *Adv. Neural Inf. Process. Syst. 27* 2672–2680 (2014). doi:10.1017/CBO9781139058452
3. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science,* **9351,** 234–241 (Springer International Publishing, 2015).
4. Maas, A. L., Hannun, A. Y. & Ng, A. Y. Rectifier Nonlinearities Improve Neural Network Acoustic Models. *Proc. 30 th Int. Conf. Mach. Learn.* (2013).
5. Xu, L., Choy, C. S. & Li, Y. W. Deep sparse rectifier neural networks for speech denoising. in *2016 International Workshop on Acoustic Signal Enhancement, IWAENC 2016* (2016). doi:10.1109/IWAENC.2016.7602891
6. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **15,** (2014).
7. Mao, X. *et al.* Least Squares Generative Adversarial Networks. in *ICCV* 2794–2802 (2017). doi:10.1109/ICCV.2017.304
8. Sermanet, P., Chintala, S. & LeCun, Y. Convolutional neural networks applied to house numbers digit classification. in *ICPR* 3288–3291 (IEEE, 2012).
9. Nair, V. & Hinton, G. E. Rectified Linear Units Improve Restricted Boltzmann Machines. *Proc. 27th Int. Conf. Mach. Learn.* 807–814 (2010). doi:10.1.1.165.6419
10. Ioffe, S. & Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *Icml* **22,** 137–141 (2015).

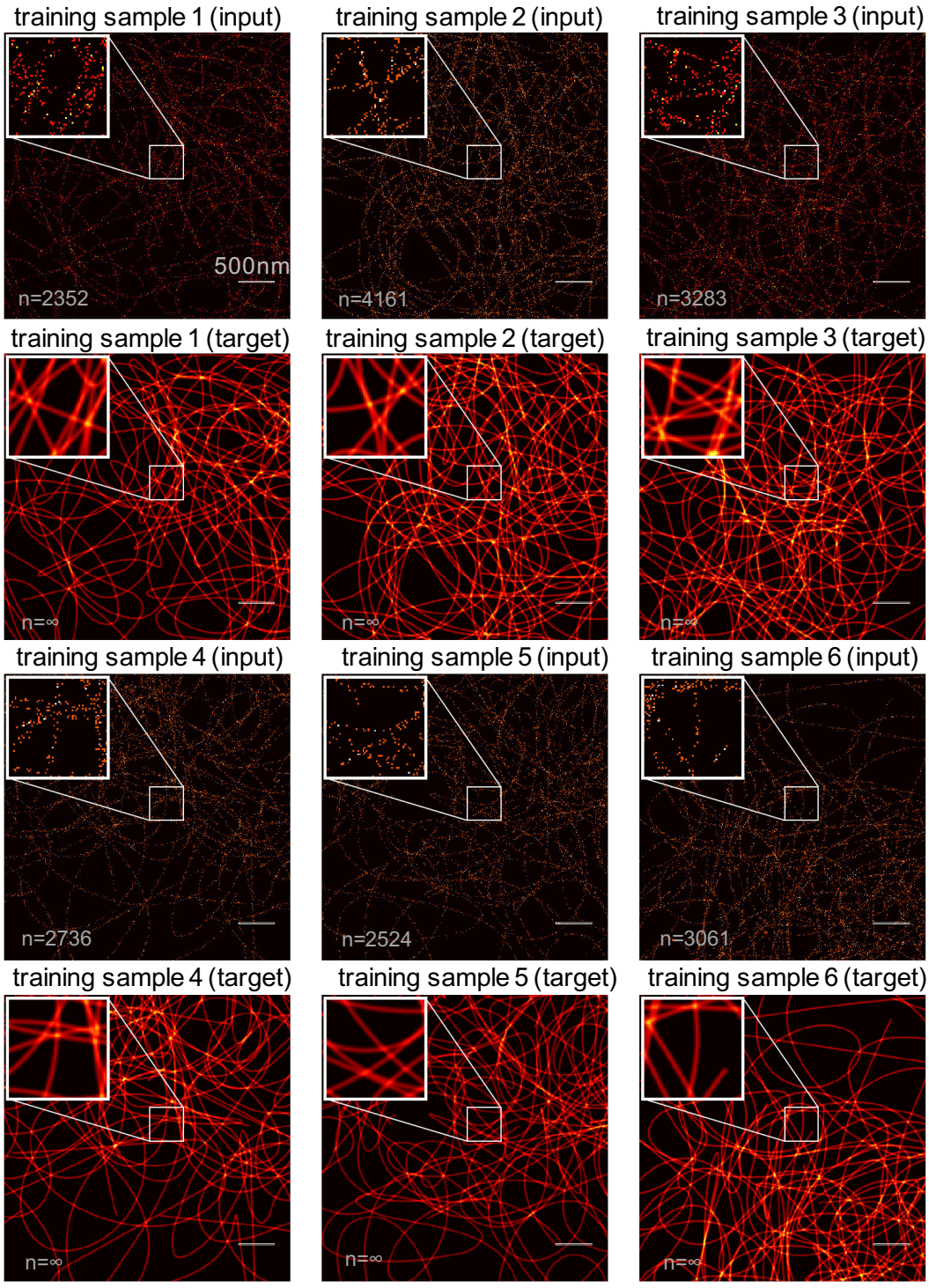
Nature Biotechnology: doi:10.1038/nbt.4106



Supplementary Figure 1 : Simulating dense and sparse PALM images of microtubules

This figure illustrates our procedure to generate synthetic PALM images of microtubules from simulated polymer chains. **a**) Multiple polymer chains are initialized as random walks (each color shows a different polymer chain). **b**) A Langevin dynamics simulation (Arbona *et al. Genome Biol,* 18, 81, 2017) is used to compute the stochastic temporal evolution of these chains. The polymer configurations are shown when the simulation is stopped. **c**) A "perfect" PALM image I is created by convolving the zero-thickness curves in (**b**) with a Gaussian kernel of standard deviation $\sigma = 1.5$ pixel. This image defines the probability density of localization events in a PALM image (with the same pixel size) and can be considered as a dense PALM image with a resolution $\approx 2.3\,\sigma$ obtained from $n = \infty$ localizations. **d**) Poisson noise and additional background noise are applied to this probability density to generate a simulated sparse PALM image with a total number of localizations n equal to the sum of all pixel values. See Online Methods section 'Image simulations' for details.

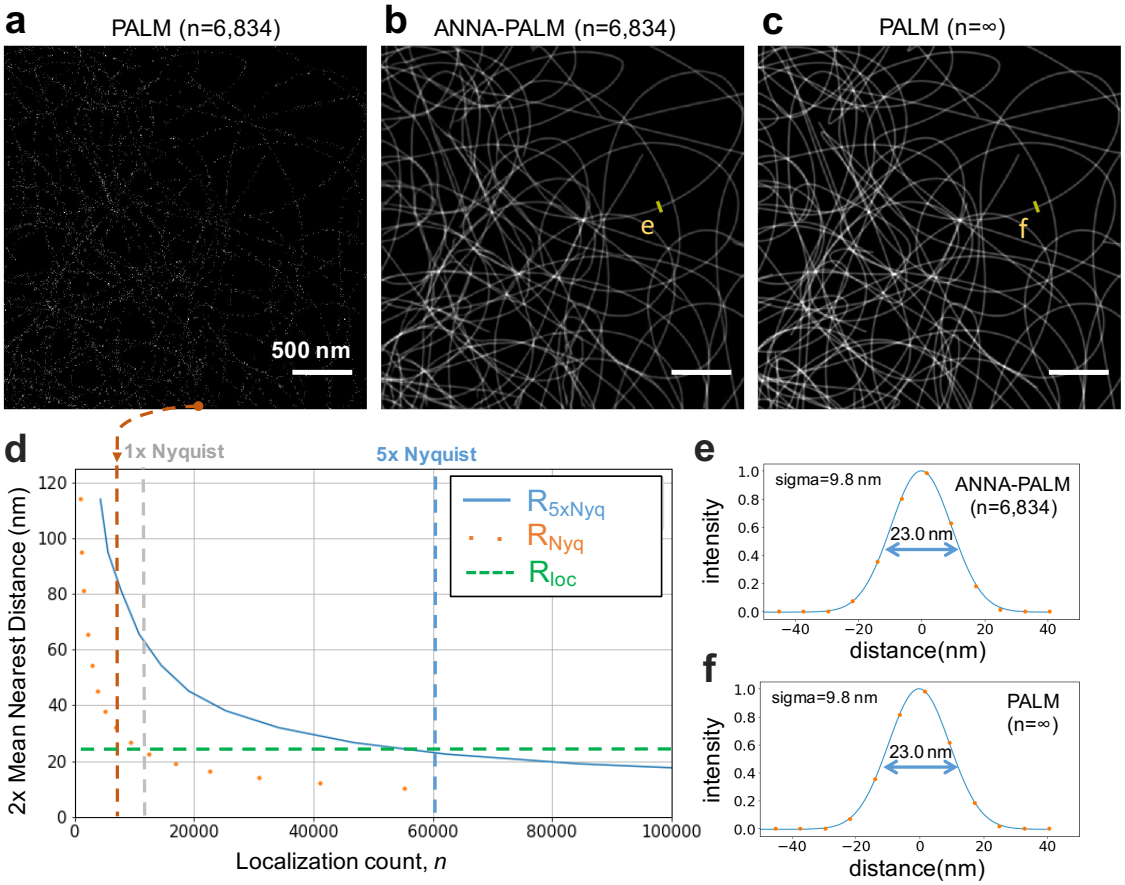
Nature Biotechnology: doi:10.1038/nbt.4106



Supplementary Figure 2 : Simulated PALM training images of microtubules

Examples of the simulated PALM images of microtubules used for ANNA-PALM training (1st and 2nd rows) and validation (3rd and 4th rows), generated as described in **Supplementary Figure 1**. Sparse PALM images with n localizations are shown in the 1st and 3rd rows, and the corresponding perfect PALM images ($n = \infty$) in the 2nd and 4th rows. The perfect PALM images in the training samples are used as training targets. Pixel values are linearly mapped to colors from the look-up table shown in **Figure 3**. Black and white correspond to values V_{min} and V_{max}, respectively, with V_{min}=0 and V_{max}=0.5 for all panels.

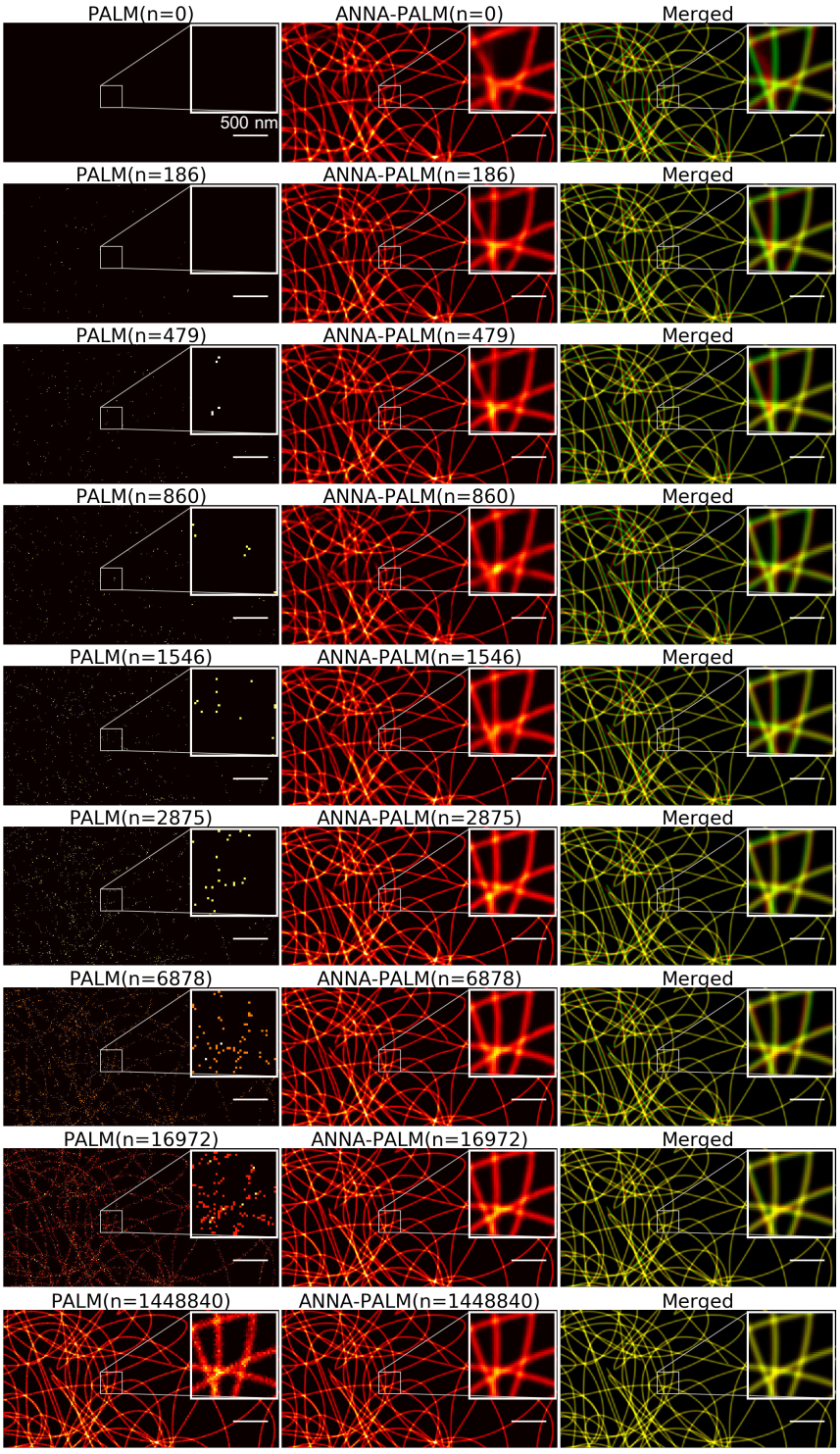
Nature Biotechnology: doi:10.1038/nbt.4106



Supplementary Figure 3 : Sampling and resolution in simulated PALM images

a,b,c) Simulated sparse PALM image of microtubules (**a**), ANNA-PALM reconstruction from the sparse PALM image (**b**) and perfect PALM image with $n = \infty$ localizations (**c**). **d**) Resolution limit due to sampling as function of localization number n, in simulated sparse PALM images obtained by random sampling of the perfect PALM image in **c**. Orange dots: the double mean distance between closest localizations, $R_{\mathrm{Nyq}} = 2\langle d \rangle$ in sparse PALM images (see Online Methods section 'Quality metrics and sampling resolution'). The blue curve shows the resolution $R_{5\times\mathrm{Nyq}}$ allowed for by the sampling density according to the five-fold Nyquist sampling criterion (Legant *et al. Nat Methods* 13, 359–365, 2016). The horizontal green dashed line shows the resolution allowed for by the localization precision σ_{loc}, namely $R_{\mathrm{loc}} = 2.3\,\sigma_{\mathrm{loc}} = 23\ \mathrm{nm}$. The dashed grey and blue lines indicate the minimum number of localizations ($N_{\mathrm{Nyq}}(R_{\mathrm{loc}})$ and $N_{5\times\mathrm{Nyq}}(R_{\mathrm{loc}}) = 5N_{\mathrm{Nyq}}(R_{\mathrm{loc}})$, respectively) needed to achieve $R_{\mathrm{Nyq}} \leq R_{\mathrm{loc}}$ and $R_{5\times\mathrm{Nyq}} \leq R_{\mathrm{loc}}$, respectively. The image resolution, as determined by both sampling density and localization precision, can be defined as $R = \max\left(R_{\mathrm{loc}}, R_{5\times\mathrm{Nyq}}\right)$. Thus, achieving a PALM image with a resolution limited only by precision and not by sampling (i.e. with $R = R_{\mathrm{loc}}$) requires a number of localizations n in excess of $N_{5\times\mathrm{Nyq}}(R_{\mathrm{loc}})$ (to the right of the blue dashed line). The red dashed line indicates the number of localizations in the sparse PALM image of panel **a**. **e,f**) Intensity profiles along the yellow segments in **b** and **c**, labelled "e" and "f", respectively. The full width half maximum (FWHM) of both intensity profiles is 23 nm.

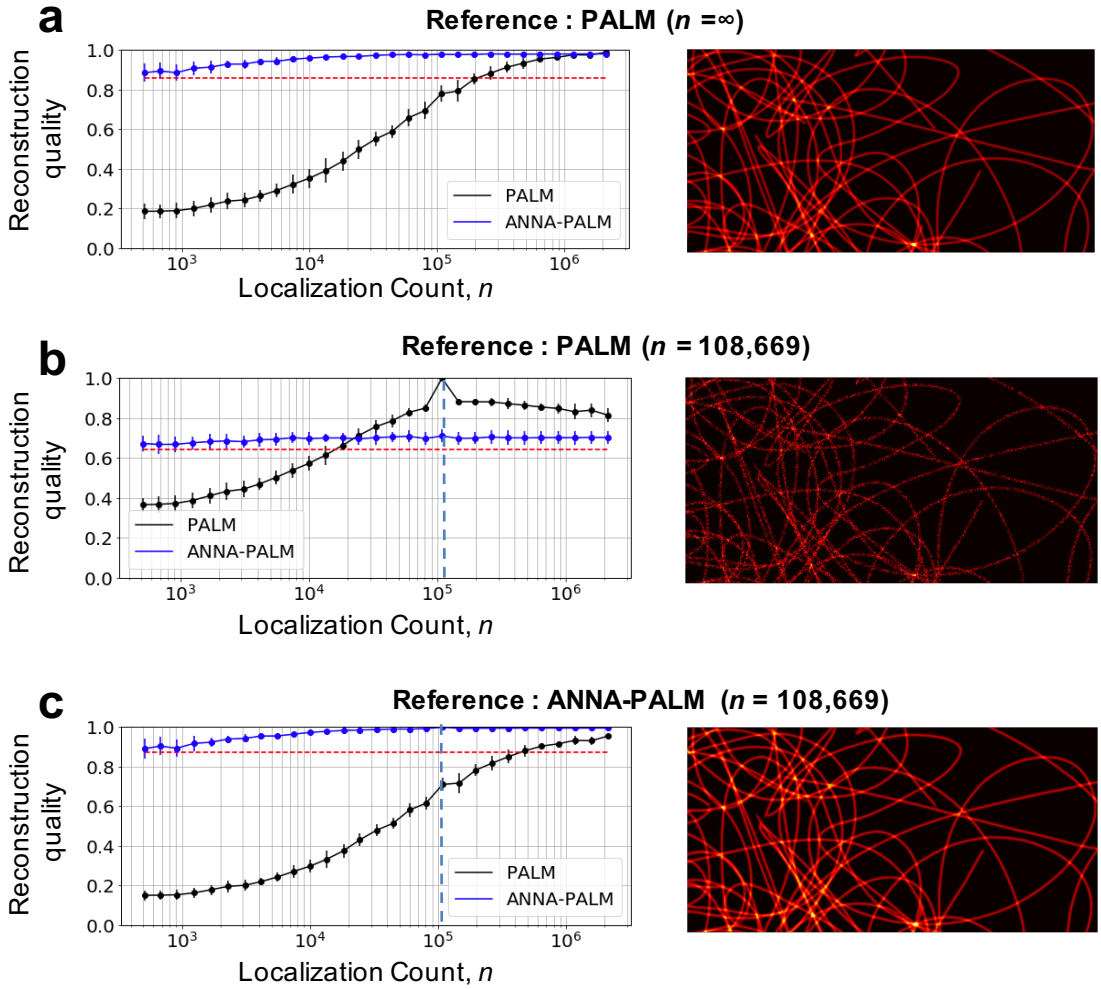
Nature Biotechnology: doi:10.1038/nbt.4106



Supplementary Figure 4 : ANNA-PALM reconstructions for simulated images of varying sparsity

The left column shows simulated sparse localization images of microtubules obtained by random subsampling of the perfect PALM image ($\boldsymbol{n} = \infty$) of **Figure 2e**, for increasing numbers of localizations $\boldsymbol{n}$, from $\boldsymbol{n} = \mathbf{0}$ (top) to $\boldsymbol{n} = \mathbf{1,448,840}$ (bottom). The middle column shows the ANNA-PALM reconstruction from the corresponding sparse PALM image and the widefield image (shown in **Figure 2a**) as inputs. The right column shows a merged image with the perfect PALM ground truth ($\boldsymbol{n} = \infty$; **Figure 2e**) in green, and the ANNA-PALM reconstruction in red.

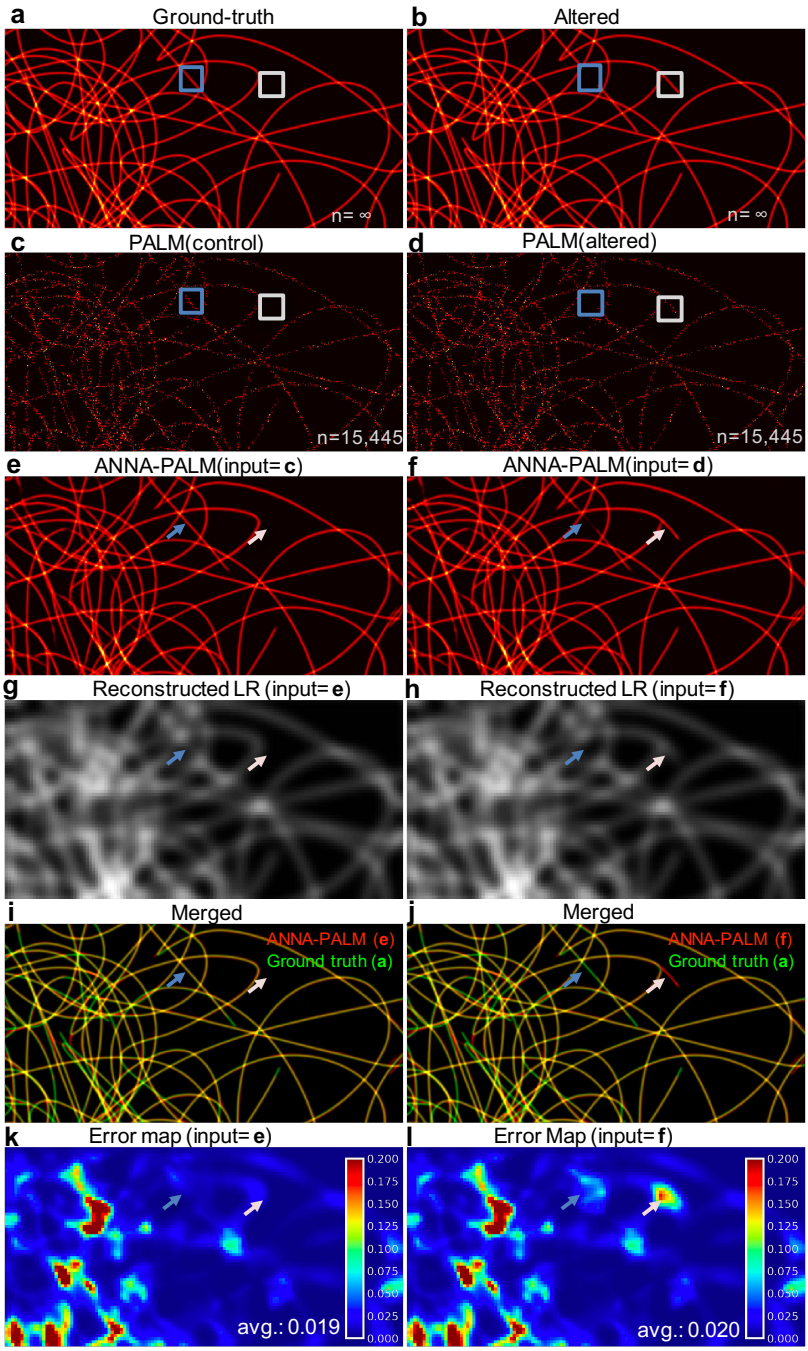
Nature Biotechnology: doi:10.1038/nbt.4106



Supplementary Figure 5 : Reconstruction quality for different ground truths

This figure shows the quality of PALM or ANNA-PALM reconstructions versus localization number n (plots on the left) as defined relative to three different ground truth images (shown on the right). The three ground truth images are, from top to bottom: the perfect PALM image corresponding to an infinite number of localizations ($n = \infty$) (**a**); a "dense" PALM image with a high number of localizations (here n = 108,669) (**b**); the ANNA-PALM reconstruction of that dense PALM image (**c**). The reconstruction quality is computed as in **Figure 2g** as the MS-SSIM of the reconstructed image with the corresponding ground truth. The black curve shows the reconstruction quality of standard PALM, the blue curve shows the reconstruction quality of ANNA-PALM (using the sparse PALM image and the widefield image as inputs). Dots are averages over n=10 simulations and error bars are standard deviations. The red dashed line shows the reconstruction quality of ANNA-PALM using the widefield image as only input. Note that the reconstruction quality curves are similar when the ground truth is defined as either the perfect PALM image (**a**) or the ANNA-PALM reconstruction of the dense PALM image (**c**). By contrast, when the ground truth is defined as the unprocessed dense PALM image, the reconstruction quality of standard PALM has incorrect behavior, as it peaks for n = 108,669 localizations instead of continuously increasing (**b**). Therefore, in absence of the perfect ground truth, i.e. if only a finite number of localizations is available (as is the case for experimental data), the ground truth is best defined as the ANNA-PALM output of the dense PALM image rather than the dense PALM image itself. We make use of this observation when quantifying the reconstruction quality of experimental data in **Supplementary Figure 9** and accordingly define the ground truth as the ANNA-PALM reconstruction of the dense PALM image.

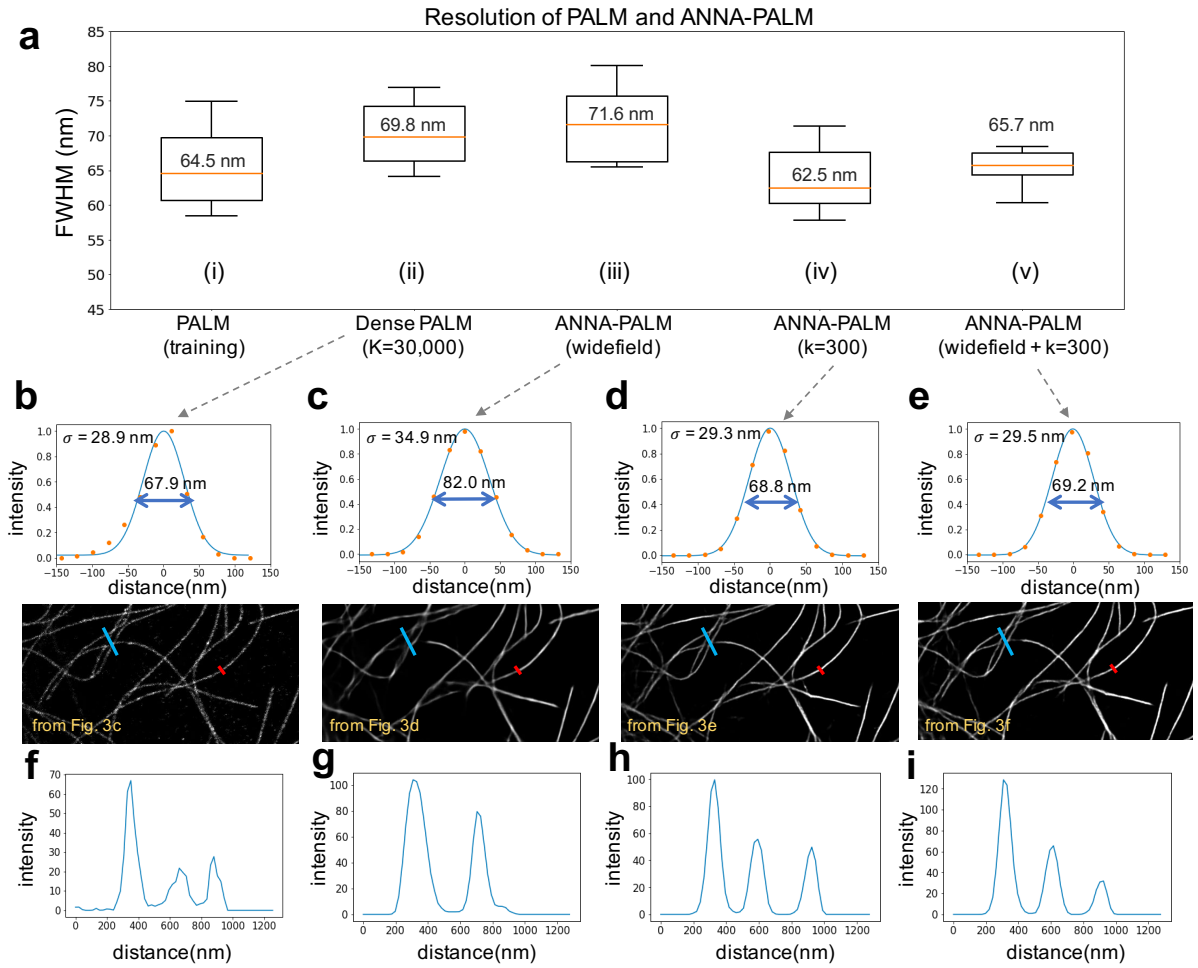
Nature Biotechnology: doi:10.1038/nbt.4106



Supplementary Figure 6 : ANNA-PALM error maps for simulated data

This figure shows how the low resolution error maps produced by ANNA-PALM (see **Figure 1b**) reflect ambiguities in the sparse localization image and/or reconstruction errors. **a,b**) Simulated "perfect" PALM images ($n = \infty$) displaying regions of high filament density (left) and regions of low filament density (right). Image **a** is identical to **Figure 2e**. Image **b** ('altered') is identical to image **a** except for the artificial displacement of a small piece of filament from the blue box to the white box. **c,d**) Sparse PALM images obtained by random subsampling of the perfect PALM images in **a,b**, respectively. **e,f**) ANNA-PALM reconstructions of the sparse PALM images **c,d,** respectively. Considering **a** as the ground truth, the blue and white arrows in **f** point to a region of incorrectly missing filament (false negative) and an incorrectly added filament (false positive), respectively. **g,h**) Low resolution images predicted by ANNA-PALM from the sparse PALM images **c,d**, respectively. **i,j**) Merged images showing the ground truth image **a** in green and the ANNA-PALM reconstructions **e** and **f**, respectively, in red. **k,l**) ANNA-PALM error maps generated by comparing the predicted low resolution images **g**,**h** to the widefield image of **Figure 2a** simulated based on the unaltered perfect PALM image **a**. Note that the high error regions pointed out by the blue and white arrow in **l** reflect the false negatives and false positives in the ANNA-PALM reconstruction **f**. The average error value is indicated and is higher for the altered image. The high error areas on the left of both **k** and **l** reflect the larger ambiguity of the sparse localization images **c**,**d** in regions of high filament density.

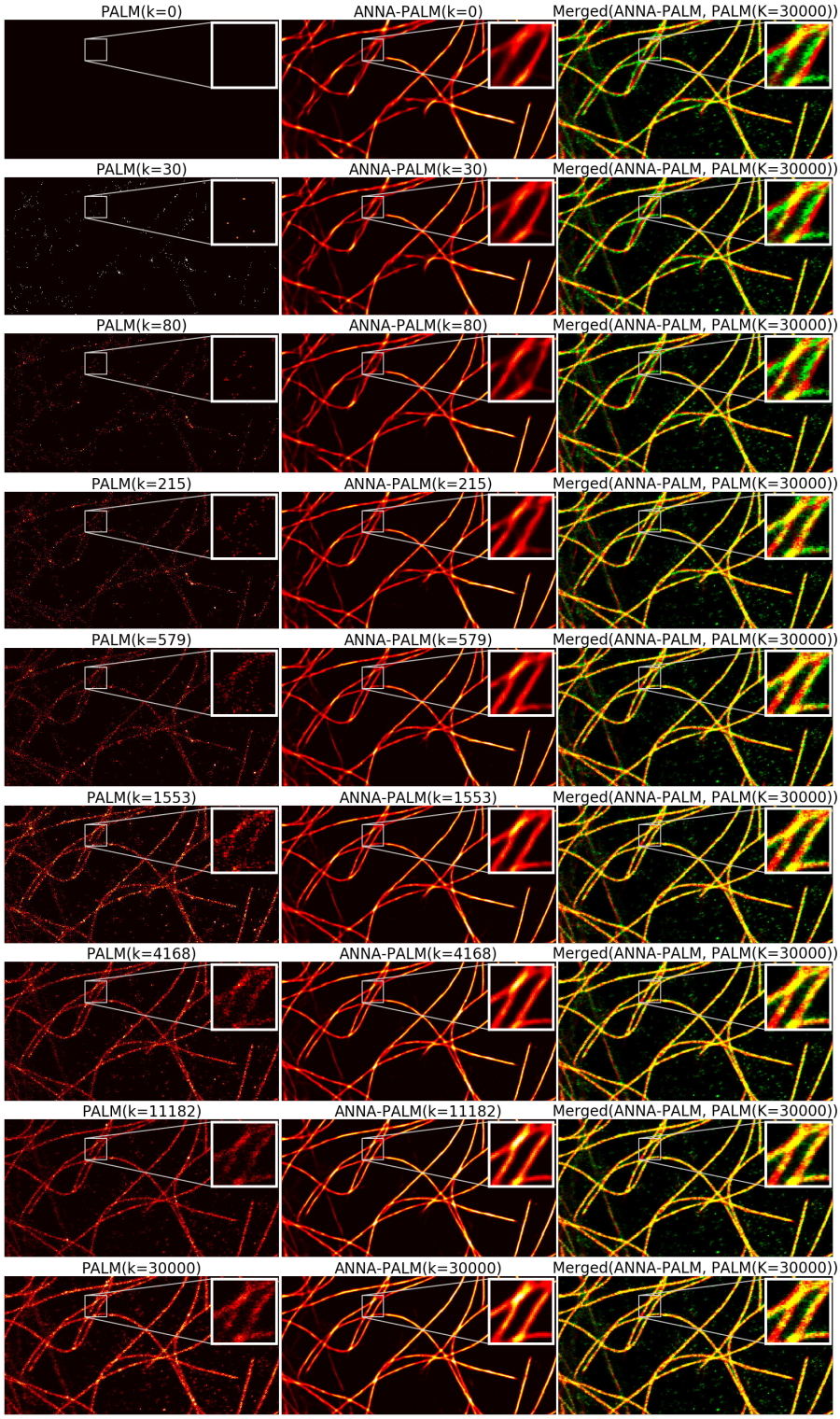
Nature Biotechnology: doi:10.1038/nbt.4106



Supplementary Figure 7: Resolution of PALM and ANNA-PALM in experiments

This figure provides measures of resolution for the PALM and ANNA-PALM microtubule images of **Figure 3,** based on the full width half maximum (FWHM) of intensity profiles across filaments. **a**) Boxplots show the distribution of FWHM measured on n = 10 different positions, as illustrated for one position in panels **b-e**. Orange lines show medians, with the value as indicated; upper and lower box limits show the upper and lower quartiles, respectively; whiskers show the full extent of the data. Each boxplot corresponds to a distinct set of images. From left to right: (i) dense PALM images used for training (not shown); (ii) dense PALM image from K=30,000 frames (**Figure 3c** and panel **b**); (iii) ANNA-PALM image reconstructed from the widefield image only (**Figure 3d** and panel **c**); (iv) ANNA-PALM image reconstructed from the sparse PALM image with k=300 frames only (**Figure 3e** and panel **d**); (v) ANNA-PALM image reconstructed from both the widefield image and the sparse PALM image with k=300 frames (**Figure 3f** and panel **e**). Except for the training images, the ten measurements were performed at the same positions in the image shown in **Figure 3c,d,e,f**. Note that the FWHM of ANNA-PALM reconstructions is similar to or lower than that of the dense PALM image, suggesting that the resolution is slightly better. Note also that the FWHM only provide an upper bound of the resolution. Assuming a microtubule diameter of ~25 nm and a combined size of the primary and secondary antibodies of ~17.5 nm, a FWHM of 70 nm is consistent with a resolution of ~20 nm. **b-e**) Intensity profiles along the red segment shown in the images below, which are the lower parts of **Figure 3 c,d,e,f**, respectively. Orange dots are measured intensities, the blue curves are fitted Gaussian functions, with standard deviation σ and FWHM (double arrow) as indicated. **f-i**) Intensity profiles along the cyan segment for the same images. Note that ANNA-PALM reconstructions can distinguish the three filaments when using the sparse PALM image as only input, but not when using the widefield image as only input.

