PERSPECTIVE

nature biotechnology

Imagining the future of bioimage analysis

Erik Meijering[1], Anne E Carpenter[2], Hanchuan Peng[3], Fred A Hamprecht[4] & Jean-Christophe Olivo-Marin[5]

Modern biological research increasingly relies on image data as a primary source of information in unraveling the cellular and molecular mechanisms of life. The quantity and complexity of the data generated by state-of-the-art microscopes preclude visual or manual analysis and require advanced computational methods to fully explore the wealth of information. In addition to making bioimage analysis more efficient, objective, and reproducible, the use of computers improves the accuracy and sensitivity of the analyses and helps to reveal subtleties that may be unnoticeable to the human eye. Many methods and software tools have already been developed to this end, but there is still a long way to go before biologists can blindly trust automated measurements. Here, we summarize the current state of the art in bioimage analysis and provide a perspective on likely future developments.

As in many other facets of life in the twenty-first century, images have come to play a major role in biology. Whether the aim is to get a first glimpse of a suspected cellular mechanism in action or to rigorously test a biological hypothesis and convince others of its validity, images are increasingly the medium of choice. Our inclination to rely on images as a key source of information should not come as a surprise, with more than half of the human brain involved in vision[1], explaining the age-old adage 'seeing is believing'. At the heart of our ability to visualize and investigate life at the cellular and molecular level with ever-increasing detail and sensitivity are technological advances in various fields. The past few decades have celebrated Nobel-prize winning inventions that have turned microscopy into nanoscopy and spurred the development of a broad arsenal of imaging modalities[2] that are now becoming widely available. Their adoption and application to important biological problems, however, would not have been possible without equally groundbreaking advances in computing technology. Not only are computers indispensable for advanced signal reconstruction during image acquisition, they are now also largely responsible for handling much of the 'big data' resulting from microscopy.

Since the first uses of computers in biological imaging about 50 years ago[3], when a single experiment typically involved just a few images of some 200 × 200 pixels taking only 40 kilobytes of memory, data sets have grown exponentially in size. Today's automated microscopes may capture information about three spatial dimensions, multiple time points, multiple viewpoints, multiple wavelengths, multiple biological parameters and more, resulting in terabytes of data. In response to the growing need for powerful automated solutions in processing and analyzing these data, a new field of biotechnology has emerged known as bioimage informatics[4]. In this rapidly evolving field, computer scientists, image analysis experts, bioinformaticians, and biologists are joining forces in creating innovative computational methods and user-friendly software tools to facilitate image-based biological experimentation. The ultimate goal is to replace human labor as much as possible with computer calculations, so that biologists can focus fully on formulating high-level hypotheses and designing increasingly sophisticated experiments, while improving the objectivity and reproducibility of these experiments. This will accelerate discoveries and breakthroughs, from basic biology all the way to clinical research.

The biggest challenge in this endeavor is to make computers 'see' and measure; that is, to enable them to automatically distinguish between relevant and irrelevant image information and to recognize and model spatial or temporal patterns of interest that could confirm or refute the hypotheses underlying the given experiment through quantitative analysis. This task of bioimage informatics is what we refer to as 'bioimage analysis'. To date, methods addressing parts of the challenge have been developed in numerous studies, and flexible software platforms[5] are available for building bioimage analysis pipelines. In practice, however, it turns out to be very hard to design pipelines that work out of the box or across applications, and substantial human effort is often spent on tuning parameters and correcting faulty output. This suggests that we are still very far from the ultimate goal. In this Perspective, we reflect on the current status of the field and ponder the future of bioimage analysis.

Virtues and vices of current methods

From our own human experience in interpreting visual information, we are easily tempted to think that it should not be too difficult to automate this task. Indeed, in the early days of artificial intelligence research (1960s), it was apparently considered not too much to ask an undergraduate student to "spend the summer linking a camera to a computer and getting the computer to describe what it saw"[6]. Since then, the enormous complexity of the problem has become apparent, and many subproblems of bioimage analysis have grown into vivid research areas (**Fig. 1**).

The first step in making sense of the image data is usually to preprocess the data to reduce noise and other imaging artifacts. A prominent example of preprocessing is deconvolution[7], which attempts to undo the signal blurring inevitably introduced by any microscope.

[1]Biomedical Imaging Group Rotterdam, Departments of Medical Informatics and Radiology, Erasmus University Medical Center, Rotterdam, the Netherlands. [2]Imaging Platform, Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts, USA. [3]Allen Institute for Brain Science, Seattle, Washington, USA. [4]Heidelberg Collaboratory for Image Processing, Heidelberg University, Heidelberg, Germany. [5]Institut Pasteur, Unité d'Analyse d'Images Biologiques, Centre National de la Recherche Scientifique Unité de Recherche Mixte, Paris, France. Correspondence should be addressed to E.M. (meijering@imagescience.org).

