The imaging tsunami: Computational opportunities and challenges

Wei Ouyang, Christophe Zimmer

The imaging tsunami: computational opportunities and challenges

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

[1] Institut Pasteur, Unité Imagerie et Modélisation, 25 rue du Docteur Roux, 75015 Paris, France
[2] UMR 3691, CNRS; C3BI, USR 3756, IP CNRS

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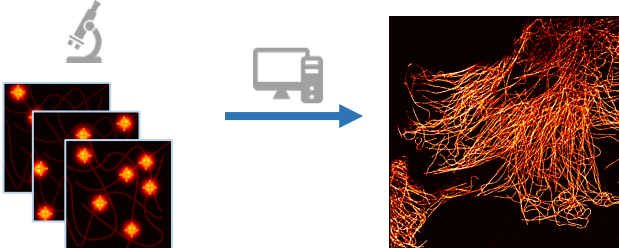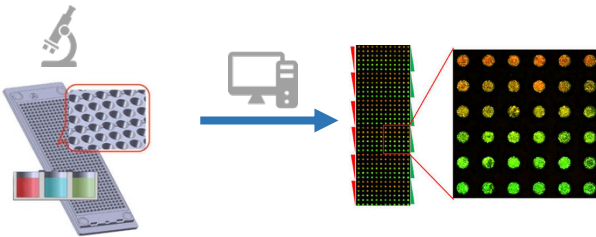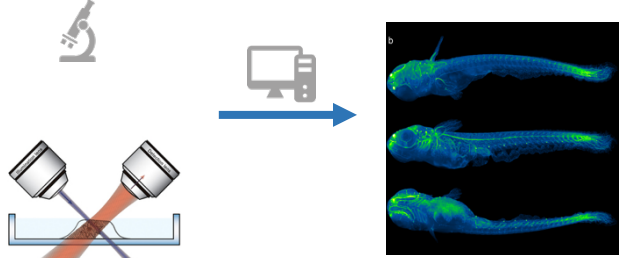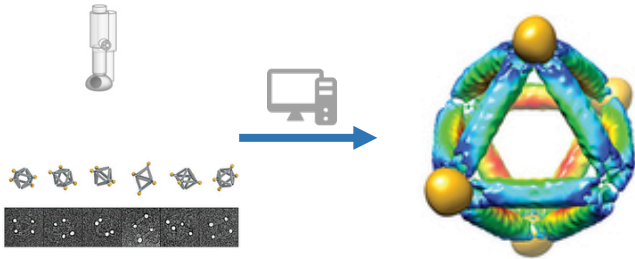
Correspondence: czimmer@pasteur.fr
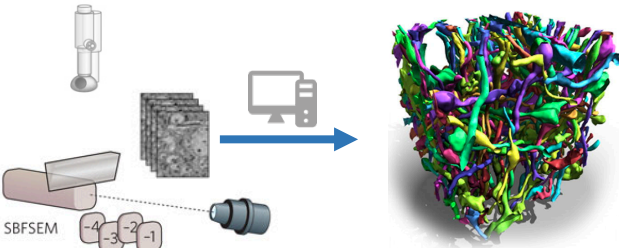
	Imaging technique			Data production rate	Example references
a	Single molecule localization microscopy (SMLM)			~10 GB/h-1TB/h	[1][2][3][4][5][6]
b	High-content image-based screening (HCS)			Up to ~100 GB/h	[7][8][9][10]
c	Light-sheet microscopy			~1-10 TB/h	[11][12][13]
d	Cryo-Electron Microscopy (cryo-EM)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

Figure 1: Five data intensive microscopy techniques.