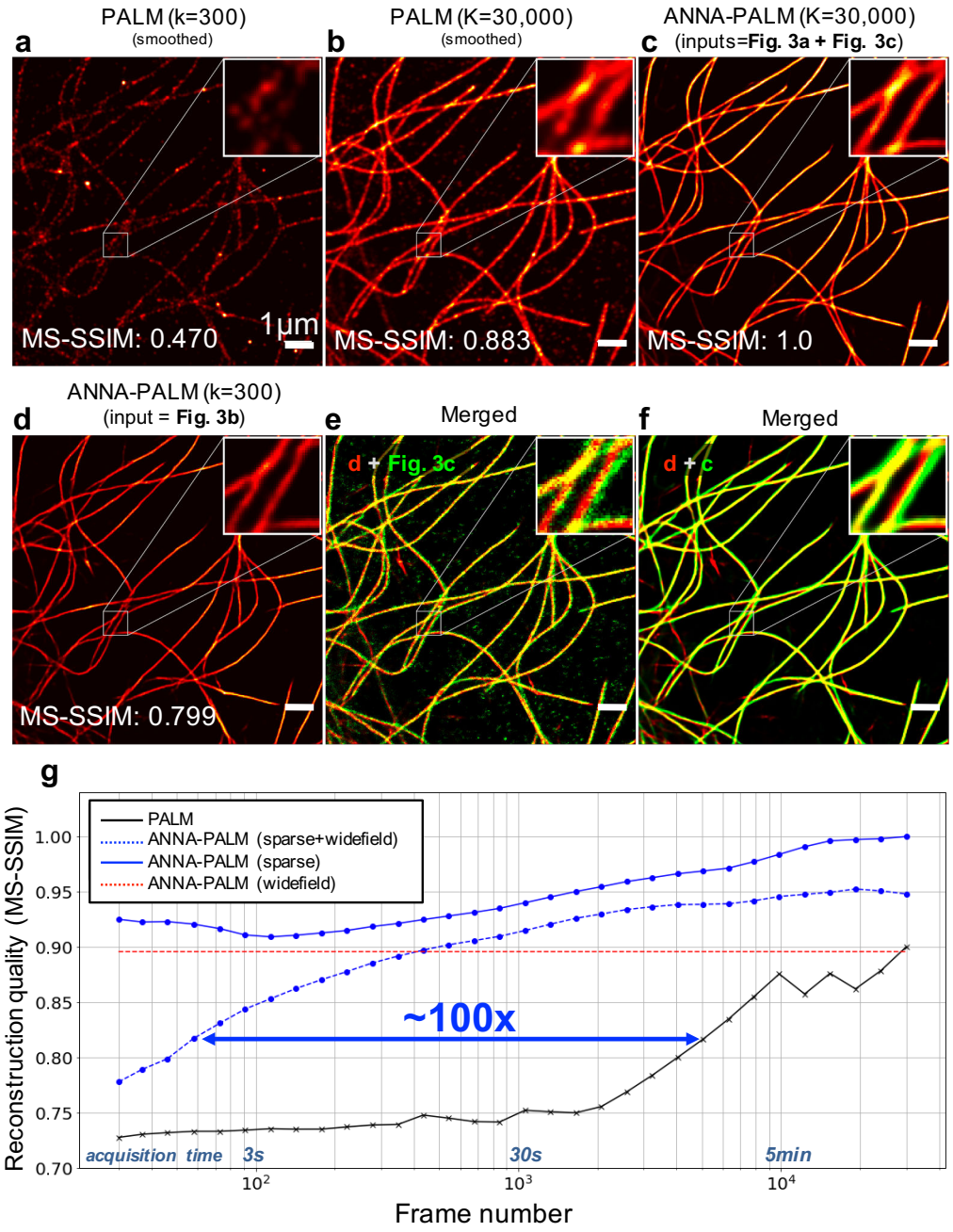
Nature Biotechnology: doi:10.1038/nbt.4106



Supplementary Figure 8: ANNA-PALM reconstructions of experimental images for varying frame numbers

The left column shows sparse PALM images for increasing numbers of frames, from k=0 (top) to k=30,000 (bottom). The region shown corresponds to the lower half of that in **Figure 3a-i.** The middle column shows the ANNA-PALM image reconstructed from the sparse PALM image and the corresponding widefield image (lower half of **Figure 3a**). The right column shows merged images with the dense PALM image obtained from K=30,000 frames (bottom left panel) in green and the ANNA-PALM reconstructions in red.

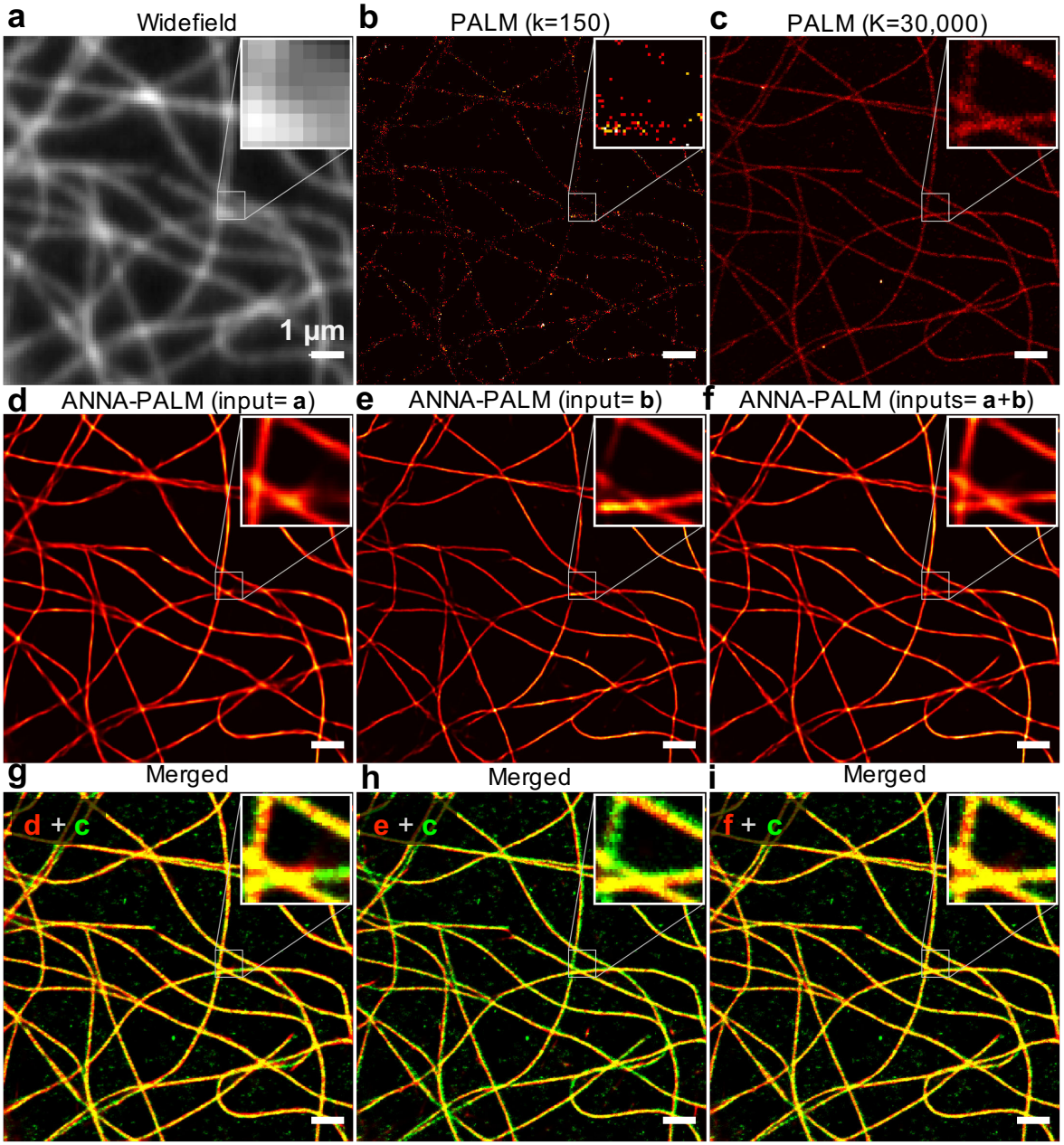
Nature Biotechnology: doi:10.1038/nbt.4106



Supplementary Figure 9: Quantifying ANNA-PALM reconstruction quality in experiments

The MS-SSIM is used to quantify the quality of PALM and ANNA-PALM images of immunolabeled microtubules. For this purpose, we defined the ground truth as the ANNA-PALM output of the dense PALM image (see **Supplementary Figure 5** for a justification). For standard PALM images, we also applied a Gaussian smoothing of standard deviation 1.5 pixels to facilitate quantification by MS-SSIM. **a,b**) Sparse (**a**) and dense (**b**) PALM images obtained from k=300 and K=30,000 frames, respectively. These images are identical to **Figure 3b,c** except for the smoothing. The MS-SSIM is indicated. **c**) ANNA-PALM reconstruction from the dense PALM image with K=30,000 frames and the widefield image of **Figure 3a**. This image is defined as the ground truth for MS-SSIM calculations. **d**) ANNA-PALM reconstruction of the sparse PALM image of **Figure 3b**. This image is identical to **Figure 3e**. Note that the MS-SSIM is much larger than for the (smoothed) sparse PALM image in **a**. **e**) Merged image showing the ANNA-PALM reconstruction of the sparse PALM image (panel **d**) in red and the (unsmoothed) dense PALM image of **Figure 3c** in green**. f)** Merged image showing the ANNA-PALM reconstruction of the sparse PALM image (panel **d**) in red and the ANNA-PALM reconstruction of the dense PALM image (panel **c**) in green. **g**) Reconstruction quality of PALM and ANNA-PALM images as function of frame number (or acquisition time), as measured by the MS-SSIM with the ANNA-PALM reconstruction of the dense PALM image. Black curve: reconstruction quality of the standard PALM image. Dashed blue curve: reconstruction quality of ANNA-PALM using the sparse PALM image as input. Solid blue curve: reconstruction quality of ANNA-PALM using both the sparse PALM and widefield images as inputs. Red dashed line: reconstruction quality of ANNA-PALM using the widefield image as only input. Note that all values were computed for a larger image (2560x2560 pixels) than the region shown in panels **a-f** or **Figure 3a-i** (512x512 pixels).

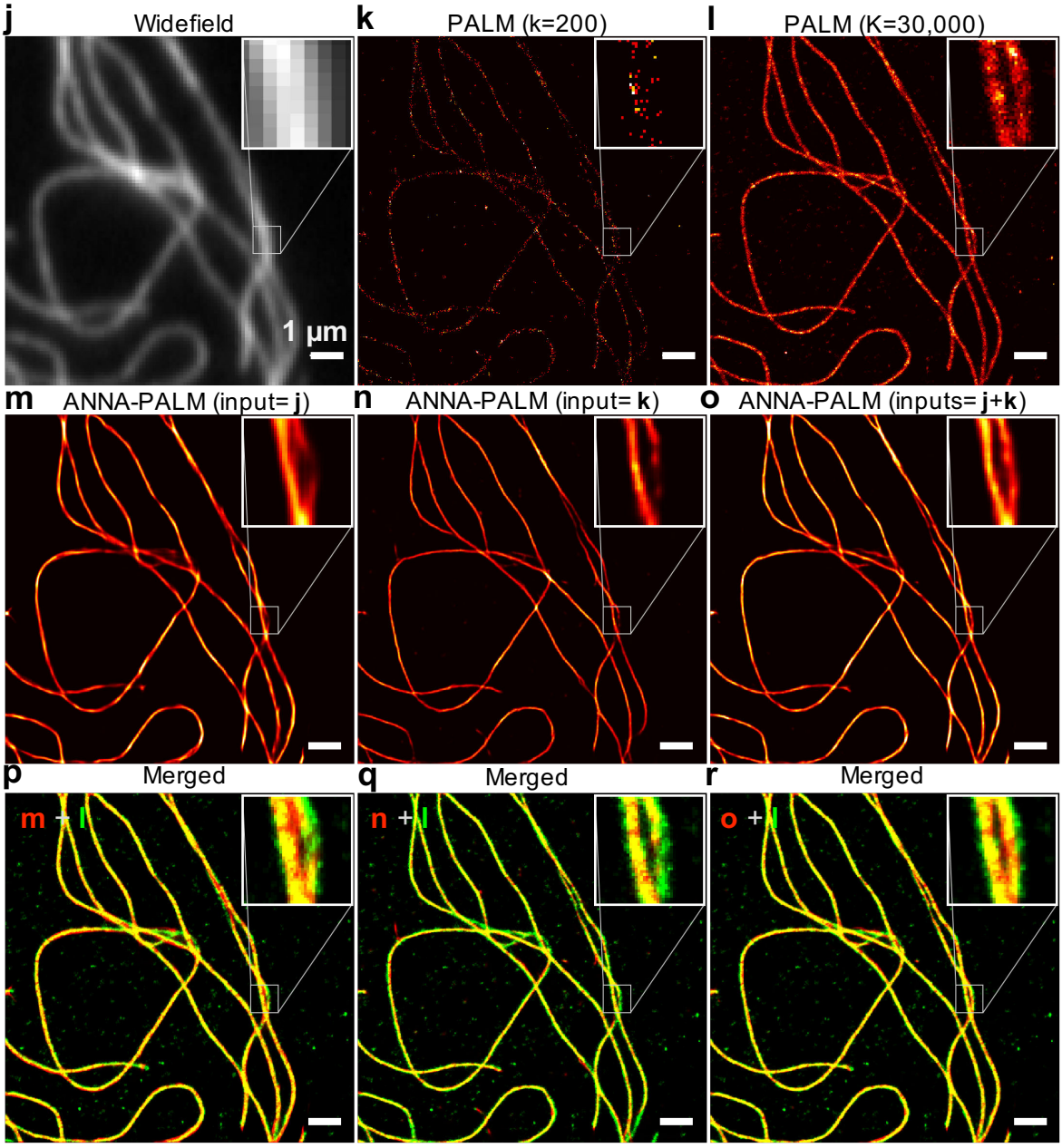
Nature Biotechnology: doi:10.1038/nbt.4106



Supplementary Figure 10 (beginning): ANNA-PALM reconstructions of immunolabeled microtubules

This figure (continued on next pages) shows additional examples of PALM images and ANNA-PALM reconstructions of immunolabeled microtubules, taken from another region within the same FoV as **Figure 3**. **a**) Widefield image; **b**) Sparse PALM image obtained from k = 150 frames with Δt= 30 ms exposure time (acquisition time: $k\Delta t$=4.5 s). **c**) Dense PALM image of the same region obtained from K=30,000 frames (acquisition time: $K\Delta t$=15 min). **d-f**) ANNA-PALM reconstructions using as input either the widefield image only (**d**), the sparse PALM image only (**e**), or both (**f**). Pixel values are linearly mapped to colors from the look-up table shown in **Figure 3**. Black and white correspond to values V_{min} and V_{max}, respectively, with V_{min}=0 for all panels, V_{max}= 3, 76.5, 102, 102, 102 for panels **b, c, d**, **e**, and **f**, respectively. **g-i**) Merged images showing the dense PALM image of **c** in green superposed to the ANNA-PALM reconstructions of panels **d**, **e**, and **f**, respectively, in red.

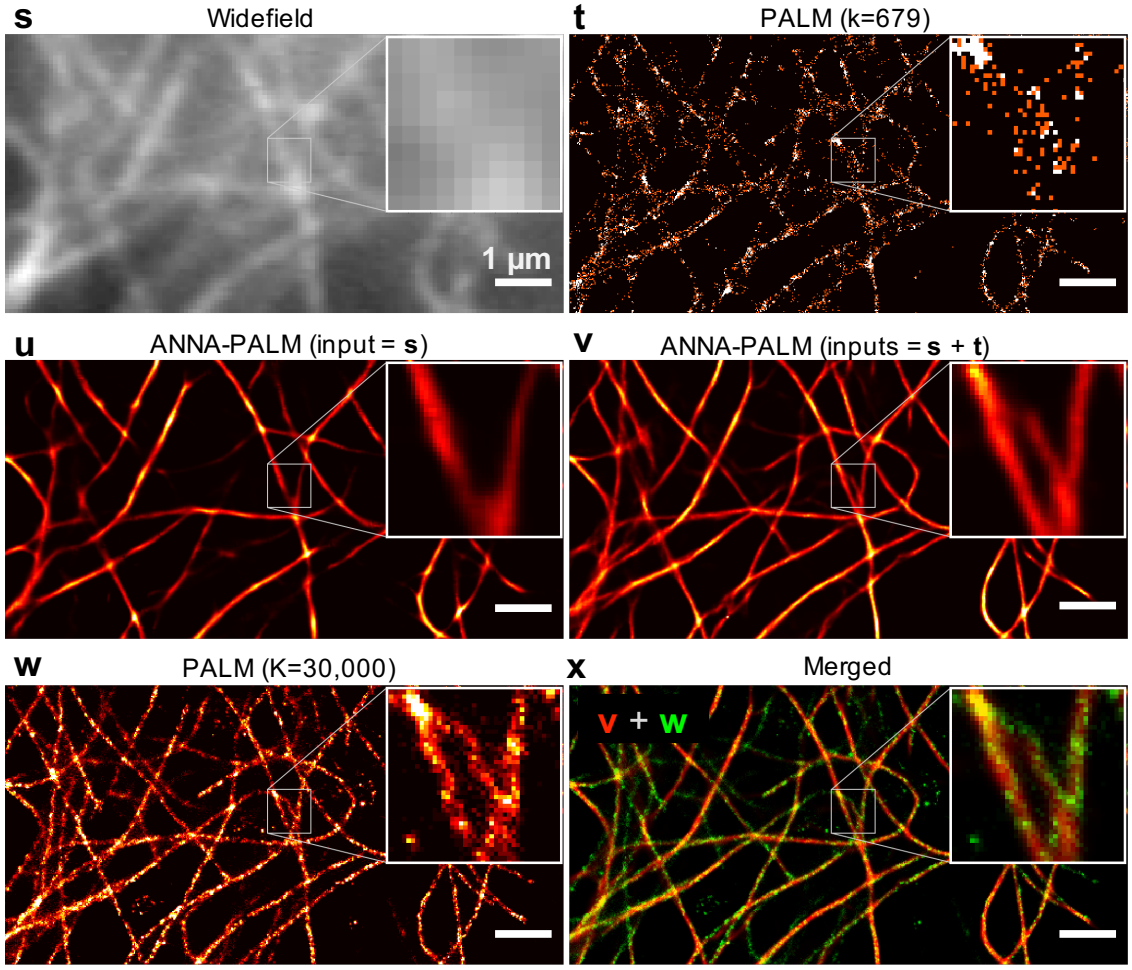
Nature Biotechnology: doi:10.1038/nbt.4106



Supplementary Figure 10 (continued): ANNA-PALM reconstructions of immunolabeled microtubules

Same as panels **a-l** for another region of the same FoV. (**j**) Widefield image; (**k**) Sparse PALM image of immunolabeled microtubules from k = 200 frames with Δt= 30 ms exposure time (acquisition time: kΔt=6 s). (**l**) Dense PALM image for the same region obtained from K=30,000 frames (acquisition time: KΔt=15 min). **m-o**) ANNA-PALM reconstructions using as input either the widefield image only (**m**), the sparse PALM image only (**n**), or both (**o**). Pixel values are linearly mapped to colors from the look-up table shown in **Figure 3**. Black and white correspond to values V_{min} and V_{max}, respectively, with V_{min}=0 for all panels, V_{max}= 3, 35.4, 102, 102, 102, for panels **k, l, m**, **n**, and **o**, respectively. **p-r**) Merged images showing the dense PALM image **l** in green superposed to the ANNA-PALM reconstructions of panels **m**, **n**, and **o**, respectively, in red.

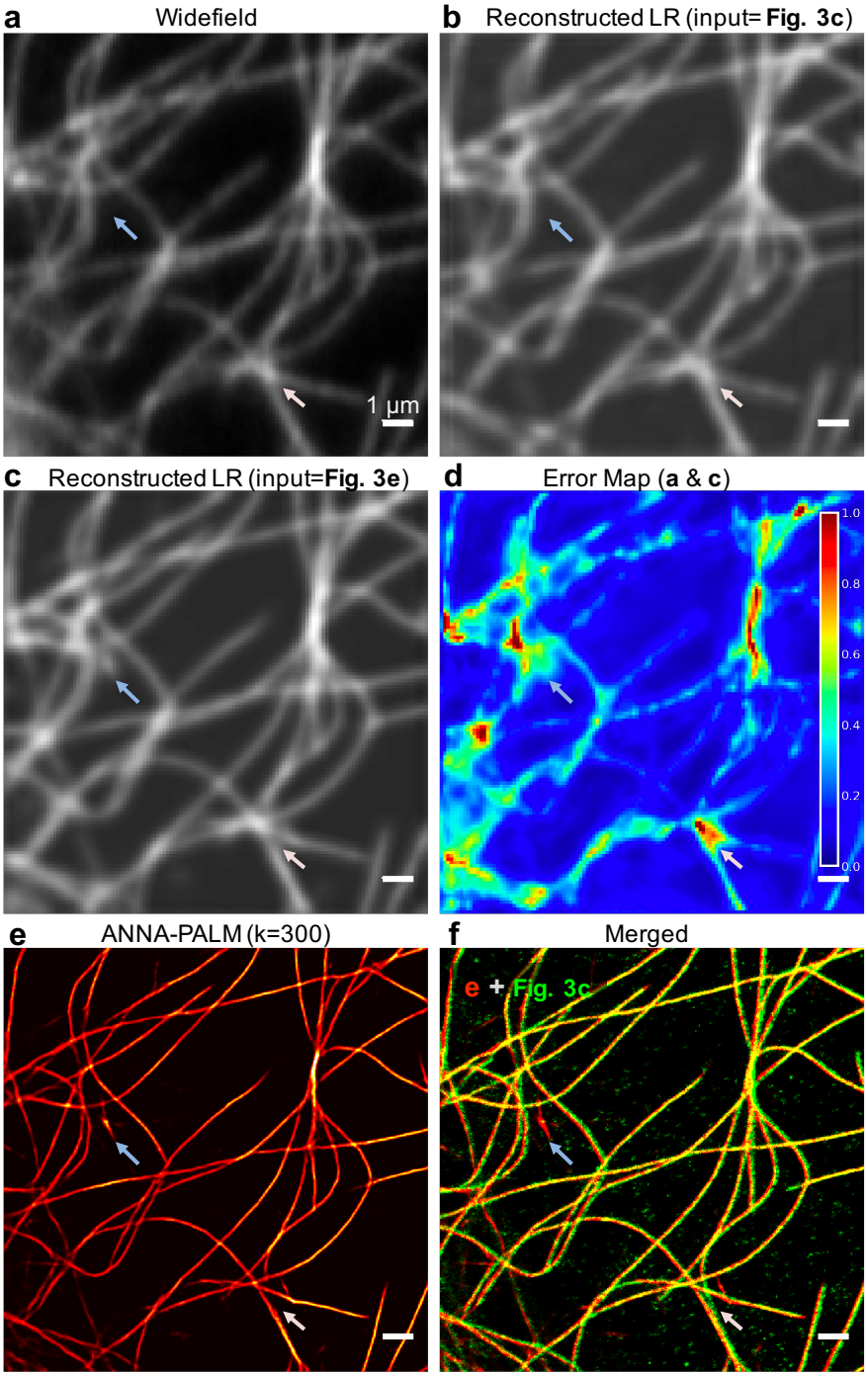
Nature Biotechnology: doi:10.1038/nbt.4106



Supplementary Figure 10 (continued): ANNA-PALM reconstructions of immunolabeled microtubules

This figure shows another example of PALM and ANNA-PALM images for immunolabeled microtubules in an experiment distinct from **Figure 3**. **s**) Widefield image. **t**) Sparse PALM image obtained from k=679 frames with Δt=20 ms exposure time (acquisition time: kΔt=13.6 s). The low quality of localizations is apparent at the intersections of distinct filaments, probably due to overlapping point spread functions caused by an excessive density of activated fluorophores (i.e. insufficient pre-bleaching). **u**) ANNA-PALM reconstruction from the widefield image **s** only. **v**) ANNA-PALM reconstruction from both the widefield image **s** and the sparse PALM image **t**. **w**) Dense PALM image obtained from K=30,000 frames (acquisition time: kΔt=10 min). **x**) Merged image showing the ANNA-PALM reconstruction of panel **v** in red superposed to the dense PALM image of **w** in green. Note how ANNA-PALM produces a high quality reconstruction despite the low quality of the sparse PALM input image.

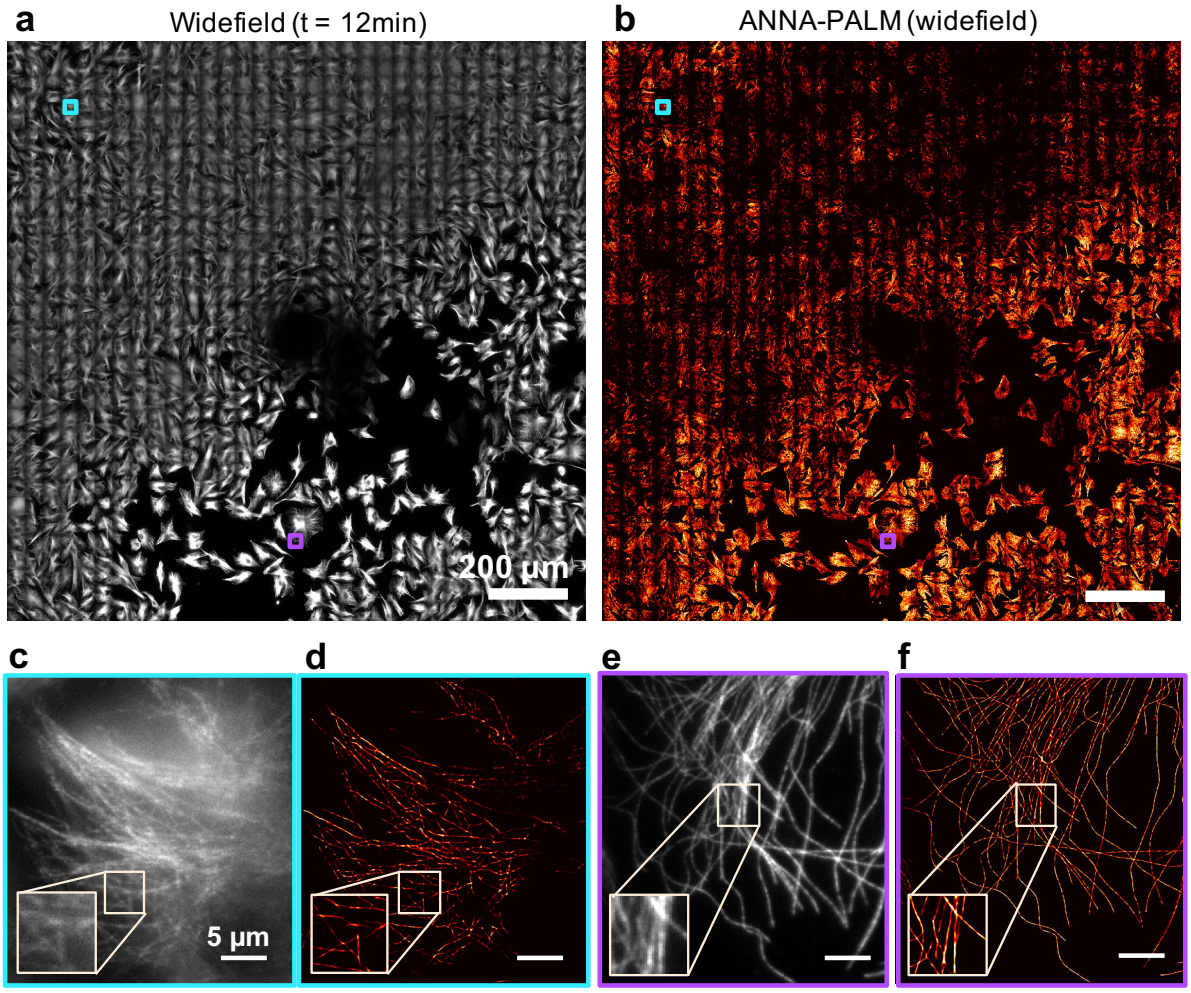
Nature Biotechnology: doi:10.1038/nbt.4106



Supplementary Figure 11: ANNA-PALM error maps for experimental images

This figure illustrates the application of ANNA-PALM error maps to the experimental data shown in **Figure 3**. **a**) Experimental widefield image (identical to **Figure 3a**). **b**) Widefield image predicted by ANNA-PALM from the dense PALM image with K=30,000 frames of **Figure 3c**. **c**) Widefield image predicted by ANNA-PALM from the sparse PALM image with k=300 frames of **Figure 3f**. **d**) Error map produced by ANNA-PALM by comparing the predicted widefield image in **c** to the observed widefield image in **a**. **e**) ANNA-PALM reconstruction from the sparse PALM image of **Figure 3b** (identical to **Figure 3e**). **f**) Merged image showing the ANNA-PALM reconstruction of **e** in red and the dense PALM image (**Figure 3c**) in green (identical to **Figure 3h**). The blue and white arrows point to a false positive and a false negative, respectively, of the ANNA-PALM reconstruction in **e**. Note the high values of the error map **d** at the corresponding locations.

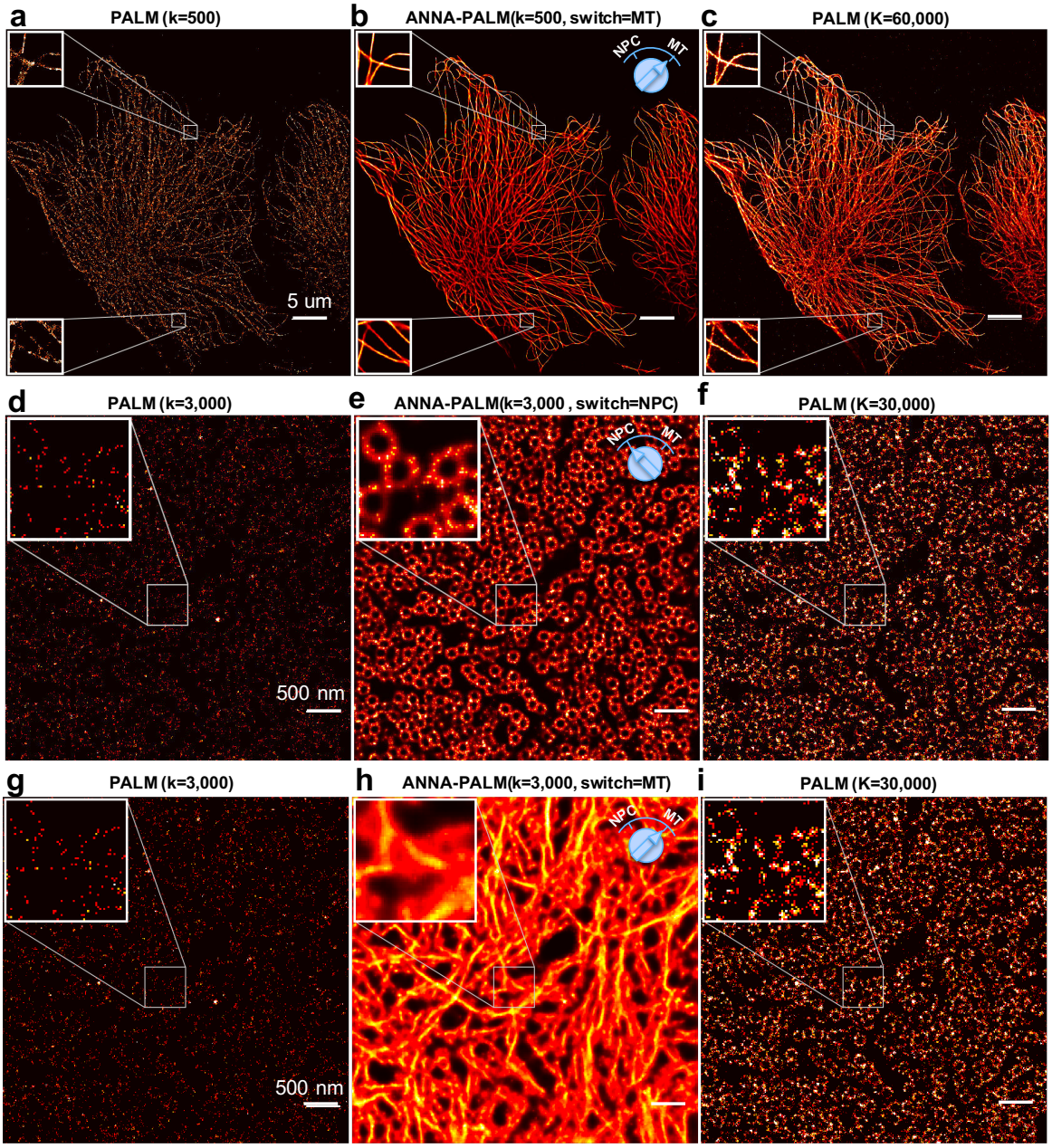
Nature Biotechnology: doi:10.1038/nbt.4106



Supplementary Figure 12: ANNA-PALM reconstruction of a large field from widefield images

a) Widefield image of the 1.8 mm x 1.8 mm field shown in **Figure 4a**, assembled as a mosaic of 1,089 individual widefield images. Total acquisition time was 12 minutes. **b**) ANNA-PALM reconstruction of the same field from the widefield images in **a** only, and assembled as a mosaic of 1,089 individual ANNA-PALM reconstructions (each obtained using a single widefield image as input). **c,e**) Magnified views of the cyan and violet boxes, respectively, in image **a**. **d,f**) Magnified views of the corresponding ANNA-PALM reconstructions in image **b** (cyan and violet boxes, respectively).

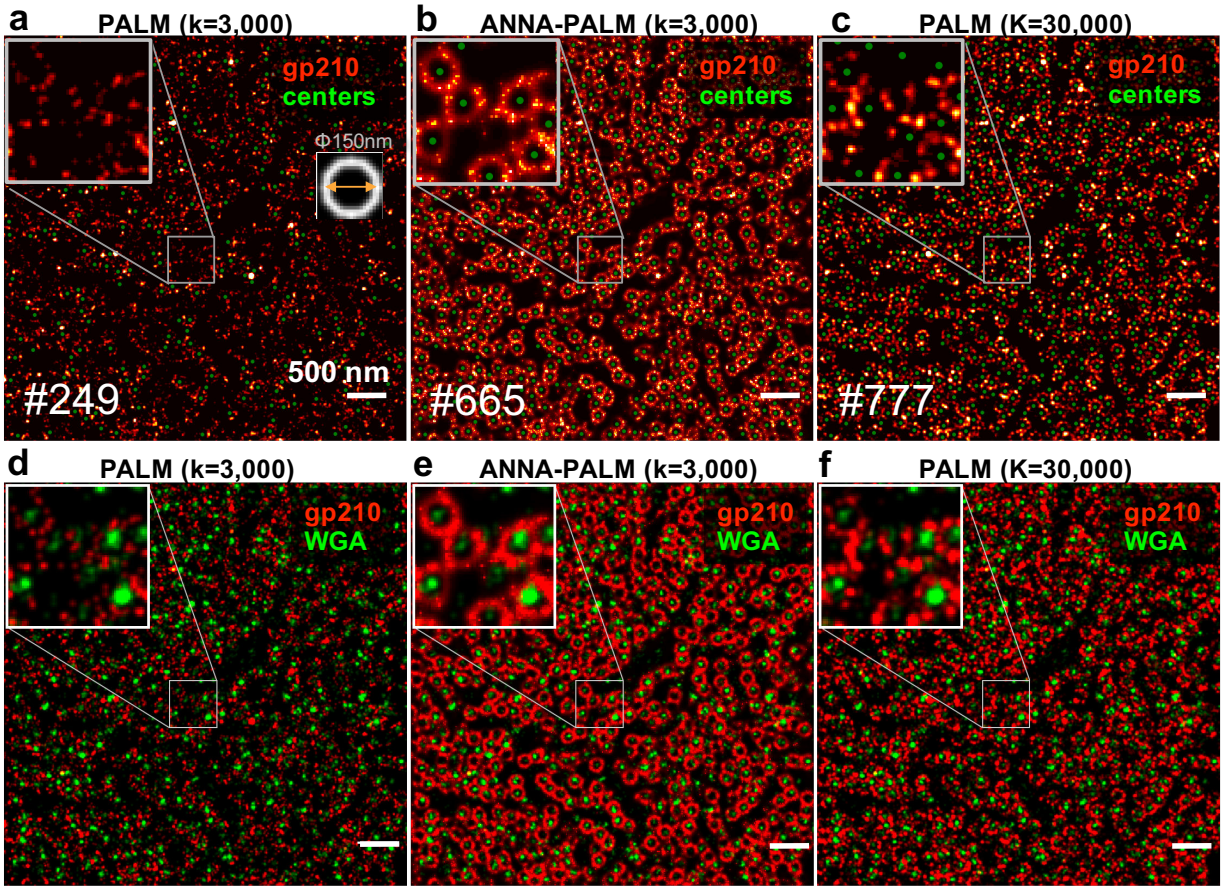
Nature Biotechnology: doi:10.1038/nbt.4106



Supplementary Figure 13: Reconstructing microtubule and nuclear pore images with a single ANN and a switch

This figure shows the application of an ANN trained on both microtubule and nuclear pore images to sparse PALM images of microtubules or nuclear pores. **a**) Sparse PALM image of microtubules obtained from k=500 frames. **b**) ANNA-PALM reconstruction of **a** with the structure switch set to microtubules ('MT'). **c**) Corresponding dense PALM image obtained from K=60,000 frames. **d**) Sparse PALM image of the nucleoporin gp210 obtained from k=3,000 frames. **e**) ANNA-PALM reconstruction of **d** with the structure switch set to nuclear pores ('NPC'). **f**) Dense PALM image obtained from K=30,000 frames. **g**) identical to **d**. **h**) ANNA-PALM reconstruction of **g** with the structure switch set to microtubules ('MT'). Note the artefactual nature of the reconstructed image **h**. **i**) identical to **f**.