Received 31 October 2015; accepted 10 November 2016; published online 7 December 2016; doi:10.1038/nbt.3722

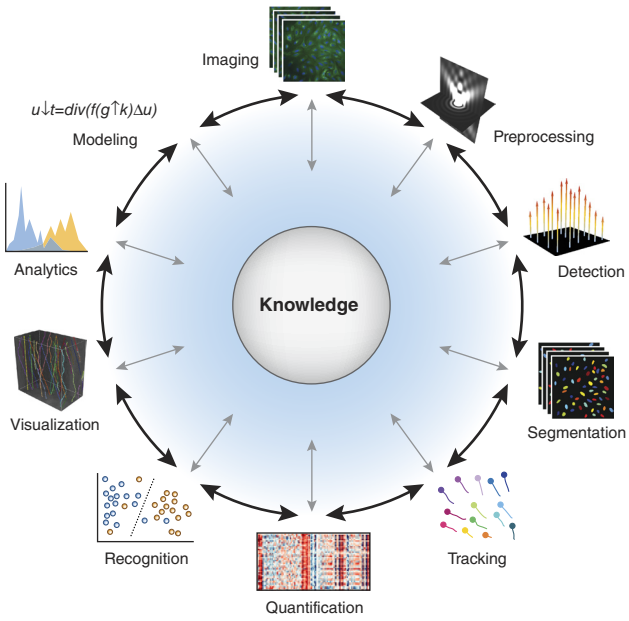

Figure 1 Common steps in bioimage analysis. The ultimate goal is to increase our knowledge of the biological process under investigation. Based on existing knowledge, new hypotheses are posed and new imaging experiments performed. Many steps are often needed to extract new knowledge from the images: preprocessing (quality control, illumination correction, denoising, deconvolution, registration); detection (determining the presence of objects based on image-filtering techniques); segmentation (grouping of pixels relating to the same object or class); tracking (linking detected objects from frame to frame in a movie); quantification (of shape, dynamics, texture, and other object properties); recognition (clustering or classifying objects and patterns); visualization (rendering high-dimensional images or results and allowing inspection, correction, or annotation); analytics (statistical processing of the extracted information); and modeling (constructing high-level descriptions of the results). Depending on the application not all steps may always be needed (shortcuts are possible). Two-headed arrows are used everywhere to indicate the interrelation and possible feedback between the steps. Any of the steps may also contribute to knowledge along the way, affecting other steps.

Another frequently needed preprocessing step is image registration[8], where images from, for example, different organisms or different time points of the same organism are aligned to allow a pixel-by-pixel comparison, fusion, or joint analysis of the images. The next step commonly taken is to detect the presence of various types of objects by extracting a range of image features using image-filtering techniques[9]. Arguably, the most challenging step in bioimage analysis is image segmentation[10], which aims to classify and group pixels as objects of interest or background. In the case of time-lapse image data, a related problem is to determine the trajectories of moving objects by frame-to-frame association of image information[11]. Once objects have been segmented and tracked, it is relatively easy to compute a host of quantitative descriptors of, for example, dynamics[11], shape[12], and texture[13], which subsequently allow the objects to be characterized and distinguished using pattern recognition methods[14]. Many experiments further require inspecting, correcting, or annotating results via interactive visualization[15]. The ultimate goal is building realistic computer models to aid understanding of the cellular or molecular process under investigation[16], which requires rigorous analytical approaches for information mining and statistical testing.

Well-designed bioimage analysis workflows based on these steps enable biologists to turn their data into biological discoveries and insights that otherwise would have proven elusive, as has been successfully shown in various areas. An obvious example is high-throughput screening (**Fig. 2a**), where the sheer volume of the data precludes manual analysis. State-of-the-art methods for cell segmentation and subsequent classification based on hundreds of morphological features extracted from nearly 200,000 time-lapse microscopy movies have been crucial in determining which of the ~22,000 human genes play a role in mitosis, via RNA interference experiments[17]. Other examples can be found in the area of neuroscience (**Fig. 2b**,**c**). Sophisticated tools for image registration and pattern recognition, though still requiring expert input, have been instrumental in reconstructing the connectome of, for example, the *Drosophila* optic medulla from electron microscopy image stacks, leading to the identification of the cell types constituting a motion detection circuit[18]. Bioimage analysis has also transformed the study of intracellular dynamic processes (**Fig. 2d**). Computational methods for particle detection, tracking, and statistical processing of their intensity fluctuations in fluorescence speckle microscopy movies revealed that the process of protrusion in cell migration is mediated by two spatially co-localized yet distinct actin networks[19]. More recently, related methods have identified the regulation of focal adhesion dynamics in cell migration by specific signaling pathways[20].