We highlight five microscopy techniques that generate vast amounts of imaging data: **a**) single molecule localization microscopy (SMLM), **b**) high content screening (HCS), **c**) light sheet microscopy, **d**) single particle electron microscopy at cryogenic temperature (cryo-EM), **e**) serial block face scanning EM (SBFSEM). 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; 1TB = 1,000 GB = 10^{12} bytes (corresponds roughly to 500 h of HD-TV video); 1PB = 1,000 TB = 10^{15} bytes (~13 years of HD-TV video). Images are reprinted with permission from Nature Publishing Group(b,

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 mutating 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 (**Fig. 1**) that exemplify the production of massive imaging data, and discuss some emerging trends and challenges (**Fig. 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 (~1,000 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], [2][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 nuclear pore complexes [23].

In order to obtain a single super-resolution image, SMLM typically requires $\sim 10^4$-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 (**Fig. 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 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 limit, such as maximum likelihood estimators (MLE). Despite this common theme, algorithms still differ in many ways, including the choice of optimization algorithm 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, 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 data 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 make possible the imaging and quantification of tens of proteins and thousands of RNA species in the same cells [34][35][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 (**Fig. 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[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 [41], [42]. Then, 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], [42]. 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 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 can be used 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 spatio-temporal 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]

(**Fig. 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] (**Fig. 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 [15] [14]. 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] (**Fig. 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 (1EB =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, i.e. artificial neural networks are best at mapping biological neural networks [59][17], [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.
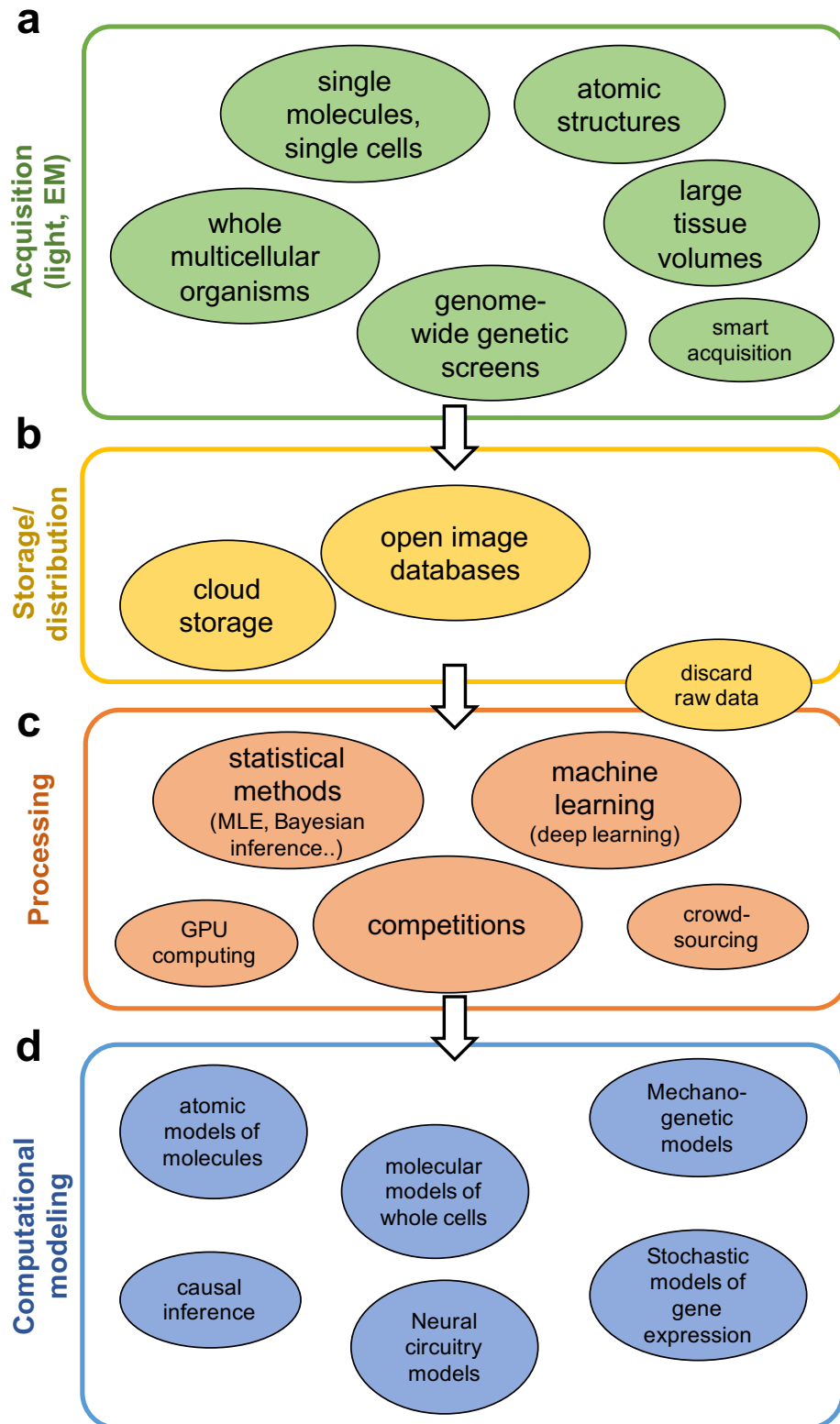
Figure 2: Trends in biological imaging and computational analysis.