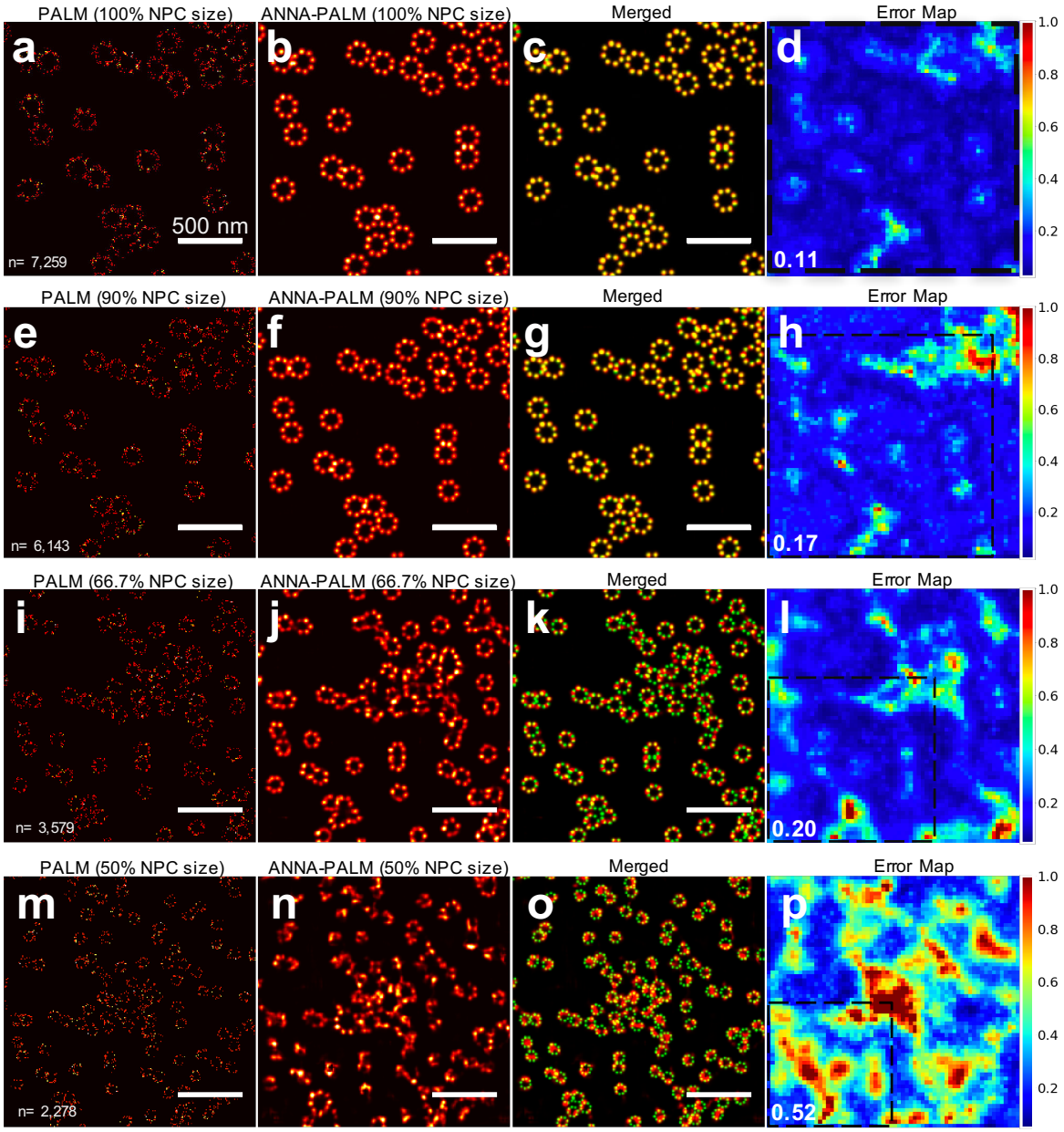
Nature Biotechnology: doi:10.1038/nbt.4106



Supplementary Figure 14: Identifying and counting nuclear pores in ANNA-PALM and PALM images

a-c) Superposed red and green images, where the red images are identical to **Figure 6a-c** (except that we applied Gaussian smoothing with a standard deviation of 1 pixel to facilitate visualization) and show the immunolabeled nucleoporin gp210 for a sparse PALM image with k=3,000 frames (**a**), its ANNA-PALM reconstruction (**b**) and the dense PALM image with K=30,000 frames (**c**). The green image ("centers") shows the positions of automatically identified nuclear pores, obtained as follows. For each of the three images, we computed the normalized cross-correlation with a template image containing a ring of radius 75 nm and thickness 20 nm convolved with a Gaussian kernel of standard deviation 10 nm (inset of panel **a**). We then thresholded the cross-correlation image at mean + 2 standard deviations and determined the peaks using the "peak_local_max" function in the scikit-image Python library. These automatically identified peaks are shown as green dots. The total number of detected peaks (#) is shown and indicates that ANNA-PALM greatly facilitates identification and counting of nuclear pores. **d-f**) Superposed red and green images, where the red images are identical to **a-c** and **Figure 3a-c** and the green image is a PALM image of the lectin WGA, which binds to the core of nuclear pores (Sellés *et al. Sci. Rep.* 7, 14732, 2017).

Nature Biotechnology: doi:10.1038/nbt.4106



Supplementary Figure 15: ANNA-PALM reconstructions and error maps for mismatch between training and testing data

This figure uses simulated images to illustrate the effect of model mismatch on ANNA-PALM reconstructions and error maps. Model mismatch occurs when A-net is used to reconstruct sparse images of a different structure than the images it has been trained on. For all results shown here, A-net has been trained once on images of simulated nuclear pores with a fixed diameter of 150 nm only. The figure shows applications of the trained A-net to simulated images of nuclear pores with diameters of 100% (**a-d**), 90% (**e-h**), 66.7% (**i-l**) or 50% (**m-p**) relative to the 150 nm of the training images. The first column (**a**,**e**,**l**,**m**) shows the sparse input images (to facilitate comparisons, images **e**,**i,m** were obtained by applying a scaling factor to the original image used to generate **a**); the second column (**b**,**f**,**j**,**n**) shows the corresponding ANNA-PALM reconstructions; the third column shows the ANNA-PALM reconstructions (red) superposed to the ground truth (perfect PALM image, green). The fourth column (**d,h,l,p)** shows the low resolution error map. The dashed black box encloses the same area (and the same number of nuclear pores) as in **a-d**. The average error inside this area is indicated in the bottom left corner of **d,h,l,p**. Note how both reconstruction errors and the error map values increase for larger model mismatch (i.e. smaller nuclear pores). For a given nuclear pore size, errors are higher in regions of larger nuclear pore density, where the sparse image is more ambiguous.

Nature Biotechnology: doi:10.1038/nbt.4106

2.2 ShareLoc: A platform for sharing localization data

2.2.1 Deep learning performance scales with data size

An important advantage of deep learning compared to traditional machine learning is its better capacity scale with the amount of data used during training. Adding more data can efficiently improve the models performance in terms of generalization and prevent over-fitting. According to an empirical study [Hestness et al., 2017], after a certain number of samples in the training dataset, there is a so called *power-law region* (Figure 2.1) where each new sample can efficiently improve the performance.

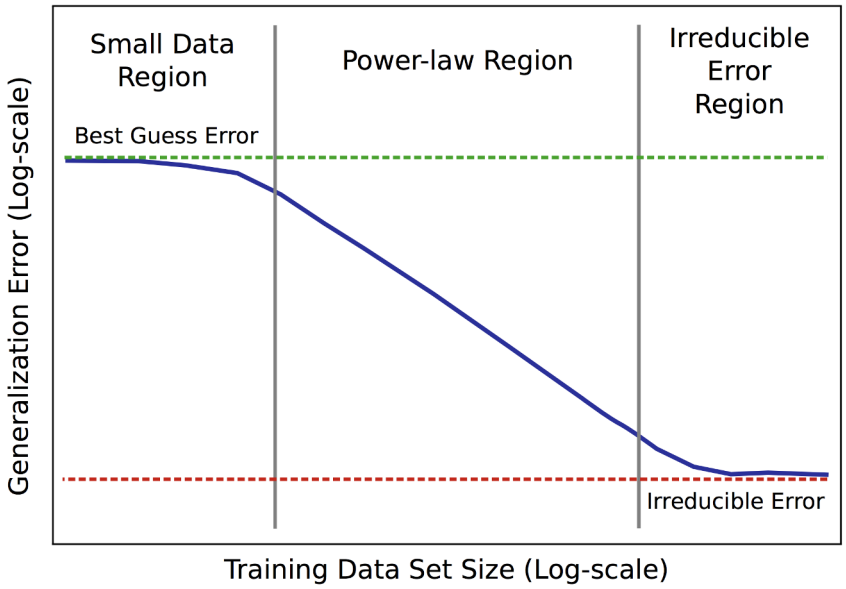


Figure 2.1: Sketch of power-law learning curves [Hestness et al., 2017]. As training dataset size grows, the generalization error drops. Depending on the dataset size range, the dropping speed varies. In the *Small Data Region*, the error drops relatively slowly (left); When the dataset size increased to the *Power-law Region*, the error drops rapidly according to a power-law (middle); After, it is the *Irreducible Error Region*, where the error is still non-zero but the model is unable to further improve because of it approaches the information theoretic lower bound.

As shown in the previous section, with the help of extensive data augmentation, A-nets with reasonable performance can be trained with data from less than 10 FoVs. Based on the sample number and experimental observations, ANNA-PALM is presumably working within the *Small Data Region*, or the *Power-law Region*. Most likely, the dataset size we used in ANNA-PALM is still far from the *Irreducible Error Region* meaning there is still a big margin to improve

the performance by adding more training data without changing the A-net algorithm itself. The advantage of using a data size within the *Power-law Region* is that the performance gain relative to the cost is higher than the *Small Data Region*, the error rate drops much faster in the *Power-law Region* while increasing the amount of data. In another words, it's much more cost effective to add data in the *Power-law Region*.

2.2.2 Maximize the usage of existing data by sharing

Adding more training data is an efficient way of achieving better generalization for training deep learning models. However, it is often difficult and costly to obtain new qualitatively high training data from imaging experiments. For ANNA-PALM for example, acquiring a new densely labeled image may take tens of minutes or even an hour due to the low throughput nature of PALM/STORM experiments.

Nevertheless, more and more data are generated by the microscopy community, and a plethora of new methods and devices such as sCMOS cameras are developed to acquire data even faster. It is also common to see that most of the acquired data are not used or under-used. Moreover, microscopy data is typically acquired and used for one research, then stored on a backup hard-drive or deleted, even though the same dataset can be potentially reused to solve different problems, or reanalyzed for different purposes [Zaritsky, 2018]. One way to increase the usage of existing data is to build a data sharing platform, with which researchers can upload their data and share with others.

Sharing data is becoming a trend [Ouyang and Zimmer, 2017] which dramatically increases the chance for existing datasets to be reused and reanalyzed by many researchers. There are many different data sharing platforms which are now freely available for researchers, including the *Image Data Resource* (IDR) [Williams et al., 2017], the *Electron Microscopy Public Image Archive* (EMPIAR) [Iudin et al., 2016], and the Data Archive for neuro-imaging [Crawford et al., 2016]. These platforms provide enormous amounts of data which can be downloaded freely through their website or programmable *Application Interface* (API). New data can be

uploaded to the platform, tagged and advertised in publications. Some sharing platform can even generate *digital object identifier* (DOI) for each dataset, which can be cited in publications.

2.2.3 Challenges in storing and sharing localization data

For SMLM, storing and transmitting data is quite challenging because of the enormous file size. In order to obtain a single super-resolution image, SMLM typically requires $\sim 10^4$-10^5 diffraction limited images. The data generated can augment to $\sim$ 100 GB (10^{11} bytes) per experiment or more [Ouyang and Zimmer, 2017]. With fast, kHz-rate cameras, the image throughput can reach terabytes (TB, or 10^{12} bytes) per hour. Even just for the localization table generated by localization softwares such as ThunderSTORM [Ovesný et al., 2014], the default data format is a widely used file format called *comma separated values format*(csv). It is easy to open 'CSV' files with text editors or to perform read and write operations in different programing languages such as Java, Python, C etc. However, it often leads to several GB in size for each localization table, especially when the user used the default setting to keep several digits after the decimal. For example, the 'CSV' file size for one DNA-PAINT image can reach up to 7GB. There are other binary data formats for localization data such as the *Tagged Spot File* (tsf) format used by Micro-Manager, which is built upon the *Google Protocol buffers* to encode the localization events. Protocol Buffers is an efficient way of encoding structured data, and uses variant encoding which can reduce the file size. Depending on the redundancy of the data, it can reduce the file size by more than 3 fold compare to the corresponding 'CSV' file.

As a result, with existing file formats, the transmission time and the storage space occupied by the files make it difficult to share localization tables for collaboration or publication. So far, there is no dedicated platform for sharing localization data.

2.2.4 The 'SMLM' file format

To address the challenges involved in storing and sharing localization data, we designed a specific data format called 'SMLM', which can reduce the file size by 6 folds or more. For

example, we can losslessly compress a 7.16 gigabytes (GB) localization table in 'CSV' format into a 1.34 GB file.

The design of '.smlm' format aims to be portable with a smaller file size, and flexible to be easily adapted for storing data for different types of imaging experiment including e.g. multi-color imaging, tiled-acquisition (similar to the high-throughput experiment mentioned in the previous section).

To reduce the file size, two strategies are adopted:

1. First, we use binary data encoding rather than text encoding, i.e. every number in the localization table is represented with data types such as integer, float or double.

2. Second, we compress the binary file using generic lossless compression algorithms such as DEFLATE [L. Peter Deutsch, 1996] which come with the standard ZIP file format.

The drawback of using only binary format to store data is the lack of human readability, which is the main reason why text formats, such as 'CSV' or xls, are widely used in the community. However, most users use a regular text editor or a spreadsheet software such as Microsoft Excel simply because the widely used operating systems associate the '.csv' or '.xls' extension to those softwares. Problems arise when the text file is too large in size causing most of the text editor to get stuck or even crash because they try to load all the content into the memory. This essentially defeats the purpose of using a text format, since the content of these localization tables in text format can not be actually read by humans easily. However, the goal of opening a'.csv' in a text editor is often not to extract information yet to get an overview of the data inside the file, and try to make sense of the localization table using information such as the number of rows, the header for each column, the min and max value for each column etc.

Based on these considerations, we add a complementary text file (here after refer to the manifest file) to describe the localization table and also to store meta information. The manifest file mainly serves two purposes:

1. To save information about the binary format such as the headers of the table, data types,

row number of the table and meta information etc., which are necessary for a reader program to load the binary table;

2. To give a summary of the localization table without loading or unzipping the whole file, which is important when summarizing or displaying a large number of files. It also provides a way for the user to know what is inside the binary file when a reader program is not available.

For the actual implementation of the manifest file, we used the 'json' format which can be easily read by human and program, and it is a widely supported text format in many programing languages such as Python or JavaScript. Figure 2.2(a) shows an example layout of a 'SMLM' file: in the manifest, there many fields including two important ones: 'formats' and 'files'. As its name suggest, the 'formats' field stores a set of formats which describe the header, the data type, and the shape etc., of each column. The 'files' field specifies which format in the 'formats' is used for reading the corresponding file, and also includes the meta information for each binary file. Figure 2.2(b) shows an example manifest file.

2.2.5 Experimental validation on file size and loading time

The major advantage of using the smlm file format is its reduced file size, which is crucial for storing data remotely with a centralized server or using cloud storage provide by companies such as Amazon and Google. To show the advantage of using the 'SMLM' file format, we used an experimentally acquired localization table from a DNA-PAINT experiment. And we used an implementation of the SMLM file format (referred as 'SMLM (ShareLoc)', where ShareLoc is the domain name of the website which serve the web application which support the 'SMLM' file format).

Two file formats supported by ThunderSTORM, the 'CSV' format and the 'TSF' format, were used to compare with the 'SMLM' file format. As shown in Figure 2.3 (a), the 'CSV' file and 'TSF' file generated from ThunderSTORM are more than 2 GB in size, yet the 'SMLM' file is

1.34 GB. Hence, the 'SMLM' file is 5.3 times smaller than the 'CSV' file and 1.6 times smaller than the 'TSF' file.

Since the 'SMLM' file is a zip compressed file, there is an extra decompression step during reading. This could potentially make the loading slower compared to regular uncompressed file formats such as 'CSV' and 'TSF'. To test this, we measured the actual loading time for each file of different formats. As shown in Figure 2.3 (b), the 'SMLM' is still 5.9 times and 2.4 times faster than the 'CSV' and 'TSF' file respectively, despite the extra decompression step. Despite the fact that the loading time is measured with different implementations (ThunderSTORM vs ShareLoc), it still reflects that the unzipping process can be effectively handled.

To demonstrate how the lossless compression ration is achieved, we show the file sizes at different steps during the conversion. In Figure 2.3 (c), the 'CSV' file is converted into a 1.72 GB binary file and then zipped into the 1.34 GB file. By comparing with the zipped files for 'CSV' and 'TSF', it means that the binary file used in 'SMLM' file format is more compressible.

Unlike the 'TSF' format which requires 'protobuf', the 'SMLM' format depends mainly on standard 'zip' and 'json' which will make it easier to implement and more ubiquitous. One important consideration when designing the 'SMLM' format is to be web-friendly, meaning that the format should be easily implemented in a web browser with JavaScript. With our JavaScript implementation which runs in the browser, one can import and export 'SMLM' without any software downloads or installation. The web application also supports filtering rows in the table based on filter expression or a range of column values. After rendering, the histogram can be visualized directly in the web application. As shown in 2.4, both 2D and 3D localization tables can be rendered directly within in the browser.

2.2.6 The flexibility of the 'SMLM' file format

It is very important for a research oriented file format to support different purposes, have the flexibility to adapt to very different applications.

The table formats in 'SMLM' file format are flexible, since the developer can define their own

header names in the manifest file, i.e. each column of the table can have a different data type. For example, a column storing values for the x coordinate can have a name such as 'x', 'X', 'x [nm]' etc. Which one can use an arbitrary name for each column, this will cause problem for the reader program to understand the actual meaning of each column. Most importantly, to render a histogram, the reader needs to parse at least x and y coordinates. For this purpose, we provide a transformation field in the table format definition where one can indicate which column name corresponds to x, and which is y and so on.

One 'SMLM' file can contain any number of binary files (table or image) with the same format or different formats, making the 'SMLM' file format very flexible in terms of the usage. Here are two example applications:

1. Multi-color imaging. For multi-color images, there is typically one table for one color, which can be added into the 'files' field of the manifest file and refers to the same format in 'formats'. Sometimes, multi-color images can include both SMLM and low resolution images. In such cases one just needs to add both, the table format and the image format, into 'formats' and refer to them in the file definition.

2. Tiled Scanning imaging or high throughput imaging. For a high throughput imaging experiment similar to that described in the ANNA-PALM section, many localization tables are reconstructed from different FoVs. Storing these separate tables into one SMLM table could make the sharing and exchange of data much easier. With the 'SMLM' file format, one can just save these tables in separate files and add them into 'files'. In order to render them correctly, each file has a field named 'offset' which can be used to store the spatial offset of each FoV.

2.2.7 Saving raw data in the 'SMLM' file

Obviously, storing and sharing only the localization table is not enough. In many cases, we still need to access the raw diffraction limited images for the purpose of reanalyzing using a different localization software or different settings. For example, the TRABI method [Franke et al., 2017]

allows for existing localization data for 2D to be reanalyzed to retrieve 3D information. Thus, including raw data in the 'SMLM' file format making it a more complete solution for storing and sharing SMLM data.

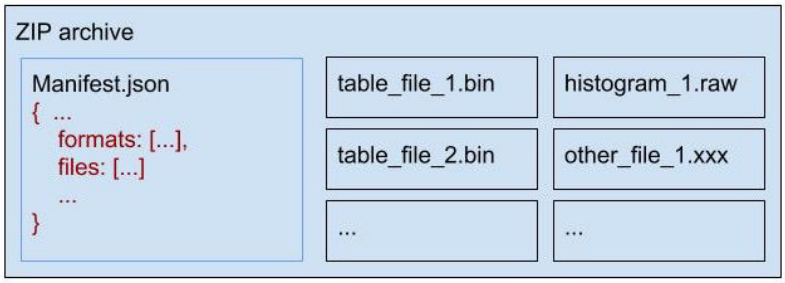
Naively, raw images could be just stored in the zip file, but this is not optimal. In a typical SMLM experiment, the images are visually sparse and could be compressed considerably by a generic image compression algorithm. Despite that using lossy compression such as JPEG can reduce the file size even more, it is typically not used to store microscopy images. In practice, when using lossless compression algorithms, the file size is still not ideal, due to the incompressible noise in the background. On the other hand, most localization algorithms have a pre-detection step which detects all the spots in the image, and exclude the background regions. One simple idea to reduce the file size is to use pre-detection to detect possible spots on the fly, and only save image patches corresponding to the detected spots. With the 'SMLM' format, image patches can be stored in a raw data column. To support this, we extended the definition of a table by allowing one element in the table to contain an image patch with a predefined size. This is implemented within the 'shape' field in the format definition. For regular columns, the shape equals 1, and the shape can have two dimensions (e.g. [10, 10]) for image patch columns.

2.2.8 Data sharing with the ShareLoc platform

With the reduction in file size and the flexibility provided by the 'SMLM' format, storing and transmitting localization data becomes much easier. We therefore built the *ShareLoc* platform based on the 'SMLM' file format and the web application, enabling easy and fast data sharing. Figure 2.5(a) shows the framework of the website. It consists of a) a messaging server called "crossbar" which is responsible for the communication, b) static web pages which provide the web interface, c) a *simple storage server* (S3) which is hosted by Amazon Web Services and d) a database which stores all the data records. The implemented features include authentication through user name and password, loading local localization files in formats including 'CSV' format and 'SMLM' formats, exporting the localization table into 'SMLM' format, and upload to the S3 server either for private sharing with a url or public sharing through a public data

repository. Figure 2.5(b) shows the home page of the web platform.

With the data sharing feature provided by the ShareLoc platform, data can be easily loaded from the server through a url, visualized and rendered within the browser. With the urls generated from the platform, one can easily send localization tables by email or share them through social media. These features will enable remote collaboration by sharing data easily and faster. With the public data repository provided by the platform, more and more public data can be added to the platform, which allows reanalysis of data, accelerates tool development, and also further promotes reproducible research by making data sharing in the localization microscopy field easier than ever.



(a) Layout of the smlm file


```
{
  "format_version": 1,
  "formats": {
    "smlm-table(binary)-0": {
      "name": "smlm-table(binary)-0",
      "type": "table",
      "mode": "binary",
      "extension": ".bin",
      "columns": 3,
      "headers": [ "frame", "x", "y" ],
      "dtype": ["float32","float32","float32"],
      "shape": [1,1,1],
      "units": ["frame", "nm", "nm"],
      "description": "the default smlm format for localization tables"
    }
  },
  "files": [
    {
      "name": "s1-c2-storm-wo-COT_1_output.bin",
      "format": "smlm-table(binary)-0",
      "channel": "default",
      "rows": 2326654,
      "offset": {"x": 0, "y": 0},
      "exposure": -1,
      "type": "table",
      "min": {"frame": 0, "x": 83, "y": 274},
      "max": {"frame": 10001, "x": 56085, "y": 56238 },
      "hash": "81f3d18cfa244ddf126e7334ea4b8572"
    }
  ],
  "tags": ["microtubules", "STORM"],
  "name": "s1-c2-storm-wo-COT_1_output.smlm",
  "description": "",
  "sample": "",
  "labeling": "",
  "hash": "a4d355c122f9f8d6cafe28dd59118444"
}
```


(b) An example manifest file

Figure 2.2: The layout of a 'SMLM' file and a example manifest file. (a) A 'SMLM' file is a zip archive which contains one or many binary files (tables or images), and a manifest file with a fixed name 'manifest.json' describes the file formats and contains meta information for each contained localization table or image files. (b) shows the content of an example manifest file contained in a 'SMLM' file with one localization table.

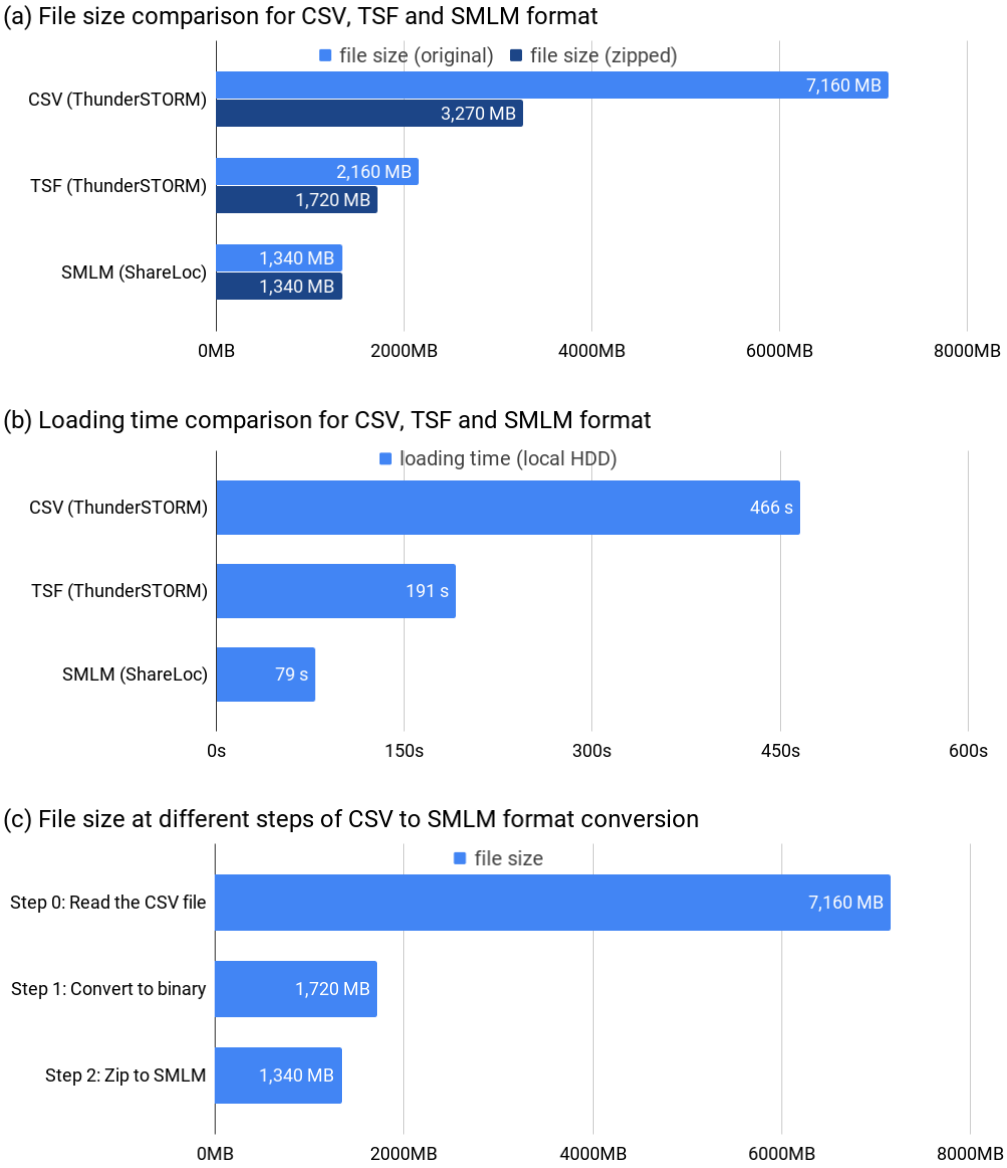


Figure 2.3: Comparison of the file size and loading of different formats, as well as file size of different conversion steps. From a DNA-PAINT (on Mouse cells with Vimentin) experiment, we generate a localization table with 60,000 frames and 47.8 million localizations. (a) The table is saved in two different formats with ThunderSTORM and compare with the corresponding 'SMLM' format. The 'CSV' file is 7,160 MB (1 GB = 1,024MB), the 'TSF' file is 2,610 MB, and the corresponding 'SMLM' file is 1,340 MB. The 'SMLM' file is 5.34x smaller than the 'CSV' file and 1.61x smaller than the 'TSF' file. Notice although both 'CSV' and 'TSF' can be zipped in order to achieve smaller file size, the sizes are still bigger than the 'SMLM' file; (b) To read the files in different formats and softwares, we measure the loading time from the software reading the file to it rendering the image. The loading time are 466s, 191s and 50s for 'CSV', 'TSF' and 'SMLM' respectively; (c) During the conversion from the 'CSV' file to the 'SMLM' file, the 'CSV' file is first converted to a binary file with 1,720 MB stored with float32 data type, and then zipped into a small 'SMLM' file with 1,340 MB. The experiment is carried out on the same computer, by using a ThunderSTROM plugin installed in Fiji to read and write the 'CSV' and 'TSF' file, and the ShareLoc web app runs in a Google Chrome browser for reading and writing the 'SMLM' file.

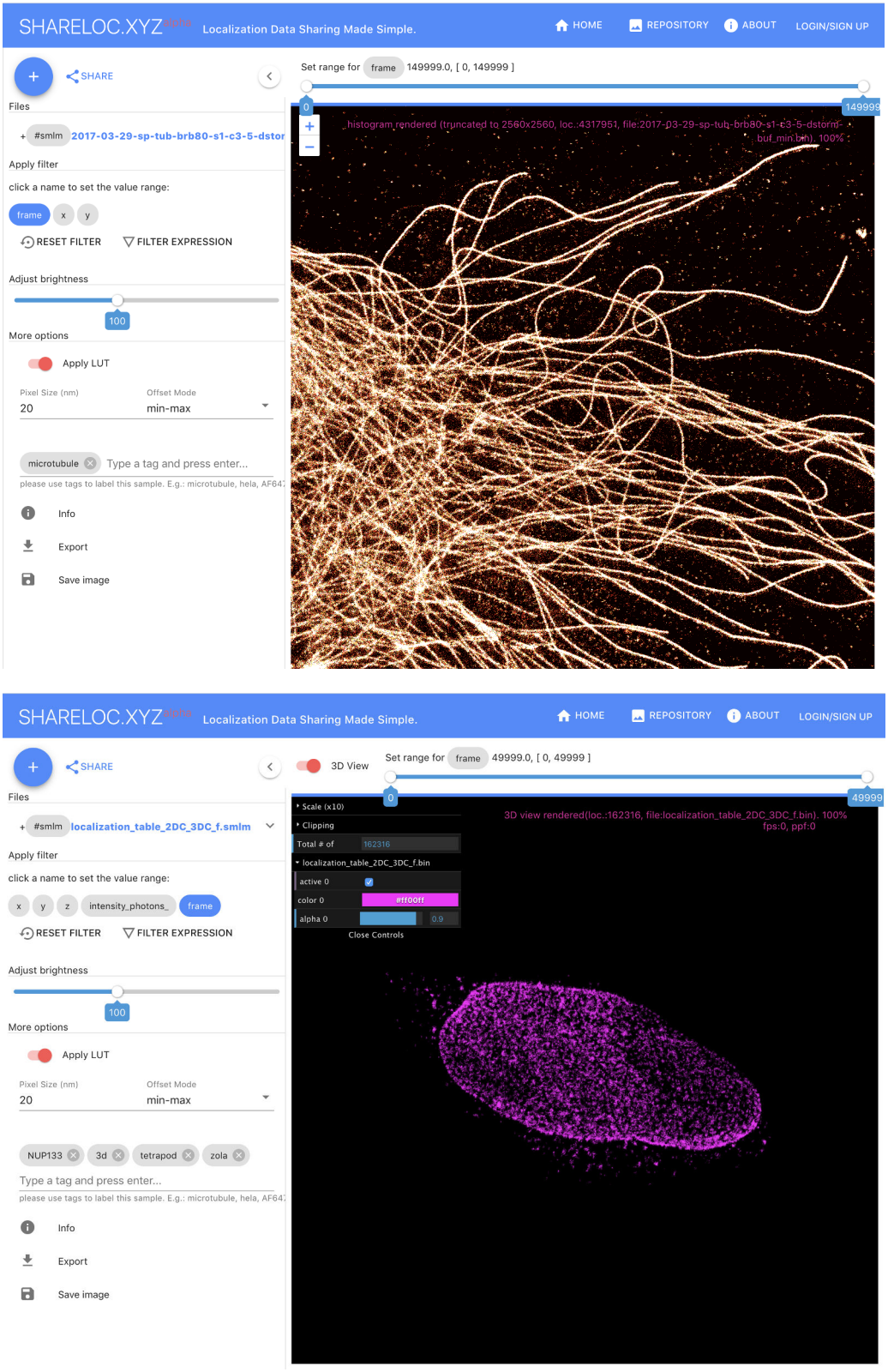
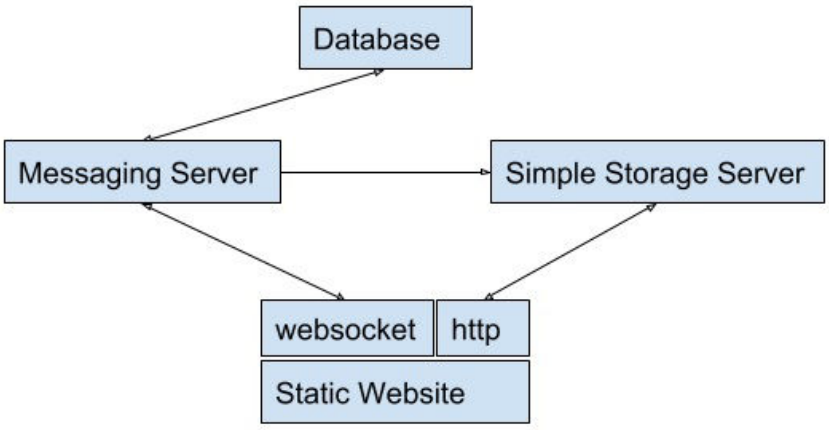
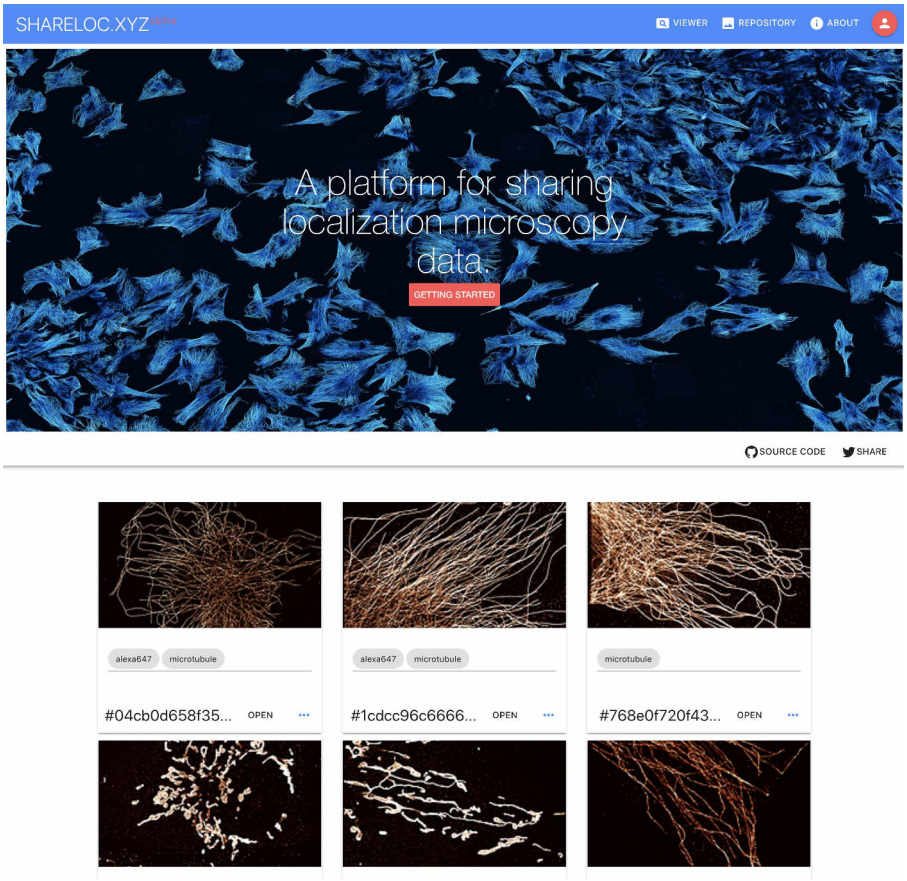


Figure 2.4: Web application for filtering, rendering and visualizing localization tables. The upper part shows a screen shot of visualizing a localization table with microtubules in 2D, and the lower part is a screenshot of visualizing a localization table with nuclear pores in 3D



(a) framework



(b) home page

Figure 2.5: Framework and home page of the Shareloc.xyz platform

Chapter 3

Discussion

3.1 Summary

Advanced microscopy such as SMLM plays an important role in biological research. For getting high resolution images with SMLM, it typically requires thousands of diffraction limited images which limits its applications to fixed biological samples. In this work, we applied deep learning methods to lower the requirements of information acquired during acquisition, which further leads to speed up of the microscopy or quality improvements. Specifically, we developed ANNA-PALM, a computational method based on deep learning that aims to reconstruct high-resolution images from very few low resolution frames. In ANNA-PALM, a deep neural network called "A-net" is trained on example images obtained from less than 10 FoVs using conventional SMLM techniques. During training, a subset of frames are used to generate sparse images which are paired with the dense images generated from a full set of frames. Sparse images are fed into the neural network and the parameters are updated with back-prorogation while minimizing the objective function computed from the corresponding output images and dense images.

We also showed that incorporating widefield images in addition to the SMLM data can dramatically reduce the artifacts in the reconstructed super-resolution images. An artificial neural network that evaluates the consistency of the super-resolution images with the corresponding widefield images, as in the SQUIRREL method [Culley et al., 2018], allows both to estimate

and reduce reconstruction artifacts. We also demonstrated how application of this approach to standard, densely sampled SMLM images, results in images with much fewer artifacts. Thus, deep learning can be used either to accelerate localization microscopy or to improve the quality of reconstructed images.

With ANNA-PALM, we demonstrate that the method can dramatically reduce the acquisition time of super-resolution localization microscopy. Less acquisition time not only saves the labor-time for the experiment, but also it can be transformed into image quality improvement by giving the same amount of acquisition time, reducing the effect of sample drift, or reducing phototoxicity for live cell imaging.

Since it is a learning method, there is no assumption about the imaging method itself, it can be extended to other types of image modalities. A similar approach as ANNA-PALM called *Content-Aware Image Restoration* (CARE) has been independently proposed in [Weigert et al., 2017]. In work of CARE, deep learning methods have been shown to work for denoising of low exposure images, improvement of resolution and so on. Compared with ANNA-PALM, the actual implementation of the technique is slightly different, but the idea under the hood remains the same. That is to say, both methods train neural network models to learn the structure information from images acquired with good condition and use the structural prior to reconstruct images acquired in less ideal conditions. While conventional microscopy techniques struggle with the trade-off between resolution, sensitivity and speed (also sample health for live cell imaging, it is also known as the *eternal triangle* [Resch-Genger and Barker, 2008]), deep learning enabled microscopy breaks the curse of the *eternal triangle*.

Furthermore, deep learning based *image-to-image translation* [Isola et al., 2016] have been used for methods such as ANNA-PALM, CARE, and extended to a spectrum of deep learning enabled microscopy techniques including [Rivenson et al., 2018, Hershko* et al., 2018, Wang et al., 2018, Nehme et al., 2018].

The performance of deep learning methods can be effectively improved by adding more training data. In fact, when the amount of training data reaches a certain level, deep learning models becomes even more data-effective [Hestness et al., 2017], the capacity of deep neural networks

allow them to handle large dataset with millions of samples (e.g. ImageNet dataset). While most experiments in microscopy are done with relatively small amount of training data, increasing the amount of data may generate even more impact in terms of performance. Given the cost of acquiring experimental data, one work-around to the issue is to increase the usage of existing data by data sharing. There have been many microscopy image sharing platform which could potentially boost the performance even further. However for SMLM data, since the localization tables alone are typically several gigabytes in size, it is challenging to store and share the data, and further prevents reanalysis and training machine learning algorithms such as ANNA-PALM. To address this issue, we developed a file format that can losslessly compress localization tables into much smaller files, and a web platform called ShareLoc which allows to easily visualize and share 2D or 3D SMLM data. We believe that this platform can greatly accelerate tool development, data reanalysis and further promote reproducible research.