Despite the successes, developing powerful bioimage analysis solutions is far from trivial for a given experiment. It requires researchers to identify the necessary processing steps to go from pixel information to biologically meaningful measurements and models, to select suitable computational algorithms for each step and tune their parameters, and to link them in a software workflow as automated as possible. The fact that one can easily find hundreds of research papers proposing image segmentation methods for even just one part of an analysis, all of them using different algorithms or parameter settings[10], suggests that, apparently, none of them really solves the segmentation problem at large, but only in a very limited context. We observe that, at present, nearly every biological experiment requires substantial effort in fine-tuning the analysis pipeline. This does not bode well, as experiments continue to grow in scale and throughput. The rapidly increasing rate of experimentation and image acquisition demands that bioimage analysis methods become more robust and generic so they can be reused in a wide range of applications with minimal effort. As biology zooms in on more complex cellular systems, the analysis task will only become harder, underscoring the need for flexible solutions. To this end, several challenges will need to be addressed, both cultural and technical, as discussed below.

Cultural issues in bioimage analysis

Broadly speaking, bioimage analysis is where computer vision meets biology[21]. As a branch of artificial intelligence, computer vision is concerned with the development of theories and algorithms to make machines interpret visual data autonomously and quantitatively. Its numerous applications include video surveillance, biometric verification, medical diagnostics, and many others[6]. The physical principles used to capture images and the physical properties of the objects of interest in these images vary widely from application to application. Thus, expert domain knowledge is critical to make computer vision tools successful. In biology, developing a tool that can, say, distinguish between normal and abnormal cell migration in a cancer screening assay, requires profound knowledge of the labeling technique, how this affects cell appearance in the images given the type of microscopy used, and what, biologically, defines a normal or abnormal phenotype.

Clearly, to optimally design and execute experiments involving bioimage analysis, expertise in both computer vision and biology

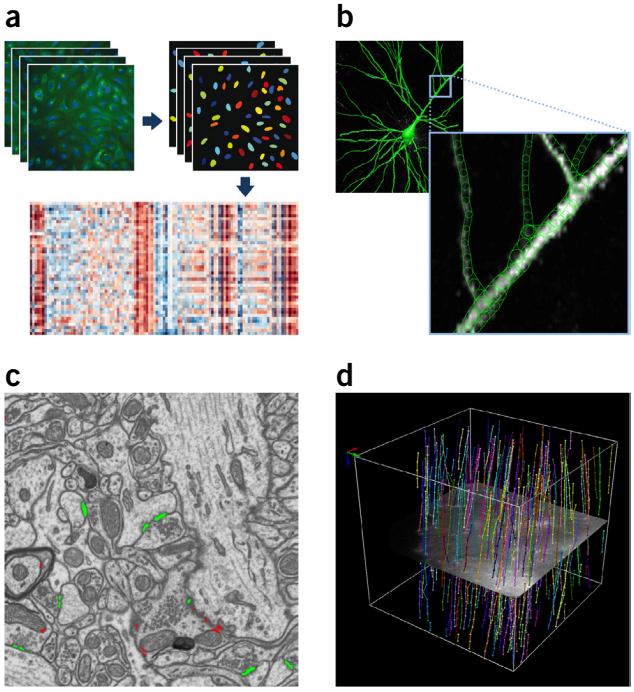

Figure 2 Examples of bioimage analysis in various applications. (**a**) In high-throughput image-based experiments, thousands to millions of images are acquired by automated microscopes, requiring fully automated segmentation algorithms that are robust to unusual cell phenotypes. Here, two-dimensional (2D) images (top left) showing DNA (blue) and a green fluorescent protein (GFP)-labeled protein (green) undergo model-based segmentation of nuclei (top right). Then, thousands of features are computed for each cell, resulting in a matrix (bottom) of cells versus feature values (shown color coded) used for morphological profiling[48]. (**b**) In neuroscience, the morphology of neuronal cells is often studied in relation to function. The example shows the result of automated three-dimensional (3D) reconstruction of a single neuron from fluorescence microscopy image stacks based on intensity thresholding followed by graph searching and pruning[49]. (**c**) Volume electron microscopy allows analysis of neural tissue at the nanometer scale. Here, automated synapse segmentation in 3D image stacks is achieved by interactive machine learning for semantic pixel classification, followed by another machine learning step at the object level to discriminate postsynaptic densities (green) from other structures (red)[50]. (**d**) In studies of intracellular dynamic processes, the trajectories of thousands of particles need to be extracted from time-lapse image data. The example shows a 2D+time spatiotemporal visualization (time runs along the vertical axis) of automatically tracked microtubule tips moving in the cytoplasm. The rendering also shows one time frame of the movie.

is needed. For researchers from the two fields to work productively together, they must not only learn each other's specialized language, but also understand differently emphasized scientific principles across these fields. Computer scientists tend to be dismayed at the irreproducibility of biological systems and the seemingly fuzzy questions asked. For their part, biologists find computer scientists tend toward overly reducing complex systems to make problems tractable. Increasing the exposure of students to the practices and philosophies of both disciplines will yield a generation of researchers who can meet closer to the middle.