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

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. [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 (**Fig. 2a**). The most immediate informatics challenge is how to reliably store such large imaging data (**Fig. 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[64]. This means that the image storage problem can no longer be dissociated from that of image processing (**Fig. 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 (**Fig. 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][66], [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 [68] [43][69]. It therefore seems safe to predict that in the coming years deep learning and related machine learning approaches will become the dominant approach 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] (**Fig. 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.

REFERENCES:

[1]   E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, "Imaging intracellular fluorescent proteins at nanometer resolution," *Science (80-.).*, vol. 313, no. 5793, pp. 1642–1645, 2006.

[2]   M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," *Nat Methods*, vol. 3, no. 10, pp. 793–795, 2006.

[3]   A. Sharonov and R. M. Hochstrasser, "Wide-field subdiffraction imaging by accumulated binding of diffusing probes.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 50, pp. 18911–6, Dec. 2006.

[*4] F. Huang, T. M. P. Hartwich, F. E. Rivera-Molina, Y. Lin, W. C. Duim, J. J. Long, P. D. Uchil, J. R. Myers, M. A. Baird, W. Mothes, M. W. Davidson, D. Toomre, and J. Bewersdorf, "Video-rate nanoscopy using sCMOS camera-specific single-molecule localization algorithms.," *Nat. Methods*, vol. 10, no. 7, pp. 653–8, Jul. 2013.

[*5] F. Huang, G. Sirinakis, E. S. Allgeyer, L. K. Schroeder, W. C. Duim, E. B. Kromann, T. Phan, F. E. Rivera-Molina, J. R. Myers, I. Irnov, M. Lessard, Y. Zhang, M. A. Handel, C. Jacobs-Wagner, C. P. Lusk, J. E. Rothman, D. Toomre, M. J. Booth, and J. Bewersdorf, "Ultra-High Resolution 3D Imaging of Whole Cells," *Cell*, vol. 166, no. 4, pp. 1028–1040, 2016.

[6]   S. Manley, J. M. Gillette, G. H. Patterson, H. Shroff, H. F. Hess, E. Betzig, and J. Lippincott-Schwartz, "High-density mapping of single-molecule trajectories with photoactivated localization microscopy," *Nat Methods*, vol. 5, no. 2, pp. 155–157, 2008.

[*7] B. Neumann, T. Walter, J. K. Heriche, J. Bulkescher, H. Erfle, C. Conrad, P. Rogers, I. Poser, M. Held, U. Liebel, C. Cetin, F. Sieckmann, G. Pau, R. Kabbe, A. Wunsche, V. Satagopam, M. H. Schmitz, C. Chapuis, D. W. Gerlich, R. Schneider, R. Eils, W. Huber, J. M. Peters, A. A. Hyman, R. Durbin, R. Pepperkok, and J. Ellenberg, "Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes," *Nature*, vol. 464, no. 7289, pp. 721–727, 2010.

