

## REVIEW

# Deep learning for bioimage analysis in developmental biology

Adrien Hallou<sup>1,2,3,\*</sup>, Hannah G. Yevick<sup>4,\*</sup>, Bianca Dumitrescu<sup>5,‡</sup> and Virginie Uhlmann<sup>6,‡</sup>

## ABSTRACT

Deep learning has transformed the way large and complex image datasets can be processed, reshaping what is possible in bioimage analysis. As the complexity and size of bioimage data continues to grow, this new analysis paradigm is becoming increasingly ubiquitous. In this Review, we begin by introducing the concepts needed for beginners to understand deep learning. We then review how deep learning has impacted bioimage analysis and explore the open-source resources available to integrate it into a research project. Finally, we discuss the future of deep learning applied to cell and developmental biology. We analyze how state-of-the-art methodologies have the potential to transform our understanding of biological systems through new image-based analysis and modelling that integrate multimodal inputs in space and time.

**KEY WORDS:** Deep learning, Neural network, Image analysis, Microscopy, Bioimaging

## Introduction

In the past decade, deep learning (DL) has revolutionized biology and medicine through its ability to automate repetitive tasks and integrate complex collections of data to produce reliable predictions (LeCun et al., 2015). Among its many uses, DL has been fruitfully exploited for image analysis. Although the first DL approaches that were successfully used for the analysis of medical and biological data were initially developed for computer vision applications, such as image database labelling (Krizhevsky et al., 2012), many research efforts have since focused on tailoring DL for medical and biological image analysis (Litjens et al., 2017). Bioimages (see Glossary, Box 1), in particular, exhibit a large variability due to the countless different possible combinations of phenotypes of interest, sample preparation protocols, imaging modalities and acquisition parameters. DL is thus a particularly appealing strategy to design general algorithms that can easily adapt to specific microscopy data with minimal human input. For this reason, the successes and promises of DL in bioimage analysis applications have been the topic of a number of recent review articles (Gupta et al., 2018; Wang et al., 2019; Moen et al., 2019; Meijering, 2020; Hoffman et al., 2021; Esteva et al., 2021).

Here, we expand upon a recent Spotlight article (Villoutreix, 2021) and tour the practicalities of the use of DL for image analysis in the context of developmental biology. We first provide a primer on key machine learning (ML) and DL concepts. We then review the use of DL in bioimage analysis and outline success stories of DL-enabled bioimage analysis in developmental biology experiments. For readers wanting to further experiment with DL, we compile a list of freely available resources, most requiring little to no coding experience. Finally, we discuss more advanced DL strategies that are still under active investigation but are likely to become routinely used in the future.

## What is machine learning?

The term machine learning defines a broad class of statistical models and algorithms that allow computers to perform specific data analysis tasks. Examples of tasks include, but are not limited to, classification, regression, ranking, clustering or dimensionality reduction (defined by Mohri et al., 2018), and are usually performed on datasets collected with or without prior human annotations.

Three main ML paradigms can be distinguished: supervised, unsupervised and reinforcement learning (Murphy, 2012; Villoutreix, 2021). The overwhelming majority of established bioimage analysis algorithms rely on supervised and unsupervised ML paradigms and we therefore focus on these two in the rest of the article. In supervised learning, existing human knowledge is used to obtain a ‘ground truth’ (see Glossary, Box 1) label for each element in a dataset. The resulting data-label pairs are then split into a ‘training’ and a ‘testing’ set (see Glossary, Box 1). Using the training set, the ML algorithm is ‘trained’ to learn the relationship between ‘input’ data and ‘output’ labels by minimizing a ‘loss’ function (see Glossary, Box 1), and its performance is assessed on the testing set. Once training is complete, the ML model can be applied to unseen, but related, input data (see Glossary, Box 1) in order to predict output labels. Classical supervised ML methods include random forests, gradient boosting and support vector machines (Mohri et al., 2018). In contrast, unsupervised learning deals with unlabelled data: ML is then employed to uncover patterns in input data without human-provided examples. Examples of unsupervised learning tasks include clustering and dimensionality reduction, routinely used in the analysis of single-cell ‘-omics’ data (Libbrecht and Noble, 2015; Argelaguet et al., 2021).

## Neural networks and deep learning

DL designates a family of ML models based on neural networks (NN) (LeCun et al., 2015). Formally, an NN aims to learn non-linear maps between inputs and outputs. An NN is a network of processing ‘layers’ composed of simple, but non-linear, units called artificial neurons (Fig. 1A–C). When composed of several layers, an NN is referred to as a deep NN. Layers of artificial neurons transform inputs at one level (starting with input data) into outputs at the next level such that the data becomes increasingly more abstract as it progresses through the different layers, encapsulating in the process the complex non-linear relationship usually existing between input

<sup>1</sup>Cavendish Laboratory, Department of Physics, University of Cambridge, Cambridge, CB3 0HE, UK. <sup>2</sup>Wellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, Cambridge, CB2 1QN, UK. <sup>3</sup>Wellcome Trust/Medical Research Council Stem Cell Institute, University of Cambridge, Cambridge, CB2 1QR, UK. <sup>4</sup>Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, 02142, USA. <sup>5</sup>Computer Laboratory, Cambridge, University of Cambridge, Cambridge, CB3 0FD, UK. <sup>6</sup>European Bioinformatics Institute, European Molecular Biology Laboratory, Cambridge, CB10 1SD, UK. \*These authors contributed equally to this work

‡Authors for correspondence (bmd39@cam.ac.uk; uhlmann@ebi.ac.uk)

ib A.H., 0000-0002-3162-7848; H.G.Y., 0000-0002-2537-4928; B.D., 0000-0001-8328-2354; V.U., 0000-0002-2859-9241

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

**Box 1. Glossary**

**Accuracy.** Ratio of correctly predicted instances to the total number of predicted instances.

**Bioimages.** Visual observations of biological structures and processes at various spatiotemporal resolutions stored as digital image data.

**Convolutional layer.** A type of layer akin to an image processing filter, the values of which are free parameters to be learnt during training. Each neuron in a convolutional layer is only connected to a few adjacent neurons in the previous layer.

**Data augmentation.** Strategy to enhance the size and quality of training sets. Typical techniques include random cropping, geometrical operations (e.g. rotations, translations, flips), intensity and contrast modifications, and non-rigid image transformations (e.g. elastic deformations).

**Dense layer or fully connected layer.** A type of layer in which all neurons are connected to all the neurons in the preceding layer.

**Ground truth.** Output known to be correct for a given input.

**Image patch.** Small, rectangular piece of a larger image (e.g. 64×64 pixel patches for 1024×1024 pixel images) used to minimize computational costs during training.

**Input data.** Data fed into an ML model.

**Loss.** Function evaluating how closely the predictions of a model match the ground truth.

**Layer.** Set of interconnected artificial neurons in an NN.

**Output data.** Data coming out of an ML model.

**Pooling.** Operation consisting of aggregating adjacent neurons with a maximum, minimum or averaging operator.

**Signal-to-noise ratio (SNR).** Measure of image quality usually computed as the ratio of the mean intensity value of a digital image to the standard deviation of its intensity values.

**Style transfer.** Method consisting of learning a specific style from a reference image, such that any input image can then be ‘painted’ in the style of the reference while retaining its specific features.

**Training.** Process through which the parameters of an ML model are optimized to best map inputs into desired outputs.

**Training/testing set.** Collections of known input-output pairs. The training set is used during training *per se*, whereas the testing set is used *a posteriori* to test the performances of the ML model on unseen data.

**Transfer learning.** Method in which an ML model developed for a task is reused for a different task. For example, an NN can be initialized with the weights of another NN pre-trained on a large unspecific image dataset, and then fine-tuned with a problem-specific training set of smaller size.

**U-net.** A highly efficient CNN architecture used for various image analysis tasks (Fig. 1B).

**Weights.** NN parameters that are iteratively adjusted during the training process.

**Voxel.** The three-dimensional equivalent to a pixel.

and output data (see Glossary, Box 1; Fig. 1A). This process allows sufficiently deep NN to learn during training some higher-level features contained in the data (Goodfellow et al., 2016). For example, for a classification problem such as the identification of cells contained in a fluorescence microscopy image, this would typically involve learning features correlated with cell contours while ignoring the noisy variation of pixel intensity in the background of the image.