Each discipline's rewards culture will also need to shift to favor collaboration. Computer vision and biology are both highly competitive fields of their own, with different primary goals and constraints[22]. Computer scientists often shy away from problems likely solvable with existing algorithms, as papers lacking mathematical novelty are less publishable in their favored venues. But novel algorithms can take so much time to develop that biologists shift to manual analysis, or more feasible problems, to secure publishable output of their own. Many biology laboratories hire their own computational technician, who typically works in isolation on the laboratory's very specific data analysis problems, often coming up with solutions that are neither optimal, reusable by others, nor publishable. Moreover, although well-engineered software is crucial to ensure long-term stability, neither computer vision researchers nor biologists generally consider software engineering their main responsibility, and do not usually look ahead to software support.

As research in biology becomes more and more multidisciplinary, it is crucial that funding bodies come to recognize and foster each and every aspect, including software engineering[23]. At present, it is not uncommon for a grant application to get rejected even in interdisciplinary programs because the research (or the principal investigator's expertise) is deemed too applied by computer scientists and not sufficiently biological by life scientists. Academia and research institutes need to become highly attractive places for professional software engineers. These experts should not be judged by the traditional measuring sticks of success in science (notably publications) but by criteria reflecting software impact and (re)usability[24].

Computer vision versus human vision

A cornerstone of bioimage analysis, computer vision research itself has traditionally been inspired by human vision. The prevailing view among vision scientists is that "there is no single equation that explains everything[25]" but visual information is processed in various stages involving increasingly sophisticated levels of representation. Upon entering the eyes, light is first probed by the retinal receptive fields operating at a wide range of scales, resulting in a 'primal sketch' of the image information, such as the presence, location, orientation, and velocity of blobs, edges, ridges, and other features. These features have been studied extensively in computer vision[26] and methods for their detection are used in virtually every bioimage analysis pipeline. But as we move from the retina into the different areas of the visual cortex and deeper into the brain, our understanding decreases of how low-level information extraction drives high-level cognition and vice versa (feedback). Therefore, it is not surprising that many computational concepts have been proposed over the years, based on different assumptions[27].

What is clear is that, for humans, seeing is a matter of learning. Our capacity to make sense of visual data so effortlessly, even when the data are very complex or incomplete, can be attributed to the brain's continuous association of new input with the memory of previous observations and interpretations. This memory is enormous[28] and accommodates highly structured (generalized) representations beyond the level of individual items[29]. A long-standing challenge in artificial intelligence is to emulate this capacity with computers, using machine learning methods[30]. The rules of operation of such methods may be learned from explicit input–output relationships provided by the user ('supervised learning'), or derived implicitly from the data itself ('unsupervised learning'), or based on a combination of both ('semi-supervised learning'). In this process, a fundamental trade-off must be made between the amount of knowledge that can be built into a system and the amount of available data to learn from. Given a well-defined problem and sufficient data, machine-learning methods can be successfully applied in biology[31], and software tools to build bioimage analysis pipelines based on them are freely available[5,14].

One important question is to what extent bioimage analysis systems should strive to mimic the human vision system. This is especially

relevant because advanced deep-learning methods are now able to perform comparably to humans in certain tasks[32]. Some strengths of human vision also imply serious weaknesses that are best avoided. For example, humans tend to focus on phenomena that support preconceived ideas, and depending on prior experiences, two observers may focus differently. Also, the association of new input with previous observations bears the risk of missing subtle differences between similarly looking yet distinct image events[21]. This lowers the objectivity, reproducibility, and sensitivity of analyses, whereas the aim of using computers should be to maximize them. Future computational bioimage analysis will greatly benefit from more research into the precise limitations of human vision and how to overcome them. This also touches on the issue of whether or not an expert biologist's observations should serve as the 'gold standard' in any part of the analysis, and in what ways humans and computers can best complement each other (man–machine interaction).

