Artificial Intelligence for Microscopy: What You Should Know

Lucas von Chamier¹, Romain F. Laine¹-³, and Ricardo Henriques¹-³-⁴

¹MRC-Laboratory for Molecular Cell Biology, University College London, London, UK
²Department of Cell and Developmental Biology, University College London, London, UK
³The Francis Crick Institute, London, UK
⁴Institute for the Physics of Living Systems, University College London, London, UK

Artificial Intelligence based on Deep Learning is opening new horizons in biomedical research and promises to revolutionize the microscopy field. Slowly, it now transitions from the hands of experts in computer sciences to researchers in cell biology. Here, we introduce recent developments in deep learning applied to microscopy, in a manner accessible to non-experts. We overview its concepts, capabilities and limitations, presenting applications in image segmentation, classification and restoration. We discuss how deep learning shows an outstanding potential to push the limits of microscopy, enhancing resolution, signal and information content in acquired data. Its pitfalls are carefully discussed, as well as the future directions expected in this field.

Introduction. Deep Learning is a type of Artificial Intelligence (AI) which has recently seen a rise in academic research and popular interest. This sudden boost has been primarily fuelled by the invention of Convolutional Neural Networks (CNNs), a novel machine learning algorithmic architecture. In the early 2010’s CNNs became increasingly prominent as tools for image classification, showing superhuman accuracy at identifying objects in images (1). Since then, Deep Learning has expanded to many research fields, showing its potential to outsmart humans in board games such as Go (2), achieve self-driving cars (3, 4) and to significantly improve biomedical image analysis (5).

Recently, a growing amount of biology studies has used CNNs to analyze microscopy data, laying the foundation for a fundamental change in how imaging data is interpreted, and how microscopy is carried out. The areas where AI has been applied include: automated, accurate classification and segmentation of microscopy images (6–8); extraction of structures from label-free microscopy imaging (artificial labelling) (9, 10); and image restoration, such as denoising and resolution enhancement (11–13). Here, we give non-specialist readers an overview of the potential of Deep Learning, specifically through CNNs, in the context of some of the major challenges of microscopy. We will also discuss some of the current limitations of the approach and give an outlook on possible future applications of Deep Learning in microscopy.

For a more in-depth view into the AI-field, we advise the reader to see the review by Lecun et al. (14), which gives an extended perspective on Deep Learning and its historical development. Additionally Litjens et al. (15), Angermueller et al. (16) and Belthangady et al. (17) comprehensively discuss the application of AI in Biomedical Sciences and Computational Biology.

How does a CNN learn? CNNs are complex networks of connected ‘neurons’ arranged in ‘layers’, nomenclature inspired by the animal visual cortex (18). Each neuron performs specific mathematical operations on its input and passes the output to the neurons of the next layer of the network. In CNNs, the operations are typically either convolutions which extract patterns from images, or so-called...
pooling operations which reduce the number of pixels in the image and therefore simplify the data representations. This combination of feature extractions and data simplification is what allows CNNs to ‘understand’ the content of the images and perform efficiently on many types of imaging data. Once several layers are stacked upon another, networks become ‘deep’ and the information they can extract from inputs becomes increasingly complex (14).

Unlike conventional computer programs which are designed to perform a set of user-defined and well-understood operations on the input data (Fig. 1a) (19), a CNN has to be trained (Fig. 1b-i) on a so-called training dataset. This set consists of paired inputs and outputs which describe the transformation of the image that the user wishes the network to learn. For instance, for an image denoising operation, noisy images and matching high signal images have to be provided for training. The training dataset therefore defines the task that the network performs. During training, the network compares its output to the provided input and adjusts the parameters of its neurons, e.g. the impact that each neuron has on the next layer, until the network is able to infer an artificial output which resembles the expected real output. After training, networks are tested on a validation dataset, using inputs previously ‘unseen’ by the network, to determine if it can generalize over new data. In this case, real outputs are not used to train the network but to determine its performance. Once trained and validated, the network can be applied to new data for which no real output exists, identically to conventional computer programs (Fig. 1b-ii). Training is computationally intensive, taking hours to days, especially for deep networks with millions of trainable parameters. In comparison, after training, the inference process is considerably fast, taking minutes to seconds.