Intuitively, a DL model can be viewed as a machine with many tuneable knobs, which are connected to one another through links. Tuning a knob changes the mathematical function that transforms the inputs into outputs. This transformation depends on the strength of the links between the knobs, and the importance of the knobs, known together as ‘weights’ (see Glossary, Box 1). A model with randomly set weights will make many mistakes, but the so-called ‘winning lottery hypothesis’ (Frankle and Carbin, 2018 preprint) assumes that an optimal configuration of knobs and weights exists.

This optimal configuration is searched for during training, in which the knobs of the DL model are reconfigured by minimizing the loss function. Although prediction with trained networks is generally fast, training deep NN *de novo* proves to be more challenging. A main difficulty in DL lies in finding an appropriate numerical scheme that allows, with limited computational power, tuning of the tens of thousands of weights contained in each layer of the networks and obtaining high ‘accuracy’ (see Glossary, Box 1) (LeCun et al., 2015; Goodfellow et al., 2016). Although the idea (McCulloch and Pitts, 1943) and the first implementations of NN (Rosenblatt, 1958) date back to the dawn of digital computing, it took several decades for the development of computing infrastructure and efficient optimization algorithms to allow implementations of practical interest, such as handwritten-digit recognition (LeCun et al., 1989).

Convolutional NN (CNN) are a particular type of NN architecture specifically designed to be trained on input data in the form of multidimensional arrays, such as images. CNN attracted particular interest in image processing when, in the 2012 edition of the ImageNet challenge on image classification, the AlexNet model outperformed by a comfortable margin other ML algorithms (Krizhevsky et al., 2017). In bioimage analysis application, the U-net architecture (see Glossary, Box 1; Fig. 1B; Falk et al., 2018) has become predominant, as discussed below.

**Deep learning for bioimage analysis**

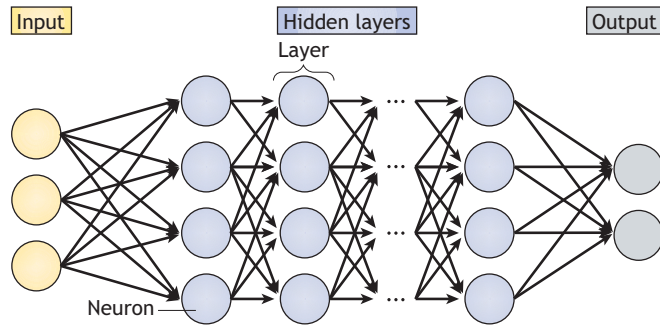
DL in bioimage analysis tackles three main kind of tasks: (1) image restoration, in which an input image is transformed into an enhanced output image; (2) image partitioning, whereby an input image is divided into regions and/or objects of interest; and (3) image quantification, whereby objects are classified, tracked or counted. Here, we illustrate each class of application with examples of DL-enabled advances in cell and developmental biology.

**Image restoration**

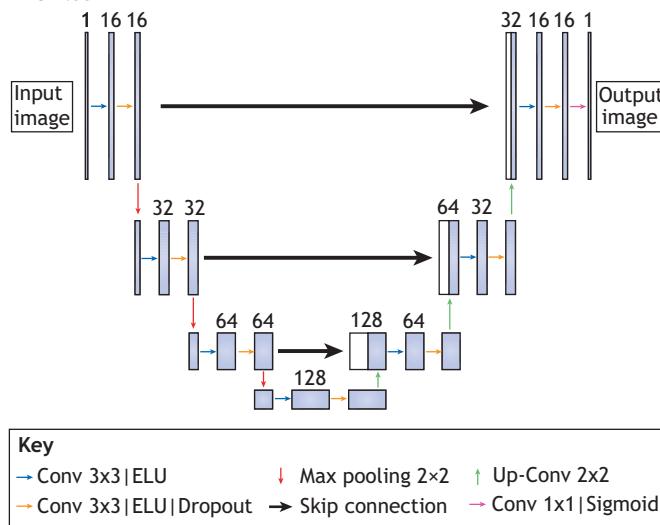
Achieving a high signal-to-noise ratio (SNR; see Glossary, Box 1) when imaging an object of interest is a ubiquitous challenge when working with developmental systems. Noise in microscopy can arise from several sources (e.g. the optics of the microscope and/or its associated detectors or camera). Live imaging, in particular, usually involves compromises between SNR, acquisition speed and imaging resolution. In addition, regions of interest in developing organisms are frequently located inside the body, far from the microscope objective. Therefore, because of scattering, light traveling from fluorescent markers can be distorted and less intense when it reaches the objective. Photobleaching and phototoxicity are also increasingly problematic deeper into the tissue, leading to low SNR as one mitigates its effect through decreased laser power and increased camera exposure or detector voltage (reviewed by Boka et al., 2021). DL has been successful at overcoming these challenges when used in the context of image restoration algorithms, which transform input images into output images with improved SNR.

Although algorithms relying on theoretical knowledge of imaging systems have made image restoration possible since the early days of bioimage analysis (Born et al., 1999; Gibson and Lanni, 1989, 1992), the competitive performance of both supervised and unsupervised forms of DL has introduced a paradigm shift. Despite lacking in theoretical guarantees, several purely data-driven DL-based approaches outshine non-DL strategies in accurate image restoration tasks. One challenge in applying supervised DL to image restoration is the need for high-quality training sets of ground truth images exhibiting a reduced

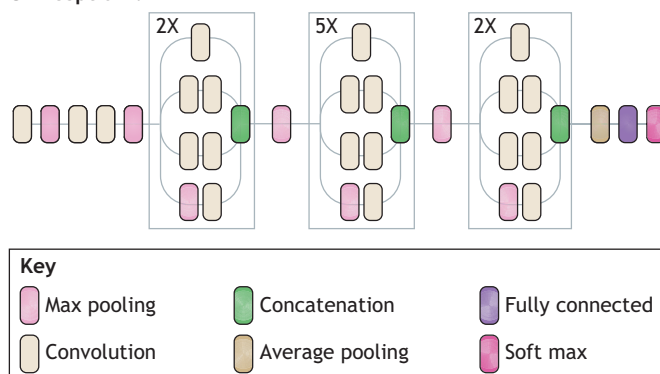
## A Neural network



## B U-Net



## C Inception V1



amount of noise. A notable example of DL-based image restoration algorithm requiring a relatively small training set [200 image patches (see Glossary, Box 1), size  $64 \times 64 \times 16$  pixels] is content-aware image restoration (CARE) (Weigert et al., 2018). To train CARE, pairs of registered low-SNR and high-SNR images must first be acquired. The high-SNR images serve as ground truth for training a DL model based on the U-net architecture (Falk et al., 2018) (Fig. 1B). The trained network can then be used to restore noiseless, higher-resolution images from unseen noisier datasets (Box 2). Often, however, high-SNR ground truth image data cannot be easily generated experimentally. In such cases, synthetic high-SNR images generated by non-DL deconvolution algorithms can be used to train the network. For example, CARE has been trained to

**Fig. 1. Neural networks and convolutional neural networks for bioimage analysis.** (A) Schematic of a typical NN composed of an input layer (green), hidden layers (blue) and an output layer (red). Each layer is composed of neurons connected to each other. (B) Schematic of a U-net architecture as used in McGinn et al. (2021) for the segmentation of cells and nuclei in mouse epithelial tissues. U-net is amongst the most popular and efficient CNN models used for bioimage analysis and is designed using 'convolutional', 'pooling' and 'dense' layers as key building blocks (see Glossary, Box 1). U-net follows a symmetric encoder-decoder architecture resulting in a characteristic U-shape. Along the encoder path, the first branch of the U, the input image is progressively compacted, leading to a representation with reduced spatial information but increased feature information. Along the decoder path, the second branch of the U, feature and spatial information are combined with information from the encoder path, enforcing the model to learn image characteristics at various spatial scales. (C) Schematic of an Inception V1 architecture, also called 'GoogleLeNet'. Inception V1 is a typical CNN architecture for image classification tasks. For example, it has been used to classify early human embryos images with very high accuracy (Khosravi et al., 2019). It is designed around a repetitive architecture made of so-called 'inception blocks', which apply several 'convolutional' and max 'pooling' layers (see Glossary, Box 1) to their input before concatenating together all generated feature maps.

resolve sub-diffraction structures in low-SNR brightfield microscopy images using synthetically generated super-resolution data (Weigert et al., 2018). More recently, the DECODE method (Speiser et al., 2020 preprint) uses a U-net architecture to address the related challenge of computationally increasing resolution in the context of single-molecule localization microscopy. The U-net model takes into account multiple image frames, as well as their temporal context. DECODE can localize single fluorophore emitters in 3D for a wide range of emitter brightnesses and densities, making it more versatile compared with previous CNN-based methods (Nehme et al., 2020; Boyd et al., 2018 preprint).