Acceleration by community efforts

Choosing an appropriate computational method for each step in a bioimage analysis pipeline requires in-depth knowledge not only of which methods are available, but especially which of these is best under which conditions. Discerning this from the literature is usually impossible. Biology-oriented papers generally aim to answer a biological question by any means possible, and thus rarely include comparative analyses of alternative methods or usable code. Conversely, papers from computer vision groups typically aim to demonstrate the superiority of a new method and provide much mathematical detail, but the claims are often based on very limited biological data, or very narrow and possibly biologically irrelevant performance criteria. In the rare case of published comparative analyses, the methods are usually (re)implemented or tuned by the authors, not by the original inventors.

The need for establishing well-defined subgoals and performing fair and consistent evaluations and comparisons of methods to accelerate progress in computer vision has been apparent for several decades[33]. Since the early 2000s, this has led to the organization of so-called challenges, in which researchers working on a particular problem are stimulated to put their methods to the test, based on standardized data and performance metrics. In biomedical imaging at large, over 100 such challenges have been held to date (http://grand-challenge.org/). Examples in bioimage analysis include challenges on neuron reconstruction[34], particle tracking[35], cell tracking[36], single-molecule localization[37], and mitosis detection[38,39] (**Table 1**). Such community efforts reduce potential sources of bias and greatly help the field by revealing the strengths and weaknesses of current methods and establishing a 'litmus test' for future developers aiming to improve on the state of the art.

A limiting factor in the current *modus operandi* of challenges is that the competing teams continue working in isolation on their own methods and in their own software environments. Not only does this often give rise to technical issues (e.g., in importing data and exporting results), it also bears the risk of duplicate work (the same basic algorithms are implemented by multiple teams), and may lower reproducibility and hamper utilization of the results. We suggest that the impact of challenges can be further increased by encouraging competing teams to port their methods to a common software platform (to be chosen by the community) and even to embark on collaborative development of new methods. An example project currently pioneering this approach is BigNeuron[40].

Organizing and participating in challenges is generally a lot of work and rarely supported by research funders. This makes many researchers (whether from computer vision or biology) reluctant to engage in them. However, these same researchers often spend enormous amounts of time (and thus project funding) on figuring out and reimplementing the best methods themselves. We argue that the return on investment will be higher through involvement in well-organized challenges. But even outside the context of such community efforts, the release of results in the form of user-friendly open-source software[22–24] and annotated image data sets[41] should be increasingly favored by funding agencies to maximize reproducibility and minimize wasting resources[42].

Future directions

Bioimage analysis has come a long way since its first steps half a century ago, yet there is a long way to go before computers can interpret data autonomously and biologists can blindly trust the findings. We have already touched upon various technical and cultural issues and hinted at possible future directions. In closing this article, we attempt to extrapolate some of the developments in the field and how they may impact bioimage analysis in the longer term.

On the technical side, much can be gained simply by better organization of the current state of the art. For example, published solutions for any of the discussed subproblems (**Fig. 1**) are increasingly becoming available as stand-alone software tools or as modules of general-purpose bioimage analysis platforms. This has generally been extremely helpful in minimizing the effort required to enable the testing of multiple approaches against each other on a problem, often suiting both computer scientists and biologists with no training in image analysis or computation. However, a failure to provide detailed specification of their design criteria, operation modes, boundary conditions, and optimal parameter settings for a range of example applications can lead to inappropriate reuse of methods, suboptimal performance, and lost time in correcting faulty results. This problem could be remedied by an online centralized database where developers and users contribute to cataloging the applicability of available tools by sharing their knowledge and experience. A first step in this direction is the BioImage Information Index (http://biii.info/) supported by Euro-BioImaging (http://www.eurobioimaging.eu/), the pan-European research infrastructure project for bioimaging technologies, and NEUBIAS (http://eubias.org/NEUBIAS/), the network of European bioimage analysts. Going one step further, interoperable versions of the tools themselves could be collected in the database, and software platforms capable of interactive construction of analysis workflows could connect to the database, enabling users to quickly identify, select, and use the best solutions for their problems. Not only could this stimulate optimal usage of existing tools by experimental biologists, but also it could reveal lacunas where new solutions are very much needed, providing highly valuable guidance to computer scientists.