[*8] N. Battich, T. Stoeger, and L. Pelkmans, "Control of Transcript Variability in Single Mammalian Cells," *Cell*, vol. 163, no. 7, pp. 1596–1610, Dec. 2015.

[**9] V. Graml, X. Studera, J. L. D. Lawson, A. Chessel, M. Geymonat, M. Bortfeld-Miller, T. Walter, L. Wagstaff, E. Piddini, and R. E. Carazo-Salas, "A Genomic Multiprocess Survey of Machineries that Control and Link Cell Shape, Microtubule Organization, and Cell-Cycle Progression," *Dev. Cell*, vol. 31, no. 2, pp. 227–239, Oct. 2014.

** 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] W. J. Godinez, I. Hossain, S. E. Lazic, J. W. Davies, and X. Zhang, "A Multi-Scale Convolutional Neural Network for Phenotyping High-Content Cellular Images," *Bioinformatics*, 2017.

[**11] W. C. Lemon, S. R. Pulver, B. Höckendorf, K. McDole, K. Branson, J. Freeman, and P. J. Keller, "Whole-central nervous system functional imaging in larval Drosophila," *Nat. Commun.*, vol. 6, p. 7924, Aug. 2015.

** 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] W. R. Legant, L. Shao, J. B. Grimm, T. A. Brown, D. E. Milkie, B. B. Avants, L. D. Lavis, and E. Betzig, "High-density three-dimensional localization microscopy across large volumes," *Nat. Methods*, vol. 13, no. 4, pp. 359–365, Mar. 2016.

** 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] L. A. Royer, W. C. Lemon, R. K. Chhetri, Y. Wan, M. Coleman, E. W. Myers, and P. J. Keller, "Adaptive light-sheet microscopy for long-term, high-resolution imaging in living organisms," *Nat. Biotechnol.*, vol. 34, no. 12, pp. 1267–1278, Oct. 2016.

[14] J. H. Lee, G. Ozorowski, and A. B. Ward, "Cryo-EM structure of a native, fully glycosylated, cleaved HIV-1 envelope trimer," *Science (80-.).*, vol. 351, no. 6277, pp. 1043–1048, Mar. 2016.

[**15] A. Merk, A. Bartesaghi, S. Banerjee, V. Falconieri, P. Rao, M. I. Davis, R. Pragani, M. B. Boxer, L. A. Earl, J. L. S. Milne, and S. Subramaniam, "Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery," *Cell*, vol. 165, no. 7, pp. 1698–1707, Jun. 2016.

** 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] A. Punjani, J. L. Rubinstein, D. J. Fleet, and M. A. Brubaker, "cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination," *Nat. Methods*, Feb. 2017.

[17] S. Dorkenwald, P. J. Schubert, M. F. Killinger, G. Urban, S. Mikula, F. Svara, and J. Kornfeld, "Automated synaptic connectivity inference for volume electron microscopy," *Nat. Methods*, Feb. 2017.

[18] N. Kasthuri, K. J. Hayworth, D. R. Berger, R. L. Schalek, J. A. Conchello, S. Knowles-Barley, D. Lee, A. Vázquez-Reina, V. Kaynig, T. R. Jones, M. Roberts, J. L. Morgan, J. C. Tapia, H. S. Seung, W. G. Roncal, J. T. Vogelstein, R. Burns, D. L. Sussman, C. E. Priebe, H. Pfister, and J. W. Lichtman, "Saturated Reconstruction of a Volume of Neocortex," *Cell*, vol. 162, no. 3, pp. 648–661, 2015.

** The authors developed an automated pipeline wherein a 80,000 µm3 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 1,700 synapses and find that connections between neurons cannot be explained by their proximity alone.

[19] A. A. Wanner, C. Genoud, T. Masudi, L. Siksou, and R. W. Friedrich, "Dense EM-based reconstruction of the interglomerular projectome in the zebrafish olfactory bulb," *Nat. Neurosci.*, vol. 19, no. 6, pp. 816–825, Apr. 2016.