Unsupervised methods for image restoration offer an alternative to the generation of dedicated or synthetic training sets. Some recent denoising approaches exploit DL to learn how to best separate signal (e.g. the fluorescent reporter from a protein of interest) from noise, in some cases without the need for any ground truth. Noise2Noise, for example, uses a U-net model to restore noiseless images after training on pairs of independent noisy images, and was demonstrated to accurately denoise biomedical image data (Lehtinen et al., 2018) (Fig. 2A). Going further, Noise2Self modifies Noise2Noise to only require noisy images split into input and target sets (Batson and Royer, 2019 preprint). In these algorithms, training is carried out on noisy images under the assumption that noise is statistically independent in image pairs, whereas the signal present is more structured. Alternatively, Noise2Void proposes a strategy to train directly on the dataset that needs to be denoised (Krull et al., 2019) (Fig. 2A). The Noise2 model family is ideal for biological applications, in which it can be challenging to obtain noise-free images.

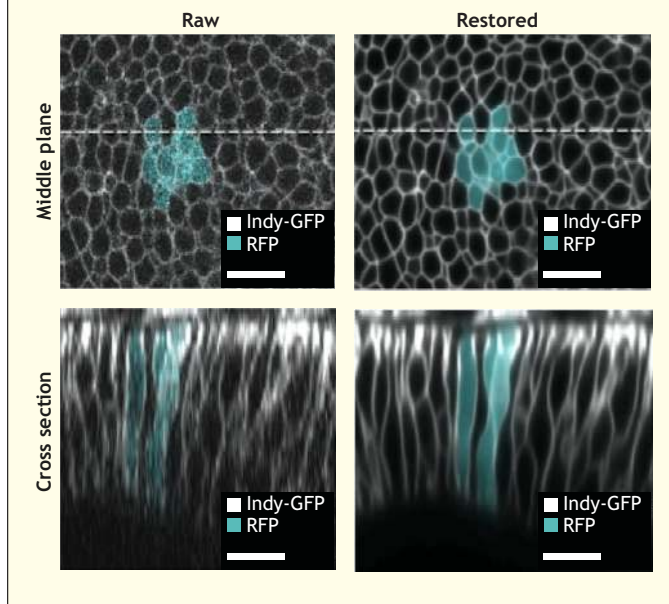
## Image partitioning

Analyzing specific objects in a biological image generally requires an image partitioning step; the separation of objects of interest from the image background. Image partitioning can either consist of detecting a bounding box around objects (object detection) or of identifying the set of pixels composing each object (segmentation). Although images featuring a few objects can be partitioned by hand, large datasets necessitate automation. DL approaches originating from computer vision have greatly enhanced the speed and accuracy of both object detection and segmentation in biological images.

### Box 2. Case study: denoising the lateral cell faces of the developing *Drosophila* wing disc with CARE

Sui and colleagues explore the role of lateral tension in the *Drosophila* wing disc in guiding epithelial folding (Sui et al., 2018; Sui and Dahmann, 2020). In this example, the plane of the fly wing is mounted facing the objective, placing the lateral sides of the wing disc cells along the z-axis. However, image resolution in the x-y plane of a microscope (top left) generally exceeds that of out-of-plane (z) resolution (bottom left). Reconstructing fluorescent signals from the lateral face also requires reconstruction of z profiles by summing together signals from multiple depths. Furthermore, a sensitivity to light exposure of the system imposes that imaging be carried out at low laser power and on a few z slices, further decreasing the resolution of the lateral face.

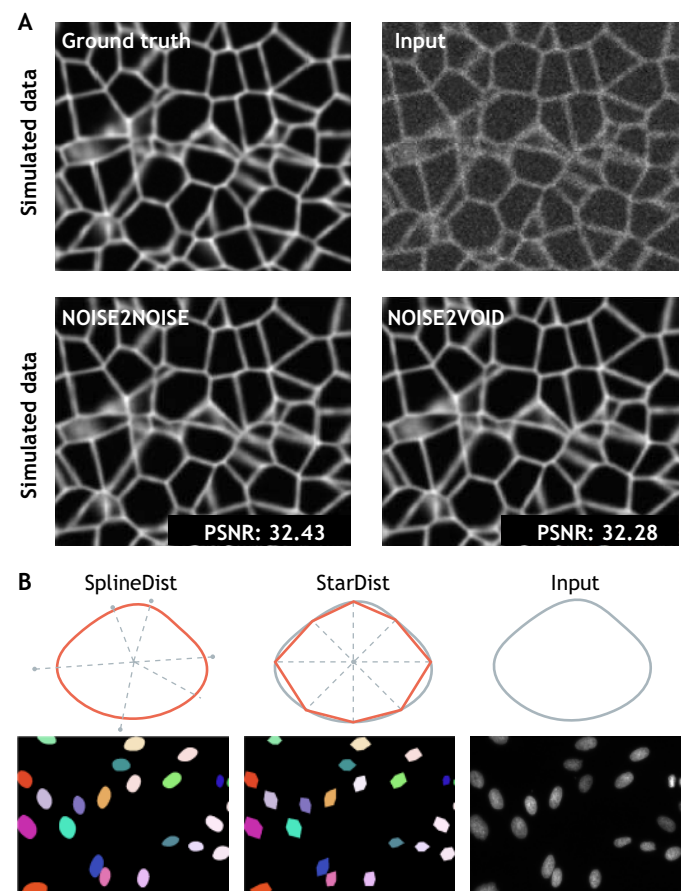
The quality of the acquired microscopy data is successfully improved relying on content-aware image restoration (CARE) (top right) (Weigert et al., 2018). The CARE network is trained on pairs of low- and high-resolution imaginal discs images. First a z-stack is acquired using low laser power and low z sampling, followed by another z-stack acquired at the same position in the sample with increased laser power and 4× more imaged focal planes ( $n=8$  stacks of average dimensions  $102 \times 512 \times 30$  with pixel size  $0.17 \times 0.17 \times 0.32 \mu\text{m}$ , for a total dataset size of  $\sim 1\text{GB}$ ). Once trained, the network is used to process low-resolution images of other lateral markers, enabling the quantitative analysis of how protein localization changes over time on lateral cell faces during and after photoactivation (bottom right). Although absolute intensity measurements extracted from images restored with DL methods should be subject to caution, restored images in this work were only used to track relative changes in apical, basal and lateral intensity over time. Image adapted from Sui et al. (2018). Scale bars:  $10 \mu\text{m}$ .



Since U-net, countless customized DL models have adapted to bioimage-specific object detection (Waithe et al., 2020; Wollmann and Rohr, 2021) and segmentation problems have been proposed (Long, 2020; Chidester et al., 2019; Tokuoka et al., 2020). A strong link to computer vision remains, as many of these methods draw from partitioning tasks in natural images. For example, algorithms initially designed to segment people and cars from crowded cityscapes can be efficiently exploited to segment challenging electron microscopy datasets (Wolf et al., 2018, 2020).

The automated segmentation of cell nuclei in various kinds of microscopy images has attracted a particular amount of attention.