Another very important technical development in dealing with the ever-increasing capacity of modern microscope systems to produce big data volumes concerns the interplay between image acquisition and image analysis. Often the paradigm in bioimaging experiments is to simply acquire as much image data as possible and afterwards let the analysis software separate the wheat from the chaff. Typically, this results in data sets that are vastly larger than needed to address the specific biological question under study and, in addition to causing unnecessarily high exposure of the sample, requires excess data storage and computational power. Treating acquisition and analysis as independent experimental steps is often a non-viable option. Biologists, microscopists, and computer scientists will need to join forces in developing 'smart imaging' systems[43] and more generally 'smart instrumentation'[44], in which sample preparation, targeting,

Table 1 Objective comparisons of bioimage analysis methods

Challenge	Task	Year(s)	Website	Refs.
DIADEM	Neuron morphology reconstruction in light microscopy	2009–2010	http://www.diademchallenge.org/	34
PTC	Particle detection and tracking in time-lapse microscopy	2012	http://bioimageanalysis.org/track/	35
SNEMI2D	Neurite segmentation in 2D electron microscopy	2012	http://brainiac2.mit.edu/isbi_challenge/	-
MITOS	Cell mitosis detection in breast cancer histopathology	2012, 2014	http://ipal.cnrs.fr/ICPR2012/	38
DM3D	3D deconvolution microscopy	2013, 2014	http://bigwww.epfl.ch/deconvolution/challenge2013/	-
CTC	Cell segmentation and tracking in time-lapse microscopy	2013–2015	http://www.codesolorzano.com/celltrackingchallenge/	36
SMLM	Single-molecule localization microscopy	2013	http://bigwww.epfl.ch/smlm/	37
SNEMI3D	Neurite segmentation in 3D electron microscopy	2013	http://brainiac2.mit.edu/SNEMI3D/	-
AMIDA	Cell mitosis detection in breast cancer histopathology	2013	http://amida13.isi.uu.nl/	39
OCCIS	Overlapping cell segmentation in cervical cytology	2014, 2015	http://cs.adelaide.edu.au/~zhi/isbi15_challenge/	-
MITOS-ATYPIA	Mitosis detection and nuclear atypia scoring in histopathology	2014	http://mitos-atypia-14.grand-challenge.org/	-
GLAS	Gland segmentation in histopathology	2015	http://www.warwick.ac.uk/bialab/GlasContest/	-
NCC	Nucleus counting in multichannel fluorescence microscopy	2015	http://isg.nist.gov/BII_2015/webPages/pages/nucleusCounting/NucleusCounting.html	-
BigNeuron	Large-scale 3D neuron morphology reconstruction in light microscopy	2015–2016	http://bigneuron.org/	40
CAMELYON	Cancer metastasis detection in histopathology	2016	http://camelyon16.grand-challenge.org/	-

These community efforts not only reveal which current methods work best for a given task, they also provide useful software tools, standardized reference data and metrics, and possible directions for future method development.

acquisition, and analysis are tightly connected through intelligent feedback mechanisms to minimize data overhead and maximize information content. This could be as simple as first acquiring images at low resolution, then identifying the areas of interest using appropriate detection methods, and subsequently acquiring high-resolution images of only those areas. A more sophisticated strategy would be to iterate over multiple resolutions and to involve higher-order analysis steps (**Fig. 1**) in the process, ultimately resulting in a fully integrated approach.

On the cultural side, the importance of fostering cross-fertilization and collaboration between the disciplines involved can hardly be overstated. The dialog between biologists and computer scientists has already been greatly improved in recent years by the organization of bioimage analysis workshops at biology conferences as well as at computer vision conferences, and by multidisciplinary meetings, such as the BioImage Informatics conference series (http://www.bioimageinformatics.org/). A more profound recent trend is the establishment of multidisciplinary education programs. For example, some technical universities and biomedical institutes are joining forces to offer students training in cell and molecular biology as well as in the computational sciences, thus resolving cultural barriers. As the new breed of professionals rises and takes on leading roles in the coming years, we can expect fundamental changes in the way research is being organized and also in how it is recognized by funding agencies. In the shorter term, a great opportunity for fruitful collaboration in taking bioimage analysis to higher levels is offered by the challenges framework. Standardization of benchmarks, including arriving at a consensus regarding representative data, performance criteria, algorithm implementation, integration, and dissemination, requires active participation of computer scientists and biologists and is probably the fastest route to improving the development, availability, and interoperability of bioimage analysis solutions. Engagement in such activities is being made considerably easier by online platforms, such as those developed by the Consortium for Open Medical Image Computing (https://grand-challenge.org/Contributors/).

We expect bioimage analysis solutions to become increasingly automated and robust. The capabilities of current solutions are already moving beyond simply automating what a biologist can do and are beginning to enable comprehensive analysis of all available information. This is especially important, given that imaging is not the only source of information in biology. Genomics, proteomics, transcriptomics, metabolomics, and other 'omics' all provide complementary views on biological processes, and their combination with imaging will yield a much more complete picture than any of these fields can offer individually[45]. The data deluge in all these fields poses major challenges in terms of standardized data storage and retrieval[46] but even more so for integrative data analysis. Although the development of methods for single-source data analysis remains important, methods that properly account for the links between data sets from multiple sources have the potential to achieve gains that go far beyond those possible when each data set is analyzed separately[47]. Thus we anticipate research in the field to become even more multidisciplinary. Meanwhile, ongoing advances in bioimage analysis are already putting more and more useful tools into the hands of biologists and will prove essential in unraveling the cellular and molecular processes of health and disease.