[20] Z. D. Stephens, S. Y. Lee, F. Faghri, R. H. Campbell, C. Zhai, M. J. Efron, R. Iyer, M. C. Schatz, S. Sinha, and G. E. Robinson, "Big Data: Astronomical or Genomical?," *PLoS Biol.*, vol. 13, no. 7, p. e1002195, Jul. 2015.

[21] P. Kanchanawong, G. Shtengel, A. M. Pasapera, E. B. Ramko, M. W. Davidson, H. F. Hess, and C. M. Waterman, "Nanoscale architecture of integrin-based cell adhesions," *Nature*, vol. 468, no. 7323, pp. 580–584, Nov. 2010.

[22] K. Xu, G. Zhong, and X. Zhuang, "Actin, Spectrin, and Associated Proteins Form a Periodic Cytoskeletal Structure in Axons," *Science (80-.).,* vol. 339, no. 6118, 2013.

[*23] A. Szymborska, A. de Marco, N. Daigle, V. C. Cordes, J. A. G. Briggs, and J. Ellenberg, "Nuclear pore scaffold structure analyzed by super-resolution microscopy and particle averaging.," *Science*, vol. 341, no. 6146, pp. 655–8, Aug. 2013.

[24] R. J. Ober, S. Ram, and E. S. Ward, "Localization Accuracy in Single-Molecule Microscopy," *Biophys. J.*, vol. 86, no. 2, pp. 1185–1200, 2004.

[**25] D. Sage, H. Kirshner, T. Pengo, N. Stuurman, J. Min, S. Manley, and M. Unser, "Quantitative evaluation of software packages for single-molecule localization microscopy," *Nat. Methods*, vol. 12, no. 8, pp. 717–724, Jun. 2015.

** 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] A. von Diezmann, Y. Shechtman, and W. E. Moerner, "Three-Dimensional Localization of Single Molecules for Super-Resolution Imaging and Single-Particle Tracking," *Chem. Rev.*, p. acs.chemrev.6b00629, Feb. 2017.

[27] S. Jia, J. C. Vaughan, and X. Zhuang, "Isotropic 3D Super-resolution Imaging with a Self-bending Point Spread Function.," *Nat. Photonics*, vol. 8, no. 4, pp. 302–306, Jan. 2014.

[28] P. Sengupta, T. Jovanovic-Talisman, D. Skoko, M. Renz, S. L. Veatch, and J. Lippincott-Schwartz, "Probing protein heterogeneity in the plasma membrane using PALM and pair correlation analysis," *Nat Methods*, vol. 8, no. 11, pp. 969–975, 2011.

[*29] S.-H. Lee, J. Y. Shin, A. Lee, and C. Bustamante, "Counting single photoactivatable fluorescent molecules by photoactivated localization microscopy (PALM).," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 43, pp. 17436–41, Oct. 2012.

[30] M. P. Backlund, M. D. Lew, A. S. Backer, S. J. Sahl, and W. E. Moerner, "The Role of Molecular Dipole Orientation in Single-Molecule Fluorescence

Microscopy and Implications for Super-Resolution Imaging," *ChemPhysChem*, vol. 15, no. 4, pp. 587–599, Mar. 2014.

[31]   N. Chenouard, I. Smal, F. de Chaumont, M. Maška, I. F. Sbalzarini, Y. Gong, J. Cardinale, C. Carthel, S. Coraluppi, M. Winter, A. R. Cohen, W. J. Godinez, K. Rohr, Y. Kalaidzidis, L. Liang, J. Duncan, H. Shen, Y. Xu, K. E. G. Magnusson, J. Jaldén, H. M. Blau, P. Paul-Gilloteaux, P. Roudot, C. Kervrann, F. Waharte, J.-Y. Tinevez, S. L. Shorte, J. Willemse, K. Celler, G. P. van Wezel, H.-W. Dan, Y.-S. Tsai, C. O. de Solórzano, J.-C. Olivo-Marin, and E. Meijering, "Objective comparison of particle tracking methods," *Nat. Methods*, vol. 11, no. 3, pp. 281–289, Jan. 2014.