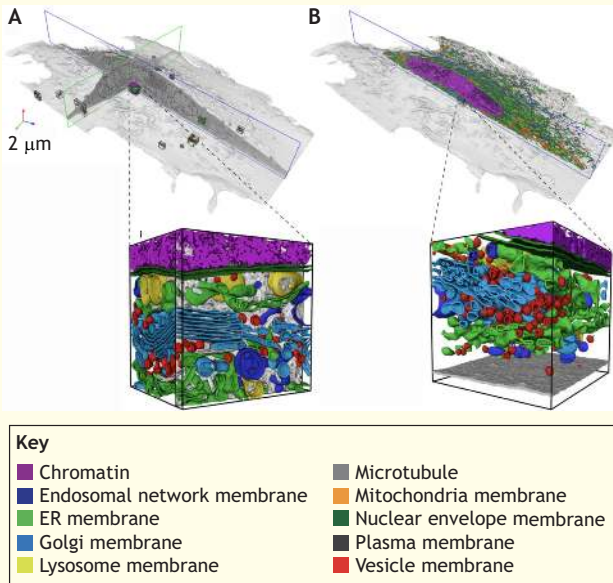
Cell nuclei can be tightly packed, making nuclei and cell bodies difficult to differentiate from neighbours. Spatial variations in marker intensity (due to local differences in staining efficacy), chromatin compaction or illumination fluctuations introduce further challenges. Mask Region-based CNN (He et al., 2017) – an extension of Fast R-CNN, first developed for the general task of object detection in natural images – has been successfully adapted to nuclei segmentation. Building on object detection methods, StarDist (Schmidt et al., 2018) adds assumptions about the geometry of nuclei shapes to improve detection performance. Relying on a U-net model, StarDist predicts a star-convex representation of individual object contours and can successfully separate overlapping nuclei in 2D images (Fig. 2B). A 3D version of StarDist is also available for volumetric (e.g. light sheet microscopy) data, which is often generated in developmental biology experiments (Weigert et al., 2020). More recently, SplineDist extends StarDist by using a more flexible representation of objects, allowing for the segmentation of more complex shapes (Mandal and Uhlmann, 2021) (Fig. 2B). For these methods, larger training sets and crowdsourced improvements on model architecture have pushed the limits of achievable accuracy



**Fig. 2. Deep learning methods applied to developmental biology applications.** (A) A simulated ground truth cell membrane image is artificially degraded with noise. Denoised outputs obtained using Noise2Noise and Noise2Void are shown at the bottom, along with their average peak signal-to-noise values (PSNR; higher values translate to sharper, less-noisy images). Image adapted from Krull et al. (2019). (B) Fluorescence microscopy cell nuclei image from the Kaggle 2018 Data Science Bowl (dataset: BBBC038v1; Caicedo et al., 2019) segmented with StarDist (Schmidt et al., 2018), in which objects are represented as star-convex polygons, and with SplineDist, in which objects are described as a planar spline curve. Image adapted from Mandal and Uhlmann (2021).

### Box 3. Case study: automatic whole cell organelle segmentation in volumetric electron microscopy

Reconstructing the shape of internal components from focused ion beam scanning electron microscopy (FIB-SEM) data is a complicated task owing to the crowded cytoplasmic environment of a cell. As a result, segmentation has been a bottleneck for understanding organelle morphologies and their spatial interactions as observed in SEM images at the nanometer scale. In OpenOrganelle (Heinrich et al., 2020 preprint), an ensemble of 3D U-nets (see Glossary, Box 1) have been trained for organelle segmentation in diverse cell types. The model is able to segment and classify up to 35 different classes of organelle, including endoplasmic reticulum (ER), microtubules and ribosomes. The network is trained with a diverse dataset of 73 volumetric regions, sampled from five different cell types, which sum up to  $\sim 635 \times 10^6$  voxels (see Glossary, Box 1). The identity of enclosed organelles in the chosen volumes are manually annotated using morphological features established in the literature (A). Achieving a manual segmentation of the dense array of organelles in a single  $\sim 1 \mu\text{m}^2$  FIB-SEM slice required 2 weeks of manual labour for an expert, meaning that manual annotation of an entire cell (2250 $\times$  larger) would take  $\sim 60$  years. In contrast, the DL model trained on these manual annotations is able to segment individual organelles on a whole cell volume in a matter of hours (B). Image adapted from Heinrich et al. (2020 preprint).



and generalization. The availability of a benchmark dataset dedicated to nuclei segmentation has played a crucial role in this success. The 2018 Kaggle Data Science Bowl dataset (Caicedo et al., 2019), hosted as part of the Broad Bioimage Benchmark Collection, was assembled to faithfully reflect the variability of nuclei appearance and 2D image types in bioimaging. This large dataset was designed to challenge the generalization capabilities of segmentation methods across these variations and has established itself as a precious resource to objectively rank and comparatively assess algorithm performances. An equivalent 3D or 3D+time benchmark dataset is yet to be assembled and would be highly valuable to developmental biology image datasets, which are often volumetric and include a temporal component.

Cell membrane segmentation poses a more complex challenge than nuclei. Cells can take on varying morphologies, ranging from

highly-stereotyped shapes to widely-varying sizes and contour roughness. DL models trained on a single dataset therefore often fail to infer accurately on different images. The true limits of the generalization capabilities of an algorithm is furthermore hard to assess in the absence of an established benchmarking dataset dedicated to whole-cell segmentation. Cellpose (Stringer et al., 2021) takes on the generalization challenge, relying on a large custom training set of microscopy images featuring cells with a wide range of diverse morphologies. This method relies on a U-net model predicting the directionality of spatial gradients in the input images and can process both 2D and 3D data. As a result, Cellpose is generalist enough to segment cells with very different morphologies and has been extensively reused (Young et al., 2021; Henninger et al., 2021). In addition, Cellpose is periodically re-trained with user-submitted data to continuously improve its performances (<https://cellpose.readthedocs.io>). Finally, many types of biological questions require organelle segmentation. Manually segmenting organelles from 3D scanning electron microscope (SEM) images is highly time consuming, with annotating a single cell estimated to take  $\sim 60$  years (Heinrich et al., 2020 preprint) (Box 3). Here, DL has been transformative as well, making it possible to automate the segmentation and classification of a wide range of cellular structures.

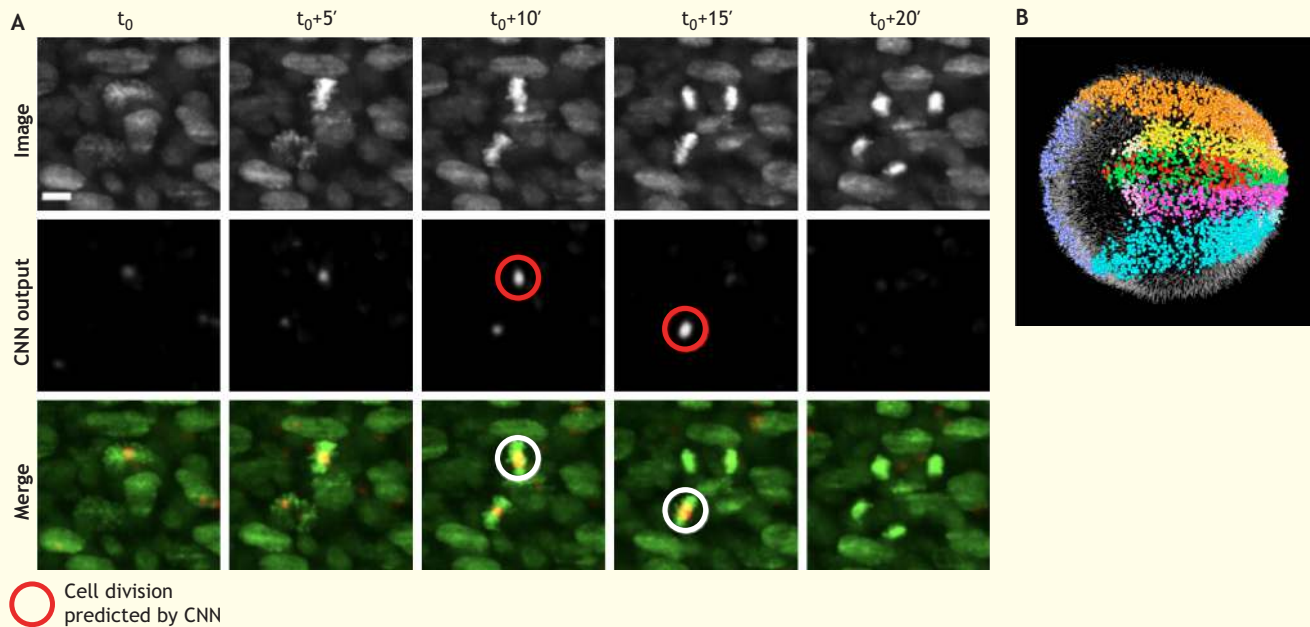
### Image quantification

Once objects have been detected in individual images, the subsequent step is their quantification. Quantification can be about the number of objects (counting), their type (categorization), their shape (morphometry) or their dynamics (tracking), among many others.