ACKNOWLEDGMENTS

The authors thank their group members and collaborators for insightful discussions that have helped shape their thoughts and research in bioimage analysis over the years. They also thankfully acknowledge support from Erasmus University Medical Center (E.M.), the National Science Foundation (CAREER DBI 1148823 to A.E.C.), the Allen Institute for Brain Science and the Janelia Research Campus of Howard Hughes Medical Institute (H.P.), the German Research Foundation (DFG SFB 1129/1134) and a Weston Visiting Professorship (F.A.H.), and Agence Nationale de la Recherche (ANR-10-INBS-04-06-France- BioImaging) and Institut Pasteur (J.C.O.M.). Raw data for the bioimage analysis examples shown are courtesy of Graham Knott (**Fig. 2c**) and Anna Akhmanova (**Fig. 2d**).

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Frisby, J.P. & Stone, J.V. *Seeing: The Computational Approach to Biological Vision* (The MIT Press, Cambridge, MA, USA, 2010).
2. Ji, N., Shroff, H., Zhong, H. & Betzig, E. Advances in the speed and resolution of light microscopy. *Curr. Opin. Neurobiol.* **18**, 605–616 (2008).
3. Prewitt, J.M.S. & Mendelsohn, M.L. The analysis of cell images. *Ann. NY Acad. Sci.* **128**, 1035–1053 (1966).
4. Peng, H. *et al.* Bioimage informatics for big data. *Adv. Anat. Embryol. Cell Biol.* **219**, 263–272 (2016).
5. Eliceiri, K.W. *et al.* Biological imaging software tools. *Nat. Methods* **9**, 697–710 (2012).
6. Szeliski, R. *Computer Vision: Algorithms and Applications* (Springer, London, UK, 2011).
7. Sarder, P. & Nehorai, A. Deconvolution methods for 3-D fluorescence microscopy images. *IEEE Signal Process. Mag.* **23**, 32–45 (2006).
8. Qu, L., Long, F. & Peng, H. 3-D registration of biological images and models: registration of microscopic images and its uses in segmentation and annotation. *IEEE Signal Process. Mag.* **32**, 70–77 (2015).
9. Wu, Q., Merchant, F.A. & Castleman, K.R. *Microscope Image Processing* (Academic Press, Burlington, MA, USA, 2008).