However, the computational performance of CNNs is incrementally improving with the development of increasingly powerful processing units, notably Graphical Processing Units (GPUs). Until the introduction of the first GPU-enabled CNN in 2012 (1) called AlexNet, CNNs were largely neglected in AI (14), because their training was too slow, sometimes requiring weeks to months of computation to complete. AlexNet greatly outperformed the competition at the ImageNet image classification challenge in 2012, a seminal breakthrough for the AI field.

The success of CNN algorithms depends on the design of the network architecture, usually carried out by computer scientists, and the availability of good training data. Generally, the training dataset should contain many different examples of the desired outputs. For example, a network designed to learn to categorize an animal should be trained with images showing the animal in different positions or environments. Generating and curating the training dataset is often the major bottleneck for the application of CNNs.

**CNNs in microscopy.** One of the first studies using GPU-enabled neural networks was able to segment neuronal membranes from electron microscopy (EM) images (20), vastly improving speed and efficiency of segmentation over the state-of-art methods. Another breakthrough came through the design of a more efficient network architecture which combines a number of convolution/pooling layers (the encoder), with a number of layers of de-convolution/up-sampling (the decoder) (7, 21). The encoder learns the main features of the image and the decoder reassigns them to different pixels of the image. Due to this conceptual down-followed by up-sampling, this network architecture was termed ‘U-net’. U-nets are therefore very powerful for image-to-image tasks (as opposed to simple classification of the image), making them one of the most important networks for microscopy applications (5, 9, 11, 22).

Researchers in life sciences face several challenges when imaging biological specimens: How can phototoxicity and bleaching of fluorescent labels be balanced against good signal or resolution? How many channels can a cell be imaged in without interfering with native processes? And how can relevant and complex information be extracted from large image datasets, without tedious manual annotation and human bias? In the following sections, we will present how AI methods have recently provided efficient solutions to these problems. While there exist some conceptual overlaps, we have separated these into four categories: image classification, image segmentation, artificial labelling and image restoration.

**Classification.** An important goal for microscopic image analysis is to recognize and assign identities to relevant features on an image (Fig. 2). For example, identifying mitotic cells in a tissue sample can be essential for cancer diagnosis. However, manual annotation is tedious, limited in throughput, and experts can introduce bias into such annotations by deciding which image features are important while ignoring others. Although several computational methods have been introduced to accelerate detection or classification tasks (23–25), these still often rely on handcrafted parameters, chosen by researchers. The advantage of CNNs is their capability to learn relevant image features automatically. CNNs have therefore been extensively used in the Biomedical imaging field, especially for cancer detection, particularly as large training sets have become more available (6, 26–30). The classification accuracy of neural networks has been successfully applied to high-throughput and high-content screens where it has shown expert-level recognition of subcellular features (31–34). Producing hand-labelled classification examples for such tasks can be a constraint for the creation of training sets. However, Lu et al. have shown that this can be overcome with unsupervised learning, where a network learns to predict protein localisations by being shown paired...
cells which were either imaged or not imaged in a specific channels, requiring no manual labelling to produce training inputs. (35).

CNNs have also shown the capacity to accurately identify cellular states from transmitted-light data, for example, differentiating cells based on cell cycle stage (36), cells affected by phototoxicity (37) or stem cell-derived endothelial cells (38). Identifying such identities, particularly in large image-sets previously often required the introduction of a label into the cell, an obstacle which CNNs can now overcome.

**Segmentation.** Segmentation is the identification of image regions that are part of specific cellular or sub-cellular structures and often is an essential step in image analysis (Fig. 3). A drawback of some existing segmentation platforms (23, 25, 39) is that they need user-based fine-tuning and manual error-removal, requiring time and expertise (40). Image segmentation can therefore be a bottleneck for research, particularly for high-throughput studies. CNNs have in multiple studies outperformed classical approaches in terms of accuracy and generalization (7, 20, 21, 40, 41), especially when performing cell segmentation in co-cultures of multiple cell types (40). In the context of histopathology, CNNs have been successfully used to segment colon glands (42–46) and breast tissues (47, 48), outperforming non-deep learning based approaches.