Categorization can either be done holistically for an entire object (e.g. wild-type versus mutant), or by looking at a specific aspect of an object (e.g. the shape of internal components). Manual object categorization is both time consuming and has the potential for bias, even when carried out by experts. In addition to speeding up the process, DL-powered image classification can limit annotation variability. Visually assessing embryo quality, for example, is subject to dispute between embryologists (Paternot et al., 2009). Khosravi and colleagues have built a DL classifier of early human embryos quality trained on 'good quality'- and 'poor quality'-labelled embryos that corresponded to the score given by the majority vote of five embryologists (Khosravi et al., 2019). Their model, based on Google's Inception-V1 architecture (Fig. 1C), can achieve a 95.7% agreement with the consensus of the embryologists. In a similar spirit, Yang and colleagues proposed a supervised DL model to assess microscopy image focus quality, providing an absolute quantitative measure of image focus that is independent of the observer (Yang et al., 2018). Eulenberg and colleagues used a different technique to learn discerning features to categorize cell cycle stages and identify cell state trajectories from high-throughput single-cell data (Eulenberg et al., 2017). Eulenberg and colleagues' DL model is trained to classify raw images into a set of discrete classes corresponding to cell cycle stages and, through the process, learn a space of features in which data are continuously organized. When visualized using the tSNE dimensionality-reduction method, feature vectors that describe image data that are temporally close in their cell cycle progression are also close in feature space. A similar strategy has been used in a medical context to classify blood cell health and avoid human bias (Doan et al., 2020). DL-based algorithms have also led to improved detection of biological events, such as cell division in the developing mouse embryo (McDole et al., 2018) (see Box 4). However, non-DL-based ML approaches do still offer a competitive alternative to DL; for example, in the automated identification of cell identities (Hailstone et al., 2020). In this example, classical ML

#### Box 4. Case study: *in toto* imaging and reconstruction of the early mouse embryo

Early development is a highly dynamic process whereby there are large changes in embryo size, shape and optical properties. Capturing the movement of cells inside the embryo and tracking cell divisions to form cell lineage maps is therefore a significant challenge, both experimentally as well as computationally. McDole and colleagues detected cell divisions using a 10-layer 4D CNN that predicts whether each voxel (see Glossary, Box 1) includes a cell division (McDole et al., 2018). The deep learning model is able to identify twice as many cell divisions as a human annotator, thus greatly increasing accuracy in addition to providing automation (A). The model has been trained on 11 image volumes, in which nuclei of both non-dividing and dividing cells have been annotated, as well as 2083 annotated divisions from the entire time series. In addition, an *in toto* picture of the entire early embryo as it grows over 250 $\times$  in volume is achieved by coupling custom adaptive light sheet microscopy with cell tracking. Tracking to retrieve cell fate maps is performed using a Bayesian framework with Gaussian mixture models and statistical vector flow analysis (B). Image from McDole et al. (2018). Scale bar: 10  $\mu$ m.



techniques trained with smaller amounts of data and requiring less computational power than DL-based ones are shown to obtain comparatively good results.

Although spatial tracking has been vastly studied in computer vision applications, biological objects present unique challenges. In addition to moving in and out of the field of view, cells divide, merge and can alter their appearance dramatically. Moen and colleagues proposed a dedicated supervised DL approach to identify matching pairs of cells in subsequent video frames by incorporating information from surrounding frames (Moen et al., 2019). The DL model thus generates a cost matrix for all possible assignments of objects in subsequent frames. The optimal tracking solution is retrieved with the Hungarian algorithm, a classical combinatorial assignment algorithm. The full pipeline (deployed at [deepcell.org](http://deepcell.org)) can thus automate tracking across entire populations of cells. The ability to track individual cells and follow their state, as well as that of their progeny, enables lineage reconstruction, a task that can rapidly become manually intractable and thus greatly benefits from DL (Lugagne et al., 2020; Cao et al., 2020). One obstacle to lineage tracing is the preparation of high-quality training sets of tracks that follow cells often over long periods of time. To address this, the DL-based lineage-tracing method ELEPHANT (Sugawara et al., 2021 preprint) incorporates annotation and proofreading in its user interface to reduce the need for time-consuming curated annotations. ELEPHANT can be trained on a large dataset in which only  $\sim 2\%$  of the data are manually annotated. The model then infers on the remaining data and its predictions are validated by

the user. Other methods, such as 3DeeCellTracker (Wen et al., 2021), rely on simulations to build large training sets with less need for human intervention.

Although cell tracking typically consists of following a single point, such as the centre of mass of a cell over time, tracking for behavioural studies requires multiple landmarks on the organism of interest. Adding landmarks to follow the movement of points on the body with respect to others is cumbersome, time consuming and not always an option. As a remedy, DeepLabCut (Mathis et al., 2018) exploits DL to automatically track points on diverse organisms from a few manual annotations. The DL model is trained on manually-annotated labels capturing striking points (e.g. left/right ear or individual digits) and learns to identify these labels on new image data without the need for added markers. DeepLabCut exploits transfer learning (see Glossary, Box 1) to achieve high accuracy tracking with small training sets of  $\sim 200$  images. Other markerless tracking algorithms with a focus on several animals have been proposed to track social interactions [SLEAP (Pereira et al., 2020 preprint); id tracker ai (Romero-Ferrero et al., 2019)] or animal posture [DeepPoseKit (Graving et al., 2019)].

#### Resources and tools

The bioimage analysis community has developed a strong culture of user-friendly open-source tool developments since its early days (Carpenter et al., 2012; Schneider et al., 2012; Eliceiri et al., 2012). Several well-established platforms, such as Weka (Arganda-Carreras et al., 2017) and ilastik (Berg et al., 2019),

provide a user-friendly interface to use conventional non-DL-based ML approaches in bioimage analysis problems. Following the rising popularity of DL in the past decade and the difficulty for non-programmers to adopt it, some of these platforms have been further developed to include DL-based algorithms, and new ones have emerged. Here, we provide an overview of selected available open-source resources developed by the bioimage analysis community that can be used to get started with DL (Table 1). Lucas and colleagues also provide an excellent in-depth discussion of open-source resources for bioimage segmentation with DL for readers wanting to explore this topic further (Lucas et al., 2021).

Several resources offer a direct point of entry into DL for bioimaging without the need for any coding expertise. The most accessible DL use-case consists of exploiting pre-trained models. This essentially means using a model that has been already trained on another image dataset to make predictions on one's own data without additional training, and requires little to no parameter tuning. Popular standalone platforms that pre-existed the DL era, such as CellProfiler (McQuin et al., 2018) and ilastik, now offer pre-trained U-net models for a variety of tasks. Both are available for all major operating systems and have their own dedicated general user interface (GUI). As these two software packages are extremely well supported and documented, and because they contain a wealth of useful methods for image analysis in addition to DL-based ones, they probably are the lowest-entry-cost options to start experimenting with DL. Several popular pre-trained models, such as the original U-net implementation (Ronneberger et al., 2015), StarDist (Schmidt et al., 2018) and Cellpose (Stringer et al., 2021), have been made available as plug-ins for ImageJ (Schindelin et al.,

2012) and napari (napari contributors, 2019). The DeepImageJ plug-in (Gómez-de-Mariscal et al., 2019 preprint), in particular, offers a unifying interface to reuse pre-trained models – a large variety of models for various image restoration and segmentation tasks is already available through it and the list is likely to grow. Searching for model implementations and pre-trained weights may be a daunting task. The reusability of most DL-based methods is significantly impacted by their custom nature, often resulting in code that is hard to distribute. The Bioimage Model Zoo (bioimage.io) is a community-driven initiative aiming to address this issue by centralizing and facilitating the reuse of published models, in DeepImageJ among others. Although it is still under development it is evolving quickly, and the Bioimage Model Zoo is poised to become a reference resource for DL models dedicated to bioimage analysis.