10. Meijering, E. Cell segmentation: 50 years down the road. *IEEE Signal Process. Mag.* **29**, 140–145 (2012).
11. Meijering, E., Dzyubachyk, O., Smal, I. & van Cappellen, W.A. Tracking in cell and developmental biology. *Semin. Cell Dev. Biol.* **20**, 894–902 (2009).
12. Pincus, Z. & Theriot, J.A. Comparison of quantitative methods for cell-shape analysis. *J. Microsc.* **227**, 140–156 (2007).
13. Depeursinge, A., Foncubierta-Rodriguez, A., Van De Ville, D. & Müller, H. Three-dimensional solid texture analysis in biomedical imaging: review and opportunities. *Med. Image Anal.* **18**, 176–196 (2014).
14. Shamir, L., Delaney, J.D., Orlov, N., Eckley, D.M. & Goldberg, I.G. Pattern recognition software and techniques for biological image analysis. *PLoS Comput. Biol.* **6**, e1000974 (2010).
15. Walter, T. *et al.* Visualization of image data from cells to organisms. *Nat. Methods* **7** (Suppl.), S26–S41 (2010).
16. Buck, T.E., Li, J., Rohde, G.K. & Murphy, R.F. Toward the virtual cell: automated approaches to building models of subcellular organization "learned" from microscopy images. *BioEssays* **34**, 791–799 (2012).
17. Neumann, B. *et al.* Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature* **464**, 721–727 (2010).
18. Takemura, S.Y. *et al.* A visual motion detection circuit suggested by *Drosophila* connectomics. *Nature* **500**, 175–181 (2013).
19. Ponti, A., Machacek, M., Gupton, S.L., Waterman-Storer, C.M. & Danuser, G. Two distinct actin networks drive the protrusion of migrating cells. *Science* **305**, 1782–1786 (2004).
20. Spanjaard, E. *et al.* Quantitative imaging of focal adhesion dynamics and their regulation by HGF and Rap1 signaling. *Exp. Cell Res.* **330**, 382–397 (2015).
21. Danuser, G. Computer vision in cell biology. *Cell* **147**, 973–978 (2011).
22. Cardona, A. & Tomancak, P. Current challenges in open-source bioimage informatics. *Nat. Methods* **9**, 661–665 (2012).
23. Prins, P. *et al.* Toward effective software solutions for big biology. *Nat. Biotechnol.* **33**, 686–687 (2015).
24. Carpenter, A.E., Kamentsky, L. & Eliceiri, K.W. A call for bioimaging software usability. *Nat. Methods* **9**, 666–670 (2012).
25. Marr, D. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (The MIT Press, Cambridge, MA, USA, 2010).
26. Ter Haar Romeny, B.M. *Front-End Vision and Multi-Scale Image Analysis* (Springer, Berlin, Germany, 2003).
27. Pridmore, T.P., French, A.P. & Pound, M.P. What lies beneath: underlying assumptions in bioimage analysis. *Trends Plant Sci.* **17**, 688–692 (2012).
28. Dudai, Y. How big is human memory, or on being just useful enough. *Learn. Mem.* **3**, 341–365 (1997).
29. Brady, T.F., Konkle, T. & Alvarez, G.A. A review of visual memory capacity: beyond individual items and toward structured representations. *J. Vis.* **11**, 4 (2011).
30. Bishop, C.M. *Pattern Recognition and Machine Learning* (Springer, New York, NY, USA, 2006).
31. Sommer, C. & Gerlich, D.W. Machine learning in cell biology - teaching computers to recognize phenotypes. *J. Cell Sci.* **126**, 5529–5539 (2013).
32. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
33. Price, K. Anything you can do, I can do better (no you can't). *Comput. Vis. Graph. Image Process.* **36**, 387–391 (1986).
34. Gillette, T.A., Brown, K.M., Svoboda, K., Liu, Y. & Ascoli, G.A. DIADEMChallenge.Org: a compendium of resources fostering the continuous development of automated neuronal reconstruction. *Neuroinform.* **9**, 303–304 (2011).
35. Chenouard, N. *et al.* Objective comparison of particle tracking methods. *Nat. Methods* **11**, 281–289 (2014).
36. Maška, M. *et al.* A benchmark for comparison of cell tracking algorithms. *Bioinformatics* **30**, 1609–1617 (2014).
37. Sage, D. *et al.* Quantitative evaluation of software packages for single-molecule localization microscopy. *Nat. Methods* **12**, 717–724 (2015).
38. Roux, L. *et al.* Mitosis detection in breast cancer histological images: An ICPR 2012 contest. *J. Pathol. Inform.* **4**, 8 (2013).
39. Veta, M. *et al.* Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Med. Image Anal.* **20**, 237–248 (2015).
40. Peng, H. *et al.* BigNeuron: large-scale 3D neuron reconstruction from optical microscopy images. *Neuron* **87**, 252–256 (2015).
41. Ljosa, V., Sokolnicki, K.L. & Carpenter, A.E. Annotated high-throughput microscopy image sets for validation. *Nat. Methods* **9**, 637 (2012).
42. Ince, D.C., Hatton, L. & Graham-Cumming, J. The case for open computer programs. *Nature* **482**, 485–488 (2012).
43. Scherf, N. & Huisken, J. The smart and gentle microscope. *Nat. Biotechnol.* **33**, 815–818 (2015).
44. Long, B., Li, L., Knoblich, U., Zeng, H. & Peng, H. 3D image-guided automatic pipette positioning for single cell experiments in vivo. *Sci. Rep.* **5**, 18426 (2015).
45. Fernández-González, R., Muñoz-Barrutia, A., Barcellos-Hoff, M.H. & Ortiz-de-Solorzano, C. Quantitative in vivo microscopy: the return from the 'omics'. *Curr. Opin. Biotechnol.* **17**, 501–510 (2006).
46. Swedlow, J.R., Zanetti, G. & Best, C. Channeling the data deluge. *Nat. Methods* **8**, 463–465 (2011).
47. Lahat, D., Adali, T. & Jutten, C. Multimodal data fusion: an overview of methods, challenges, and prospects. *Proc. IEEE* **103**, 1449–1477 (2015).
48. Ljosa, V. *et al.* Comparison of methods for image-based profiling of cellular morphological responses to small-molecule treatment. *J. Biomol. Screen.* **18**, 1321–1329 (2013).
49. Peng, H., Ruan, Z., Long, F., Simpson, J.H. & Myers, E.W. V3D enables real-time 3D visualization and quantitative analysis of large-scale biological image data sets. *Nat. Biotechnol.* **28**, 348–353 (2010).
50. Kreshuk, A., Koethe, U., Pax, E., Bock, D.D. & Hamprecht, F.A. Automated detection of synapses in serial section transmission electron microscopy image stacks. *PLoS One* **9**, e87351 (2014).