Naturally, there is overlap between the challenges of classification and segmentation since both require the network to learn about specific regions of an image. Hence, segmentation is often used with subsequent classification and can even improve the accuracy of classification (28, 49).

**Artificial Labelling.** The analysis of specific structures in cells, especially in light microscopy, typically requires the introduction of labels, either by genetic labelling or chemical staining, which can disturb the biological system. Additionally, fluorescence microscopy, especially when using laser illumination, is inherently more phototoxic to cells than transmitted-light imaging (50, 51). With this in mind, two studies using CNN methods have shown that specific cellular structures, such as nuclear membrane, nucleoli, plasma membranes and mitochondria, can be extracted by neural networks from label-free images (9, 10). The networks used here were trained to predict a fluorescent label from transmitted-light or EM images, alleviating the need to label and acquire the corresponding fluorescence images (Fig. 4). This capacity is especially useful when tracking individual structures over long periods of time. It was also shown that the networks can achieve high accuracy using a training dataset of only 30 to 40 images (9), and can simultaneously identify dying cells or distinguish different cell-types and subcellular structures (10). Christiansen et al. (10) also demonstrated their network’s ability for transfer learning, a method allowing a pre-trained network to perform a new task with minimal additional training. The network showed promising results for transfer learning between different microscopes and labels, highlighting the versatility of these networks’ performance. While the task of artificial labelling is similar to ‘classic’ segmentation, the main difference in this approach lies in the creation of the training set which does not require to be hand-labelled. The networks are simply trained from paired frames obtained from cells imaged in bright-field and fluorescence modalities.

**Image Restoration: resolution and signal.** The amount and quality of features which can be extracted from a microscopic image is limited by fundamental constraints inherent to all optical set-ups: signal-to-noise ratio (SNR) and resolution. Therefore, overcoming these limitations constitutes a central research goal in microscopy. In particular, super-resolution microscopy (SRM) (52–56) now allows imaging of cellular structures at the nanoscale using light microscopy. However, phototoxicity, bleaching and low temporal resolution still limit the capacity to achieve high-resolution long-term imaging in living specimens. Several research groups have proposed CNN methods addressing some of these issues in the last two years (11, 13?). Training approaches for networks published to date fall into two categories which could be described as naïve and informed. In the naïve approach, training datasets consist, for instance, of pairs of images acquired at low and high SNR and the network learns to predict a denoised image from a noisy input. The training dataset can be obtained from acquiring images at short and long exposure time or at different laser intensities. This approach was demonstrated on the highly photosensitive organism *Schmidtea mediterranea* and allowed a 60-fold decrease in illumination dose, enabling longer and more detailed observation of this organism *in vivo* (11). The authors also used the same approach to obtain SRM images from diffraction-limited image data, using the SRRF method as a reference (56, 57). Similarly, SRM images can be obtained from conventional confocal microscopy images using STED to acquire the high-resolution training dataset (?)
CNNs in Single-Molecule Localization Microscopy. CNNs have also recently generated interest in the Single-molecule Localization Microscopy (SMLM) field. All available studies were published within the last year, by independent groups, suggesting that the potential of AI for SRM is increasingly recognized in the community (13, 58–61). Applying sophisticated network architectures, with combinations of widefield and SMLM data as inputs (13, 58), it is possible to train a CNN for SMLM reconstruction. Here, the networks do not learn to localize individual fluorophores as typical single-molecule localization algorithms do, but instead to map sparse SMLM data of either microtubules, mitochondria or nuclear pores into SRM output images. This demonstrates the strength of CNNs for pattern recognition in redundant data, like SMLM data where only a small number of frames may suffice to reconstruct a Super-Resolution image. Interestingly, some of these algorithms require no parameter tuning or specific knowledge about the imaged structures (58). Especially, for high emitter density, this is advantageous over conventional SMLM reconstruction algorithms which can be time-consuming and require sample-dependent optimisation of imaging parameters.

Two other groups have used a different approach to SMLM reconstruction which could be termed ‘informed’ as it makes use of the intrinsic properties of SMLM data (59, 60). Here, networks are trained to learn the positions of fluorophores from SMLM input images. In contrast to naïve training, this means that such CNNs can only be used for data acquired with SMLM techniques. However, the advantage is that the output of these networks contains the positional information of the fluorophores whereas the output of naïve networks does not. The outstanding question about how neural networks manage to produce Super Resolution images from sparse or widefield data, is circumvented by this approach. The reconstructed images are therefore more similar to standard SMLM reconstructions making the resolution improvement easier to interpret.