Although pre-trained models are a good starting point, their use may not suffice to obtain good results, or worse, it may cause serious underperformance and poses a risk of generating artefacts due to dataset shift (discussed below). A more reliable, yet more involved, strategy consists of training an existing model with one's own data, either from scratch or by fine-tuning a pre-trained model, which is a particular feature of transfer learning. Although several recent tools facilitate the annotation of 2D and 3D image datasets (Hollandi et al., 2020a,b; Borland et al., 2021), the process of manually producing high-quality ground-truth annotations for training remains tedious, in particular for 3D+time datasets. The web-based platform ImJoy (Ouyang et al., 2019) hosts a large collection of plug-ins that provide interactive interfaces to generate ground-truth annotations on multi-dimensional images and train various DL algorithms. From ImJoy, algorithms can be run directly

**Table 1. Open-source tools for deep learning in bioimaging**

Tool	Reference	URL	Type	Use-case	Prerequisites
<b>CellProfiler</b>	McQuin et al., 2018	cellprofiler.org	GUI-based standalone general bioimage analysis software	Inference with pre-trained models and model training from existing ground truth for classification	None
<b>ilastik</b>	Berg et al., 2019	ilastik.org	GUI-based standalone general bioimage analysis software	Inference with pre-trained models (fully supported) and model training from scratch (debug mode) for segmentation	None
<b>DeepImageJ</b>	Gómez-de-Mariscal et al., 2019 preprint	deepimagej.github.io	ImageJ/Fiji plug-in enabling the use of pre-trained DL-based bioimage analysis algorithms	Inference with pre-trained models for various tasks	Experience with Fiji/ImageJ
<b>ImJoy</b>	Ouyang et al., 2019	imjoy.io	Online computing platform for deploying DL bioimage analysis pipelines	Inference with pre-trained models and model training from existing ground truth for various tasks	None
<b>ZeroCostDL4Mic</b>	von Chamier et al., 2021	github.com/HenriquesLab/ZeroCostDL4Mic/wiki	Google Colab Python notebooks implementing DL-based bioimage analysis algorithms	Inference with pre-trained models and model training from existing ground truth for various tasks	None
<b>Bioimage Model Zoo</b>	N.A.	bioimage.io	Community-driven online repository facilitating reuse and access to pre-trained DL models	Retrieve models architecture for various tasks, along with pre-trained weights	Dependent on the considered model
<b>CSBDeep</b>	N.A.	csbdeep.bioimagecomputing.com	Python DL toolbox for general bioimage analysis	Model training from existing ground truth for image restoration	Experience with Python

DL, deep learning; GUI, general user interface; N.A., not applicable.

in the browser on a local host, remotely, or on a cloud server. The ZeroCostDL4Mic (von Chamier et al., 2021) toolbox also provides an excellent user-friendly solution for training DL models through guided notebooks, requiring no programming knowledge.

For the experienced programmer wishing to go further, many DL models are freely available as Python libraries. However, the level of user support may vary dramatically and can range from undocumented code on GitHub repositories to dedicated webpages with thorough user manuals and example data. CSBDeep (csbdeep.bioimagecomputing.com) offers one of the best examples of one such well-maintained resource, providing a wealth of documentation facilitating the reuse and adoption of DL models.

### Going further with deep learning

DL offers a plethora of exciting possibilities that go far beyond automating classical bioimage analysis tasks. Here, we discuss some DL avenues that look promising in the analysis of quantitative biological data beyond images and for modelling.

#### Transfer learning

DL models usually require large amounts of data for training, which requires significant annotation efforts. In many cases, such ground truth sets cannot be easily generated, either because of technical limitations (e.g. in the context of image restoration) or owing to the sheer amount of manual curation required. Transfer learning thus holds huge potential to enable the creation of all-rounder deep NN (DNN) that can then be fine-tuned to many specific applications relying on a few annotated data only.

In the context of image restoration, Jin and colleagues have illustrated the benefit of transfer learning in a DL pipeline by improving structured illumination microscopy image quality at low light levels (Jin et al., 2020). The DNN trained with transfer learning have been shown to perform equally well as their equivalents trained from scratch, but require 90% fewer ground truth samples and 10× fewer iterations to converge. Strategies aimed at reducing the number of training samples are particularly relevant to developmental biology experiments, which often rely on costly protocols to harvest few numbers of samples and results in scarce datasets. In spite of encouraging demonstrations of the benefits of transfer learning, several questions around trust in DL-generated results remain to be answered. Pre-trained NN must be used with caution, because they may be subject to dataset shift when dealing with data that are too dissimilar to what they have been trained on. Dataset shift refers to the general problem of how information can be transferred from a variety of different previous environments to help with learning, inference and prediction in a new environment (Storkey, 2008). Understanding dataset shift thus translates to characterizing how the information held in several closely related environments (e.g. data collected in other laboratories) can help with prediction in new environments. Dataset shift in bioimaging can have several origins, such as batch effects, different sample preparation protocols or different imaging systems. Different mitigation strategies should be used to address dataset shift depending on its nature (Quiñonero-Candela et al., 2008), and the topic is being actively investigated. However, because dataset shift is a complex phenomenon that may be hard to fully characterize, one must exercise the utmost caution when using DL models outside of their training domain. When relying on pre-trained models, practitioners hold responsibility to understand the strategy and type of data that have been used to train the NN, identify the type of shift they may be facing, and remain aware of the existing mitigation

strategies or lack thereof. In situations where the discrepancy between the data to be processed and those used for the initial training of the NN cannot be clearly characterized, preference should be given to conventional ML and non-DL image processing techniques.

#### Style transfer

Style transfer (see Glossary, Box 1) has been famously applied in the context of artistic illustrations, allowing any photographs to be turned into van Gogh paintings (Gatys et al., 2016). Similarly, it can be used to learn the image style of different microscopy modalities, with numerous applications from synthetic data generation to image enhancement. For example, this strategy has been successfully employed to adapt a nucleus segmentation task to unseen microscopy image types (Hollandi et al., 2020a,b). In this example, style transfer is used to synthesize different types of artificial microscopy images from a single training set of ground-truth labelled images. The style learned from unlabelled image samples, which are drawn from a different distribution than the training samples, is transferred to the labelled training samples. Thus, for the same set of labels, new images with realistic-looking texture, colouration and background pattern elements can be generated. This approach outperforms fine-tuning the network with a small set of additional labelled data and, in contrast, does not require any extra labelling effort. The approach has been shown to perform well on various types of microscopy images, including haematoxylin and eosin histological staining and fluorescence. Although this work focuses on nucleus segmentation, the possibility to augment difficult-to-obtain data with style transfer has enormous potential in many bioimage analysis applications beyond this specific problem.

Although a lot of the enthusiasm for style transfer can be attributed to its potential as a data augmentation technique (see Glossary, Box 1), it is equally stimulating to envision it as a computational alternative to image acquisition. Style transfer can be exploited for synthetic image generation, for example to produce microscopy images from different modalities, such as inferring phase-contrast microscopy images from differential interference contrast images and vice versa (Han and Yin, 2017). This kind of approach is appealing for many reasons; for example, it reduces the equipment needed and reduces image acquisition time. However, current style transfer methods are oblivious to the physical properties of the specimens being imaged. Despite the visually realistic and convincing images generated for relatively simple specimens, further investigations are needed to assess how such a method would perform for more complex specimens. One study perfectly summarizes the underlying dilemma: ‘the more we rely on DL the less confident we can be’ (Hoffman et al., 2021). When doing style transfer, one specifically trains networks to lie plausibly. As such, the resulting DNN will be able to turn any object into a realistic-looking biological structure and will do so, regardless of the input; happily turning cat pictures into plausible microscopy images of cells, for example. Subsequent work is required to define confidence and uncertainty metrics for style transfer to be used in the context of scientific discovery. When exploiting style transfer strategies, biologists should make sure to follow recent developments in these directions and, most importantly, remain fully aware that a consensus of best practices on this matter has yet to be reached.