3.2 Limitations and possible solutions

Despite the success of deep learning methods, practitioners using the method need to be aware of the limitations, including some inherent limitations of deep learning and task specific issues. These limitations affect the quality of the reconstruction. If not controlled properly, the performance can drop dramatically, lead to artifacts in the resulting image which may generate misinterpretation of the underlying phenomenon.

3.2.1 Generalization and memorization

In ANNA-PALM, problems such as artifacts are closely related to the generalization capability of deep learning. One of the most important property of learning algorithms such as ANNA-PALM is their capability to generalize to unseen datasets in the same problem domain. In the context of machine learning, two terms are used to describe the generalization capability to new data, namely overfitting and underfitting. Overfitting means that a learning algorithm performs well on training data but fails on similar unseen data. If the algorithm failed to perform well on both training data and unseen data, it is called underfitting.

A good ANNA-PALM model needs to generalize well from training images to any unseen images with the same type. For example, an ideal ANNA-PALM model trained on microtubule images should perform well on any microtubule images, despite the differences in imaging conditions such as the labeling technique, noise level etc.

Compared to other machine learning approaches, deep learning has been recognized for its good generalization capability despite its large capacity, complexity, instability and sharp minima. Generally speaking, the larger number of parameters the algorithm has, the more it is likely to overfit. However, due to the use of various techniques such as SGD, dropout, ReLU, batch normalization and optimizers, deep neural networks tend to have good generalization capability in practice. For example, ANNA-PALM models trained on STORM images can make good prediction on DNA-PAINT images.

However, why deep learning can generalize well is still an open question. In computer vision, it is widely believed that the generalization is a result of "inductive bias" – it is common to see a CNN discover several levels of features hierarchically similar to the V1 system of the mammalian brain. During training, deep neural network tries to find a hierarchical representation of the input image, which leads to a generalized representation and abstraction of objects or concepts within the input image. Mathematically, [Kawaguchi et al., 2017] gave some insights on why and how deep learning can generalize well. Meanwhile, the flatness of minima of the loss function found by SGD is proposed to be a indicator for generalization. [Keskar et al., 2016] shows that large batch sizes in SGD method can lead to sharp minima which further yields poor generalization. However, [Dinh et al., 2017] suggests that using the flatness of minima to analyze generalization is still problematic, because the minima can be made arbitrarily sharp without affecting generalization properties. Besides that, other intriguing properties may also related to the generalization or failures of generalization. For example, experiments shows that the input-output mappings learned by deep neural networks are significantly discontinuous [Szegedy et al., 2013]. It also inspired attacking methods which aims to design adversarial examples to hijack deep learning systems [Yuan et al., 2017], which could potentially cause security issue when deep learning is deployed in critical applications.

Meanwhile, deep learning is also known for its capability to memorize data. In [Zhang et al., 2016a], the authors show that by randomizing the labels in an image classification task, deep neural networks can easily achieve zero training error. This means deep neural networks are capable of memorizing the labels which can potentially lead to overfitting. However, this does not mean that neural networks only try to memorize the label. When noise is present in the input data or the label, deep neural networks try to capture the signal in the data and fit the noisy part using brute-force memorization [Zhang et al., 2016a, Krueger et al., 2017].

Since ANNA-PALM is powered by deep neural networks, both generalization and memorization play their role in models trained on real-world data. Because the training data is experimentally acquired, both the input images and target images typically contain significant amounts of noise. On one hand, the network will be able to capture the meaningful feature from the signal part which can lead to generalization of certain types of structures. On the other hand, since there

are considerably amounts of noise in the target images, the neural network will have to memorize the random structure or noise in the training images, which may also hurt the generalization. It could be also argued that memorization is not always bad for deep neural networks because it help the network to better approximate the target. Moreover, since ANNA-PALM uses an adversarial loss term, the network will be encouraged to learn the noise pattern in the target image. Due to the randomness of the noise, the network will be forced to memorize the noise distribution in the target image.

3.2.2 Systematic extrapolation

Related to the memorization effect of neural networks, it is known that neural networks do not generally exhibit systematic generalization [Fodor and Pylyshyn, 1988, Marcus, 2001]. They tend to memorize at least some parts of the data rather than to generate systematic abstractions. As a result, neural networks can easily fail to extrapolate data which lay outside the training data distribution, even for very simple numerical arithmetic tasks [Trask et al., 2018].

In [Trask et al., 2018], various neural networks (with different activation functions) are investigated in a task of learning the identity function – given a scalar value, the neural network is asked to output the same scalar value. To test whether neural networks are capable of extrapolation in this simple setting, training values are limited to the range of [-5, 5] while values used for inference are sampled from the range of [-20, 20]. As shown in Figure 3.1, most of the neural networks gave small errors for the training range, but failed outside the training range with large prediction errors. This experiment highlights the limitation of neural networks in terms of systematic extrapolation capability. Even though theoretically neural networks are capable of representing functions which can systematically generalize well, presumably because the training algorithms or objective functions we used in practice failed to encourage systematic generalization.

For higher level tasks, deep learning models used in computer vision tasks including ANNA-PALM exhibit similar behavior on unseen data, i.e., lack of systematic generalization capability. In our experiments, we observed that ANNA-PALM models can easily generate artifacts when a

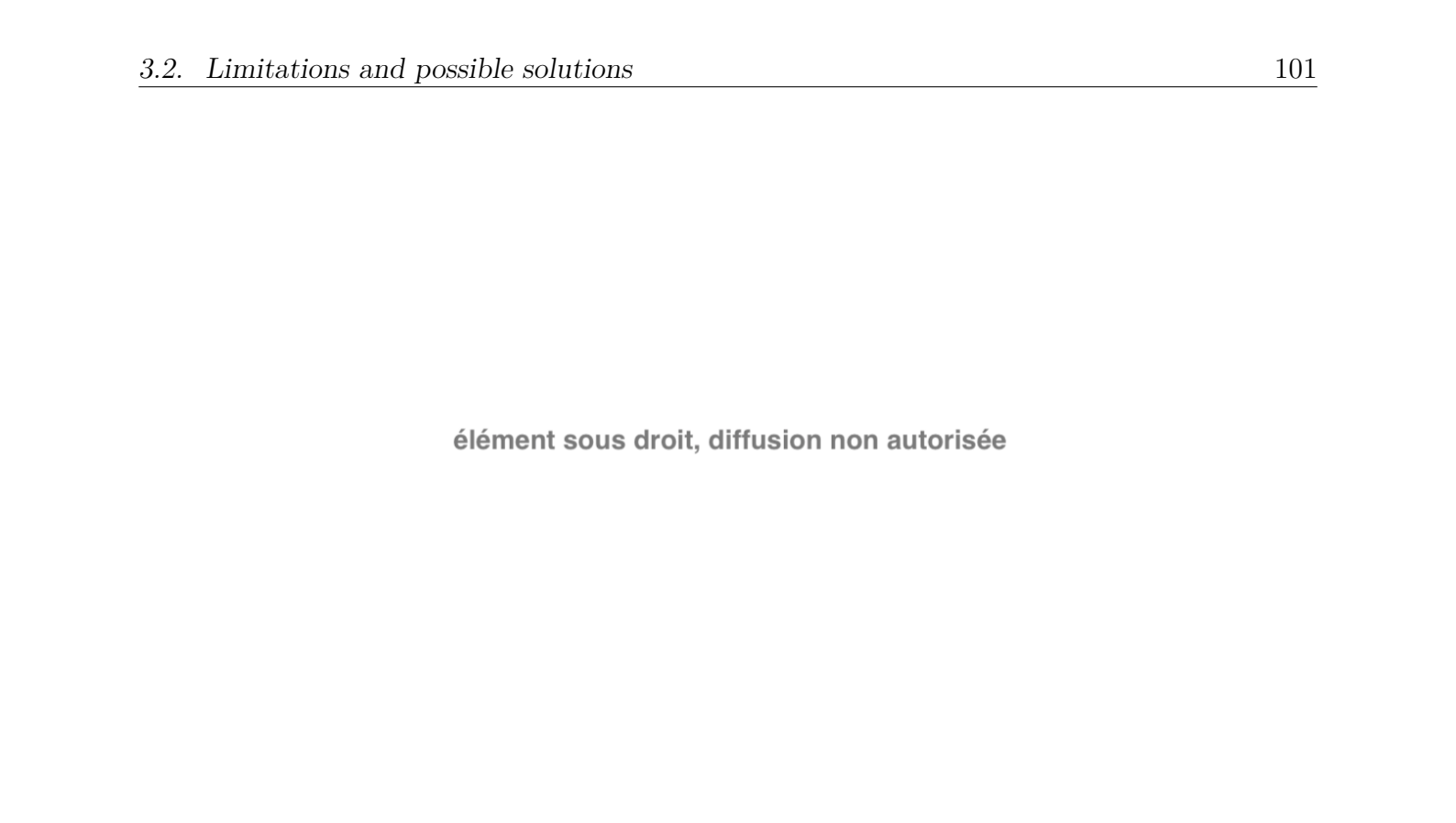


Figure 3.1: Numerical Extrapolation Failure in Neural Networks [Trask et al., 2018]. For the task of learning the identity function, experiments are carried out with MLPs (with various types of activation functions). Small mean absolute errors can only be achieved when the input number falls in the range of values they were trained on (in this case from -5 to 5).

certain type of input image never appeared in the training dataset. For example, with a mean squared error loss function, an ANNA-PALM model trained with very sparse input images failed to reconstruct images with very dense input images, even though the latter task should be much easier.

While solving the extrapolation problem may need more profound changes to deep learning, different ways of building learning models which can systematically generalize have been proposed [Trask et al., 2018, Sahoo et al., 2018]. Most of them define a set of operations or syntax which the model can use for composing a solution for the task. By doing that, systematic generalization can be guaranteed by the operations or syntax provided, for example, [Trask et al., 2018] uses a set of primitive arithmetic operators controlled by learned gates. This type of approach is also called *program induction* or [Devlin et al., 2017] which can be seen as a form of meta-learning [Finn et al., 2017, Bengio et al., 1991]. An even more challenging and more generalized field is called *program synthesis* [Manna and Waldinger, 1980], which tries to generate programs using generic programing languages such as C/C++ and Python. Be-

cause the possible program space is enormously large, the problem itself is extremely difficult. However, there have been successful designs such as Neural Turing Machine [Graves et al., 2014], Differentiable Neural Computer [Graves et al., 2016], Neural Stacks [Grefenstette et al., 2015, Joulin and Mikolov, 2015], and Differentiable Forth Interpreters [Bošnjak et al., 2016] etc. These abstract machines and program induction models shows a promising direction to solve the systematic generalization problem of deep learning.

3.2.3 The complexity of the structure

Another limitation of the deep learning method used in ANNA-PALM relates to the complexity or randomness of the biological structure in the image.

One explanation for why ANNA-PALM can recover missing information is that neural networks capture information about the structure during training. For a successful reconstruction, the neural network needs to encode the structure information into its weights during training. As described previously, both generalization and memorization play their role during the weight update process with SGD. As seen in our experiments, images with simpler structure (e.g.: nuclear pores) and less structural randomness can be efficiently represented and encoded into its weights, which leads to generalization. However for the non-structured part (the noise) or structure with more randomness (e.g.: randomly distributed viral particles without any structure or pattern), it is hard to generalize and the model will need to memorize them by brute-force.

The randomness of different types of image can be characterized with information entropy, which is also known as surprisal and the expected self-information. The information entropy reflects how much information is contained in an image. Figure 3.2 shows an experiment from [Krueger et al., 2017] to measure how many hidden units are needed to achieve the same accuracy for different levels of noise in a classification task. It illustrates the fact that when the data is more random, more weights are needed to capture it. Presumably, this is mainly due to the memorization effect of the neural networks. Information entropy of images without random noise reflects the overall complexity of the structure in the image. Since there is no

Figure 3.2: Performance as a function of capacity in 2-layer MLPs for an image classification task [Krueger et al., 2017]. The plots illustrate that when there is noise in the dataset (inputs or labels), more weights are needed to achieve the same level of classification accuracy.

clear division between noise and signal, a larger capacity of the neural networks is required if the image has higher information entropy. Even without any noise, for a certain type of biological structure images, if the average information entropy is higher, more weights will be needed to encode the structural information.

Since the capacity and the number of weights are limited for a given neural network in ANNA-PALM, it cannot deal with any type of biological image. Disregarding the noise, if the image has higher information entropy, meaning it looks more random, it will be more difficult for the ANNA-PALM algorithm to perform well. In extreme cases, for example, ANNA-PALM can easily fail when used for random viral particle images.

While this limits the range of image types ANNA-PALM can work with, it also affects the size of the training dataset required for achieving a certain level of performance. As shown previously, we only used 1 FoV for training nuclear pore models, 7 FoVs for microtubules models, and 9 FoVs for mitochondrial models. This can be explained intuitively by the differences in the complexity of these structures: nuclear pores have a fixed and repetitive pattern at a small

scale, microtubules and mitochondrial networks are more diverse and random across different scales.

3.2.4 Reliability and uncertainty

Given limitations of deep learning methods, real-life applications such as microscopy image reconstruction require a measure for inspecting whether the model is working in reliable conditions. Moreover, it is often a desired feature to get indications on how much the user can trust the results of inference. When applied for scientific discovery for example, a user of ANNA-PALM would like to know which part of the result image is more likely to contain true structures, or artifacts.

ANNA-PALM uses a method inspired by the SQUIRREL method [Culley et al., 2018] which compares the consistency of the reconstructed imge with an additional widefield image. While this method does provide a way for highlighting possible artifacts (as we showed in Supplementary Figure 11), it has the following limitations:

1. it can only detect large-scale errors. By design, the SQUIRREL approach makes a comparison with a diffraction limited image, which has a resolution of around 250nm. The sensitivity of the method is therefore inherently limited by the resolution of the reference image, which is around 150nm according to [Culley et al., 2018]. In our experiments, we observed that the sensitivity also depends on the background noise level, i.e., the higher the noise level, the less accurate the error map.

2. When a widefield image is used as input to the model (with or without the sparse PALM image), the error which SQUIRREL can detect is minimal. As shown in Section 2 , by using the widefield image, ANNA-PALM can efficiently reduce the artifacts which also reduce their appearance on the error map. Especially when we use the Low-res Reconstruction Error term in the objective function, the network is forced to minimize the inconsistency between the result image and the widefield image. This further makes the error detected from the inconsistency minimized or even not detectable. In other

words, when the widefield is used as input to the model, we cannot generate a useful error map with the widefield image.

Based on these considerations, it is often more practical to use the widefield to image reduce the artifacts rather than keeping it for error estimation. In such cases, the model uncertainty estimation remains as an open question.

In terms of the uncertainties of the model encountered during inference time, there are mainly two types of uncertainties which can be potentially modeled: *aleatoric uncertainty* and *epistemic uncertainty* [Der Kiureghian and Ditlevsen, 2008]. While *aleatoric uncertainty* reflects the noise inherent in the observations which is fed into the model, *epistemic uncertainty* captures the amounts of uncertainty in the model parameters. Given a trained deep learning model, it is very hard to actually model these uncertainties and give a quantitative measurement of them. There have been different approaches which try to model the uncertainty, including Bayesian deep learning, which aims to combine Bayesian approaches with deep learning [Neal, 1996, MacKay, 1992].

While the field is still evolving, a recent work by Kendall et al. proposed a simple alternative way to model these uncertainties [Kendall and Gal, 2017]. The idea is to add a loss attenuation term to the mean squared error loss which explicitly represents the *aleatoric uncertainty*. Meanwhile, this paper proposes to turn on the dropout layer during inference time and use the standard deviation of the output values as *epistemic uncertainty*. This has been formulated as a simple way of doing variational Bayesian approximation in the work. In practice, the ideas from this paper could be implemented into ANNA-PALM as follows:

1. double the output channels, and treat one of them as the result image $\hat{y}_i$, and the other one as loss attenuation term $\hat{\sigma}_i$, which represents the *aleatoric uncertainty*.

2. change the MSE loss from

$$L = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

to the loss with the attenuation term

$$L = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\hat{\sigma}_i^2(y_i - \hat{y}_i)^2 + \frac{1}{2}log\hat{\sigma}_i^2$$

where n is the number of pixels in the output image, and $y_i$is the target image used for training.

3. turn on the dropout layer during inference time, run the inference T times (e.g. $T = 100$) and get the per-pixel variance which represents the *epistemic uncertainty*:

$$\hat{e}_i^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{y}_{it}^2 - (\frac{1}{T}\sum_{t=1}^{T}\hat{y}_{it})^2$$

4. optionally, summarizing these two terms with $\hat{u}_i = \sqrt{\hat{e}_i^2 + \hat{\sigma}_i^2}$, and the final results will be the reconstructed image $\hat{y}_i$ and the corresponding uncertainty map $\hat{u}_i$.

We performed some preliminary experiments with this method using simulated microtubule images. Figure 3.3 shows some results of our experiments, which shows that we can produce an uncertainty map that roughly agrees with the true error map.

However, since the true uncertainty is model dependent and hard to know, it is difficult to evaluate the uncertainty map produced. Therefore, comparing with true error in simulation can only give qualitative impression on the performance of the uncertainty map. When applying to real images, we noticed that the produced uncertainty map can be easily affected by the high frequency noise in the SMLM image and we have no ground-truth uncertainty map to evaluate the performance.

While this preliminary result shows the possibility to apply this approach, future investigations are required to better evaluate this method.

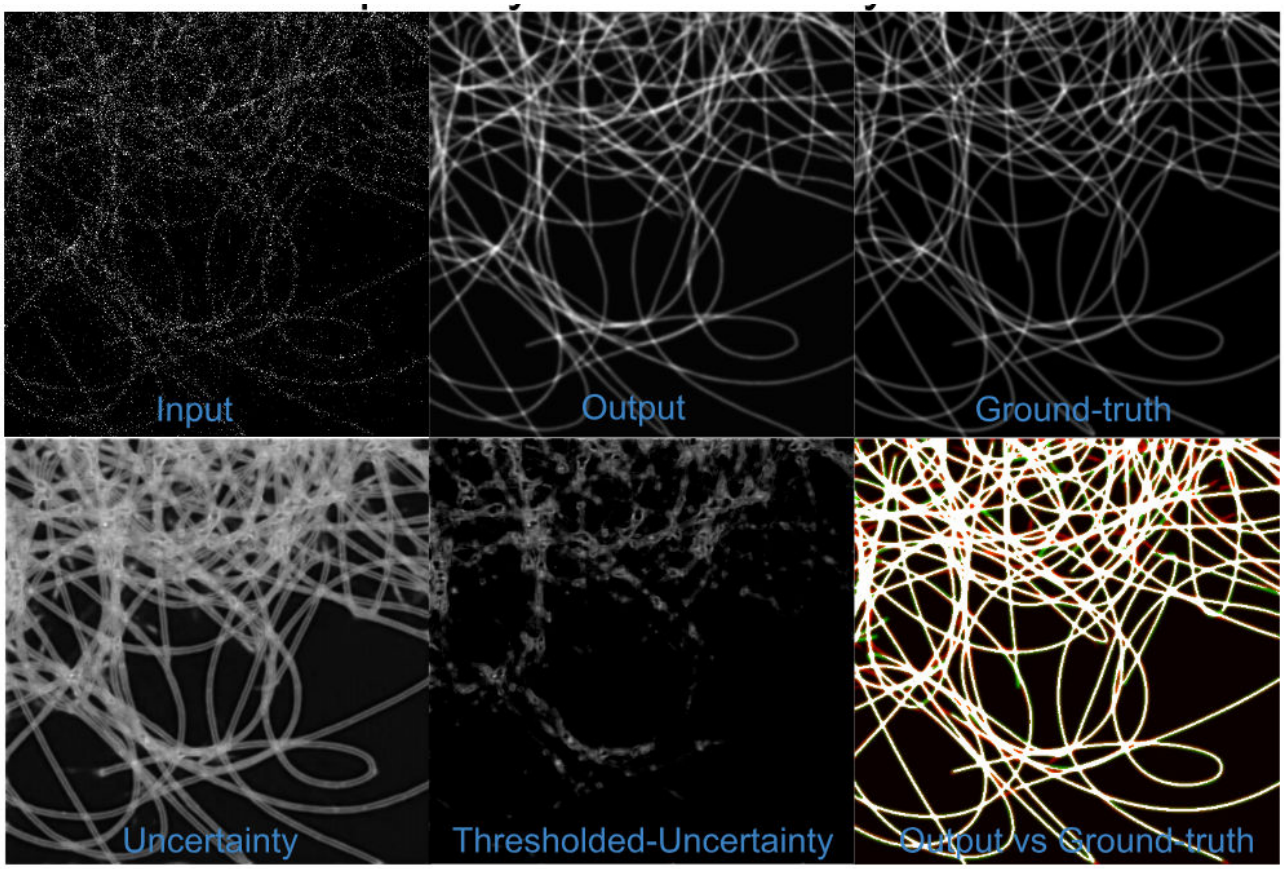


Figure 3.3: An uncertainty map produced with the method described in [Kendall and Gal, 2017] on simulated microtubule images. After applying the method, the neural network can produce both the output image(labeled as Output) and the uncertainty map(labeled as Uncertainty). Removing pixel values below a threshold yields the image labeled with Thresholded-Uncertainty, which agrees with the Output (red) and Ground-truth (green) regions in the overlay images (labeled as Output vs Ground-truth).

3.3 Practical considerations

Despite the limitations of deep learning methods and their blackbox nature, there are practical guidelines supported by extensive experiments in the field. In practice, these guidelines provide considerations which help the user to train deep learning models with good performance.

3.3.1 The impact of training datasets

Given a deep neural network and its training strategy, the performance is mostly dependent on the training dataset. It is well known that data size is crucial for deep neural networks to perform well, i.e. the larger the data size, the better the performance. As discussed in Section 2.2.1, there is a so-called *Power-law Region* where the performance can scale very well with data size [Hestness et al., 2017]. As discussed, deep learning models with large capacity can not only capture abstract information by feature representation, but also memorize the random part of the data. This makes deep learning models able to scale better with the amount of data when compared to most other learning algorithms such a support vector machines.

Besides the dataset size, there are many other properties of the training dataset which can also affect the performance of deep learning methods.

1. Including images with different modalities(e.g. widefield, confocal). As discussed in Section 2, providing an extra widefield image as input to the model can dramatically reduce artifacts and increase the performance. Whenever possible, providing images from another modality can improve the performance. In practice, these images may need to be scaled or registered before stacking them together as different image channels, similar to what we did for the widefield image in ANNA-PALM.

2. The diversity in the training data. Samples in a training dataset can often differ from each other due to variations in parameters such as imaging condition, sample condition, source of sample and acquisition device etc. For example, microtubule images can be acquired with different exposure time, laser power or different microscopes. It is important to

ensure the diversity of these parameters in the training samples, such that the neural networks can sample more variations which can further improve generalization, reduce overfitting and provide more robustness during inference. The overall design principle is to cover those situations which may appear during inference time. On the other hand, for a given model trained with a known dataset, we can roughly tell whether the model can be used for images acquired with a certain condition by inspecting the conditions appearing in the training samples.

3. Sample balance. A point related to diversity is that training samples are typically acquired with a set of conditions that are non-uniformly sampled, and the numbers of samples for each condition are unbalanced. For instance, considering the exposure time in a STORM experiment, a training dataset may contain 9 super-resolution images acquired with 60ms exposure time, but only 2 images with 20ms exposure time. If the training dataset contains more training samples from one condition than another, the trained model will perform better and more sensitivity for that condition. This is a big challenge for real-life applications because training data are always unbalanced, which makes the trained model biased towards conditions that appear more often in the training dataset. This can become a problem especially when we want to train models to capture rare but critical conditions. Nevertheless, there are training strategies to reduce the model bias for imbalanced training datasets, for example, a different step size for weights update in SGD can be used when training with different sample types, or we can feed those rare samples more frequently to train the model.

4. Similarity between the training dataset and the test dataset. Intuitively, if the unseen data is more similar to those used for training, the neural network have higher chance to solve the task with high performance. For a trained ANNA-PALM model, if the input images used during inference are statistically different from those used for training, the performance can drop significantly and result in faulty reconstructions with artifacts. However, obtaining direct similarity measurement is non-trivial in practice, because that will depend on the generalized representation or memorized features within the neural network. It also makes it hard to evaluate the reliability or uncertainty during inference,

which becomes another constraint for real-life applications.

5. Training sequence order. Neural network models trained with back-propagation such as SGD are sensitive not only to the actual content of the dataset, but also the order of the training sequence used for training. For example, instead of randomly choosing training examples at each iteration, a curriculum can be defined by sorting the training sequence according to their difficulties. Specifically, input images with low levels of noise could be fed to ANNA-PALM during the initial stages of training, and the level of noise gradually increased thereafter. This technique is also known as *curriculum learning* [Elman, 1993, Bengio et al., 2009] and is inspired by how humans acquire knowledge. It has been shown that curriculum learning has a positive effect on the performance.

3.3.2 Data augmentation strategies

Data augmentation is an effective training strategy for applying deep learning methods to real-life applications such as SMLM. Because acquiring new data is costly and time consuming, cheap computational methods to increase the amount of data make the method more affordable.

The idea of data augmentation is to generate variations of new data based on existing experimental data, under the constraint that the newly generated data should be still statistically similar to the experimental data. In ANNA-PALM we use data augmentation extensively to reduce the number of training samples required for training. For most of the microscopy images, data augmentation methods such as random rotation, translation, flipping etc. can be applied. Some other data augmentation methods may need to carefully inspected, for example elastic transformation, addition of random noise etc. For example, in Section 2 we showed that elastic transformations can increase the robustness of the model when applied to images with slightly varying nuclear pore ring size. However, extensive elastic transformation may introduce unwanted variations to the image structure. For example, for nuclear pore images, the nuclear pore rings should not be stretched too much, otherwise the model may generate highly stretched and artefactual nuclear pore rings in real applications. A positive example of

successful data augmentation is shown in Figure 3.4, which demonstrates that adding random noise in the input images during training can reduce artifacts encountered at inference time .

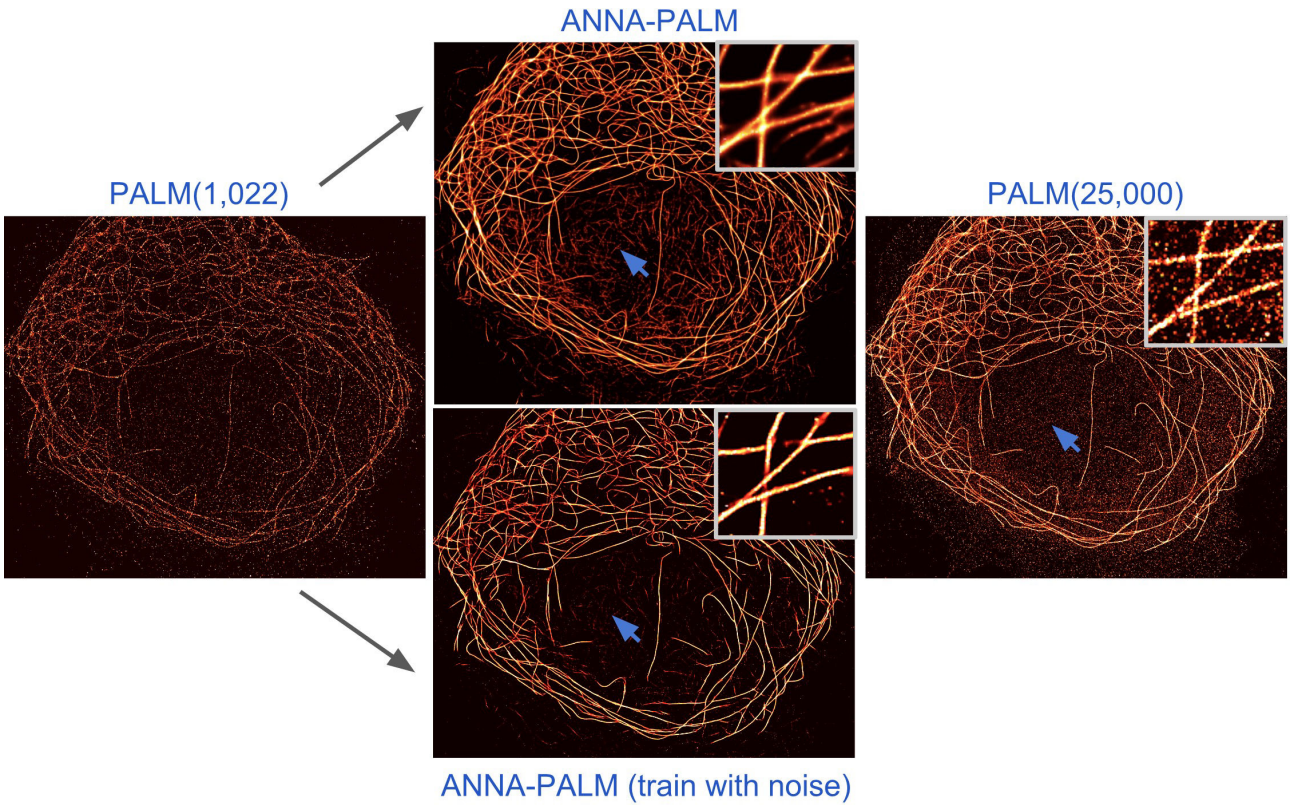


Figure 3.4: Training ANNA-PALM with noise reduces reconstruction artifacts in the background region. An ANNA-PALM model is first trained with images with small amounts of background noise. When applied to images acquired under CA-Rac1 conditions described in [Zhang et al., 2017b], significant amount of artifacts can be seen in the background(the upper image in the second column, shown by the blue arrow). Another model was trained with data augmentation strategy which artificially adds random noise into the background during training. As evident when comparing with the dense PALM image(the right column), the artifacts in the reconstructed image produced with aded noise during training (the lower image of the second column) are dramatically reduced.

3.3.3 Training objective functions

The objective function determines how to measure the error between the output image and the target image. For example, for image reconstruction or regression tasks, one can use *mean squared error* (L2), *mean absolute error* (L1), *structural similarity index* (SSIM), *multi-scale structural similarity index* (MS-SSIM) or *generative adversarial loss* (GAN) etc. The choice of objective function can change the loss landscape dramatically, and have big impacts on the

performance. For example, in Figure 3.5 (e) and (f) the output images look differently when trained with different objective functions, the sensitivity to small structures are different from each other. Therefore, it is important to carefully choose a proper objective function or use the right weights for different terms if the objective function has multiple terms. While it is hard to predict how to design a good objective function, we can use validation datasets to compare the performance for different choices.

There are also some known properties which can be used as a guideline for choosing different objective functions. For example, L2 loss captures the average of all plausible explanations of the data, which produces blurry output images [Mathieu et al., 2015, Pathak et al., 2016, Isola et al., 2016]; using a GAN loss term can improve the sharpness of the output image [Isola et al., 2016]; we also observed that combining SSIM or MS-SSIM with L1 or L2 losses instead of L2 loss alone increases the robustness when the model is applied to images with very different noise levels.

3.3.4 Better reconstructions than the training data

Another observation when applying the ANNA-PALM method is that it is capable of generating results which are substantially better than any of the training images. In terms of signal-to-noise ratio or overall visual quality, it is even possible to generate results better than the best dense image we could possibly acquire experimentally. For example, in Section 2 Figure 6 (b) and (c), nuclear pore images produced by ANNA-PALM are visually better than the experimental results with less noise, the quality improvement was confirmed with *Wheat Germ Agglutinin* (WGA) signal and the quantitative measurement of the number of nuclear pore rings in Supplementary Figure 14. In Figure 3.6, a densely sampled microtubule image is processed with an ANNA-PALM model and produces an image with better visual quality in terms of noisiness (d) than the original version (a). When using both for automated post analysis with SIFNE, a software package for tracking microtubules [Zhang et al., 2017b], using the ANNA-PALM result (d) can significantly improve the robustness of the analysis. ANNA-PALM balances the pixel intensities across different regions and removes the ambiguity in noisy

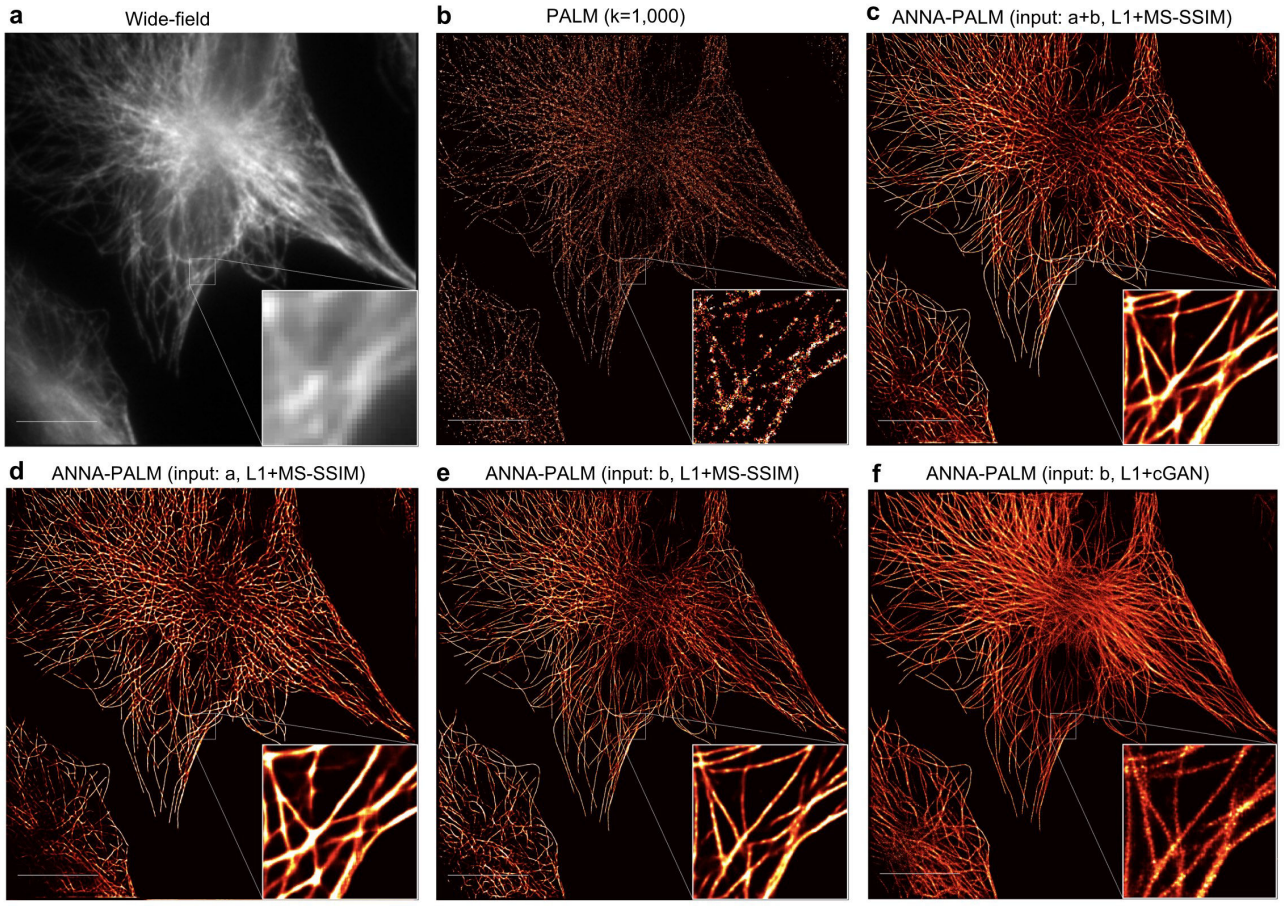


Figure 3.5: ANNA-PALM results with different inputs and training objectives. For images acquired during the high throughput experiment in Section2, different variations of ANNA-PALM methods are applied: (c) combining L1 and MS-SSIM as training loss, use the widefield(a) and the PALM image(b) as inputs; (d) use the same loss as (c) but only use the widefield image(a) as input; (e) use the same loss as (c) but only use the PALM image(b) as input; (f) use L1 and conditional GAN as loss function and the PALM image (b) as input.

regions, allowing SIFNE to recognize more microtubule filaments as shown in (c) and (f).

The improvement over training data implies that the neural network we used is not trying to produce exactly the images provided in the training dataset, but that there are side effects such as denoising during the process. Intuitively, this can be explained by the fact that it is hard to reproduce noise in the output image. Despite the memorization effect of neural networks, most of the incompressible noise is averaged out during training. Theoretically, this behavior is an effect induced by the L1 (or L2) loss terms [Lehtinen et al., 2018]. Specifically, when both the inputs and outputs of the training dataset are corrupted with Gaussian or Poisson noise, by using L2 or L1 in the loss, the neural network model will be able to recover the mean or median of the noisy targets, i.e. the clean targets.

However, in ANNA-PALM, there are other terms in the loss balanced with weights. Among them, the cGAN loss term has an opposite effect to L2 or L1 since it encourages the result image to contain noise, for the sake of realism. Practically, this means that if the training data is more noisy, a bigger weight should be applied to the L1 term rather than the cGAN term.