While achieving similar accuracy to state of the art SMLM algorithms (62), a main achievement of deep learning for SMLM is the speed with which super-resolved images can be produced. In several studies this was increased by several orders of magnitude compared to conventional reconstruction algorithms (13, 58, 59, 61).

Discussion. The use of neural networks is transforming microscopy both by allowing human or super-human performances for a number of image analysis tasks and as an automated high-performance tool for big-data analysis (33, 34, 40) (Table 1). However, while performance, versatility and speed of neural networks is likely to continue increasing, there are significant challenges which will not be solved by improved processing units. A frequently raised concern in the microscopy community over AI is how much machine outputs can be trusted to represent data. This is a real concern since CNNs have been observed to cause image hallucinations (67) or to fail catastrophically simply as a result of minute changes in the image (68). To address this issue, several groups have assessed the presence of artefacts in their network output images, notably using SQUIRREL (11–13, 61, 69, 70). While this may identify the presence of artefacts, it does not address the underlying problem that it is difficult to interpret how CNN architectures produce their output from the image input. This lack of interpretability of network outputs is particularly concerning in the case of resolution enhancement, where it is not clear what information a CNN can extract from a diffraction limited image to achieve a non-diffraction limited image and how deep learning algorithms achieve this without producing significantly more artefacts than standard algorithms (11, 12). Another consequence of this is that the design of CNN architectures has been referred to as ‘notorious as an empirical endeavour’ (10). Choosing network parameters such as network depth, number of neural connections, learning rate and other hand-coded features of neural networks, also termed hyperparameters, is therefore based on evaluating the respective choices by performance (33, 60, 71, 72).

Beside issues of interpretability, there are other anecdotal examples where networks have ‘cheated’ their way to high performance, e.g. by using undesirable features such as empty space (37) to identify dead cells or by identifying patterns in the ordering of the training set, but not in the images themselves (73). This shows how much of the performance of Deep Learning methods relies on the choice and curation of training data sets, which can be a significant challenge for the application of deep learning for microscopy.

Despite these issues, AI has great enabling potential for microscopy, given super-human performance in classification tasks and image reconstruction. Hence, the issues discussed above should not discourage the use of CNNs as a research tool but be reason for caution when interpreting the performance of neural networks, as for any computational analysis tool.

Outlook. Most of the methods reviewed here were published within the past 24 months, demonstrating the massive interest in deep learning as a versatile and powerful tool for microscopy applications. However, the delay between developments and their applications means that some areas of AI research have not yet been widely translated to microscopy. For example, transfer learning is an area which will likely become more widely investigated, allowing the use of pre-trained networks to carry out a new task, forms of which are only starting to become available (8). Finding methods to reuse neural networks robustly on multiple different tasks, different image sizes or images taken on different microscopes would make deep learning a more flexible and usable approach for image analysis than is currently possible. Importantly, it would reduce the need for large training datasets and shorten the training needed for new tasks. It could therefore lower the accessibility barrier of the approach and minimize the need for biologists to be fully familiar with neural network specifics. In turn, this would allow deep learning to become a tool, rather than a method that needs expert knowledge. This will likely open the door for deep learning to become even more widely used within life science.
Table 1. Major publications from the recent years applying or developing deep learning to microscopy. This table covers the main four themes where AI has provided solutions to some of the major limitations of microscopy in the recent years.