[32]   N. Monnier, S.-M. Guo, M. Mori, J. He, P. Lénárt, and M. Bathe, "Bayesian Approach to MSD-Based Analysis of Particle Motion in Live Cells," *Biophys. J.*, vol. 103, no. 3, pp. 616–626, 2012.

[33]   M. El Beheiry, M. Dahan, and J.-B. Masson, "InferenceMAP: mapping of single-molecule dynamics with Bayesian inference," *Nat. Methods*, vol. 12, no. 7, pp. 594–595, 2015.

[34]   Z. Zhang, S. J. Kenny, M. Hauser, W. Li, and K. Xu, "Ultrahigh-throughput single-molecule spectroscopy and spectrally resolved super-resolution microscopy," *Nat. Methods*, vol. 12, no. 10, pp. 935–938, Aug. 2015.

[35]   R. Jungmann, M. S. Avendaño, J. B. Woehrstein, M. Dai, W. M. Shih, and P. Yin, "Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT," *Nat. Methods*, vol. 11, no. 3, pp. 313–318, Feb. 2014.

[**36] K. H. Chen, A. N. Boettiger, J. R. Moffitt, S. Wang, and X. Zhuang, "Spatially resolved, highly multiplexed RNA profiling in single cells.," *Science*, vol. 348, no. 6233, p. aaa6090, Apr. 2015.

**This study uses a combinatorial labeling strategy to visualize up to 1,000 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]   B. Munsky, G. Neuert, and A. van Oudenaarden, "Using Gene Expression Noise to Understand Gene Regulation," *Science (80-.).*, vol. 336, no. 6078, pp. 183–187, Apr. 2012.

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

[39]   M. Mattiazzi Usaj, E. B. Styles, A. J. Verster, H. Friesen, C. Boone, and B. J. Andrews, "High-Content Screening for Quantitative Cell Biology," *Trends Cell Biol.*, vol. 26, no. 8, pp. 598–611, 2016.

[40]   N. Battich, T. Stoeger, and L. Pelkmans, "Control of Transcript Variability in Single Mammalian Cells," *Cell*, vol. 163, no. 7, pp. 1596–1610, Dec. 2015.

[41]   T. R. Jones, A. E. Carpenter, M. R. Lamprecht, J. Moffat, S. J. Silver, J. K. Grenier, A. B. Castoreno, U. S. Eggert, D. E. Root, P. Golland, and D. M. Sabatini, "Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning.," *Proc. Natl. Acad. Sci. U. S. A.*, vol.

106, no. 6, pp. 1826–31, Feb. 2009.

[42] N. Battich, T. Stoeger, and L. Pelkmans, "Image-based transcriptomics in thousands of single human cells at single-molecule resolution," *Nat. Methods*, vol. 10, no. 11, pp. 1127–1133, Oct. 2013.

[43] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[44] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion," *J. Mach. Learn. Res.*, vol. 11, no. Dec, pp. 3371–3408, 2010.

[45] C. Kandaswamy, L. M. Silva, L. A. Alexandre, and J. M. Santos, "High-Content Analysis of Breast Cancer Using Single-Cell Deep Transfer Learning," *J. Biomol. Screen.*, vol. 21, no. 3, pp. 252–259, 2016.

[**46] K. Lee, H. L. Elliott, Y. Oak, C.-T. Zee, A. Groisman, J. D. Tytell, and G. Danuser, "Functional Hierarchy of Redundant Actin Assembly Factors Revealed by Fine-Grained Registration of Intrinsic Image Fluctuations," *Cell Syst.*, vol. 1, no. 1, pp. 37–50, Jul. 2015.

[47] E. S. Welf and G. Danuser, "Using Fluctuation Analysis to Establish Causal Relations between Cellular Events without Experimental Perturbation," *Biophys. J.*, vol. 107, no. 11, pp. 2492–2498, 2014.