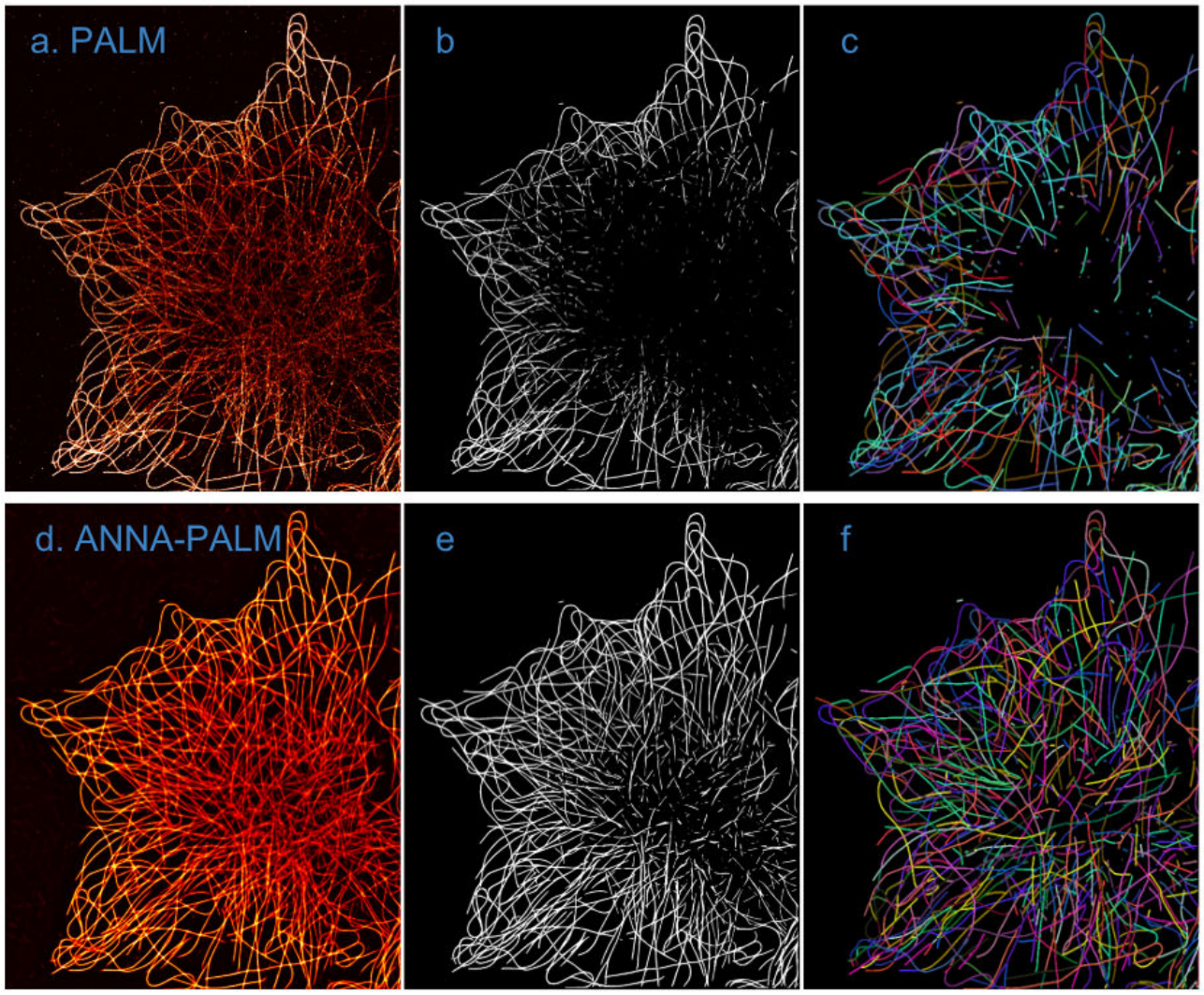


Figure 3.6: ANNA-PALM improves post image analysis with microtubule network extraction software SIFNE. A densely reconstructed PALM image (a) was processed with SIFNE, and it first produce a binary thresholded images (b) and then generate extracted individual microtubules with different colors shown in (c). Meanwhile, using a trained A-net to process image (a) produced a reconstructed image (d), similarly (e) and (f) are the image after thresholding and microtubule extraction with SIFNE. By comparing (b) and (e), more filaments are preserved after the thresholding step with the ANNA-PALM processed image, which further result in more filament extracted in (f) than (c)

3.4 Perspectives

Deep learning has been used to empower many successful applications in different fields despite the limitations discussed above. For research in life science, it enables faster and gentler microscopy as shown with ANNA-PALM. We envision that in the near future, more and more deep learning powered techniques will be developed for advanced microscopy. As potential future directions of the project, we aim short-term goals which could potentially boost the performance of the method or scale it to a larger variety of applications.

3.4.1 Improving the method

Increasing training dataset

In terms of performance, the most promising and straightforward way of improving the method is to increase the amount of training data. As we only used less than ten FoVs in the training dataset, we envision that the performance of ANNA-PALM can be pushed even further by increasing the number of training samples to hundreds of thousands of images. While acquiring that amount of data would not be possible for individuals, with the help of data sharing platforms such as ShareLoc, crowd sourcing large amounts of SMLM data becomes possible.

In ANNA-PALM, we have demonstrated that an additional numerical 'switch' input to the neural networks, we can handle multiple types of biological structures with one single model. An extension of this would be scaling the number of sample types to tens or more. By preparing a single training dataset with larger variations of biological structures and imaging conditions, we expect the generalization capability and robustness of the method to be improved greatly before we hit the capacity limit of the neural network.

Self-supervised learning

As observed in our experiments, training the ANNA-PALM network with noisy images can eventually lead to results as if it was trained with clean data. While the dense images used for

training contain a significant amount of sampling noise, due to the pointillist nature of SMLM, the quality of the images is well above average. That is to say, real-life SMLM image can be much worse in terms of quality (e.g. SNR). Presumably, using average-quality training data would increase the requirements of the amount of data needed to obtain the same performance. It would be worth pursuing the question how much average-quality data is required to obtain a good model. This could lead to the idea of a self-supervised training strategy with only noisy training data.

Giving a set of noisy images, one could construct an input-output image pair by using various methods such as adding Gaussian and/or Poisson noise, introducing random occlusions and mixing up the [Zhang et al., 2017a] samples – all these would encourage the model to generate abstractions or generalizations of the underlying structure. While it is possible that this method could already be used to eliminate the requirement of a relatively clean training image, it could be used as a way to bootstrap or pre-train a neural network before a regular training. This would essentially further reduce the requirement of the big amount of training data.

Domain adaptation

Another attractive field is *Domain Adaptation*, which aims to adapt a model trained on one dataset to another one which is similar to the former. For example, training the ANNA-PALM on synthetic/simulated microtubule images and then use it as a pre-trained model for real data. Domain adaptation, sometimes also called as *transfer learning*, is also of interest because it maximizes the use of existing data when a certain type of training data is not sufficient to train a model or add up to an existing dataset to make a bigger dataset. This method has been successful previously and is commonly used in the image classification tasks– where one often takes a model trained on large dataset such as ImageNet [Deng et al., 2009] and then fine-tune it with a relatively small training dataset. While transferring between similar datasets make sense, because the representations can be similar, it shows that it also can improve the performance when the images are substantially different, e.g. from natural images of the ImageNet to skin cancer images [Esteva et al., 2017]. There are other works which propose a more advanced

form of domain adaptation. For example, [Ganin and Lempitsky, 2015] use backpropagation to make use of unlabeled data. Such domain adaptation strategies can also be adopted within the ANNA-PALM method possibly further improving it.

Meta-learning

Another potential direction which also tries to exploit the learning capability across multiple domains is *meta-learning*. This approach indicates whether the learning algorithm has the ability to acquire knowledge versatility. Many different types of meta-learning models exist: few shots meta-learning [Santoro et al., 2016], optimizer meta-learning [Andrychowicz et al., 2016], metric meta-learning [Snell et al., 2017], recurrent model meta-learning [Duan et al., 2016] and *model-agnostic meta-learning* (MAML) [Finn et al., 2017]. While all these methods dramatically reduce the number of training data required, many of them can utilize data across different domains efficiently. Among them, there is a more scalable meta-learning algorithm called *Reptile* [Nichol et al., 2018], which works by repeatedly sampling a task, performing a stochastic gradient descent on it, and updating the initial parameters towards the final parameters learned on that task. Reptile can be useful for ANNA-PALM when many types of images with different biological structures are available; it may help to train the model to a state which can be quickly adapted to different types of structures. As it algorithmically simple to implement, it provides a promising way to help the ANNA-PALM method to scale for more types of images. Potentially meta-learning can be integrated with the ShareLoc platform and benefit from the versatility of images that this platform will provide.

3.4.2 Extended applications

Improving standard SMLM images

Standard SMLM imaging experiments typically need considerable efforts to optimize the sample preparation or imaging protocol in order to achieve high-quality SMLM images. That is because to obtain a super-resolution image involves many steps including sample preparation

(fixation and labeling), image acquisition, detecting localization events with a software, filtering of the localization events and rendering. Since this is a chained process, any step in the workflow will have an effect on the final image. The difference in quality of the final images therefore result from different fixation protocols, labeling strategies, optical setups, cameras or different configurations of the localization software. For example, when imaging microtubules, addition of glutaraldehyde in the PHEMO fixation buffer provides better preservation of the structure without compromising antigenicity [Goldspink et al., 2017]. A successful SMLM imaging experiment also depends on many factors such as the sample condition [Whelan and Bell, 2015], photophysics of the fluorophore and the imaging condition. Although the influence of these factors on SMLM experiments have been extensively described, due to practical constraints such as limited time and resources, it is often difficult to optimize all of these factors. As a result, many SMLM experiments suffer from common issues such as unspecific binding of the antibody, out-of-focus background, drift and misconfiguration of the localization software. All these issues ultimately lead to images with less resolution, lower quality, or, even worse, can produce artifacts that can cause misinterpretation of the underlying biological structure or phenomenon [Culley et al., 2018].

As discussed in Section 3.3.4, ANNA-PALM, which uses a *L1* or *L2* term in the training objective could potentially be trained on a relatively noisy dataset and yield a better image quality on the output during inference. It is worth exploring the capability of ANNA-PALM to process standard densely sampled SMLM images, not only for denoising but also for solving many different problems that can occur in a standard SMLM experiment, including the removal of the artifacts caused by sample preparation, camera noise, drift etc. As an example, ANNA-PALM could be used to train on SMLM images acquired with cheap cameras or even with a mobile phone [Diederich et al., 2018] and achieve better image quality.

3.4.3 Live cell imaging

In general, live cell imaging is challenging, especially for super-resolution. To keep the cells alive and in a physiologically relevant state, the imaging conditions have to be strictly maintained

during the experiment [Tinevez et al., 2012]. This makes live cell imaging even harder because it adds another constraint to the *eternal triangle* (as shown in Figure 1.7). Live cell imaging experiments need to deal with the trade-off between spatial resolution, sensitivity, sample health and speed. However, if strong prior information can be acquired and applied to the imaging, the curse of the *eternal triangle* can possibly be broken.

With deep learning, prior information about the biological structure can be learned from training data and encoded into the weights of the neural network. By applying these pieces of information implicitly during inference time, the amount of information required from the actual biological sample being imaged can be strongly reduced. Consequently, acquisition time, laser power, exposure time, and thus sample health can all be improved.

Until now, in the ANNA-PALM approach, the training images are required to be from the same type of biological objects in the same or similar conditions. By applying data augmentation on the acquired data to simulate cases of low illumination power, short exposure time and quick motion, the network could be trained to recover the signal. This process will encourage the network to generalize to more stringent imaging conditions, in the meantime memorization will also occur. Nevertheless, the trained network then can be used for recovering images or movies acquired with low laser power, short exposure time etc.

In addition to that, if the same object can be imaged with two or multiple image modalities such as widefield and SMLM, the deep learning methods can also be used to improve spatial resolution. To pursue this idea, we have performed a preliminary experiment. As shown in Figure 3.7, the deep neural network(A-net) was able to reconstruct the super-resolution image from a noisy widefield image. However, due to the low SNR and large uncertainty within the widefield image, it is an under-determined inverse problem, i.e. one widefield image could correspond to a large number of super-resolution images. When the reconstructed super-resolution images are assembled into a movie, filaments are frequently appearing and disappearing when playing the movie. This phenomenon can also be seen in the lower part of Figure 3.7, it shows a *region of interest* (ROI) with 5 consecutive frames of the A-net reconstructions. This indicates that these results are highly uncertain and many artifacts are produced by the network. There-

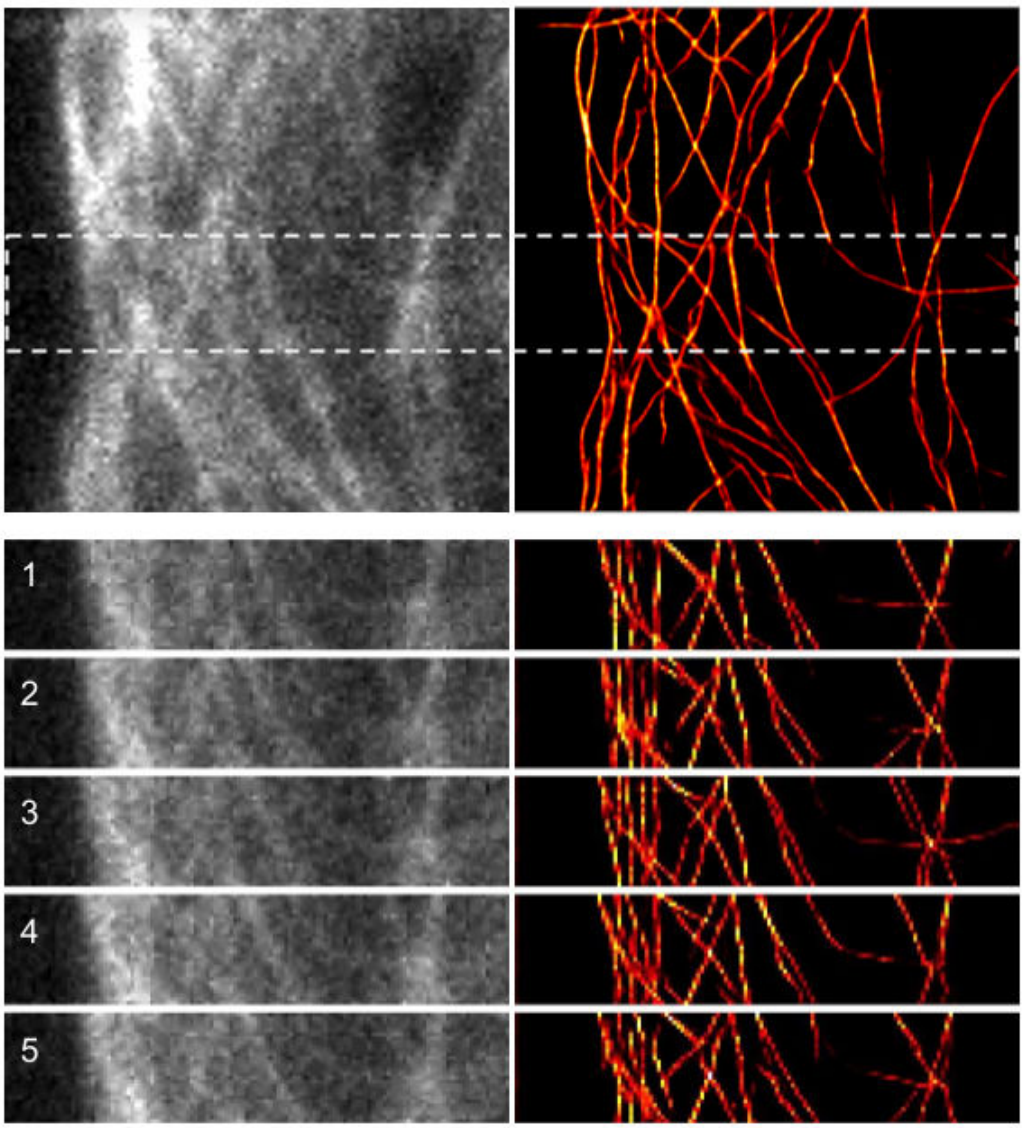


Figure 3.7: Preliminary result on live cell image reconstruction from widefield movies using A-net. For training an A-net for live cell image reconstruction, U2OS cells were stained with TMR and an antibody coupled with Alexa 647 targeting the same structure for acquiring widefield and STORM image pairs. An A-net was trained on these widefield-STORM image pairs in order to recover super-resolution images from low resolution images. For testing, a movie consists of 500 frames with 200 ms exposure time were taken from live U2OS cells labeled with TMR. The upper part of the figure shows the A-net prediction of one frame. When the widefield frame (left) was fed into the trained A-net, it was able to produce the microtubule image (right) with much higher resolution. A-net predictions of a region of interest (dash line in the upper part) for another five consecutive frames are shown in the lower part of the figure, the numbers on the left (1 to 5) are the frame numbers with an interval of 200 ms.

fore, more work is needed to improve the performance of this method. The potential directions for solving this problem are:

1. reduce the target resolution by under-sampling the super-resolution image with binning.

This would make the problem easier to solve or less under-determined, and hopefully yield more stable images across the frames.

2. redesign the neural network architecture by making use of the temporal continuity of the biological structure in a movie. Currently, the A-net used is based on a fully convolutional neural network, which cannot make decisions based on information across different frames. Introducing recurrent neural network layers such as long short-term memory units [Hochreiter and Schmidhuber, 1997] could transform to a recurrent convolutional neural network which can handle spatio-temporal data, such as a widefield movie.

3. including different new image modalities such as confocal, STED or RESOLFT [Hell, 2005] microscopy. STED or RESOLFT microscopy make it easier to acquire live cell super-resolution images than the PALM/STORM approach. One could therefore acquire confocal-RESOLFT movie pairs to train the network, and then use these for improving the resolution of standard confocal microscopy in live cell experiments.

3.5 Conclusion

In the context of advanced microscopy, we have explored deep learning based methods for Super-resolution Localization Microscopy. In this thesis, we have used the deep learning based method ANNA-PALM which learns structural information from densely sampled examples and reconstructs high quality images from sparsely acquired images. Our method significantly accelerates the acquisition time, and so reduces the requirements for drift control, laser power and buffer conditions. With validation experiments, We have shown that ANNA-PALM can achieve up to a hundred fold acceleration, which significantly reduces the expense of imaging experiments and enables high throughput imaging and screening with SMLM. We envision that ANNA-PALM can be applied to live cell imaging experiments, which commonly can be extremely difficult for conventional SMLM. Meanwhile, the neural network architecture A-net and its training strategy that was developed for ANNA-PALM can be directly used for other image restoration or inverse problems. In a longer term, the approach of using a deep neural network for transforming image(s) to image(s) can be established in a more systematic way, such that it can be applied to a broader range of applications. Without changing the deep learning algorithm itself, different ways to generate training data can be extended for more applications. For example, using training data with different combinations of imaging modalities (e.g. SIM, STED, EM etc.) can lead to improvements in resolution and speed; using data from *Correlative light-electron microscopy* (CLEM) can assist in colorizing EM images (similar to *in-silico* labeling [Christiansen et al., 2018]) or producing images with EM resolution based on light microscopy data. Recent works, including ANNA-PALM [Ouyang et al., 2018], CARE [Weigert et al., 2017] and others already moved in this direction. Although these efforts are still at an early stage, they will likely converge into a more established framework for processing microscopy images with deep learning. Such a framework will bring current microscopy to a new era of *Augmented Microscopy*, and enabling augmented view of the cellular environment with colorized cellular elements, enhanced speed and resolution.

Although most of the current deep learning models are lacking of systematic generalization capability, efforts are made in various directions. Some of these are neural abstract machines,

probabilistic programing [McCallum et al., 2009] and program induction [Lake et al., 2015]. These efforts try to exploit the capability of deep neural networks to perform tasks using trainable memory abstractions. Along with many successful designs, these methods show a great potential for supporting long-term memory, improving the interpretability, and providing systematically generalized solutions. They can be applied to microscopy imaging as a powerful, robust and interpretable alternative to current deep learning algorithms. Furthermore, bridging deep learning with more flexible and transparent analytical methods or physical models [Geffner, 2018] can also potentially lead to a more robust and intelligent Augmented Microscopy in the near future.

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Systems Biology

The imaging tsunami: Computational opportunities and challenges

Wei Ouyang[1,2] and Christophe Zimmer[1,2]

Abstract

Microscopy has fueled biological discoveries for centuries, but technical progress during the last decades has greatly expanded the type and quantity of biological information that can be revealed by imaging. Advances in instrumentation, labeling and computation are driving an imaging data tsunami on par with that of DNA sequencing. We illustrate this trend on five data intensive microscopy techniques and discuss some of the challenges and opportunities that massive imaging data raise for our understanding of biological systems.

Addresses

[1] Institut Pasteur, Unité Imagerie et Modélisation, 25 rue du Docteur Roux, 75015 Paris, France

[2] UMR 3691, CNRS, C3BI, USR 3756, IP CNRS, France

Corresponding author: Zimmer, Christophe (czimmer@pasteur.fr)

Current Opinion in Systems Biology 2017, **4**:105–113

This review comes from a themed issue on **Big data acquisition and analysis (2017)**

Edited by **Pascal Falter-Braun and Michael A. Calderwood**

For a complete overview see the Issue and the Editorial

Available online 3 August 2017

http://dx.doi.org/10.1016/j.coisb.2017.07.011

2452-3100/© 2017 Elsevier Ltd. All rights reserved.

Introduction

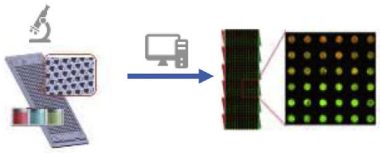
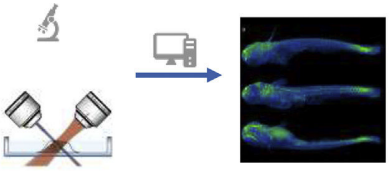
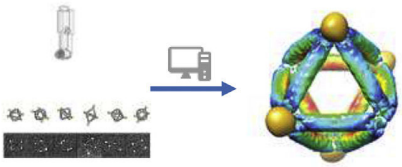
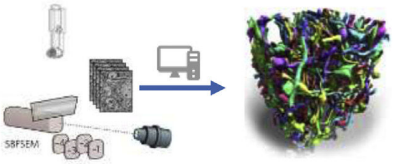
Microscopy has been at the forefront of biological research since centuries, and has enabled major breakthroughs, such as the discovery of microorganisms in the 17th century or that of chromosomes in the 19th century. In those times, analyzing microscopy data meant that an investigator looked through the microscope and drew sketches to report the observations. The digital images produced by modern microscopes are still most often analyzed by visual inspection, sometimes augmented by manual, computer assisted, quantification. However, powerful imaging technologies are now on the rise, that outgrow, often vastly, the capacity of manual analyses and human inspection. These technologies allow us to observe the ultrastructure of cells at molecular or atomic resolution, study the dynamics of single molecules, analyze the effect of silencing every single gene in the genome, or monitor the development of entire organisms at subcellular detail. The amount and complexity of these data scream for automated analysis, and harvesting the quantitative information that they contain has much to offer for computational models of biological systems. In this perspective, we highlight five microscopy approaches (Figure 1) that exemplify the production of massive imaging data, and discuss some emerging trends and challenges (Figure 2).

Single molecule localization microscopy

A major advance in light microscopy over the past decades has been the advent of super-resolution methods, which have pushed the limit of resolution due to diffraction from ~200–300 nm down to ~20 nm or less. For a given field of view, super-resolution images contain at least ~100 times more pixels than conventional microscopy images (~1000 times more for 3D images). An important and widely popular class of super-resolution methods include those called photoactivated localization microscopy (PALM), stochastic optical reconstruction microscopy (STORM) or points accumulation for imaging in nanoscale topography (PAINT) [1–3], These methods are based on the accurate localization of individual molecules undergoing stochastic blinking or binding and are often collectively designated as single molecule localization microscopy (SMLM). In recent years, SMLM has driven a range of discoveries in cell biology, such as deciphering the highly organized architecture of focal adhesions into distinct protein-specific layers [21], identifying a periodic structure of cytoskeletal proteins in neuronal axons [22], or discriminating among structural models of the nuclear pore complex [23].

In order to obtain a single super-resolution image, SMLM typically requires ~10^4–10^5 diffraction limited (i.e. low resolution) images, thereby increasing the data volume to ~100 GB (10^{11} bytes) per experiment; with fast, kHz-rate cameras, the image throughput can reach Terabytes (TB, or 10^{12} bytes) per hour (Figure 1a). These raw, low resolution images, each of which only shows a few fluorescent spots, are of little use on their own. In order to obtain super-resolution visualizations, the precise, subpixelic coordinates of individual molecules must be determined from these spots, which can easily number in the tens of millions. Manually determining these coordinates with high precision is impossible, and therefore algorithms have been used from the beginning of SMLM [1–3]. Although computing the position of a molecule from a fluorescent spot might seem like a mundane task, performing this precisely and reliably is critical for the quality and resolution of the

Figure 1

	Imaging technique		Data production rate	Example references
a	Single molecule localization microscopy (SMLM)		~10 GB/h-1TB/h	[1][2][3][4][5][6]
b	High-content image-based screening (HCS)		Up to ~100 GB/h	[7][8][9][10]
c	Light-sheet microscopy		~1-10 TB/h	[11][12][13]
d	Cryo-Electron Microscopy (cryo-EM)		Up to ~1 TB/h	[14][15][16]
e	Serial block face scanning electron microscopy (SBFSEM)		~1 TB/h	[17][18][19]

Five data intensive microscopy techniques. We highlight five microscopy techniques that generate vast amounts of imaging data: **a**) single molecule localization microscopy (SMLM) [1–6], **b**) high content screening (HCS) [7–10], **c**) light sheet microscopy [11–13], **d**) single particle electron microscopy at cryogenic temperature (cryo-EM) [14–16], **e**) serial block face scanning EM (SBFSEM) [17–19]. The indicated data production rates are orders of magnitude estimates of current state-of-the art techniques. For comparison, a high-end DNA sequencing instrument (Illumina X Ten) can produce ~60 GB/h of sequence data [20] and the rate of videos uploaded to YouTube is on the order of ~1 TB/h. 1 MB = 10^6 bytes; 1 GB = 10^9 bytes; 1 TB = 1,000 GB = 10^{12} bytes (corresponds roughly to 500 h of HD-TV video); 1 PB = 1,000 TB = 10^{15} bytes (~13 years of HD-TV video). Images are reprinted with permission from Nature Publishing Group (b, d, e left), the American Association for the Advancement of Science (c) and Elsevier (c, e right).

final image. The intrinsic stochasticity of single molecule images, which arises from photon counting noise and other factors, makes absolute precision impossible. While a variety of ad-hoc algorithms have been developed over the years, state-of-the art methods usually frame single molecule localization as a statistical detection and estimation problem. This framework provides a sound basis to derive fundamental performance bounds, such as the Cramer-Rao limit to localization precision [24], and to design statistically optimal algorithms that approach the fundamental limits, such as maximum likelihood estimators (MLE). Despite this common theme, algorithms still differ in many ways, including the choice of optimization procedure to calculate the MLE, the assumptions about the microscope's optical properties, image noise, the background, drift, fluorophore photophysics, etc. Several tens of competing software solutions have been proposed so far, creating the need for an objective and unbiased comparison. As is now common practice for many applications of image processing, an online challenge was organized (as part of the ISBI 2013 conference), where different algorithms could be quantitatively compared against a common simulated ground truth, according to criteria such as localization precision, detection recall, speed, and ease of use [25]. This competition is a useful

milestone for practitioners of SMLM, facilitating the choice of the best performing software solutions. However, this competition is by no means a final word, since many challenges remain to be addressed, such as multicolor or 3D SMLM, and as new algorithms steadily continue to appear, in part to accompany the ever expanding palette of optical systems [5,26,27]. New installments of this competition are therefore, already underway.

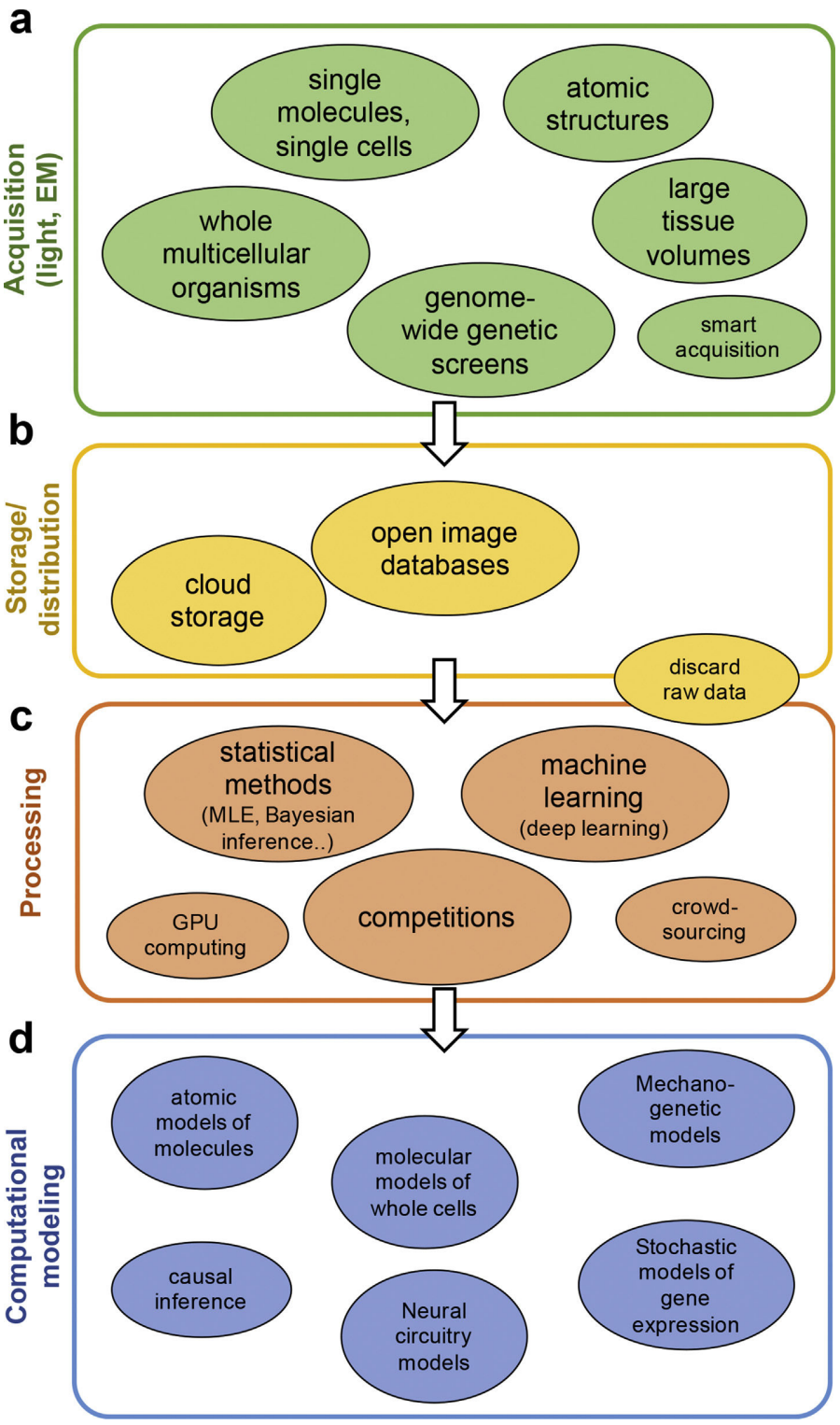
Processing SMLM data does not necessarily end with the display of a super-resolution image. Molecular localizations can be further analyzed in various ways to yield more quantitative information. For example, pair-correlation analyses allow to test subdiffraction protein clusters for spatial randomness and measure their size [28]. Algorithms based on kinetic photoswitching models can count single proteins, or measure molecular orientations [29,30]. An even larger realm for downstream analyses is opened by live cell experiments, when molecules are followed over time, as in sptPALM, where thousands of molecular trajectories are obtained [6]. Tracking moving molecules is more difficult than localizing immobilized molecules, and various methods have been proposed, often based on different ways to link detected molecules across time points [31]. The ability to extract many thousands of individual molecular trajectories from an sptPALM experiment has inspired the development of statistical methods, often based on Bayesian inference, to determine biophysical properties such as transport states, diffusion maps or potential energy landscapes [32,33]. Such analyses can provide useful biophysical information on molecular dynamics, for example by determining how frequently mRNP cargos switch between directed transport along microtubules and free diffusion [32], or by measuring the interaction energies between molecules [33]. These parameters can then inform stochastic simulations of these biophysical processes. Combinatorial labeling and spectral discrimination techniques begin to enable the imaging and quantification of tens of proteins and thousands of RNA species in the same cells [34–36]. Such methods open the door to spatially resolved transcriptomics and the construction of models of stochastic gene expression and gene regulatory networks in single cell level [29,37,38]. Future developments may allow to image multiple protein or RNA species simultaneously at the single molecule level. When this will be extended to live cells, it should become possible to infer causal relationships and to build predictive dynamic models of molecular pathways in their cellular context.

Image based screening

A mainstay approach in genetics is to mutate or silence an individual gene, and characterize the resulting phenotype by microscopy. High content screening (HCS) can scale up such studies to entire genomes using automated microscopes and sample preparation in combination with tools to systematically manipulate gene expression, e.g. RNA interference or CRISPR/Cas9 genome editing (Figure 1b) [9,39]. HCS projects can generate tens of TB of data, especially when dynamic imaging is performed on live cells. This was exemplified by a massive study, in which each of 21,000 genes was individual silenced by siRNA, and cells were monitored by video microscopy for 2 days. The resulting data set contained almost 20 million cell division events and was used to identify hundreds of genes, more than doubling the number of genes known to be involved in this fundamental cellular function [7]. Visual inspection of all resulting movies would have taken hundreds or thousands of hours and lack objectivity and reproducibility. Similar difficulties can arise for static data, e.g. images of fluorescently labelled RNA [8,40]. Therefore, images from HCS studies are usually processed by algorithms that transform them into large sets of quantitative descriptors called features. These features may include biologically meaningful quantities such as nuclear size or transcript numbers, or less readily interpretable measures such as image texture or entropy. The features are then typically fed to machine learning algorithms that seek to associate images to particular phenotypes. In supervised machine learning, a set of example images are manually assigned to previously defined phenotypic classes, such as cell division defects [7] or a polarized distribution of RNA [8,41]. Next, the algorithm learns to automatically determine the phenotype from appropriate combinations of the features, which are then used to phenotype images taken under new experimental conditions. While support vector machines and random forests have been used successfully, methods based on deep artificial neural networks (deep learning) have gained renewed prominence in recent years [43]. A key advantage of deep learning methods is their ability to automatically learn suitable features directly from the images ('end-to-end learning'), which generally outperform standard preimposed feature sets. This was illustrated by a recent study where deep convolutional neural nets were employed to assign cell phenotypes in response to drug treatments. This deep learning method was shown to outperform previous methods in 5 out of 8 benchmark data sets and to correctly quantify drug potency [10]. In unsupervised machine learning, features can be learned on training data without phenotypic annotations and used to cluster similar images together, thereby defining groups of functionally related genes, or compounds with similar phenotypic effects, or detect anomalies [7,8]. Deep neural nets such as autoencoders provide a powerful means to learn appropriate features for such tasks [44,45].

A common criticism of statistical approaches and machine learning is their supposed inability to address

Figure 2



Trends in biological imaging and computational analysis. Bubbles highlight selected trends in biological imaging, from image acquisition (**a**), to storage (**b**), processing (**c**), and modeling (**d**), most of which are discussed in the text. Arrows indicate the direction of information flow.

causal links, e.g. between molecules and phenotypes. While correlation does not imply causation, the reverse implication holds true, hence screening for correlations is at the very least an efficient way to filter out potential causal links that can be tested by other experimental means. In addition, tools such as belief networks can be used to determine conditional links between variables and hence approach causal understanding [9]. Furthermore, video-microscopy data allows to dissect the temporal hierarchy of molecular factors involved in a given cellular process by analyzing correlations between stochastic image fluctuations. This was demonstrated for example in a study of actin assembly at the cell membrane, that determined the formin mDia1 as the initiator of cell protrusions, and Arp2/3 recruitment as a downstream step [46]. Such analyses do not require experimental perturbations and can reveal the contribution of individual molecules to redundant pathways in a more truthful way than standard genetic perturbation experiments, whose interpretation can be obscured by adaptation mechanisms [47]. It will be interesting to see how kindred approaches can be combined with state-of-the-art machine learning methods to help uncover mechanistic links between genotypes and phenotypes in image-based screening.

Light sheet microscopy

The study of developing organisms demands dynamic imaging of living multicellular systems at high spatiotemporal resolution in physiological conditions. This has long been hampered by photobleaching, i.e. the rapid decay of fluorescence due to intense laser illumination of the whole sample, and the associated toxicity due to photodamage. Light sheet illumination techniques (also known as single plane illumination microscopy, or SPIM), which restrict laser excitation to the focal plane have essentially removed this hurdle and enabled the 3D observation of living multicellular organisms in physiological conditions over several days or more [48,49], (Figure 1c). SPIM has allowed, for example, to monitor developing zebrafish embryos with sufficient spatiotemporal resolution to follow the lineage of thousands of cells [50]. Modern light sheet microscopes can produce imaging data at rates of ~10 TB/h [51]. In order to make sense of these data, investigators are developing methods to computationally align and fuse the raw images, and to automatically segment and track individual cells in 3D and in real time through their divisions, thereby generating dynamic 'digital embryos' [50,52]. These data are invaluable to inform or test predictive biomechanical and/or genetic 3D models of development [53,54].

Electron microscopy

Despite the advent of super-resolution light microscopy, electron microscopy (EM) still offers views of the cellular ultrastructure at unequaled spatial resolution. Freezing the sample to cryogenic temperatures maintains structures in their native state, minimizing the artefacts that plagued older fixation protocols. Recent advances in detector technology and sample preparation have pushed the resolution even closer to the atomic level, and it is now possible to reconstruct 3D density maps of molecules from cryo-EM data with Angstrom-scale detail, a feat hitherto reserved to X-ray crystallography [55] (Figure 1d). To achieve this, very noisy 2D EM images of thousands or millions of copies of a particle are processed in order to computationally reconstruct a single 3D structure (or a small number of conformations). As in SMLM, the determination of these structures is nowadays usually performed using MLE algorithms [16]. The selection of single particle images as input to the MLE reconstruction is still often done manually, although computational strategies including template matching are also being employed. Automated template matching however needs to be carefully controlled for biases, or else arbitrary patterns may emerge from pure noise [56]. When properly controlled, cryo-EM now provides a very attractive tool to determine 3D structures with near atomic resolution, which is applicable to molecules that are hard to crystallize (e.g. membrane or nuclear proteins) or that are present in multiple, even short-lived, conformations. Recent highlights demonstrating the power of cryo-EM include the elucidation of the glutamate dehydrogenase structure at <2 Å resolution or the structure of the HIV envelope trimer [14,15]. Advances in data analysis and sample preparation will likely further improve this resolution and extend applicability to more complex cases, such as highly flexible domains.