#### Natural texture generation

Related to style transfer, natural texture synthesis is an emerging and less studied research direction with broad implications in bioimage



analysis. Texture synthesis is a well-studied problem in computer graphics where, broadly, one wants to algorithmically generate a larger image from smaller parts by exploiting geometric regularly occurring motifs (Niklasson et al., 2021). Whether concerning histological images (Ash et al., 2021), early developmental patterning (McDole et al., 2018) or man-made textures (such as textiles), local interactions between smaller parts (cells, morphogens or threads, respectively) can give rise to larger, emerging structures. Rather than encoding the large dimensional spaces of pixels and colours, recent DL approaches aim to describe these images in a generative way, through feed-forward stochastic processes (Reinke et al., 2020; Pathak et al., 2019). The most recent approaches to generative modelling of texture synthesis are systems of partial differential equations aimed at modelling reaction-diffusion equations (Chan, 2020), cellular automata (Niklasson et al., 2021; Mordvintsev et al., 2020) and oscillator-based multi-agent particle systems (Ricci et al., 2021 preprint). Among these, the neural cellular automata takes its inspiration from reaction-diffusion models of morphogenesis by modelling a system of locally communicating cells that evolve and self-organize to form a desired input pattern. In the dynamic process of learning a pattern, the cells learn local rules that exhibit global properties. However, these rules are abstract and not readily interpretable biologically. Mapping them to explicit gene modules involved in signalling pathways or intra cellular communication is essential to making these models useful to the developmental biological community.

Alternatively, non-black box approaches (Zhao et al., 2021) aim to identify the physical properties that can be accurately inferred from full images. Once parameters and physical properties are inferred, we can naturally wonder whether these 2D abstractions generalize, not only to 3D settings (Sudhakaran et al., 2021 preprint), but also to systems reminiscent of cellular self-organization (Gilpin, 2019). In this context, an experimental counterpart is provided by studies in which the problem of pattern formation is addressed synthetically by creating morphogen systems that yield patterns reminiscent of those observed *in vivo* (Toda et al., 2020; Zhang et al., 2017).

#### Lessons from statistical genetics: the need for proper null models

When trying to establish whether a molecular event, such as change in gene expression, affects cells in an observable manner, one faces the statistical challenge of assessing significance. In past decades, statistical genetics has developed an arsenal of tools for assessing statistical significance in high dimensional problems, where hypothesis correction is essential for distinguishing between true correlation and spurious events (Barber and Candès, 2019; Stephens, 2017). Despite focusing on prediction, DL architectures for computer vision do offer many recipes for probing the interpretability of a classifier [e.g. saliency maps (Adebayo et al., 2018)]. Additional approaches aim to identify meaningful perturbations in training data that can lead to misclassifications (Fong and Vedaldi, 2017), whereas others detect subparts or prototypical parts of an image that could impact classification (Chen et al., 2019 preprint). A good resource and point of access into this vast community is the Computer Vision and Pattern Recognition Conference (CVPR) series of workshops ([https://interpretablevision.github.io/index\\_cvpr2020.html](https://interpretablevision.github.io/index_cvpr2020.html)) on interpretable ML. However, all these approaches require a large amount of training data, often unavailable in biological settings, as previously discussed.

In low-data regimes, one falls back into statistician territory and typically relies on the existence of a properly chosen null model to evaluate significance (Schäfer and Strimmer, 2005). When testing

for the significance of an effect variable (e.g. gene expression change) on a quantity of interest (e.g. image feature), null models represent a way to formalize how data might look in the absence of the effect. By comparing statistical estimates from data to statistical estimates generated through appropriate null models, one can assess whether an effect is spurious or real. In simple regression models, for example, null datasets are often generated through permutation tests where, as the name suggests, data (input and outputs) are shuffled and their correlation contrasted with correlation from unshuffled data. However, permutation is not always an appropriate baseline, as illustrated in cases where data is not independently identically distributed (Dumitrascu et al., 2019; Elsayed and Cunningham, 2017). To mediate these issues, approaches for creating ‘fake data’ that can represent null models have been proposed with false discovery correction in mind (Barber and Candès, 2019). Although fake data is easier to generate when dealing with Gaussian variables, it is not clear what a proper, highly-structured, fake image would look like. For example, randomly shuffling the pixels of an image will just create noise. Despite the inherent difficulties, designing appropriate visual counterfactual and null models for bioimage data is essential in augmenting studies that aim to relate genomics with morphology, and is a promising area of research.

#### Multimodal learning

Bioimages are typically collected across multiple conditions spanning, for example, different replicates, cell types and time scales, as well as various perturbations, such as mechanical, genetic or biochemical. Layering in additional molecular information, such as gene expression, cell lineage or chromatin accessibility (Dries et al., 2021) from high-throughput sequencing experiments, brings the challenge of integrating data from multiple modalities, and the challenge of quantifying how predictive of one another the different modalities can be (Pratapa et al., 2021). Common tasks in multi-modal transfer learning particularly relevant to bioimage analysis include integrating and visualizing data from different sources (data fusion), translating between different modes (transfer) and aligning data collected across multiple modes (alignment). Data fusion is the challenge of aggregating modalities in a manner that improves prediction, especially when data might be missing or noisy. Data fusion has been thoroughly explored in the context of single-cell batch correction, where computational methods allow the integration of datasets of the same kind of modality, namely single-cell gene expression data (Argelaguet et al., 2021). However, a similar problem can be framed for microscopy data collected using different imaging modalities or in different laboratories.

Data from fundamentally different modalities, such as image data accompanied by single-cell gene expression or chromatin packing (Clark et al., 2018; Gundersen et al., 2019), poses additional challenges. Integrating them can help in understanding whether changes in gene expression have a direct consequence not only on how individual cells look, but also on how they interact with their neighbours. However, the resolution of these different data types may be significantly different, making a direct correspondence between modalities hard to achieve (Vergara et al., 2021). In these situations, modality alignment becomes paramount (Lopez et al., 2019). A recent work integrating single-cell RNA-sequencing data and single-cell nuclear-imaging of naive T-cells (Yang et al., 2021) has shown that DL representation of images contain signals predictive of true fold change of gene expression between different classes of cells.

## Conclusion

The interaction between DL and developmental biology is in its nascent stages, but will continue to grow. As a result, it is of increasing importance for biologists to become aware of applications in which DL can be exploited, but also to discern its limitations and potential pitfalls when analysing and interpreting biological data. Although undeniably powerful in some settings, the use of DL comes at a cost in resources (e.g. large amount of labelled data required for training, high computational demands) and incurs risks (e.g. the black-box nature of the algorithms and dataset shift). For these reasons, conventional ML and non-DL-based image processing methods should always be tried first and chosen whenever possible. As a community, cell and developmental biologists can make a conscious effort to support a scientifically sound and informed use of DL through the standardization of data acquisition, archival protocols, annotation conventions and metadata describing image processing pipelines. There is an opportunity for the developmental biology community to centralize published data and analysis pipelines in an open source and curated manner, to promote a healthy use of DL in scientific discovery. As DL is rapidly pushing the limits of what is achievable in science, it calls on us to reflect on our common goals and re-evaluate how we share data and collaborate.

## Acknowledgements

A.H. thanks J.-B. Lugagne and N. Lawrence for insightful discussions on DL-based image analysis and acknowledges the support of Benjamin Simons and his lab. H.G.Y. thanks Juan Caicedo for his generous DL mentorship. The authors thank Martin Weigert for providing helpful details on CARE.

## Competing interests

The authors declare no competing or financial interests.