<table>
<thead>
<tr>
<th>Classification/Segmentation</th>
<th>Authors</th>
<th>Year</th>
<th>Details</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image Restoration/Super Resolution</td>
<td>Nehme et al.</td>
<td>2018</td>
<td>SMLM images from diffraction limited input</td>
<td>(58)</td>
</tr>
<tr>
<td></td>
<td>Boyd et al.</td>
<td>2018</td>
<td>Identifies localisation of fluorophores from STORM single frames</td>
<td>(59)</td>
</tr>
<tr>
<td></td>
<td>Ouyang et al.</td>
<td>2018</td>
<td>SMLM reconstruction using very small number of frames to predict SR image</td>
<td>(13)</td>
</tr>
<tr>
<td></td>
<td>Nelson &amp; Hess</td>
<td>2018</td>
<td>PALM reconstruction network trained directly on the image to be analysed</td>
<td>(60)</td>
</tr>
<tr>
<td></td>
<td>Weigert et al.</td>
<td>2018</td>
<td>Denoising and resolution enhancement in different organisms and cell types</td>
<td>(11)</td>
</tr>
<tr>
<td></td>
<td>Li et al.</td>
<td>2018</td>
<td>SMLM localisation by deep learning and artefact removal by statistical inference</td>
<td>(61)</td>
</tr>
<tr>
<td></td>
<td>Wang et al.</td>
<td>2019</td>
<td>Conversion of low NA to high NA or diffraction limited to STED resolution images</td>
<td>(12)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Artificial Labelling</th>
<th>Authors</th>
<th>Year</th>
<th>Details</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Christiansen et al.</td>
<td>2018</td>
<td>Label Prediction in fixed and live cells</td>
<td>(10)</td>
</tr>
<tr>
<td></td>
<td>Ounkomol et al.</td>
<td>2018</td>
<td>3D label prediction in live-cell, IF and EM images</td>
<td>(9)</td>
</tr>
</tbody>
</table>

ACKNOWLEDGEMENTS

We'd like to thank Simon F. Norrelykke (ETH Zürich, Switzerland) and Alex Lu (University of Toronto, Canada) for kindly suggesting corrections to this manuscript, see here. This work was funded by grants from the UK Biotechnology and Biological Sciences Research Council (BB/M022374/1; BB/P027431/1; BB/R000067/1; BB/S507532/1) (R.H.), the Wellcome Trust (203276/Z/16/Z) (R.H.), Core fundings (BB/K015826/1; BB/P027431/1) (R.H. and R.F.L.), the UK Medical Research Council (BB/M022374/1; BB/P027431/1; BB/R00067/1; BB/S507532/1) (R.H.), the MRC Laboratory for Molecular Cell Biology, University College London (MC_UU12018/7). L.C supported by a 4-year MRC Research Studentship.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Bibliography

2. David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur
19. Ann Wheeler and Ricardo Henriques. Standard and super-resolution bioimaging data anal-
17. Chinmay Belthangady and Loic A Royer. Applications, promises, and pitfalls of deep learn-
35. Alex Lu, Oren Z Kraus, Sam Cooper, and Alan M Moses. Learning unsupervised feature
32. Oren Z Kraus, Jimmy Lei Ba, and Brendan J Frey. Classifying and segmenting microscopy
26. Zheng Xu and Junzhou Huang. Detecting 10,000 cells in one second. In
15. Le Quoc, Xuan Le Tran, and Thang Dinh. Deep learning for cell recognition and indenfication.
14. Chawin Ounkomol, Sharmishtaa Seshamani, Mary M Maleckar, Forrest Collman, and Gre-
13. Martin Weigert, Uwe Schmidt, Tobias Boothe, Andreas Müller, Alexander Dibrov, Akanksha
12. Jan van der Schans, Wiel Coenen, and Mibe van Ginneken. Learning to detect cell nuclei in
10. Eric Betzig, George H Patterson, Rachid Sougrat, O Wolf Lindwasser, Scott Olenych,
9. Chawin Ounkomol, Sharmishtaa Seshamani, Mary M Maleckar, Forrest Collman, and Gre-
8. Philipp Kainz, Michael Pfeiffer, and Martin Urschler. Semantic segmentation of colon glands
7. Siân Culley, Kalina L Tosheva, Pedro Matos Pereira, and Ricardo Henriques. Srrf: Univer-
5. Eric M Christiansen, Margaret J Wang, O D Mitchel, Ashkan Jahangirian, Gaia Sibkinesis, Scott
4. Alex Lu, Oren Z Kraus, Sam Cooper, and Alan M Moses. Learning unsupervised feature
2. Eric Betzig, George H Patterson, Rachid Sougrat, O Wolf Lindwasser, Scott Olenych,