While single particle cryo-EM aims to characterize the atomic structures of molecules, other EM approaches are being employed to systematically explore the organization of large tissue volumes. This is most prominent in the field of connectomics, which seeks to determine the wiring diagram of the brain [57]. Scanning electron microscopy (SEM) combined with automated cutting or ablation methods is used to image pieces of the fly or mouse brain slice by slice, taking thousands of images at the nanometer-scale resolution needed to identify individual synapses [18,57], (Figure 1e). Such projects are on track to deliver staggering amounts of data. It has been estimated, for example, that imaging a full rat cortex will generate ~1 Exabyte of data (1 EB = 1,000 PB = 10^{18} bytes), and a human cortex ~1 Zetabyte (1 ZB = 10^{21} bytes), an order of magnitude approaching that of the annual Internet traffic [57]. Storing and handling such data volumes is the most immediate, but by no means trivial issue: with current technology, entire rooms might be required to store images from a single mouse brain. Beyond this, a major computational challenge is to reliably and efficiently trace individual neurons and their connections in

3D. It has been estimated that manually determining the wiring diagram of a whole mouse brain would take tens of millions of years [17]. To spur the development of automated tracing algorithms, competitions such as ISBI 2012 or MICCAI CREMI provide manually annotated training data and blind test data [58]. As in many other imaging applications, the current winners of these challenges are segmentation algorithms based on deep learning, indicating that artificial neural networks are best at mapping biological neural networks [17,58,60]. These wiring diagrams, once accurately reconstructed, can yield unprecedented insights into the architecture and function of neuronal circuits, e.g. in the insect visual system [61] or the mouse neocortex [18] and instruct computational models of neuronal systems.

Concluding remarks

As we have illustrated, automated high throughput and high resolution microscopy methods are generating a data tsunami that rivals or exceeds that delivered by DNA sequencing. Although smart microscopy methods may somewhat mitigate this trend by focusing image acquisition where it is most informative (e.g. Ref. [13]), the data deluge will surely accelerate. Further improvements in spatio-temporal resolution and molecular multiplexing will automatically increase the data output, as will combinations of methods mentioned above, e.g. SMLM with light sheets [12] or super-resolution HCS (Figure 2a). The most immediate informatics challenge is how to reliably store such large imaging data (Figure 2b). Unlike in DNA sequencing, where the raw data used to identify base pairs are routinely discarded, most microscopy projects still keep the original images. Keeping the raw data or uploading it to public data bases is necessary to reproduce analysis results, and is therefore increasingly required by journals. Moreover, this allows to extract new information by computational reanalysis of the data using improved algorithms, for example to determine 3D localizations from 2D SMLM images [62] or obtain higher resolution structures in single particle EM [63]. Nevertheless, storing might become too expensive in some applications, forcing researchers to discard the raw data and replace them by strongly condensed representations, e.g. molecule coordinates instead of the low resolution images in SMLM, or cell positions and trajectories instead of raw SPIM images. For some applications, it might even be more convenient to repeat the experiment rather than storing and reanalyzing old data [51]. This means that the image storage problem can no longer be dissociated from that of image processing (Figure 2c).

As discussed for some examples above, the community is developing increasingly powerful computational tools to turn images into biologically relevant information. These algorithms are also taking advantage of advances in computing power, and in particular increasingly make use of graphical processing units (GPUs), which enable considerable speed-ups for highly parallel calculations compared to CPUs. To this date, no single standard algorithm has emerged for processing even some of the apparently simplest images, as those of single molecules in SMLM. There will likely be no definitive consensus on how to best analyze biological imaging data in the near future. Nevertheless, within the wide landscape of computational methods, we see two strong trends: probabilistic inference methods, in particular MLE, and machine learning approaches, in particular deep learning (Figure 2c). Each of these two approaches is suited to a different class of problems. MLE is a favored approach whenever the image analysis problem can be formulated as extracting a finite set of deterministic parameters and a procedure is available to calculate the probability of an image given these parameters. This generally requires a computational model of the image formation, including optics and noise, and assumptions about the underlying objects, for example that they consist of single point light sources in SMLM, or of projections along various angles of identical copies of the same structure in single particle EM [16,24]. Bayesian inference methods can also be applied in this case, and provide several added benefits, including the determination of entire parameter distributions rather than a single estimation, a reduced need for tuning technical parameters, and conversely the option to incorporate prior knowledge about the structures, such as photoswitching probabilities in SMLM, or degrees of symmetry in cryo-EM [65–67].

For many applications, however, such as the segmentation of neurons and dendrites in 3D EM stacks, the complexity of the biological structures implies that explicit probabilistic models are not available. This difficulty is well known in the field of general computer vision, and indeed much work in biological image analysis is directly drawn from this larger research area. Until recently, however, computer vision was arguably characterized by rather incremental progress and a myriad of competing approaches, which usually reached state-of-the-art performance only in narrow application domains. On this backdrop, we consider the recent breakthroughs achieved by deep learning algorithms as refreshing, since these methods appear as both much more generic and powerful, and provide a seamless means to transfer human knowledge into an algorithm that can often outperform human experts [43,68,69]. It therefore seems safe to predict that in the coming years deep learning and related machine learning approaches will become the dominant technique for analyzing complex biological images, from the segmentation of cells in EM or light sheet microscopy to the automated annotation of gene phenotypes in HCS. Because deep learning thrives on training data, the success of future image processing efforts will largely depend on open access to vast annotated image data bases. Projects such as the Image Data Resource (IDR), the Electron

Microscopy Public Image Archive (EMPIAR) or the Open Connectome [70,71], which aim to compile such data, are therefore likely to become crucial resources and should be encouraged. Efficient reanalysis of these data, e.g. using deep learning methods, requires flexible allocation of large computing resources, such as GPU clusters. In this context, cloud computing offers a potentially attractive solution, provided that data transfer is not prohibitive [72,73].

Another important trend, cutting through all methodological approaches, are competitions. Well organized competitions such as those mentioned above for SMLM or connectomics are a compelling means to guide users towards the best software solutions, and to spur the development of better solutions. Among the important challenges in the future will be to design test data that more fully mimic the complexities of real experiments, in the context of rapidly evolving instrumentation and applications, and to automate the setting of hyperparameters, such that algorithms can be used by others as successfully as by their authors.

If properly analyzed, images from modern microscopy techniques can provide invaluable quantitative information on cellular and molecular structures and their dynamic interactions in living multicellular organisms. Together with data acquired by other technologies such as genomics and mass spectrometry, imaging data will empower the construction of computational models of cells and entire organisms and help move biology from a largely descriptive science to a predictive one [74] (Figure 2d). Rather than drowning us, we expect that the imaging tsunami, when channeled into quantitative descriptions and computational models, will lift our understanding of biological systems to new heights.

Acknowledgements

We thank F. Mueller and J. Krijnse Locker for helpful comments on the manuscript. W.O. is a scholar in the Pasteur - Paris University (PPU) International PhD program. C.Z. acknowledges funding by Institut Pasteur, Fondation pour la Recherche Médicale (Equipe FRM, DEQ 20150331762), Agence Nationale de la Recherche (ANR 14 CE10 0018 02, ANR 11 MONU 020 02) and the Région Ile de France (DIM Malinf).

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Betzig E, Patterson GH, Sougrat R, Lindwasser OW, Olenych S, Bonifacino JS, Davidson MW, Lippincott-Schwartz J, Hess HF: **Imaging intracellular fluorescent proteins at nanometer resolution**. *Science* 2006, **313**:1642–1645.

2. Rust MJ, Bates M, Zhuang X: **Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)**. *Nat Methods* 2006, **3**:793–795.

3. Sharonov A, Hochstrasser RM: **Wide-field subdiffraction imaging by accumulated binding of diffusing probes**. *Proc Natl Acad Sci U S A* Dec. 2006, **103**:18911–18916.

4. * Huang F, Hartwich TMP, Rivera-Molina FE, Lin Y, Duim WC, Long JJ, Uchil PD, Myers JR, Baird MA, Mothes W, Davidson MW, Toomre D, Bewersdorf J: **Video-rate nanoscopy using sCMOS camera-specific single-molecule localization algorithms**. *Nat Methods* Jul. 2013, **10**:653–658.

5. * Huang F, Sirinakis G, Allgeyer ES, Schroeder LK, Duim WC, Kromann EB, Phan T, Rivera-Molina FE, Myers JR, Irnov I, Lessard M, Zhang Y, Handel MA, Jacobs-Wagner C, Lusk CP, Rothman JE, Toomre D, Booth MJ, Bewersdorf J: **Ultra-high resolution 3D imaging of whole cells**. *Cell* 2016, **166**:1028–1040.

6. Manley S, Gillette JM, Patterson GH, Shroff H, Hess HF, Betzig E, Lippincott-Schwartz J: **High-density mapping of single-molecule trajectories with photoactivated localization microscopy**. *Nat Methods* 2008, **5**:155–157.

7. * Neumann B, Walter T, Heriche JK, Bulkescher J, Erfle H, Conrad C, Rogers P, Poser I, Held M, Liebel U, Cetin C, Sieckmann F, Pau G, Kabbe R, Wunsche A, Satagopam V, Schmitz MH, Chapuis C, Gerlich DW, Schneider R, Eils R, Huber W, Peters JM, Hyman AA, Durbin R, Pepperkok R, Ellenberg J: **Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes**. *Nature* 2010, **464**:721–727.

8. * Battich N, Stoeger T, Pelkmans L: **Image-based transcriptomics in thousands of single human cells at single-molecule resolution**. *Nat Methods* Oct. 2013, **10**:1127–1133.

9. ** Graml V, Studera X, Lawson JLD, Chessel A, Geymonat M, Bortfeld-Miller M, Walter T, Wagstaff L, Piddini E, Carazo-Salas RE: **A genomic multiprocess survey of machineries that control and link cell shape, microtubule organization, and cell-cycle progression**. *Dev Cell* Oct. 2014, **31**:227–239.

This paper uses HCS with 3D microscopy to systematically analyze the effect of 3000 non essential fission yeast genes on cell shape, cell cycle and microtubule organization. Almost 2 million images are analyzed and cellular features are analyzed using Bayesian networks to infer causal relations, e.g. that cell length influences microtubule length but not vice-versa, which was experimentally confirmed.

10. Godinez WJ, Hossain I, Lazic SE, Davies JW, Zhang X: **A multi-scale convolutional neural network for phenotyping high-content cellular images**. *Bioinformatics* 2017, **33**:2010–2019.

11. ** Lemon WC, Pulver SR, Höckendorf B, McDole K, Branson K, Freeman J, Keller PJ: **Whole-central nervous system functional imaging in larval Drosophila**. *Nat Commun* Aug. 2015, **6**:7924.

Using a fast multi-view light-sheet microscope, the authors imaged the entire central nervous system (CNS) of a fly larva at 2–5 Hz temporal resolution. Together with improved image analysis tools for Terabyte-size data, this allowed them, for the first time, to follow the propagation of neuronal activity throughout the CNS, opening the door to a systems level analysis of neural network activity.

12. ** Legant WR, Shao L, Grimm JB, Brown TA, Milkie DE, Avants BB, Lavis LD, Betzig E: **High-density three-dimensional localization microscopy across large volumes**. *Nat Methods* Mar. 2016, **13**:359–365.

This work combines an advanced type of light sheet illumination (a lattice light sheet) with PAINT, a type of SMLM based on transient binding of dyes. This system enabled the authors to localize up to ~1 billion molecules with high precision and create super-resolution (<100 nm) images of cells and small organs up to 20 μm thick.

13. Royer LA, Lemon WC, Chhetri RK, Wan Y, Coleman M, Myers EW, Keller PJ: **Adaptive light-sheet microscopy for long-term, high-resolution imaging in living organisms**. *Nat Biotechnol* Oct. 2016, **34**:1267–1278.

14. Lee JH, Ozorowski G, Ward AB: **Cryo-EM structure of a native, fully glycosylated, cleaved HIV-1 envelope trimer**. *Science* Mar. 2016, **351**:1043–1048.

15. ** Merk A, Bartesaghi A, Banerjee S, Falconieri V, Rao P, Davis MI, Pragani R, Boxer MB, Earl LA, Milne JLS, Subramaniam S: **Breaking cryo-EM resolution barriers to facilitate drug discovery**. *Cell* Jun. 2016, **165**:1698–1707.

This study reports the first structure with 1.8 Å resolution obtained by cryo-EM single particle analysis. This resolution allowed to visualize

the densities of each amino-acid in the core of the protein, even though a large fraction of the protein is conformationally mobile.

16. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA: **cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination**. *Nat Methods* Feb. 2017, **14**:290–296.

17. Dorkenwald S, Schubert PJ, Killinger MF, Urban G, Mikula S, Svara F, Kornfeld J: **Automated synaptic connectivity inference for volume electron microscopy**. *Nat Methods* Feb. 2017, **14**:435–442.

18. ** Kasthuri N, Hayworth KJ, Berger DR, Schalek RL, Conchello JA, Knowles-Barley S, Lee D, Vázquez-Reina A, Kaynig V, Jones TR, Roberts M, Morgan JL, Tapia JC, Seung HS, Roncal WG, Vogelstein JT, Burns R, Sussman DL, Priebe CE, Pfister H, Lichtman JW: **Saturated reconstruction of a volume of neocortex**. *Cell* 2015, **162**:648–661.
The authors developed an automated pipeline wherein a 80,000 μm^3 piece of mouse brain was automatically cut into thousands of thin sections by an ultramicrotome and imaged by scanning electron microscopy at nanometer resolution. Using manual and automated segmentation they reconstructed cellular and subcellular structures and created an online data base of 1700 synapses and find that connections between neurons cannot be explained by their proximity alone.

19. Wanner AA, Genoud C, Masudi T, Siksou L, Friedrich RW: **Dense EM-based reconstruction of the interglomerular projectome in the zebrafish olfactory bulb**. *Nat Neurosci* Apr. 2016, **19**:816–825.

20. Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, Efron MJ, Iyer R, Schatz MC, Sinha S, Robinson GE: **Big data: astronomical or genomical?** *PLoS Biol* Jul. 2015, **13**:e1002195.

21. Kanchanawong P, Shtengel G, Pasapera AM, Ramko EB, Davidson MW, Hess HF, Waterman CM: **Nanoscale architecture of integrin-based cell adhesions**. *Nature* Nov. 2010, **468**:580–584.

22. Xu K, Zhong G, Zhuang X: **Actin, spectrin, and associated proteins form a periodic cytoskeletal structure in axons**. *Science* 2013, **339**.

23. * Szymborska A, de Marco A, Daigle N, Cordes VC, Briggs JAG, Ellenberg J: **Nuclear pore scaffold structure analyzed by super-resolution microscopy and particle averaging**. *Science* Aug. 2013, **341**:655–658.

24. Ober RJ, Ram S, Ward ES: **Localization accuracy in single-molecule microscopy**. *Biophys J* 2004, **86**:1185–1200.

25. ** Sage D, Kirshner H, Pengo T, Stuurman N, Min J, Manley S, Unser M: **Quantitative evaluation of software packages for single-molecule localization microscopy**. *Nat Methods* Jun. 2015, **12**:717–724.
References [25] and [58] illustrate how challenges can be used to objectively compare competing algorithms against simulated or manually defined ground truth. Ref. [25] addresses image analysis software for SMLM, ref. [58] software for EM images in the context of connectomics.

26. von Diezmann A, Shechtman Y, Moerner WE: **Three-Dimensional localization of single molecules for super-resolution imaging and single-particle tracking**. *Chem Rev* Feb. 2017. p. acs.chemrev.6b00629.

27. Jia S, Vaughan JC, Zhuang X: **Isotropic 3D super-resolution imaging with a self-bending point spread function**. *Nat Photonics* Jan. 2014, **8**:302–306.

28. Sengupta P, Jovanovic-Talisman T, Skoko D, Renz M, Veatch SL, Lippincott-Schwartz J: **Probing protein heterogeneity in the plasma membrane using PALM and pair correlation analysis**. *Nat Methods* 2011, **8**:969–975.

29. * Lee S-H, Shin JY, Lee A, Bustamante C: **Counting single photoactivatable fluorescent molecules by photoactivated localization microscopy (PALM)**. *Proc Natl Acad Sci U S A* Oct. 2012, **109**:17436–17441.

30. Backlund MP, Lew MD, Backer AS, Sahl SJ, Moerner WE: **The role of molecular dipole orientation in single-molecule fluorescence microscopy and implications for super-resolution imaging**. *ChemPhysChem* Mar. 2014, **15**:587–599.

31. Chenouard N, Smal I, de Chaumont F, Maška M, Sbalzarini IF, Gong Y, Cardinale J, Carthel C, Coraluppi S, Winter M, Cohen AR, Godinez WJ, Rohr K, Kalaidzidis Y, Liang L, Duncan J, Shen H, Xu Y, Magnusson KEG, Jaldén J, Blau HM, Paul-Gilloteaux P, Roudot P, Kervrann C, Waharte F, Tinevez J-Y, Shorte SL, Willemse J, Celler K, van Wezel GP, Dan H-W, Tsai Y-S, de Solórzano CO, Olivo-Marin J-C, Meijering E: **Objective comparison of particle tracking methods**. *Nat Methods* Jan. 2014, **11**:281–289.

32. Monnier N, Guo S-M, Mori M, He J, Lénárt P, Bathe M: **Bayesian approach to MSD-based analysis of particle motion in live cells**. *Biophys J* 2012, **103**:616–626.

33. El Beheiry M, Dahan M, Masson J-B: **InferenceMAP: mapping of single-molecule dynamics with Bayesian inference**. *Nat Methods* 2015, **12**:594–595.

34. Zhang Z, Kenny SJ, Hauser M, Li W, Xu K: **Ultrahigh-throughput single-molecule spectroscopy and spectrally resolved super-resolution microscopy**. *Nat Methods* Aug. 2015, **12**:935–938.

35. Jungmann R, Avendaño MS, Woehrstein JB, Dai M, Shih WM, Yin P: **Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT**. *Nat Methods* Feb. 2014, **11**:313–318.

36. ** Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X: **Spatially resolved, highly multiplexed RNA profiling in single cells**. *Science* Apr. 2015, **348**, aaa6090.
This study uses a combinatorial labeling strategy to visualize up to 1000 distinct mRNA species in the same cells at the single molecule level. The authors use their data to determine covarying genes and propose new functional annotations for ~100 genes. This paper is a milestone towards single cell transcriptomics.

37. Munsky B, Neuert G, van Oudenaarden A: **Using gene expression noise to understand gene regulation**. *Science* Apr. 2012, **336**:183–187.

38. Tantale K, Mueller F, Kozulic-Pirher A, Lesne A, Victor J-M, Robert M-C, Capozi S, Chouaib R, Bäcker V, Mateos-Langerak J, Darzacq X, Zimmer C, Basyuk E, Bertrand E: **A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting**. *Nat Commun* Jul. 2016, **7**:12248.

39. Mattiazzi Usaj M, Styles EB, Verster AJ, Friesen H, Boone C, Andrews BJ: **High-content screening for quantitative cell biology**. *Trends Cell Biol* 2016, **26**:598–611.

40. Battich N, Stoeger T, Pelkmans L: **Control of transcript variability in single mammalian cells**. *Cell* Dec. 2015, **163**:1596–1610.

41. Jones TR, Carpenter AE, Lamprecht MR, Moffat J, Silver SJ, Grenier JK, Castoreno AB, Eggert US, Root DE, Golland P, Sabatini DM: **Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning**. *Proc Natl Acad Sci U S A* Feb. 2009, **106**:1826–1831.

43. LeCun Y, Bengio Y, Hinton G: **Deep learning**. *Nature* May 2015, **521**:436–444.

44. Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol P-A: **Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion**. *J Mach Learn Res* 2010, **11**:3371–3408.

45. Kandaswamy C, Silva LM, Alexandre LA, Santos JM: **High-content analysis of breast cancer using single-cell deep transfer learning**. *J Biomol Screen* 2016, **21**:252–259.

46. ** Lee K, Elliott HL, Oak Y, Zee C-T, Groisman A, Tytell JD, Danuser G: **Functional hierarchy of redundant actin assembly factors revealed by fine-grained registration of intrinsic image fluctuations**. *Cell Syst* Jul. 2015, **1**:37–50.

47. Welf ES, Danuser G: **Using fluctuation analysis to establish causal relations between cellular events without experimental perturbation**. *Biophys J* 2014, **107**:2492–2498.

48. Huisken J, Swoger J, Del Bene F, Wittbrodt J, Stelzer EHK: **Optical sectioning deep inside live embryos by selective plane illumination microscopy**. *Science* 2004, **305**:1007.

49. Chen B-C, Legant WR, Wang K, Shao L, Milkie DE, Davidson MW, Janetopoulos C, Wu XS, Hammer JA, Liu Z,

English BP, Mimori-Kiyosue Y, Romero DP, Ritter AT, Lippincott-Schwartz J, Fritz-Laylin L, Mullins RD, Mitchell DM, Bembenek JN, Reymann A-C, Bohme R, Grill SW, Wang JT, Seydoux G, Tulu US, Kiehart DP, Betzig E: **Lattice light-sheet microscopy: imaging molecules to embryos at high spatiotemporal resolution**. *Science* Oct. 2014, **346**. 1257998–1257998.

50. Keller PJ, Schmidt AD, Wittbrodt J, Stelzer EHK: **Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy**. *Science* 2008, **322**.

51. Reynaud EG, Peychl J, Huisken J, Tomancak P: **Guide to light-sheet microscopy for adventurous biologists**. *Nat Methods* Dec. 2014, **12**:30–34.

52. Faure E, Savy T, Rizzi B, Melani C, Stašová O, Fabrèges D, Špir R, Hammons M, Čúnderlík R, Recher G, Lombardot B, Duloquin L, Colin I, Kollár J, Desnoulez S, Affaticati P, Maury B, Boyreau A, Nief J-Y, Calvat P, Vernier P, Frain M, Lutfalla G, Kergosien Y, Suret P, Remešíková M, Doursat R, Sarti A, Mikula K, Peyriéras N, Bourgine P: **A workflow to process 3D+time microscopy images of developing organisms and reconstruct their cell lineage**. *Nat Commun* Feb. 2016, **7**:8674.

53. ** Delile J, Herrmann M, Peyriéras N, Doursat R: **A cell-based computational model of early embryogenesis coupling mechanical behaviour and gene regulation**. *Nat Commun* Jan. 2017, **8**:13929.

This study introduces an open-source software platform for simulating complex collective dynamics of thousands of cells in morphogenesis by coupling models of mechanical behaviour and gene regulation. The method was used to simulate the massive dynamic reorganization of cells during early zebrafish embryogenesis and quantitatively compared to experimental data.

54. Rauzi M, Krzic U, Saunders TE, Krajnc M, Ziherl P, Hufnagel L, Leptin M: **Embryo-scale tissue mechanics during Drosophila gastrulation movements**. *Nat Commun* Oct. 2015, **6**:8677.

55. Egelman EH: **The current revolution in cryo-EM**. *Biophys J* Mar. 2016, **110**:1008–1012.

56. Henderson R: **Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise**. *Proc Natl Acad Sci U S A* Nov. 2013, **110**:18037–18041.

57. Lichtman JW, Pfister H, Shavit N: **The big data challenges of connectomics**. *Nat Neurosci* Oct. 2014, **17**:1448–1454.

58. ** Arganda-Carreras I, Turaga SC, Berger DR, Cireşan D, Giusti A, Gambardella LM, Schmidhuber J, Laptev D, Dwivedi S, Buhmann JM, Liu T, Seyedhosseini M, Tasdizen T, Kamentsky L, Burget R, Uher V, Tan X, Sun C, Pham TD, Bas E, Uzunbas MG, Cardona A, Schindelin J, Seung HS: **Crowdsourcing the creation of image segmentation algorithms for connectomics**. *Front Neuroanat* Nov. 2015, **9**:142.

60. Beier T, Pape C, Rahaman N, Prange T, Berg S, Bock DD, Cardona A, Knott GW, Plaza SM, Scheffer LK, Koethe U, Kreshuk A, Hamprecht FA: **Multicut brings automated neurite segmentation closer to human performance**. *Nat Methods* Jan. 2017, **14**:101–102.

61. Takemura S, Bharioke A, Lu Z, Nern A, Vitaladevuni S, Rivlin PK, Katz WT, Olbris DJ, Plaza SM, Winston P, Zhao T, Horne JA, Fetter RD, Takemura S, Blazek K, Chang L-A, Ogundeyi O, Saunders MA, Shapiro V, Sigmund C, Rubin GM, Scheffer LK, Meinertzhagen IA, Chklovskii DB: **A visual motion detection circuit suggested by Drosophila connectomics**. *Nature* Aug. 2013, **500**:175–181.

62. Franke C, Sauer M, van de Linde S: **Photometry unlocks 3D information from 2D localization microscopy data**. *Nat Methods* Nov. 2016, **14**:41–44.

63. Scheres S H w: **Beam-induced motion correction for sub-megadalton cryo-EM particles**. *Elife* 2014, **3**:e03665.

65. Cox S, Rosten E, Monypenny J, Jovanovic-Talisman T, Burnette DT, Lippincott-Schwartz J, Jones GE, Heintzmann R: **Bayesian localization microscopy reveals nanoscale podosome dynamics**. *Nat Methods* 2012, **9**:195–200.

66. Joubert P, Habeck M: **Bayesian inference of initial models in cryo-electron microscopy using pseudo-atoms**. *Biophys J* Mar. 2015, **108**:1165–1175.

67. Scheres SHW: **A bayesian view on cryo-EM structure determination**. *J Mol Biol* Jan. 2012, **415**:406–418.

68. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, Kim R, Raman R, Nelson PC, Mega JL, Webster DR: **Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs**. *JAMA* Dec. 2016, **316**:2402.

69. * Ronneberger O, Fischer P, Brox T: **"U-net: convolutional networks for biomedical image segmentation," in medical image computing and computer-assisted intervention – MICCAI 2015**. *Lect Notes Comput Sci* 2015, **9351**:234–241.

70. Williams E, Moore J, Li SW, Rustici G, Tarkowska A, Chessel A, Leo S, Antal B, Ferguson RK, Sarkans U, Brazma A, Salas REC, Swedlow JR: **Image Data Resource: a bioimage data integration and publication platform**. *Nat Methods* Aug. 2017, **14**: 775–781.

71. Iudin A, Korir PK, Salavert-Torres J, Kleywegt GJ, Patwardhan A: **EMPIAR: a public archive for raw electron microscopy image data**. *Nat Methods* Mar. 2016, **13**:387–388.

72. Cianfrocco MA, Leschziner AE: **Low cost, high performance processing of single particle cryo-electron microscopy data in the cloud**. *Elife* May 2015, **4**:1–10.

73. Hu YS, Nan X, Sengupta P, Lippincott-Schwartz J, Cang H: **Accelerating 3B single-molecule super-resolution microscopy with cloud computing**. *Nat Methods* 2013, **10**:96–97.

74. * Sanghvi JC, Regot S, Carrasco S, Karr JR, V Gutschow M, Bolival B, Covert MW: **Accelerated discovery via a whole-cell model**. *Nat Methods* Nov. 2013, **10**:1192–1195.

nature COMMUNICATIONS

ARTICLE

DOI: 10.1038/s41467-018-06868-w OPEN

A computational framework to study sub-cellular RNA localization

Aubin Samacoits[1,2], Racha Chouaib[3,4], Adham Safieddine[3,4], Abdel-Meneem Traboulsi[3,4], Wei Ouyang[1,2], Christophe Zimmer[1,2], Marion Peter[3,4], Edouard Bertrand[3,4], Thomas Walter[5,6,7] & Florian Mueller[1,2]

RNA localization is a crucial process for cellular function and can be quantitatively studied by single molecule FISH (smFISH). Here, we present an integrated analysis framework to analyze sub-cellular RNA localization. Using simulated images, we design and validate a set of features describing different RNA localization patterns including polarized distribution, accumulation in cell extensions or foci, at the cell membrane or nuclear envelope. These features are largely invariant to RNA levels, work in multiple cell lines, and can measure localization strength in perturbation experiments. Most importantly, they allow classification by supervised and unsupervised learning at unprecedented accuracy. We successfully validate our approach on representative experimental data. This analysis reveals a surprisingly high degree of localization heterogeneity at the single cell level, indicating a dynamic and plastic nature of RNA localization.

[1] Unité Imagerie et Modélisation, Institut Pasteur and CNRS UMR 3691, 28 rue du Docteur Roux, 75015 Paris, France. [2] C3BI, USR 3756 IP CNRS, 28 rue du Docteur Roux, 75015 Paris, France. [3] Institut de Génétique Moléculaire de Montpellier, University of Montpellier, CNRS, Montpellier, France. [4] Equipe labellisée Ligue Nationale Contre le Cancer, Paris, France. [5] MINES ParisTech, PSL-Research University, CBIO-Centre for Computational Biology, 75006 Paris, France. [6] Institut Curie, PSL Research University, 75005 Paris, France. [7] INSERM, U900, 75005 Paris, France. Correspondence and requests for materials should be addressed to E.B. (email: edouard.bertrand@igmm.cnrs.fr) or to T.W. (email: thomas.walter@mines-paristech.fr) or to F.M. (email: muellerf.research@gmail.com)

ARTICLE NATURE COMMUNICATIONS | DOI: 10.1038/s41467-018-06868-w

Non-random sub-cellular RNA localization is important for cellular function and its misregulation is linked to a number of diseases[1,2]. Initially observed in highly polarized cells such as oocytes or embryonic fibroblasts, more recent studies revealed diverse and wide-spread RNA localization in other systems[3], including bacteria[4], yeast[5], and developing embryos of fruitfly, ascidians and zebrafish[3,6]. RNA localization also occurs in cultured mammalian cell[7–9]. Besides the particular case of neurons where a large number of mRNAs localize in cellular processes, mRNA localization also occurs in regular cell lines to regulate gene expression at the spatial level. Secreted and mitochondrial proteins are often translated at the endoplasmic reticulum and mitochondria, respectively, while mRNA repressed for translation can accumulate in P-bodies or stress granules. More specific examples of localization include mRNAs that accumulate at the tip of cellular extensions[9], localize at the cell periphery[10], or DYNC1H1 mRNA that accumulates in foci representing dedicated translation factories[11]. With the rapid development of high-throughput techniques, it is likely that many more localized RNAs will be discovered. However, validated analysis tools to identify and classify such RNA localization patterns are currently lacking.

Imaging technologies, especially single molecule FISH[7,12,13] (smFISH), allow to observe single RNA molecules in their native cellular environment. This technique is now easy to implement and can be performed at low cost[13]. It provides unique quantitative spatial information[2,7] and thanks to recent advances, can be performed at large scale in cell lines and embryos[7,10,12,14,15]. Image analysis then allows to discover genes displaying non-random localization patterns. While many localization patterns are distinguishable by visual inspection[3,8], manual annotation can be biased, is often not quantitative and influenced by confounding factors such as RNA expression level. In addition, comprehensive manual annotation at the single cell level hardly seems an option for larger scale studies where thousands of cells are imaged in a single experiment. Indeed, the benefits of automatic analysis of smFISH data[7,16] include scalability and reproducibility, allowing for an accurate and quantitative description of the spatial aspects of gene expression.

In smFISH images, individual RNA molecules appear as bright diffraction-limited spots, which can be localized in 3D with published image analysis tools[12,14]. In contrast to the analysis of cellular phenotypes[17] and protein localization[18], smFISH data can be treated as point clouds. The smFISH signal inside a cell can thus be represented by features describing this spatial distribution of points, such as the mean nearest neighbor distance between spots or their average distance to the nuclear envelope. These features can then be used to group cells based on similarity in their RNA localization patterns, using supervised or unsupervised machine learning methods[7]. However, one of the main difficulty in this approach is the absence of a ground truth for RNA localization in smFISH data, making it impossible to assess usefulness of features and performance of the classification workflow. Hence, as of today, there is no rigorously validated method to analyze smFISH data at the cellular level.

Here, we present a simulation framework to create a synthetic ground-truth data set to perform this validation. Such simulated ground-truth data provide a number of key advantages to the traditional strategy relying exclusively on manual annotation[17–21]. Manual annotation of 3D point clouds irrespective of their number and reference volume is time consuming, difficult, error prone and tends to be subjective, in particular for subtle differences. In addition, we can only annotate already observed patterns from already identified example genes. This encouraged us to build a simulation framework in order to complement or replace manual annotation. We generated point patterns from known localization rules to create large amounts of ground-truth data. This allowed us to also control the parameters of the generative model in order to study robustness and limitations of the automatic algorithms. We show that the simulation of a large set of images enables designing and validating workflows for unsupervised and supervised analysis of smFISH data, which are capable of detecting a large variety of localization classes. We applied this approach to experimental data and successfully detected the different manually annotated localization classes. We also implemented a metric to quantitatively analyze heterogeneity of localization patterns. Application of this metric to our experimental data set revealed a surprisingly high degree of localization heterogeneity.

Results

Simulating realistic ground-truth data. We designed a simulation framework capable of generating ground-truth data closely mimicking experimental smFISH images. First, we acquired fluorescence microscopy data to determine the 3D volume of cells and nuclei, as well as the typical smFISH background signal. We then placed individual RNAs with realistic signals according to pre-defined localization rules inside these cellular volumes. This set of simulated images allowed us to validate the entire analysis workflow from RNA detection to the identification of localization classes. Our approach to generate such synthetic smFISH images consists of four main steps (Supplementary Note 1):

First, we inferred accurate 3D cellular shapes from experimental data[22]. To do so, we performed an smFISH experiment in HeLa cells using Cy5-labeled probes against the highly expressed *GAPDH* mRNA. We defined the 3D cellular outline as the conforming boundary containing all detected mRNAs (Fig. 1a, b). We further used our observation that *GAPDH* mRNA was largely absent from the nucleus (Fig. 1a) to determine the average height and position of nuclei in the cells (Fig. 1b). This provided us with a collection of cellular and nuclear volumes, where RNAs can be placed. For each cell, we also acquired another channel with realistic background from a mock smFISH experiment, using Cy3-labeled probes against a not-expressed reporter gene. In addition, we acquired images for 2D segmentation of cells and nuclei (CellMask^{TM} and DAPI, respectively), as they are used in standard screening applications. The open design of our workflow makes it possible to add additional marker channels, e.g., P-bodies, for more specific screening applications.

Second, we defined different RNA levels. It is well known that expression levels can vary greatly between genes and even between clonal cells for a single gene. Ongoing efforts identify stochastic noise in transcription and extrinsic factors such as cellular microenvironment or cell volume as the molecular and environmental origins of these cell-to-cell variations[22,23]. In contrast, an analysis workflow of RNA localization has to be independent of RNA levels, because different cells should be grouped together based on similarity in their RNA localization patterns, and not in their expression levels. We defined RNA density as a free parameter, and the absolute number of RNAs was assumed to be proportional to the cell volume, as shown in recent studies[22,23] (Supplementary Note 1). We simulated four regimes of expression, each regime with constant RNA density modulated with an additional Poisson noise term, which when pooled cover a large range of expressions levels in agreement with a recently observed large-scale screen (Fig. 1c)[7]. Lastly, we also considered a scenario with very high expression levels to test the potential limitations of the classification approaches.

Third, we simulated realistic images of individual RNA molecules. Each RNA was simulated as a point-spread-function (PSF) with sub-pixel localization and intensities from an

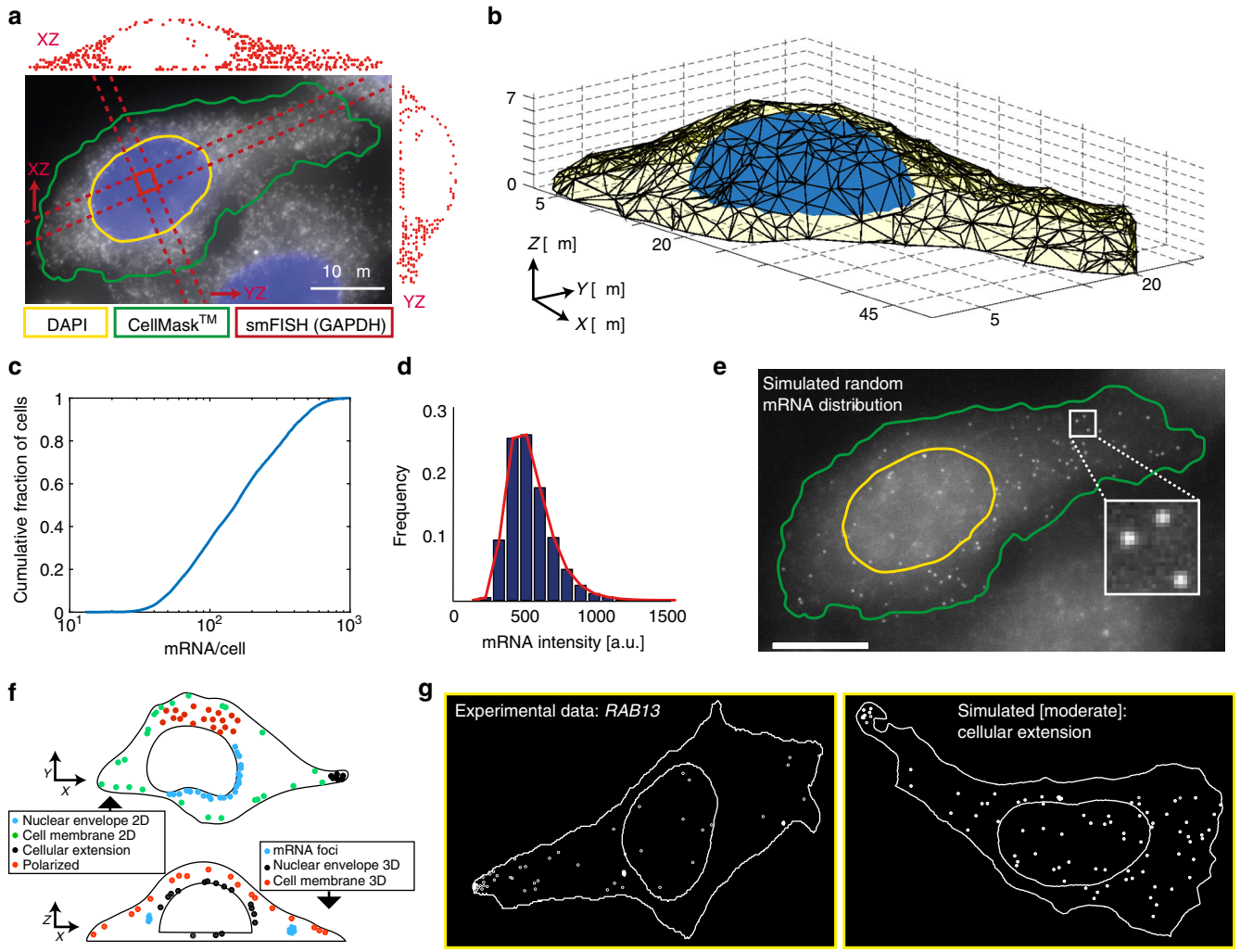


Fig. 1 Simulation of smFISH images. **a** smFISH experiment against GAPDH (probes labeled with Cy5) in HeLa cells shown as a maximum intensity projection along Z (MIP). Cell segmentation (green) was performed with CellMask™, nuclear segmentation (yellow) with DAPI signal. Projections of detected GAPDH mRNAs located between the dashed lines are displayed as indicated by the arrows as side panels. **b** Final 3D polygon of a cell (yellow) and its nucleus (blue). **c** Cumulative histogram of pooled expression level. **d** mRNA intensity distribution extracted from experimental data (*KIF1C* mRNA). Histogram is fitted with a skewed normal distribution (red). **e** Example of a simulated cell with random mRNA localization shown as a MIP. Outline of cell and nucleus in green and yellow, respectively. Scale bars 10μm. **f** Cartoon illustrating the simulated seven non-random localization patterns. **g** Outlines of a cell and nucleus with mRNA positions of the pattern cell extension: experimental data (*RAB13* mRNA) (left), simulated data with low mRNA density and moderate pattern strength (right)

experimentally measured distribution (Fig. 1d, Supplementary Note 1). Together with the realistic smFISH background, this simulation strategy also allows testing the performance of RNA detection algorithms [12,14] (Supplementary Note 2).