[48] J. Huisken, J. Swoger, F. Del Bene, J. Wittbrodt, and E. H. K. Stelzer, "Optical sectioning deep inside live embryos by selective plane illumination microscopy," *Science (80-.).*, vol. 305, no. 5686, p. 1007, 2004.

[49] B.-C. Chen, W. R. Legant, K. Wang, L. Shao, D. E. Milkie, M. W. Davidson, C. Janetopoulos, X. S. Wu, J. A. Hammer, Z. Liu, B. P. English, Y. Mimori-Kiyosue, D. P. Romero, A. T. Ritter, J. Lippincott-Schwartz, L. Fritz-Laylin, R. D. Mullins, D. M. Mitchell, J. N. Bembenek, A.-C. Reymann, R. Bohme, S. W. Grill, J. T. Wang, G. Seydoux, U. S. Tulu, D. P. Kiehart, and E. Betzig, "Lattice light-sheet microscopy: Imaging molecules to embryos at high spatiotemporal resolution," *Science (80-.).*, vol. 346, no. 6208, pp. 1257998–1257998, Oct. 2014.

[50] P. J. Keller, A. D. Schmidt, J. Wittbrodt, and E. H. K. Stelzer, "Reconstruction of Zebrafish Early Embryonic Development by Scanned Light Sheet Microscopy," *Science (80-.).*, vol. 322, no. 5904, 2008.

[51] E. G. Reynaud, J. Peychl, J. Huisken, and P. Tomancak, "Guide to light-sheet microscopy for adventurous biologists," *Nat. Methods*, vol. 12, no. 1, pp. 30–34, Dec. 2014.

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

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

** 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] M. Rauzi, U. Krzic, T. E. Saunders, M. Krajnc, P. Ziherl, L. Hufnagel, and M. Leptin, "Embryo-scale tissue mechanics during Drosophila gastrulation movements," *Nat. Commun.*, vol. 6, p. 8677, Oct. 2015.

[55] E. H. Egelman, "The Current Revolution in Cryo-EM," *Biophys. J.*, vol. 110, no. 5, pp. 1008–1012, Mar. 2016.

[56] R. Henderson, "Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 45, pp. 18037–41, Nov. 2013.

[57] J. W. Lichtman, H. Pfister, and N. Shavit, "The big data challenges of connectomics," *Nat. Neurosci.*, vol. 17, no. 11, pp. 1448–1454, Oct. 2014.

[**58] I. Arganda-Carreras, S. C. Turaga, D. R. Berger, D. Cireşan, A. Giusti, L. M. Gambardella, J. Schmidhuber, D. Laptev, S. Dwivedi, J. M. Buhmann, T. Liu, M. Seyedhosseini, T. Tasdizen, L. Kamentsky, R. Burget, V. Uher, X. Tan, C. Sun, T. D. Pham, E. Bas, M. G. Uzunbas, A. Cardona, J. Schindelin, and H. S. Seung, "Crowdsourcing the creation of image segmentation algorithms for connectomics," *Front. Neuroanat.*, vol. 9, p. 142, Nov. 2015.

[59] I. Arganda-Carreras, S. C. Turaga, D. R. Berger, D. Cireşan, A. Giusti, L. M. Gambardella, J. Schmidhuber, D. Laptev, S. Dwivedi, J. M. Buhmann, T. Liu, M. Seyedhosseini, T. Tasdizen, L. Kamentsky, R. Burget, V. Uher, X. Tan, C. Sun, T. D. Pham, E. Bas, M. G. Uzunbas, A. Cardona, J. Schindelin, and H. S. Seung, "Crowdsourcing the creation of image segmentation algorithms for connectomics," *Front. Neuroanat.*, vol. 9, p. 142, Nov. 2015.

[60] T. Beier, C. Pape, N. Rahaman, T. Prange, S. Berg, D. D. Bock, A. Cardona, G. W. Knott, S. M. Plaza, L. K. Scheffer, U. Koethe, A. Kreshuk, and F. A. Hamprecht, "Multicut brings automated neurite segmentation closer to human performance," *Nat. Methods*, vol. 14, no. 2, pp. 101–102, Jan. 2017.