Fourth, we placed RNAs within the 3D cellular volume either randomly (Fig. 1e) or according to one of seven specific localization patterns (Fig. 1f–g). We designed these patterns to either mimic experimental data or to generate so far unobserved, yet plausible localizations. The patterns correspond to RNAs localizing to the tip of cellular extensions, around the cell membrane in 2D or 3D, with a polarized pattern, in foci, and around the nuclear envelope in 2D or 3D. Each pattern corresponds to a set of rules defining the spatial distribution of RNA molecules. For each pattern, there is a set of fixed parameters that define the nature of the localization pattern. In addition, we define one parameter (pattern strength) to control how extreme a pattern is (e.g., for the nuclear envelope localization, the pattern strength corresponds to the fraction of RNA in proximity of the envelope). For low pattern strength, the distribution approaches spatial randomness. For more details we refer to Supplementary Note 1. For this study, we defined three strengths, with the intermediate corresponding to a typically observed experimental pattern as assessed by a human observer (for an example see Fig. 1g, where the localization of *RAB13* mRNAs in cellular extensions[9] is compared to simulations). The pattern strength can be used to study the sensitivity of the workflow in discriminating localized and non-localized RNAs. Taken together, we can simulate realistic images with different RNA expression levels and pattern strength.

Analysis of simulated data with existing approaches. We first generated data with little heterogeneity for both the pattern strength and expression level. We used simulated images with moderate pattern strength and an average of 200 RNAs per cell (100 cells per condition), to assess the performance of previously published analysis methods aiming at identifying RNA localization patterns. Our first analysis was based on a recent publication

presenting a very comprehensive analysis workflow[7,24]. Here, 32 localization features are calculated from a mRNA detection performed on a maximum-intensity projection of the smFISH images (Supplementary Note 4).

In order to visualize the feature distributions for the simulated patterns, we projected these features onto two dimensions by t-Distributed Stochastic Neighbor Embedding (t-SNE)[25]. While half of the patterns were reasonably well separated, the other half was inseparable (Fig. 2a). The inseparable patterns were RNA localizing in cellular extensions, in foci, at the cell membrane in 3D or randomly. k-means clustering yielded the same set of indistinguishable patterns (50% accuracy, Fig. 2b). This encouraged us to improve RNA detection in 3D and to design new features describing their spatial distribution.

RNA detection and localization features. Large-scale studies have shown that accumulation of RNA in cytoplasmic foci is a predominant localization pattern[3]. Despite their biological relevance, previous analysis approaches were not designed to correctly identify these foci. We hypothesized that this might be due to problems during RNA detection in existing methods, where an accumulation of very close RNAs is detected as a single molecule. We hence designed an analysis method based on Gaussian Mixture Models (GMM), where these foci are decomposed into individual RNA molecules. We detail the validation of this approach on simulations and experimental data in Supplementary Note 2.

We then designed new feature families based on established concepts from spatial statistics and image analysis. First, we used Ripley's L-function, which provides information about homogeneity of spatial density. We assumed that this would be particularly useful for detection of patterns like foci and polarized RNAs. Second, we used morphological operators to extract cellular extensions from the 2D mask of the cells and used the enrichment ratio of RNA counts in these extensions as a feature. Third, we developed features capturing RNA localization with respect to the cell membrane. This is challenging, since usually no information about the 3D cellular shape is available in standard smFISH experiments. By using our mock smFISH experiment, we could show that the estimated background of wide-field smFISH images was correlated with the cell height (Supplementary Note 3). We thus defined as a feature the correlation of the measured z-positions of RNAs and the background intensity (approximation for cell height). Lastly, we normalized the features describing distances between RNAs and cellular structures by the expectation of those distances under complete spatial randomness. Using our simulated data, we found that this normalization reduced the impact of cellular shape on the feature distributions, as compared to previously proposed normalization schemes[7] (Supplementary Note 3). We also added two previously published features for the description of polarized mRNA localization (Polarization and dispersion index)[16]. Altogether, we defined a new set of features (14 new and 9 previously published, Table 1).

We then analyzed the simulated data with this new feature set and the new RNA detection scheme in 3D including the GMM. This greatly improved the identification of localization patterns: all localization patterns were distinguished in an unsupervised setting with an overall accuracy of 90% (Fig. 2c, d, Supplementary Note 4).

Robustness of classification toward heterogeneity. We next investigated the robustness of this classification toward increased heterogeneity. First, we pooled the three pattern strengths defined above and found that the classification results remain good (accuracy 79%; Supplement Note 4). We then pooled data simulated with RNA densities from an average of 50–400 RNAs per cell (Supplement Note 4), corresponding to the range recently measured in a large-scale smFISH study [7]. This test is crucial, since the analysis should identify cells with similar RNA localization patterns independently of variations in their expression level. We found that in a t-SNE plot, cells remained separated based on their localization pattern (Fig. 2e), showing that the features are stable with respect to expression heterogeneity. We also analyzed simulated cells with a very high expression level (average of 800 RNAs per cell). Here, we found that 7 out of 8 patterns could still be correctly identified (Supplementary Note 4). Only RNA foci were confounded with random, which can be explained by the increased local density of genes expressed at very high levels. However, we expect such extreme levels to be rare.

We then compared various clustering strategies (Supplementary Note 4) and found that k-means applied to a six-dimensional t-SNE analysis gave the best results (88% accuracy, Fig. 2f). If the number of clusters is not known, it can be inferred from the data with traditional methods such as the silhouette score or an analysis of the within-class variability (Supplementary Note 4).

Lastly, we tested whether our simulations could also be used to analyze the extent by which two spatial distributions must differ to be still detectable as being different. This is particularly interesting in the context of drug perturbation experiments. Specifically, we can use our simulation framework to perform a sensitivity analysis. We simulated data with a large range of pattern strength and investigate the impact of the pattern strength on the accuracy of the analysis (Supplementary Note 4).

Overall, we designed a workflow for the unsupervised analysis of localization patterns, that we validated on simulated data with varying degrees of both pattern strength and expression heterogeneity. Our results indicate that, while some localization patterns are easier to detect than others, all these localization patterns could be detected with good accuracy.

Detection of localized mRNAs in experimental data. Motivated by these results, we analyzed experimental data from 10 genes in HeLa cells (150–400 cells per gene, 2600 cells in total; Supplementary Note 5). Three mRNAs were manually annotated as random (*KIF20B*, *MYO18A*, *PAK2*), and seven were chosen because of their non-random localization. *RAB13*[9] and KIF1C mRNA accumulate in cell extensions. *DYNC1H1*[11] and BUB1 mRNA concentrate in cytoplasmic foci. The last three mRNAs display a localization associated with the nucleus: the *ATP6A2* protein is synthesized on the endoplasmic reticulum and its mRNA concentrate in the perinuclear region; *SPEN* mRNA form a rim that decorates the nuclear edge; while *CEP192* mRNA concentrates inside the nucleus. Some of these patterns were not explicitly included in our simulated classes (e.g., perinuclear and intranuclear), but they were nevertheless included to test whether our feature set would be general enough to enable classification of novel patterns. It is also important to note that, for any of these genes, we typically observe different cellular subpopulations with different spatial distributions of the encoded transcripts. Gene-level annotation therefore describes only a tendency and cannot be considered as a single-cell annotation.

We applied the workflow that we had benchmarked on the simulated data to the experimental data and extracted localization features for each cell. We represented the extracted localization features as a t-SNE plot (Fig. 3a, Supplementary Note 5). We also created a more detailed version of this plot where each point was replaced by a thumbnail representation of the cell, showing outlines of cells and nuclei and the detected RNAs. The plot can

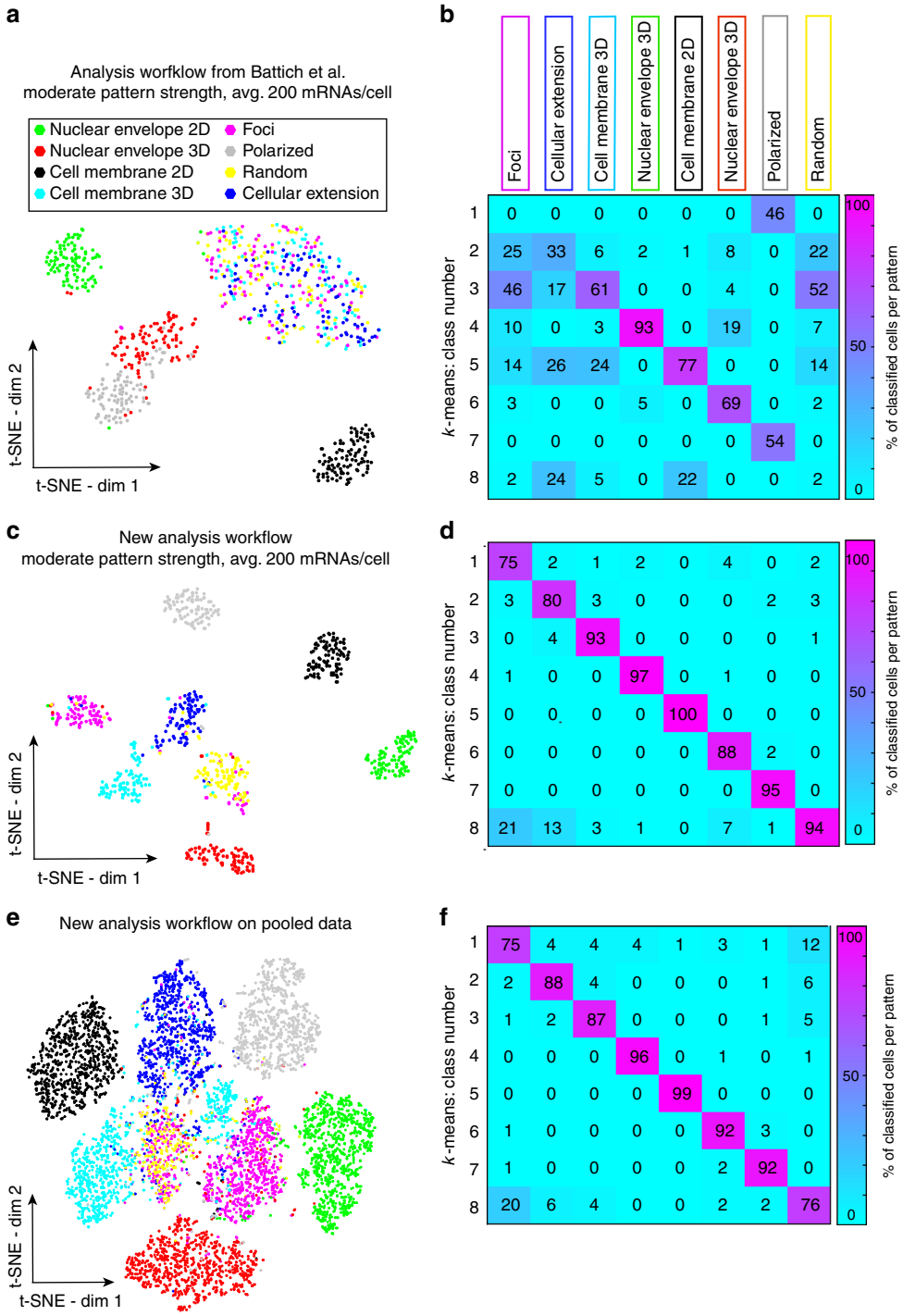


Fig. 2 Analysis of simulated data. **a** Analysis with previously published mRNA detection approach and localization features. In the t-SNE plot, each dot is a cell colored with its localization pattern. **b** Confusion matrix for k-means clustering with 8 classes for data in **a**. Rows show identified classes, columns the known localization pattern. Numbers in each square indicate percentage of cells for a localization pattern (column) that was assigned to a given class (row). Off-diagonal elements are mis-classifications. **c**, **d** t-SNE projection and confusion matrix as in **a** and **b** but determined with our 3D spot detection with GMM and the new localization features. **e** Analysis of pooled simulated data (different pattern strength and expression levels). Analysis and t-SNE projection as in **c**. **f** Confusion matrix for data in **e** with k-means clustering performed on t-SNE projected data in six dimensions

be viewed at https://muellerflorian.github.io/locFISH_deepzoom/#results/tsne_exp (for examples of the thumbnails see Fig. 3a). The tools to generate such plots are part of the provided Matlab source code. This zoomable t-SNE plot gives a clear picture of what the localization patterns in different regions of the feature space look like.

We found that cells from different genes with the same or similar manually annotated localization pattern showed an important overlap in the t-SNE plot. Conversely, we found that cells from genes with different manual annotations tended to populate distinct regions (Fig. 3a). This shows that the localization features are well suited to describe mRNA

Table 1 List of localization features used in Figs. 3 and 4

Feature ID	Feature description
1	Ripley: maximum
2	Ripley: max gradient [0,max]
3	Ripley: min gradient [max,end]
4	Ripley: value at mid-point between center and boundary
5	Ripley: Spearman correlation between Ripley and radius
6	Ripley: radius of max value
7	Polarization index
8	Dispersion index
9–12	Morph opening-enrichment ratio: 15, 30, 45, 60 pixels
13	Cell height: Spearman correlation with Z_{mRNA}
14	Cell height: R^2 with Z_{mRNA}
15	Cell membrane: distance - mean
16-19	Distance membrane: quantile 5%, 10%, 20%, 50%
20	Nucleus: distance - mean
21	Cell centroid: distance - mean
22	Nucleus centroid: distance - mean
23	Ratio: mRNAs inside nucleus/outside nucleus

localization from experimental data. This result was confirmed by the analysis of a different cell line (C2C12) (Supplementary Note 5). These cells are larger than HeLa cells (600 vs. 200 μm^2) and also more elongated. We found that C2C12 cells with three different manually annotated localization patterns (*DYNC1H1* - foci, *KIF1C* - cellular extension, *ACTN1* - random and polarized) were well separated.

Interestingly, the t-SNE plot also captures the heterogeneity we had observed during visual inspection of the data. First, for genes with non-random localization, not all cells have a distinct mRNA localization. For instance, most *CEP192* cells form a clearly separated cluster with a strong intra-nuclear localization (Fig. 3a, region 2), while other *CEP192* cells are random and grouped together with genes also showing a random localization pattern (Fig. 3a, region 1). Similar observations can be made for all other genes annotated as having non-random RNA localization. Interestingly, the intra-nuclear localization pattern of *CEP192* transcripts was not among the patterns used for simulation. This shows that the features we have designed are informative beyond the simulated patterns and in principle capable of adequately describing other patterns as well. Second, some spatial RNA distributions appear to be pattern mixtures. For instance, *DYNC1H1* mRNAs form foci[11], which are displayed in the upper part of the t-SNE plot (Fig. 3a, region 3). However, in some cells these foci are located toward the nuclear envelope, and these cells are positioned in close proximity with other cells displaying nuclear localization in the t-SNE plot (Fig. 3a, region 4). We conclude that our features are capable of capturing the complex structure in the data, and that t-SNE visualization allows to explore the heterogeneous localization patterns displayed by different cell populations.

Supervised classification of RNA localization. Next, we investigated whether we can infer different clusters of subcellular localization at the single-cell level. k-means with 4 classes correctly separates the manually annotated localization classes (random, cellular extension, foci, and nuclear-associated) but failed to isolate a small cluster of cells with strong intra-nuclear localization (Fig. 3b). In contrast, spectral clustering was able to find this cluster but fused larger localization patterns (Fig. 3b, Supplementary Note 5). Thus, unsupervised learning with fixed number of clusters can be used in order to identify localization patterns. These results further suggest that it is important to explore the data structure in detail and in particular the large clusters. This then allows to tailor the level of detail in the analysis to the biological question.

Next, we applied hierarchical clustering (Supplementary Note 5). This approach arranges the data based on similarity and does not require a pre-definition or inference of the number of classes. Importantly, this allows to visually explore the substructure of large clusters. It also permits to inspect which features are similar among sub-groups, providing a basis for a more informed bio-physical interpretation of clusters and sub-clusters. Hierarchical clustering revealed diverse and subtle localization patterns. For instance, we observed highly specific sub-clusters of nuclear-associated patterns corresponding to intra-nuclear localization (Fig. 3c). We also observed distinct pattern mixtures for RNA foci, with localization either towards the cell membrane or the nuclear envelope (Fig. 3d).

In summary, we show that the developed workflow is capable of describing the different manually annotated localization patterns in real experiments. We also illustrate how different non-supervised clustering methods can be used to explore the data.

Supervised classification of experimental data. We next addressed the question of whether the simulation framework can be used in a supervised setting. Unlike unsupervised methods, supervised learning allows to impose prior knowledge in the form of a training set, i.e., to give the algorithm the opportunity to learn which feature combinations are relevant for a biologically meaningful distinction between patterns. We found that due to the realistic nature of our simulations (Supplementary Note 5), we could train a classifier (here: Random Forests[26] on simulated data, and successfully detected the correct localization patterns in experimental data (Fig. 4a, Supplementary Note 5). As a further validation, we compared the performance of this classifier and a classifier trained on manually annotated data and found nearly identical performance (Supplementary Note 5). These results show that the feature distributions of simulated and real data are close enough to allow to train a classifier with simulated data without a notable loss in performance.

Heterogeneity in mRNA localization. Interestingly, with this analysis we can study heterogeneity in RNA localization in detail. As described before, such heterogeneity can correspond to the co-existence of several cellular subpopulations with pure patterns, or a mixture of patterns within individual cells. To distinguish between these two scenarios, we turned to the posterior probabilities for single cells, i.e., the probabilities of a cell to belong to each of the patterns (Fig. 4b, c, Supplementary Note 5). From these posterior probabilities we can calculate the Gini impurity at the single-cell level, which gives an indication of the purity of the pattern for that particular cell (low values indicating a very pure pattern). The Gini impurity can also be calculated on the population level, i.e., on the average posterior probabilities, indicating the heterogeneity at the population level. By plotting the Gini impurity for the population against the intracellular Gini impurity (Fig. 4d), we can further investigate the nature of the heterogeneity. This reveals that some genes are characterized by rather low heterogeneity both at the cellular and the population level, such as *RAB13* and *KIF1C*, the two mRNAs accumulating in cell extensions. For other genes, we observed low heterogeneity at the cellular level and high heterogeneity at the population level. This is the case for *CEP192*, and it indicates the existence of pure subpopulations and suggests the co-existence of different localization states (Fig. 4c). Other genes are characterized by a high intra-cellular heterogeneity, i.e., a mixture of patterns, such as *DYNC1H1* (Fig. 4c). Both population and intra-cellular

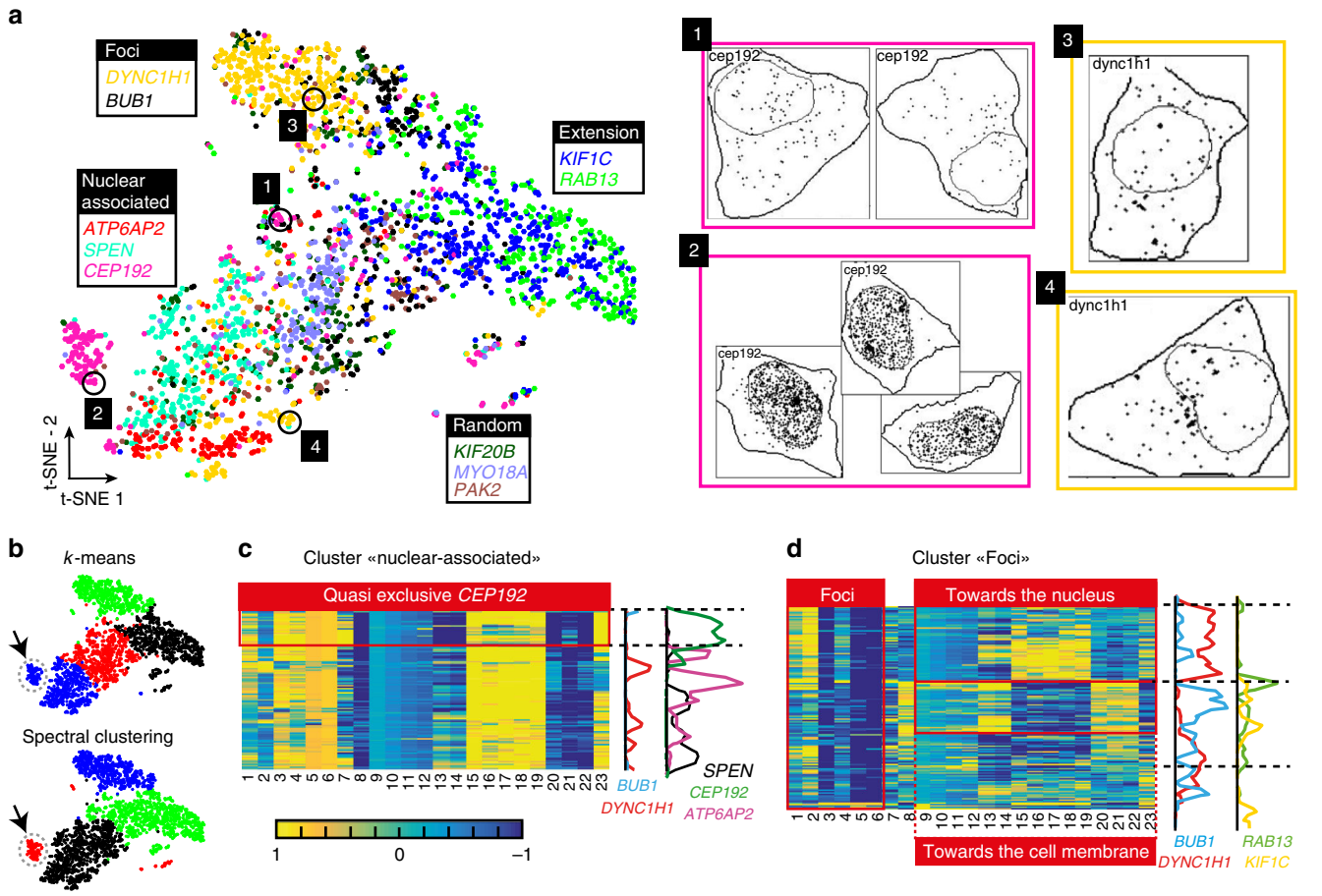


Fig. 3 Analysis of experimental data with unsupervised approaches. **a** t-SNE projection of the 23 localization features for smFISH experiments against 10 different genes. Each dot is one cell and is color-coded according to the gene. Images on the right are examples of *DYNC1H1* and *CEP192* cells with different localization patterns from the numbered regions in the t-SNE indicated with circles. **b** t-SNE plot as in **a**, but color-coded according to results of a k-means classification or spectral clustering. The latter finds a small cluster with strong intranuclear localization (lower plot, black arrow). **c, d** Examples of hierarchical clustering results. Each row is a cell, each column a localization feature (see list of features in Table 1). Plots on the right show the smoothed distribution of the most enriched genes in the clusters

heterogeneity could imply a high degree of plasticity in mRNA localization mechanisms, possibly because of the need to rapidly adapt to changes in cell shape and cellular micro-environment. This also shows that cell-to-cell variations in gene expression occur not only on the expression level as often noted[23], but also extends to the spatial dimension.

mRNA localization in different experimental conditions. We finally tested how our workflow could be used to investigate changes in RNA localization. Such an analysis can allow to infer the biological relevance of an observed pattern, for instance by performing perturbation experiments. In a recent study, we reported that *DYNC1H1* displays RNA foci[11]. We further showed that these foci do not overlap with P-bodies or stress granules, but are translation factories. To test the role of translation of the formation of foci, we treated cells with Puromycin, a terminator of translation elongation. Visual examination indicates that the foci disappeared after a brief puromycin treatment (30 min). When analyzing these data with our GMM RNA detection approach, we could successfully quantify this change in localization pattern, and also show a change in the localization classes with both unsupervised and supervised approaches (Supplementary Note 5).

Discussion

In this study, we present a framework for the simulation of smFISH images with non-random mRNA localization patterns and methods for their computational analysis. The realistic simulations allow to: (i) benchmark computational workflows aiming at studying RNA localization; (ii) assess the descriptive power of features; and (iii) evaluate the overall performance of clustering and classification methods. We developed a set of 23 localization features that could group cells based on seven simulated localization patterns with unprecedented accuracy. We further analyzed experimental data and correctly identified manually annotated localization classes. Importantly, we also demonstrate that simulated smFISH data can be used to train classifiers in a supervised setting, which when applied on real smFISH data allow to observe and quantify heterogeneity in RNA localization.

Manual annotation of smFISH images is complicated, time-consuming and subjective. When manually annotating RNA localization patterns, one typically would like to label cells with respect to the rule according to which RNAs are placed in the cell (e.g., localization close to a membrane or polarized localization). While visually solving this problem can be challenging, it is straightforward to generate images using this rule by simulation. With this, we do not only overcome the shortcomings of manual

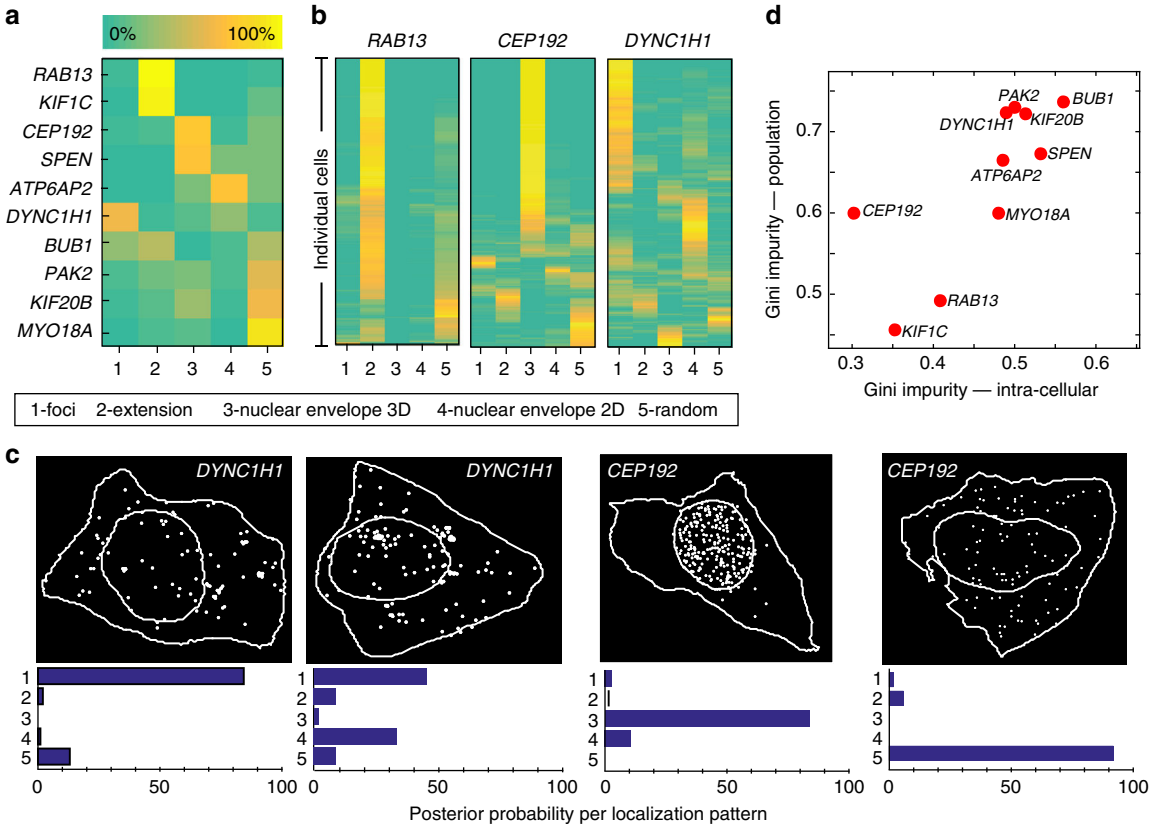


Fig. 4 Supervised analysis of experimental data reveals localization heterogeneity. **a** Results of supervised random forest classification trained on 5 classes (mRNA foci, nuclear envelope in 2D or 3D, cell extension and random) from simulated data. Heatmap shows the majority voting results for all cells of each gene. Among the simulation classes, "Nuclear envelope 3D" is closest to the experimental class intranuclear. **b** Posterior probabilities for single cells of the indicated genes. **c** Example of individual DYNC1H1 cells and their posterior probability. **d** Scatter plot of Gini impurity calculated on average probability of a gene (population) against the average Gini impurity of individual cells of a gene (intra-cellular)

annotation, but we can also use these simulations to identify limitations of the methods and to evaluate the robustness with respect to confounders, such as the density of transcripts or the shape of cells. While we cannot exclude differences between synthetic and real images, we demonstrate that synthetic and real images are similar enough to be used for benchmarking and training.

We explore several different ways of analyzing simulated and experimental data and illustrate how the richness of RNA localization can be explored and quantified by different established statistical approaches. First, we used a qualitative exploration with t-SNE. This analysis not only provides important insights about predominant RNA localization patterns but is also informative about the subtleties in these patterns. For instance, it can show the existence of small sub-groups of cells with a very specific localization pattern, or mixtures of different patterns. Second, we used different unsupervised classification approaches (k-means, spectral clustering, and hierarchical clustering). k-means provides a good global view over the predominant localization classes but is not well suited to detect small sub-populations, even if they have a strong signature in their features. Here, spectral-clustering provides better sensitivity. Complementary to this, hierarchical clustering allows performing a detailed inspection of the data. In this analysis, the number of classes is not pre-defined or inferred and the structure of the data can be investigated to discover sub-clusters. An additional benefit is the interpretability of the features, allowing attribution of biophysical properties to the identified clusters. Lastly, we applied supervised classification with Random Forests. We used the high quality of the simulated data to train a classifier, which could then be applied to experimental data. Supervised classification allows imposing prior biological knowledge and an advantage is to find rare localization classes that can be missed by unsupervised approaches. We also propose a way to quantitatively analyze the localization heterogeneity at the population and intra-cellular level. Which approach is used for a particular data set depends on the underlying biological question and the required level of detail necessary to answer this question.

Automatic analysis of an experimental data set consisting of 10 genes revealed different levels of heterogeneity in terms of RNA localization: heterogeneity at the population and intra-cellular level have different origins and have to be interpreted accordingly. The existence of different localization patterns in different cellular subpopulations for a given gene could be indicative of different biological states, and the localization pattern could be the approximation of these cellular states. Interpretation of pattern mixtures within a single cell is more complicated. If the patterns are mutually exclusive (for instance, if a fraction of RNA localizes at the nuclear envelope and another subset of RNA at the cell membrane), this may point to a double function of the encoded protein or a very high localization dynamic. If the patterns are not mutually exclusive (for instance organization into foci, where foci can be at different locations in the cell), this rather indicates the existence of subclasses, i.e., subtleties inside the defined classes, that have not been defined as separate classes. These subtleties could potentially bare important information, as

they might be indicative of different biological functions, e.g., by interaction with different cellular components. An important case corresponds to mRNA accumulating in foci. This generic localization class represents one of the most frequently observed patterns in *Drosophila*[3], but occurs in other model systems as well. Such foci are likely implicated in different biological processes, such as storage in P-bodies[27], vesicles[28], or specialized translation factories[11]. Their function can then be investigated by perturbation experiments, additional markers, or secondary criteria such as sub-cellular position.

While we validated and applied the workflow mainly in HeLa cells, we also successfully identified RNA localization patterns in a different cell line (C2C12 cells, Supplementary Note 5). The presented methodology should thus be applicable to other model systems. Recent studies show wide-spread RNA localization in organisms such as *Drosophila* embryos[3]. While patterns occurring at the scale of the entire embryo (or tissue) would require an extension of the proposed workflow, namely the addition of other landmarks and probably other localization features, there are number of intra-cellular localization patterns where our methodology could essentially complement these studies.

In this work, we integrated two generic cellular landmarks, namely the cytoplasm and the nucleus, which we believe are important for any large-scale localization screen. Such a screen could then reveal interesting localization patterns that are investigated in targeted follow-up experiments employing other markers tailored to the hypothesis inferred from the large-scale screen. For instance, RNA foci have been observed for a number of genes[3], but their functional role still needs to be elucidated. Such an analysis requires perturbation experiments and/or the inclusion of additional makers (Supplementary Note 5). We showed previously that *DYNC1H1* foci are neither P-bodies nor stress granules, but act as specialized translation factories. The open design of our approach allows integration of additional markers to discriminate cellular compartments.

In summary, we provide a framework for simulation of synthetic smFISH images and validated methods and tools for the analysis of intracellular RNA localization. We validated this workflow on simulated data and demonstrated how different statistical approaches can be used to investigate the complexity and richness of smFISH images and therefore provide methods and tools to explore the spatial dimension of gene expression inside cells.

Methods

Cell culture. HeLa and C2C12 cell lines were obtained from ATCC and grown in DMEM medium (Gibco) supplemented with 10% FCS (Sigma). HeLa cell lines stably transfected with GFP-tagged BAC[29] were a kind gift from Anthony Hyman laboratory (MPI-CBG, Dresden, Germany) and were grown in DMEM medium supplemented with 10% FCS and 0.4 mg/ml G418 (Gibco).

smFISH probes. smiFISH probes were purchased from Integrated DNA Technologies. smFISH probes were synthesized by J.M. Escudier (SPCMIB, Toulouse, France) and labeled with Cy3 mono-reactive dye pack (GE Healthcare). Probe sequences are available in Supplementary Data 1.

smFISH experiments for cell shape. Single molecule FISH (smFISH) and our recently published inexpensive variant (smiFISH[13]) were performed as follows. Briefly, cells were fixed with 4% paraformaldehyde (Electron Microscopy Sciences) for 20 min at RT, and permeabilised in 70% ethanol overnight at 4 °C.

smiFISH against GAPDH, *RAB13*, *KIF1C*, *DYNC1H1*, *ACTN1* was performed in HeLa or C2C12 cells against the endogenous mRNAs, using two types of probes: (i) 24 unlabeled primary probes containing both a mRNA targeting sequence and a shared sequence (FLAP); (ii) a secondary probe conjugated to two Cy5 moieties (for *GAPDH*) or two Cy3 moieties (for *RAB13*, *KIF1C*, *DYNC1H1*, *ACTN1*), pre-hybridized in vitro at 65 °C to the primary probes via the FLAP sequence. For all other mRNAs, HeLa cell lines stably expressing a GFP tagged version of the gene of interest, which was expressed from a bacterial artificial chromosome (BAC)[29], were used. These BACs carry the entire gene regulatory sequences and in particular the mRNA untranslated regions often required for proper mRNA localization. smFISH was performed against the GFP-IRES-Neo sequence of the BAC, with a pool of 40 oligonucleotide probes, carrying up to four Cy3 fluorophores each. For *BUB1*, *RAB13*, *KIF1C*, a CellMask^{TM} channel was additionally recorded. Mock-smFISH was performed in HeLa cells with the above cited pool of probes, targeting an artificial sequence (GFP-IRES-Neo), not expressed in the cells.

All FISH experiments were performed overnight at 37 °C in a buffer containing 1X SSC, 15% formamide (Sigma), 10% dextran sulfate (Sigma), 0.34 mg/ml tRNA (Sigma), 0.2 mg/ml BSA (Roche Diagnostics), 2 mM VRC (Sigma). The next day, the samples were washed at 37 °C in 1× SSC, 15% formamide. In the specified cases, the cells were labeled with HCS CellMask^{TM} Green Stain (Molecular Probes) diluted to 50 ng/ml in PBS, for 5 min following FISH. Cells were subsequently washed with PBS, and mounted with Vectashield mounting medium with DAPI (Vector Laboratories).

Imaging to infer cell shape. Three-dimensional image stacks were captured on a wide-field microscope (Zeiss Axioimager Z1) equipped with a 63 × 1.4 NA objective and a scMOS camera (Zyla 4.2 MP, Andor Technology) and controlled with Metamorph (Version 7.8.8.0; Molecular Devices).

Image analysis to infer cell shape. 2D segmentation of nuclei and cells was performed with the DAPI and CellMask^{TM} channels with the open-source software CellCognition[19] using a standard segmentation workflow: Otsu and watershed separation for nuclei in the DAPI channel. Each nucleus then serves as a seed for a watershed segmentation to obtain the cells in the CellMask^{TM} channel. Individual GAPDH mRNA molecules were localized with FISH-quant[14] in 3D. The Matlab function boundary was used to determine the conforming 3D boundary around the mRNAs, corresponding to the 3D cellular outline. GAPDH is predominately excluded from the nuclei. To infer the positioning and height of the nuclei, the z-positions of mRNAs localized in small region in the center of the nuclei were analyzed. mRNAs were automatically separated into two groups above and below the nucleus by k-means clustering. This allowed inference of the average 3D height and position of the nuclei. For more information, see Supplementary Note 1.

mRNA detection and localization in simulations. mRNA detection in 2D was performed with the MATLAB script provided in Battich et al.[7].

mRNA detection in 3D was performed with a standard spot detection approach in FISH-quant[14]. In short, an approximation of the second derivative of the images was first calculated with a LoG filter. mRNAs were pre-detected with a local maximum detection with a manually determined intensity threshold. After detection, spots are fitted with a 3D Gaussian function and the GMM approach applied as detailed in Supplementary Note 2. In short, the algorithm reconstructs the mRNA foci in multiple single mRNA by using the average signal of a single mRNA as a reference.

For each cell, the different localization features are calculated in Matlab, and saved as a text file that can be further analyzed. Ripley-K function calculation was implemented from scratch. The Matlab functions pdist and p_poly_dist (from Matlab file exchange) were used to compute the distance between mRNAs and a reference point (cell centroid, nucleus centroid) and mRNA and a polygon (cell membrane, nuclear envelope). The Matlab function imopen was used to perform a morphological opening on the cellular outlines. The normalization coefficients were calculated using the bwdist function that compute the distance map of a binary mask.

Statistical methods to analyze localization features. t-SNE projections were performed using the tsne function implemented in MATLAB available at https://lvdmaaten.github.io/tsne/. k-means classification was performed using the Matlab function kmeans using 50 replicates; spectral clustering with the function SpectralClustering from Matlab file exchange; hierarchical clustering with the Matlab function clustergram with linkage method ward. Random forest (RF) classification was performed with the Matlab function treebagger. 100 trees were trained on simulated cells with moderate and high pattern strength and mRNA densities corresponding to an average of 100 and 200 mRNAs/cell. Gini impurity was calculated from the posterior probabilities for the RF classification: either from the mean probability per mRNA (population impurity) or the mean impurity for all cells of a mRNA (intracellular impurity).

Cell segmentation for mRNA localization. Automated nuclear and cell segmentation was performed with a custom algorithm based on the U-net[30] deep convolutional network. Nuclear segmentation was performed with the DAPI channel, cell segmentation was performed either with CellMask^{TM} (if present) or with the actual smFISH image. For segmentation, 3D images were projected into 2D images with a recently described approach[13].

mRNA detection and feature calculation. Analysis of experimental smFISH data was performed as described for simulated data. Except that the detection intensity threshold was determined for each experiment (mRNA) separately. We implemented a simple Matlab user interface FQ_detect, which is now part of FISH-

quant, to facilitate the determination of this threshold. It proposes an automatically calculated threshold based on Otsu thresholding, based on empirically determined criteria. We refer to the FISH-quant documentation for more details.

Code availability. We provide the entire Matlab code for the analysis described in this paper, which we integrated into our smFISH analysis package FISH-quant[14] available at https://bitbucket.org/muellerflorian/fish_quant

Data availability

Data used to simulate images are available on Zenodo https://doi.org/10.5281/zenodo.1413488. Additional experimental data supporting the findings of this study are available from the corresponding authors upon reasonable request.

Received: 30 November 2017 Accepted: 1 October 2018
Published online: 02 November 2018

References

1. Chin, A. & Lécuyer, E. RNA localization: making its way to the center stage. *Biochim. Biophys. Acta Gen. Subj.* **1861**, 2956–2970 (2017).
2. Buxbaum, A. R., Haimovich, G. & Singer, R. H. In the right place at the right time: visualizing and understanding mRNA localization. *Nat. Rev. Mol. Cell Biol.* **16**, 95–109 (2015).
3. Wilk, R., Hu, J., Blotsky, D. & Krause, H. M. Diverse and pervasive subcellular distributions for both coding and long noncoding RNAs. *Genes Dev.* **30**, 594–609 (2016).
4. Buskila, A. A., Kannaiah, S. & Amster-Choder, O. RNA localization in bacteria. *Rna. Biol.* **11**, 1051–1060 (2014).
5. Ni, L. & Snyder, M. A genomic study of the bipolar bud site selection pattern in Saccharomyces cerevisiae. *Mol. Biol. Cell* **12**, 2147–2170 (2001).
6. Jambor, H. *et al.* Systematic imaging reveals features and changing localization of mRNAs in Drosophila development. *eLife* **4**, e05003 (2015).
7. Battich, N., Stoeger, T. & Pelkmans, L. Image-based transcriptomics in thousands of single human cells at single-molecule resolution. *Nat. Methods* **10**, 1127–1133 (2013).
8. Cabili, M. N. et al. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol.* **16**, 20 (2015).
9. Mili, S., Moissoglu, K. & Macara, I. G. Genome-wide screen reveals APC-associated RNAs enriched in cell protrusions. *Nature* **453**, 115–119 (2008).
10. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
11. Pichon, X. et al. Visualization of single endogenous polysomes reveals the dynamics of translation in live human cells. *J. Cell. Biol.* **214**, 769–781 (2016).
12. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).
13. Tsanov, N. et al. smiFISH and FISH-quant - a flexible single RNA detection approach with super-resolution capability. *Nucleic Acids Res.* **44**, e165 (2016).
14. Mueller, F. et al. FISH-quant: automatic counting of transcripts in 3D FISH images. *Nat. Methods* **10**, 277–278 (2013).
15. Shah, S., Lubeck, E., Zhou, W. & Cai, L. In situ transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron* **92**, 342–357 (2016).
16. Park, H. Y. et al. Visualization of dynamics of single endogenous mRNA labeled in live mouse. *Science* **343**, 422–424 (2014).
17. Glory, E. & Murphy, R. F. Automated subcellular location determination and high-throughput microscopy. *Dev. Cell* **12**, 7–16 (2007).
18. Conrad, C. & Gerlich, D. W. Automated microscopy for high-content RNAi screening. *J. Cell. Biol.* **188**, 453–461 (2010).
19. Held, M. et al. CellCognition: time-resolved phenotype annotation in high-throughput live cell imaging. *Nat. Methods* **7**, 747–754 (2010).
20. Graml, V. et al. A genomic multiprocess survey of machineries that control and link cell shape, microtubule organization, and cell-cycle progression. *Dev. Cell* **31**, 227–239 (2014).
21. Neumann, B. et al. Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature* **464**, 721–727 (2010).
22. Padovan-Merhar, O. et al. Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. *Mol. Cell* **58**, 339–352 (2015).
23. Battich, N., Stoeger, T. & Pelkmans, L. Control of transcript variability in single mammalian cells. *Cell* **163**, 1596–1610 (2015).
24. Stoeger, T., Battich, N., Herrmann, M. D., Yakimovich, Y. & Pelkmans, L. Computer vision for image-based transcriptomics. *Methods* (2015). https://doi.org/10.1016/j.ymeth.2015.05.016
25. Maaten, Lvander & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
26. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).
27. Hubstenberger, A. et al. P-Body purification reveals the condensation of repressed mRNA regulons. *Mol. Cell* **68**, 144–157.e5 (2017).
28. Basyuk, E. et al. Retroviral genomic RNAs are transported to the plasma membrane by endosomal vesicles. *Dev. Cell* **5**, 161–174 (2003).
29. Poser, I. et al. BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nat. Methods* **5**, 409–415 (2008).
30. Ronneberger, O., Fischer, P. & Brox, T. U-Net: convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention,* 234–241 (2015).

Acknowledgements

We thank Anthony Hyman for providing the HeLa cell lines transfected with BAC. We thank Maria Vera for critical reading of the manuscript. This research was supported by the Agence Nationale de la Recherche (ANR-11-BSV8-018-02 and ANR-14-CE10-0018-01), Institut Pasteur, Fondation pour la Recherche Médicale (FRM DBI20131228556), and La Ligue Nationale Contre le Cancer.

Author contributions

F.M., T.W., and E.B. conceived the project. F.M. and T.W. supervised Au.S. Au.S. and F.M. developed the Matlab package to simulate and analyze RNA localization data. R.C., Ad.S., AM.T., M.P. and F.M. performed the smFISH experiments and analyzed the localization patterns with E.B. Au.S., F.M. and T.W. analyzed the data. C.Z. provided suggestions on data analysis. W.O. implemented the cell segmentation with neural networks. F.M. and T.W. wrote the manuscript.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-06868-w.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

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

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

© The Author(s) 2018

Bibliography

[Abadi et al., 2016] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X. (2016). TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *arXiv:1603.04467*.

[Ahonen et al., 2006] Ahonen, T., Hadid, A., and Pietikainen, M. (2006). Face Description with Local Binary Patterns: Application to Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):2037–2041.

[Andrychowicz et al., 2016] Andrychowicz, M., Denil, M., Gomez, S., Hoffman, M. W., Pfau, D., Schaul, T., Shillingford, B., and de Freitas, N. (2016). Learning to learn by gradient descent by gradient descent. *arXiv:1606.04474*.

[Axelrod, 1981] Axelrod, D. (1981). Cell-substrate contacts illuminated by total internal reflection fluorescence. *The Journal of cell biology*, 89(1):141–5.

[Baker and Leeuwenhoek, 1739] Baker, H. and Leeuwenhoek, M. (1739). An Account of Mr. Leeuwenhoek's Microscopes; By Mr. Henry Baker, F. R. S. *Philosophical Transactions of the Royal Society of London*, 41(452-461):503–519.

[Balzarotti et al., 2017] 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 (New York, N.Y.)*, 355(6325):606–612.

[Bates et al., 2007] Bates, M., Huang, B., Dempsey, G. T., and Zhuang, X. (2007). Multicolor Super-Resolution Imaging with Photo-Switchable Fluorescent Probes. *Science*, 317(5845):1749–1753.

[Bengio et al., 1991] Bengio, Y., Bengio, S., and Cloutier, J. (1991). Learning a synaptic learning rule. In *IJCNN-91-Seattle International Joint Conference on Neural Networks*, volume ii, page 969. IEEE.

[Bengio et al., 2009] Bengio, Y., Louradour, J., Collobert, R., and Weston, J. (2009). Curriculum learning. In *Proceedings of the 26th Annual International Conference on Machine Learning - ICML '09*, pages 1–8, New York, New York, USA. ACM Press.

[Bergstra et al., 2010] Bergstra, J., Breuleux, O., Bastien, F., Lamblin, P., Pascanu, R., Desjardins, G., Turian, J., Warde-Farley, D., and Bengio, Y. (2010). Theano: A CPU and GPU Math Compiler in Python. Technical report.

[Betzig et al., 2006] 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 (New York, N.Y.)*, 313(5793):1642–5.

[BEWERSDORF et al., 2006] BEWERSDORF, J., SCHMIDT, R., and HELL, S. W. (2006). Comparison of I5M and 4Pi-microscopy. *Journal of Microscopy*, 222(2):105–117.

[Bošnjak et al., 2016] Bošnjak, M., Rocktäschel, T., Naradowsky, J., and Riedel, S. (2016). Programming with a Differentiable Forth Interpreter. *arXiv:1605.06640*.

[Burnette et al., 2011] Burnette, D. T., Sengupta, P., Dai, Y., Lippincott-Schwartz, J., and Kachar, B. (2011). Bleaching/blinking assisted localization microscopy for superresolution imaging using standard fluorescent molecules. *Proceedings of the National Academy of Sciences*, 108(52):21081–21086.

[Carpenter et al., 2006] Carpenter, A. E., Jones, T. R., Lamprecht, M. R., Clarke, C., Kang, I., Friman, O., Guertin, D. A., Chang, J., Lindquist, R. A., Moffat, J., Golland, P., and Sabatini, D. M. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biology*, 7(10):R100.

[Caruana et al., 2000] Caruana, R., Lawrence, S., and Giles, L. (2000). Overfitting in Neural Nets: Backpropagation, Conjugate Gradient, and Early Stopping. *Advances in Neural Information Processing Systems*, pages 402–408.

[Chen et al., 2014] Chen, B.-C., Legant, W. R., Wang, K., Shao, L., Milkie, D. E., Davidson, M. W., Janetopoulos, C., Wu, X. S., Hammer, J. A., Liu, Z., English, B. P., Mimori-Kiyosue, Y., Romero, D. P., Ritter, A. T., Lippincott-Schwartz, J., Fritz-Laylin, L., Mullins, R. D., Mitchell, D. M., Bembenek, J. N., Reymann, A.-C., Böhme, R., Grill, S. W., Wang, J. T., Seydoux, G., Tulu, U. S., Kiehart, D. P., and Betzig, E. (2014). Lattice light-sheet microscopy: imaging molecules to embryos at high spatiotemporal resolution. *Science (New York, N.Y.)*, 346(6208):1257998.

[Chen and Tao, 2017] Chen, K. and Tao, W. (2017). Once for All: a Two-flow Convolutional Neural Network for Visual Tracking. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1–1.

[Cho et al., 2014] Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *arXiv:1406.1078*.

[Christiansen et al., 2018] Christiansen, E. M., Yang, S. J., Ando, D. M., Javaherian, A., Skibinski, G., Lipnick, S., Mount, E., O'Neil, A., Shah, K., Lee, A. K., Goyal, P., Fedus, W., Poplin, R., Esteva, A., Berndl, M., Rubin, L. L., Nelson, P., and Finkbeiner, S. (2018). In Silico Labeling: Predicting Fluorescent Labels in Unlabeled Images. *Cell*, 173(3):792–803.

[Cireşan et al., 2012] Cireşan, D., Meier, U., and Schmidhuber, J. (2012). Multi-column Deep Neural Networks for Image Classification. *arXiv:1202.2745*.

[Ciressan et al., 2012] Ciressan, D. C., Giusti, A., Gambardella, L., and Schmidhuber, J. (2012). Neural networks for segmenting neuronal structures in EM stacks. *IN ISBI SEGMENTATION CHALLENGE COMPETITION: ABSTRACTS*.

[Cong et al., 2013] Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., and Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*, 339(6121):819–823.

[Coons et al., 1941] 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):200–202.

[Cox and Sheppard, 2004] Cox, G. and Sheppard, C. J. R. (2004). Practical Limits of Resolution in Confocal and Non-Linear Microscopy. *Microsc. Res. Tech*, 63:18–22.

[Cox et al., 2012] Cox, S., Rosten, E., Monypenny, J., Jovanovic-Talisman, T., Burnette, D. T., Lippincott-Schwartz, J., Jones, G. E., and Heintzmann, R. (2012). Bayesian localization microscopy reveals nanoscale podosome dynamics. *Nature Methods*, 9(2):195–200.

[Crawford et al., 2016] Crawford, K. L., Neu, S. C., and Toga, A. W. (2016). The Image and Data Archive at the Laboratory of Neuro Imaging. *NeuroImage*, 124(Pt B):1080–3.

[Culley et al., 2018] Culley, S., Albrecht, D., Jacobs, C., Pereira, P. M., Leterrier, C., Mercer, J., and Henriques, R. (2018). Quantitative mapping and minimization of super-resolution optical imaging artifacts. *Nature Methods*, 15(4):263–266.

[De Cao and Kipf, 2018] De Cao, N. and Kipf, T. (2018). MolGAN: An implicit generative model for small molecular graphs. *arXiv:1805.11973*.

[Deng et al., 2009] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Li, F.-F. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255.

[Der Kiureghian and Ditlevsen, 2008] Der Kiureghian, A. and Ditlevsen, O. (2008). Aleatory or epistemic? Does it matter? *Structural Safety*, 31:105–112.

[Dertinger et al., 2009] Dertinger, T., Colyer, R., Iyer, G., Weiss, S., and Enderlein, J. (2009). Fast, background-free, 3D super-resolution optical fluctuation imaging (SOFI). *Proceedings of the National Academy of Sciences*, 106(52):22287–22292.

[Devlin et al., 2017] Devlin, J., Bunel, R., Singh, R., Hausknecht, M., and Kohli, P. (2017). Neural Program Meta-Induction. *arXiv:1710.04157v1*.

[Diederich et al., 2018] Diederich, B., Then, P., Jügler, A., Förster, R., and Heintzmann, R. (2018). cellSTORM - Cost-effective Super-Resolution on a Cellphone using dSTORM. *arXiv:1804.06244*.

[Dinh et al., 2017] Dinh, L., Pascanu, R., Bengio, S., and Bengio, Y. (2017). Sharp Minima Can Generalize For Deep Nets. *arXiv:1703.04933*.

[Dong et al., 2016a] Dong, C., Loy, C. C., He, K., and Tang, X. (2016a). Image Super-Resolution Using Deep Convolutional Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):295–307.

[Dong et al., 2016b] Dong, C., Loy, C. C., and Tang, X. (2016b). Accelerating the Super-Resolution Convolutional Neural Network. In *ECCV*, pages 391–407, Amsterdam.

[Duan et al., 2016] Duan, Y., Schulman, J., Chen, X., Bartlett, P. L., Sutskever, I., and Abbeel, P. (2016). RLˆ2: Fast Reinforcement Learning via Slow Reinforcement Learning. *arXiv:1611.02779*.

[Duchi JDUCHI and Singer, 2011] Duchi JDUCHI, J. and Singer, Y. (2011). Adaptive Subgradient Methods for Online Learning and Stochastic Optimization * Elad Hazan. *Journal of Machine Learning Research*, 12:2121–2159.

[Eigen et al., 2014] Eigen, D., Puhrsch, C., and Fergus, R. (2014). Depth Map Prediction from a Single Image using a Multi-Scale Deep Network. *arXiv:1406.2283*.

[Elman, 1993] Elman, J. L. (1993). Learning and development in neural networks: the importance of starting small. *Cognition*, 48(1):71–99.

[Esteva et al., 2017] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., and Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118.

[Finn et al., 2017] Finn, C., Abbeel, P., and Levine, S. (2017). Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. *arXiv:1703.03400*.

[Fodor and Pylyshyn, 1988] Fodor, J. A. and Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28(1-2):3–71.

[Franke et al., 2017] Franke, C., Sauer, M., and van de Linde, S. (2017). Photometry unlocks 3D information from 2D localization microscopy data. *Nature Methods*, 14(1):41–44.

[Ganin and Lempitsky, 2015] Ganin, Y. and Lempitsky, V. (2015). Unsupervised Domain Adaptation by Backpropagation. *Proceedings of The 32nd International Conference on Machine Learning*, (i):1180–1189.

[Ge et al., 2015] Ge, R., Huang, F., Jin, C., and Yuan, Y. (2015). Escaping From Saddle Points — Online Stochastic Gradient for Tensor Decomposition. *arXiv:1503.02101*.

[Geffner, 2018] Geffner, H. (2018). Model-free, Model-based, and General Intelligence. *arXiv:1806.02308*.

[Ghahramani and Jordan, 1997] Ghahramani, Z. and Jordan, M. I. (1997). Factorial Hidden Markov Models. *Machine Learning*, 29(2/3):245–273.

[Goldspink et al., 2017] Goldspink, D. A., Matthews, Z. J., Lund, E. K., Wileman, T., and Mogensen, M. M. (2017). Immuno-fluorescent Labeling of Microtubules and Centrosomal Proteins in Ex Vivo Intestinal Tissue and 3D In Vitro Intestinal Organoids. *Journal of visualized experiments*, (130).

[Goodfellow et al., 2014] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative Adversarial Nets. *Advances in Neural Information Processing Systems 27*, pages 2672–2680.

[Graves et al., 2014] Graves, A., Wayne, G., and Danihelka, I. (2014). Neural Turing Machines. *arXiv:1410.5401*.

[Graves et al., 2016] Graves, A., Wayne, G., Reynolds, M., Harley, T., Danihelka, I., Grabska-Barwińska, A., Colmenarejo, S. G., Grefenstette, E., Ramalho, T., Agapiou, J., Badia, A. P., Hermann, K. M., Zwols, Y., Ostrovski, G., Cain, A., King, H., Summerfield, C., Blunsom, P., Kavukcuoglu, K., and Hassabis, D. (2016). Hybrid computing using a neural network with dynamic external memory. *Nature*, 538(7626):471–476.

[Grefenstette et al., 2015] Grefenstette, E., Hermann, K. M., Suleyman, M., and Blunsom, P. (2015). Learning to Transduce with Unbounded Memory. *arXiv:1506.02516*.

[Gustafsson et al., 1999] Gustafsson, M. G., Agard, D. A., and Sedat, J. W. (1999). I5M: 3D widefield light microscopy with better than 100 nm axial resolution. *Journal of microscopy*, 195(Pt 1):10–6.

[Gustafsson, 2005] Gustafsson, M. G. L. (2005). Nonlinear structured-illumination microscopy: Wide-field fluorescence imaging with theoretically unlimited resolution. *Proceedings of the National Academy of Sciences*, 102(37):13081–13086.

[Gustafsson et al., 2016] Gustafsson, N., Culley, S., Ashdown, G., Owen, D. M., Pereira, P. M., and Henriques, R. (2016). Fast live-cell conventional fluorophore nanoscopy with ImageJ through super-resolution radial fluctuations. *Nature Communications*, 7:12471.

[He et al., 2015] He, K., Zhang, X., Ren, S., and Sun, J. (2015). Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *arXiv:1502.01852*.

[Hell et al., 2003] Hell, S., Jakobs, S., and Kastrup, L. (2003). Imaging and writing at the nanoscale with focused visible light through saturable optical transitions. *Applied Physics A: Materials Science & Processing*, 77(7):859–860.

[Hell, 2005] Hell, S. W. (2005). Fluorescence nanoscopy: Breaking the diffraction barrier by the RESOLFT concept. *NanoBiotechnology*, 1(3):296–297.

[Hell, 2007] Hell, S. W. (2007). Far-field optical nanoscopy. *Science (New York, N.Y.)*, 316(5828):1153–8.

[Hershko* et al., 2018] Hershko*, E., Weiss*, L. E., Michaeli, T., and Shechtman, Y. (2018). Multicolor localization microscopy by deep learning. *arXiv:1807.01637*.

[Hestness et al., 2017] Hestness, J., Narang, S., Ardalani, N., Diamos, G., Jun, H., Kianinejad, H., Patwary, M. M. A., Yang, Y., and Zhou, Y. (2017). Deep Learning Scaling is Predictable, Empirically. *arXiv:1712.00409*.

[Hinton et al.,] Hinton, G., Srivastava, N., and Swersky, K. Neural Networks for Machine Learning Lecture 6a Overview of mini--batch gradient descent. Technical report.

[Hochreiter and Schmidhuber, 1997] Hochreiter, S. and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780.

[Holden et al., 2011] Holden, S. J., Uphoff, S., and Kapanidis, A. N. (2011). DAOSTORM: an algorithm for high- density super-resolution microscopy. *Nature Methods*, 8(4):279–280.

[Hsu et al., 2014] Hsu, P. D., Lander, E. S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, 157(6):1262–78.

[Huang et al., 2013] Huang, F., Hartwich, T. M. P., Rivera-Molina, F. E., Lin, Y., Duim, W. C., Long, J. J., Uchil, P. D., Myers, J. R., Baird, M. A., Mothes, W., Davidson, M. W., Toomre, D., and Bewersdorf, J. (2013). Video-rate nanoscopy using sCMOS camera-specific single-molecule localization algorithms. *Nature methods*, 10(7):653–8.

[Huang et al., 2011] Huang, F., Schwartz, S. L., Byars, J. M., and Lidke, K. A. (2011). Simultaneous multiple-emitter fitting for single molecule super-resolution imaging. *Biomedical optics express*, 2(5):1377–93.

[Ioffe and Szegedy, 2015] Ioffe, S. and Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv:1502.03167*.

[Isola et al., 2016] Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2016). Image-to-Image Translation with Conditional Adversarial Networks. *arXiv:1611.07004*.

[Iudin et al., 2016] Iudin, A., Korir, P. K., Salavert-Torres, J., Kleywegt, G. J., and Patwardhan, A. (2016). EMPIAR: a public archive for raw electron microscopy image data. *Nature Methods*, 13(5):387–388.

[Jianchao Yang et al., 2010] Jianchao Yang, Wright, J., Huang, T. S., and Yi Ma (2010). Image Super-Resolution Via Sparse Representation. *IEEE Transactions on Image Processing*, 19(11):2861–2873.

[Jones et al., 2011] Jones, S. A., Shim, S.-H., He, J., and Zhuang, X. (2011). Fast, three-dimensional super-resolution imaging of live cells. *Nature methods*, 8(6):499–508.

[Joulin and Mikolov, 2015] Joulin, A. and Mikolov, T. (2015). Inferring Algorithmic Patterns with Stack-Augmented Recurrent Nets. *arXiv:1503.01007*.

[Kawaguchi et al., 2017] Kawaguchi, K., Kaelbling, L. P., and Bengio, Y. (2017). Generalization in Deep Learning. *arXiv:1710.05468*.

[Kendall and Gal, 2017] Kendall, A. and Gal, Y. (2017). What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? pages 1–19. *arXiv:1703.04977*.

[Keskar et al., 2016] Keskar, N. S., Mudigere, D., Nocedal, J., Smelyanskiy, M., and Tang, P. T. P. (2016). On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima. *arXiv:1609.04836*.

[Kim et al., 2015] Kim, J., Lee, J. K., and Lee, K. M. (2015). Deeply-Recursive Convolutional Network for Image Super-Resolution. *arXiv:1511.04491*.

[Kim et al., 2016] Kim, J., Lee, J. K., and Lee, K. M. (2016). Accurate Image Super-Resolution Using Very Deep Convolutional Networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1646–1654. IEEE.

[Kingma and Ba, 2014] Kingma, D. P. and Ba, J. (2014). Adam: A Method for Stochastic Optimization. *arXiv:1412.6980*.

[Kingma and Welling, 2013] Kingma, D. P. and Welling, M. (2013). Auto-Encoding Variational Bayes. *arXiv:1312.6114*.

[Kolb et al., 2001] 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):2004–2021.

[Krueger et al., 2017] Krueger, D., Ballas, N., Jastrzebski, S., Arpit, D., Kanwal, M. S., Maharaj, T., Bengio, E., Fischer, A., and Courville, A. (2017). Deep Nets Don't Learn via Memorization. *ICLR*, pages 1–4.

[L. Peter Deutsch, 1996] L. Peter Deutsch (1996). DEFLATE Compressed Data Format Specification version 1.3. Technical report.

[Lai et al., 2017] Lai, W.-S., Huang, J.-B., Ahuja, N., and Yang, M.-H. (2017). Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution. *IEEE Conference on Computer Vision and Pattern Recognition*.

[Lake et al., 2015] Lake, B. M., Salakhutdinov, R., and Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science (New York, N.Y.)*, 350(6266):1332–8.

[LeCun et al., 2015] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

[Lecun et al., 1998] Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.

[Lee et al., 2009] Lee, H., Grosse, R., Ranganath, R., and Ng, A. Y. (2009). Convolutional Deep Belief Networks for Scalable Unsupervised Learning of Hierarchical Representations. Technical report.

[Legant et al., 2016] 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(4):359–365.

[Lehtinen et al., 2018] Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., and Aila, T. (2018). Noise2Noise: Learning Image Restoration without Clean Data. *arXiv:1803.04189*.

[Li et al., 2018] Li, Y., Mund, M., Hoess, P., Deschamps, J., Matti, U., Nijmeijer, B., Sabinina, V. J., Ellenberg, J., Schoen, I., and Ries, J. (2018). Real-time 3D single-molecule localization using experimental point spread functions. *Nature Methods*, 15(5):367–369.

[Long et al., 2014] Long, J., Shelhamer, E., and Darrell, T. (2014). Fully Convolutional Networks for Semantic Segmentation. *arXiv:1411.4038*.

[Lotter et al., 2015] Lotter, W., Kreiman, G., and Cox, D. (2015). Unsupervised Learning of Visual Structure using Predictive Generative Networks. *arXiv:1511.06380*.

[MacKay, 1992] MacKay, D. J. C. (1992). A Practical Bayesian Framework for Backpropagation Networks. *Neural Computation*, 4(3):448–472.

[Manna and Waldinger, 1980] Manna, Z. and Waldinger, R. (1980). A Deductive Approach to Program Synthesis. *ACM Transactions on Programming Languages and Systems*, 2(1):90–121.

[Marcus, 2001] Marcus, G. F. G. F. (2001). *The algebraic mind : integrating connectionism and cognitive science*. MIT Press.

[Marsh et al., 2018] Marsh, R. J., Pfisterer, K., Bennett, P., Hirvonen, L. M., Gautel, M., Jones, G. E., and Cox, S. (2018). Artifact-free high-density localization microscopy analysis. *Nature Methods*, 15(9):689–692.

[Mathieu et al., 2015] Mathieu, M., Couprie, C., and LeCun, Y. (2015). Deep multi-scale video prediction beyond mean square error. *arXiv:1511.05440*.

[McCallum et al., 2009] McCallum, a., Schultz, K., and Singh, S. (2009). Factorie: Probabilistic programming via imperatively defined factor graphs. *Advances in Neural Information Processing Systems*, 22:1249–1257.

[Minsky, 1988] Minsky, M. (1988). Memoir on inventing the confocal scanning microscope. *Scanning*, 10(4):128–138.

[Mishkin and Matas, 2015] Mishkin, D. and Matas, J. (2015). All you need is a good init. *arXiv:1511.06422*.

[Mnih et al., 2015] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., and Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533.

[Neal, 1996] Neal, R. M. (1996). *Bayesian Learning for Neural Networks*. Springer New York.

[Nehme et al., 2018] Nehme, E., Weiss, L. E., Michaeli, T., and Shechtman, Y. (2018). Deep-STORM: super-resolution single-molecule microscopy by deep learning. *Optica*, 5(4):458.

[Nichol et al., 2018] Nichol, A., Achiam, J., and Schulman, J. (2018). On First-Order Meta-Learning Algorithms. *arXiv:1803.02999*.

[Nollmann and Georgieva, 2015] Nollmann, M. and Georgieva, M. (2015). Superresolution microscopy for bioimaging at the nanoscale: from concepts to applications in the nucleus. *Research and Reports in Biology*, 6:157.

[Osokin et al., 2017] Osokin, A., Chessel, A., Salas, R. E. C., and Vaggi, F. (2017). GANs for Biological Image Synthesis. In *IEEE International Conference on Computer Vision*.

[Ouyang et al., 2018] Ouyang, W., Aristov, A., Lelek, M., Hao, X., and Zimmer, C. (2018). Deep learning massively accelerates super-resolution localization microscopy. *Nature Biotechnology*, 36(5):460–468.

[Ouyang and Zimmer, 2017] Ouyang, W. and Zimmer, C. (2017). The imaging tsunami: Computational opportunities and challenges. *Current Opinion in Systems Biology*, 4:105–113.

[Ovesný et al., 2014] Ovesný, M., Křížek, P., Borkovec, J., Švindrych, Z., and Hagen, G. M. (2014). ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. *Bioinformatics*, 30(16):2389–2390.

[Paszke et al., 2017] Paszke, A., Chanan, G., Lin, Z., Gross, S., Yang, E., Antiga, L., and Devito, Z. (2017). Automatic differentiation in PyTorch. *Advances in Neural Information Processing Systems 30*, (Nips):1–4.

[Pathak et al., 2016] Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., and Efros, A. A. (2016). Context Encoders: Feature Learning by Inpainting. *arXiv:1604.07379*.

[Plautz et al., 1996] Plautz, J. D., Day, R. N., Dailey, G. M., Welsh, S. B., Hall, J. C., Halpain, S., and Kay, S. A. (1996). Green fluorescent protein and its derivatives as versatile markers for gene expression in living Drosophila melanogaster, plant and mammalian cells. *Gene*, 173(1 Spec No):83–7.

[Pohl et al., 1984] Pohl, D. W., Denk, W., and Lanz, M. (1984). Optical stethoscopy: Image recording with resolution $\lambda/20$. *Applied Physics Letters*, 44(7):651–653.

[Przybylski et al., 2017] Przybylski, A., Thiel, B., Keller-Findeisen, J., Stock, B., and Bates, M. (2017). Gpufit: An open-source toolkit for GPU-accelerated curve fitting. *Scientific Reports*, 7(1):15722.

[Ram et al., 2006] Ram, S., Ward, E. S., and Ober, R. J. (2006). Beyond Rayleigh's criterion: a resolution measure with application to single-molecule microscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 103(12):4457–62.

[Resch-Genger and Barker, 2008] Resch-Genger, U. and Barker, P. E. (2008). *Standardization and quality assurance in fluorescence measurements. II Bioanalytical and biomedical applications*. Springer-Verlag.

[Rivenson et al., 2018] Rivenson, Y., Liu, T., Wei, Z., Zhang, Y., and Ozcan, A. (2018). PhaseStain: Digital staining of label-free quantitative phase microscopy images using deep learning. *arXiv:1807.07701*.

[Ronneberger et al., 2015] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. pages 234–241. Springer, Cham.

[Rumelhart et al., 1986] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088):533–536.

[Rust et al., 2006] Rust, M. J., Bates, M., and Zhuang, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods*, 3(10):793–796.

[Sahoo et al., 2018] Sahoo, S. S., Lampert, C. H., and Martius, G. (2018). Learning Equations for Extrapolation and Control. *arXiv:1806.07259*.

[Santoro et al., 2016] Santoro, A., Bartunov, S., Botvinick, M., Wierstra, D., and Lillicrap, T. (2016). Meta-learning with memory-augmented neural networks. In *International conference on machine learning*, pages 1842–1850.

[Santoro et al., 2017] Santoro, A., Raposo, D., Barrett, D. G. T., Malinowski, M., Pascanu, R., Battaglia, P., and Lillicrap, T. (2017). A simple neural network module for relational reasoning. Technical report, *arXiv:arXiv:1706.01427v1*.

[Schneider et al., 2012] Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9(7):671–675.

[Shannon, 1949] Shannon, C. (1949). Communication in the Presence of Noise. *Proceedings of the IRE*, 37(1):10–21.

[Shroff et al., 2008] Shroff, H., Galbraith, C. G., Galbraith, J. A., and Betzig, E. (2008). Live-cell photoactivated localization microscopy of nanoscale adhesion dynamics. *Nature Methods*, 5(5):417–423.

[Silver et al., 2016] Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., and Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489.

[Silver et al., 2017] Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., and Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature*, 550(7676):354–359.

[Simonson et al., 2011] Simonson, P. D., Rothenberg, E., and Selvin, P. R. (2011). Single-Molecule-Based Super-Resolution Images in the Presence of Multiple Fluorophores. *Nano Letters*, 11(11):5090–5096.

[Snell et al., 2017] Snell, J., Swersky, K., and Zemel, R. S. (2017). Prototypical Networks for Few-shot Learning. *arXiv:1703.05175*.

[Srivastava et al., 2014] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15:1929–1958.

[Su et al., 2018] Su, M., Zhang, H., Schawinski, K., Zhang, C., and Cianfrocco, M. A. (2018). Generative adversarial networks as a tool to recover structural information from cryo-electron microscopy data. *bioRxiv:256792*.

[Szegedy et al., 2013] Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., and Fergus, R. (2013). Intriguing properties of neural networks. *arXiv:1312.6199*.

[Takeshima et al., 2018] Takeshima, T., Takahashi, T., Yamashita, J., Okada, Y., and Watanabe, S. (2018). A multi-emitter fitting algorithm for potential live cell super-resolution imaging over a wide range of molecular densities. *Journal of Microscopy*, 271(3):266–281.

[Tinevez et al., 2012] Tinevez, J.-Y., Dragavon, J., Baba-Aissa, L., Roux, P., Perret, E., Canivet, A., Galy, V., and Shorte, S. (2012). A Quantitative Method for Measuring Phototoxicity of a Live Cell Imaging Microscope. In *Methods in enzymology*, volume 506, pages 291–309.

[Trask et al., 2018] Trask, A., Hill, F., Reed, S., Rae, J., Dyer, C., and Blunsom, P. (2018). Neural Arithmetic Logic Units. *arXiv:1808.00508*.

[Tsien, 1998] Tsien, R. Y. (1998). The green fluorescent protein. *Annual Review of Biochemistry*, 67(1):509–544.

[Wang et al., 2018] Wang, H., Rivenson, Y., Jin, Y., Wei, Z., Gao, R., Gunaydin, H., Bentolila, L., and Ozcan, A. (2018). Deep learning achieves super-resolution in fluorescence microscopy. *bioRxiv:309641*.

[Weigert et al., 2017] Weigert, M., Schmidt, U., Boothe, T., Müller, A., Dibrov, A., Jain, A., Wilhelm, B., Schmidt, D., Broaddus, C., Culley, S., Rocha-Martins, M., Segovia-Miranda, F., Norden, C., Henriques, R., Zerial, M., Solimena, M., Rink, J., Tomancak, P., Royer, L., Jug, F., and Myers, E. W. (2017). Content-Aware Image Restoration: Pushing the Limits of Fluorescence Microscopy. *bioRxiv:236463*.

[Westphal and Hell, 2005] Westphal, V. and Hell, S. W. (2005). Nanoscale Resolution in the Focal Plane of an Optical Microscope. *Physical Review Letters*, 94(14):143903.

[Whelan and Bell, 2015] Whelan, D. R. and Bell, T. D. M. (2015). Image artifacts in Single Molecule Localization Microscopy: why optimization of sample preparation protocols matters. *Scientific Reports*, 5(1):7924.

[Williams et al., 2017] Williams, E., Moore, J., Li, S. W., Rustici, G., Tarkowska, A., Chessel, A., Leo, S., Antal, B., Ferguson, R. K., Sarkans, U., Brazma, A., Carazo Salas, R. E., and Swedlow, J. R. (2017). Image Data Resource: a bioimage data integration and publication platform. *Nature Methods*, 14(8):775–781.

[Yuan et al., 2017] Yuan, X., He, P., Zhu, Q., and Li, X. (2017). Adversarial Examples: Attacks and Defenses for Deep Learning. *arXiv:1712.07107*.

[Zaritsky, 2018] Zaritsky, A. (2018). Sharing and reusing cell image data. *Molecular Biology of the Cell*, 29(11):1274–1280.

[Zernike, 1955] Zernike, F. (1955). How I discovered phase contrast. *Science (New York, N.Y.)*, 121(3141):345–9.

[Zhang et al., 2007] Zhang, B., Zerubia, J., and Olivo-Marin, J.-C. (2007). Gaussian approximations of fluorescence microscope point-spread function models. *Applied Optics*, 46(10):1819.

[Zhang et al., 2016a] Zhang, C., Bengio, S., Hardt, M., Recht, B., and Vinyals, O. (2016a). Understanding deep learning requires rethinking generalization. *arXiv:1611.03530*.

[Zhang et al., 2017a] Zhang, H., Cisse, M., Dauphin, Y. N., and Lopez-Paz, D. (2017a). mixup: Beyond Empirical Risk Minimization. *arXiv:1710.09412*.

[Zhang et al., 2016b] Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., and Metaxas, D. (2016b). StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks. *arXiv:1612.03242*.

[Zhang et al., 2017b] Zhang, Z., Nishimura, Y., and Kanchanawong, P. (2017b). Extracting microtubule networks from superresolution single-molecule localization microscopy data. *Molecular biology of the cell*, 28(2):333–345.

[Zhu et al., 2017] Zhu, S., Fidler, S., Urtasun, R., Lin, D., and Loy, C. C. (2017). Be Your Own Prada: Fashion Synthesis with Structural Coherence. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 1689–1697. IEEE.

[Zipfel et al., 2003] Zipfel, W. R., Williams, R. M., and Webb, W. W. (2003). Nonlinear magic: multiphoton microscopy in the biosciences. *Nature Biotechnology*, 21(11):1369–1377.