[61] S. Takemura, A. Bharioke, Z. Lu, A. Nern, S. Vitaladevuni, P. K. Rivlin, W. T. Katz, D. J. Olbris, S. M. Plaza, P. Winston, T. Zhao, J. A. Horne, R. D. Fetter, S. Takemura, K. Blazek, L.-A. Chang, O. Ogundeyi, M. A. Saunders, V. Shapiro, C. Sigmund, G. M. Rubin, L. K. Scheffer, I. A. Meinertzhagen, and D. B. Chklovskii, "A visual motion detection circuit suggested by Drosophila connectomics," *Nature*, vol. 500, no. 7461, pp. 175–181, Aug. 2013.

[62] C. Franke, M. Sauer, and S. van de Linde, "Photometry unlocks 3D information from 2D localization microscopy data," *Nat. Methods*, vol. 14, no. 1, pp. 41–44, Nov. 2016.

[63] S. H. w Scheres, "Beam-induced motion correction for sub-megadalton cryo-EM particles," *Elife*, vol. 3, p. e03665, 2014.

[64] E. G. Reynaud, J. Peychl, J. Huisken, and P. Tomancak, "Guide to light-sheet microscopy for adventurous biologists," *Nat. Methods*, vol. 12, no. 1, pp. 30–34, Dec. 2014.

[65] S. Cox, E. Rosten, J. Monypenny, T. Jovanovic-Talisman, D. T. Burnette, J. Lippincott-Schwartz, G. E. Jones, and R. Heintzmann, "Bayesian localization microscopy reveals nanoscale podosome dynamics," *Nat Methods*, vol. 9, no. 2, pp. 195–200, 2012.

[66] P. Joubert and M. Habeck, "Bayesian inference of initial models in cryo-

electron microscopy using pseudo-atoms.," *Biophys. J.*, vol. 108, no. 5, pp. 1165–75, Mar. 2015.

[67]	S. H. W. Scheres, "A Bayesian View on Cryo-EM Structure Determination," *J. Mol. Biol.*, vol. 415, no. 2, pp. 406–418, Jan. 2012.

[68]	V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, R. Kim, R. Raman, P. C. Nelson, J. L. Mega, and D. R. Webster, "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs," *JAMA*, vol. 316, no. 22, p. 2402, Dec. 2016.

[*69]	O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science,* 2015, vol. 9351, pp. 234–241.

[70]	E. Williams, J. Moore, S. W. Li, G. Rustici, A. Tarkowska, A. Chessel, S. Leo, B. Antal, R. K. Ferguson, U. Sarkans, A. Brazma, R. E. Carazo-Salas, and J. Swedlow, "The Image Data Resource: A Scalable Platform for Biological Image Data Access, Integration, and Dissemination," *bioRxiv*, 2016.

[71]	A. Iudin, P. K. Korir, J. Salavert-Torres, G. J. Kleywegt, and A. Patwardhan, "EMPIAR: a public archive for raw electron microscopy image data," *Nat. Methods*, vol. 13, no. 5, pp. 387–388, Mar. 2016.

[72]	M. A. Cianfrocco and A. E. Leschziner, "Low cost, high performance processing of single particle cryo-electron microscopy data in the cloud," *Elife*, vol. 4, no. MAY, pp. 1–10, 2015.

[73]	Y. S. Hu, X. Nan, P. Sengupta, J. Lippincott-Schwartz, and H. Cang, "Accelerating 3B single-molecule super-resolution microscopy with cloud computing," *Nat. Methods*, vol. 10, no. 2, pp. 96–97, 2013.

[*74]	J. C. Sanghvi, S. Regot, S. Carrasco, J. R. Karr, M. V Gutschow, B. Bolival, and M. W. Covert, "Accelerated discovery via a whole-cell model," *Nat. Methods*, vol. 10, no. 12, pp. 1192–1195, Nov. 2013.