## Funding

A.H. gratefully acknowledges the support of the Wellcome Trust through a Junior Interdisciplinary Research Fellowship (098357/Z/12/Z) and of the University of Cambridge through a Herchel Smith Postdoctoral Research Fellowship. He also acknowledges the support of the core funding to the Wellcome Trust/CRUK Gurdon Institute (203144/Z/16/Z and C6946/A24843). H.G.Y. gratefully acknowledges that the research reported in this publication was partly supported by the National Institute of General Medical Sciences under award number K99GM136915. B.D. acknowledges support from the Accelerate Programme for Scientific Discovery. V.U. acknowledges support from European Molecular Biology Laboratory core funding. Deposited in PMC for immediate release.

## References

- Adebayo, J., Gilmer, J., Muelly, M., Goodfellow, I., Hardt, M. and Kim, B. (2018). Sanity checks for saliency maps. In *Proceedings of NeurIPS 2018*.
- Arganda-Carreras, I., Kaynig, V., Rueden, C., Eliceiri, K. W., Schindelin, J., Cardona, A. and Sebastian Seung, H. (2017). Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification. *Bioinformatics* **33**, 2424–2426. doi:10.1093/bioinformatics/btx180
- Argelaguet, R., Cuomo, A. S. E., Stegle, O. and Marioni, J. C. (2021). Computational principles and challenges in single-cell data integration. *Nat. Biotechnol.* [Epub ahead of print]. doi:10.1038/s41587-021-00895-7
- Ash, J. T., Darnell, G., Munro, D. and Engelhardt, B. E. (2021). Joint analysis of expression levels and histological images identifies genes associated with tissue morphology. *Nat. Commun.* **12**, 1609. doi:10.1038/s41467-021-21727-x
- Barber, R. F. and Candès, E. J. (2019). A knockoff filter for high-dimensional selective inference. *Annals of Statistics* **47**, 2504–2537. doi:10.1214/18-AOS1755
- Batson, J. and Royer, L. (2019). Noise2Self: blind denoising by self-supervision. In *Proceedings of ICML 2019*, pp. 524–533.
- Berg, S., Kutra, D., Kroeger, T., Straehle, C. N., Kausler, B. X., Haubold, C., Schiegg, M., Ales, J., Beier, T., Rudy, M. et al. (2019). Ilastik: interactive machine learning for (bio)image analysis. *Nat. Methods* **16**, 1226–1232. doi:10.1038/s41592-019-0582-9
- Boka, A. P., Mukherjee, A. and Mir, M. (2021). Single-molecule tracking technologies for quantifying the dynamics of gene regulation in cells, tissue and embryos. *Development* **148**, dev199744. doi:10.1242/dev.199744
- Borland, D., McCormick, C. M., Patel, N. K., Krupa, O., Mory, J. T., Beltran, A. A., Farah, T. M., Escobar-Tomliénovich, C. F., Olson, S. S., Kim, M. et al. (2021). Segmentor: a tool for manual refinement of 3D microscopy annotations. *BMC Bioinformatics* **22**, 260. doi:10.1186/s12859-021-04202-8
- Born, M., Wolf, E., Bhatia, A. B., Clemmow, P. C., Gabor, D., Stokes, A. R., Taylor, A. M., Wayman, P. A. and Wilcock, W. L. (1999). *Principles of Optics*. Cambridge University Press.
- Boyd, N., Jonas, E., Babcock, H. and Recht, B. (2018). DeepLoco: Fast 3D localization microscopy using neural networks. *bioRxiv*. doi:10.1101/267096
- Caicedo, J. C., Goodman, A., Karhohs, K. W., Cimini, B. A., Ackerman, J., Haghghi, M., Heng, C., Becker, T., Doan, M., Mcquin, C. et al. (2019). Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl. *Nat. Methods* **16**, 1247–1253. doi:10.1038/s41592-019-0612-7
- Cao, J., Guan, G., Ho, V. W. S., Wong, M.-K., Chan, L.-Y., Tang, C., Zhao, Z. and Yan, H. (2020). Establishment of a morphological atlas of the *Caenorhabditis elegans* embryo using deep-learning-based 4D segmentation. *Nat. Commun.* **11**, 6254. doi:10.1038/s41467-020-19863-x
- Carpenter, A. E., Kametsky, L. and Eliceiri, K. W. (2012). A call for bioimaging software usability. *Nat. Methods* **9**, 666–670. doi:10.1038/nmeth.2073
- Chan, B. W.-C. (2020). Lenia and Expanded Universe. In *ALIFE 2020: The 2020 Conference on Artificial Life* pp. 221–229.
- Chen, C., Li, O., Tao, C., Barnett, A. J., Su, J. and Rudin, C. (2019). This looks like that: deep learning for interpretable image recognition. In *Proceedings of NeurIPS 2019*.
- Chidester, B., Ton, T.-V., Tran, M.-T., Ma, J. and Do, M. N. (2019). Enhanced rotation-equivariant U-net for nuclear segmentation. In *Proceedings of CVPRW 2019*. doi:10.1109/cvprw.2019.00143
- Clark, S. J., Argelaguet, R., Kapourani, C.-A., Stubbs, T. M., Lee, H. J., Alda-Catalinas, C., Krueger, F., Sanguinetti, G., Kelsey, G., Marioni, J. C. et al. (2018). scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nat. Commun.* **9**, 781. doi:10.1038/s41467-018-03149-4
- Doan, M., Sebastian, J. A., Caicedo, J. C., Siegert, S., Roch, A., Turner, T. R., Mykhailova, O., Pinto, R. N., Mcquin, C., Goodman, A. et al. (2020). Objective assessment of stored blood quality by deep learning. *Proc. Natl. Acad. Sci. USA* **117**, 21381–21390. doi:10.1073/pnas.2001227117
- Dries, R., Zhu, Q., Dong, R., Eng, C.-H. L., Li, H., Liu, K., Fu, Y., Zhao, T., Sarkar, A., Bao, F. et al. (2021). Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol.* **22**, 78. doi:10.1186/s13059-021-02286-2
- Dumitrascu, B., Darnell, G., Ayroles, J. and Engelhardt, B. E. (2019). Statistical tests for detecting variance effects in quantitative trait studies. *Bioinformatics* **35**, 200–210. doi:10.1093/bioinformatics/bty565
- Eliceiri, K. W., Berthold, M. R., Goldberg, I. G., Ibáñez, L., Manjunath, B. S., Martone, M. E., Murphy, R. F., Peng, H., Plant, A. L., Roysam, B. et al. (2012). Biological imaging software tools. *Nat. Methods* **9**, 697–710. doi:10.1038/nmeth.2084
- Elsayed, G. F. and Cunningham, J. P. (2017). Structure in neural population recordings: an expected byproduct of simpler phenomena? *Nat. Neurosci.* **20**, 1310–1318. doi:10.1038/nn.4617
- Esteve, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., Liu, Y., Topol, E., Dean, J. and Socher, R. (2021). Deep learning-enabled medical computer vision. *Npj Digital Medicine* **4**, 5. doi:10.1038/s41746-020-00376-2
- Eulenberg, P., Köhler, N., Blasi, T., Filby, A., Carpenter, A. E., Rees, P., Theis, F. J. and Wolf, F. A. (2017). Reconstructing cell cycle and disease progression using deep learning. *Nat. Commun.* **8**, 463. doi:10.1038/s41467-017-00623-3
- Falk, T., Mai, D., Bensch, R., Çiçek, Ö., Abdulkadir, A., Marrakchi, Y., Böhm, A., Deubner, J., Jäckel, Z., Seiwald, K. et al. (2018). U-net: deep learning for cell counting, detection, and morphometry. *Nat. Methods* **16**, 67–70. doi:10.1038/s41592-018-0261-2
- Fong, R. C. and Vedaldi, A. (2017). Interpretable explanations of black boxes by meaningful perturbation. In *Proceedings of ICCV*, pp. 3429–3437. doi:10.1109/ICCV.2017.371
- Frankle, J. and Carbin, M. (2018). The lottery ticket hypothesis: Finding sparse, trainable neural networks. In *Proceedings of ICLR 2019*.
- Gatys, L. A., Ecker, A. S. and Bethge, M. (2016). Image style transfer using convolutional neural networks. In *Proceedings of ICCV 2016*, pp. 2414–2423. doi:10.1109/cvpr.2016.265
- Gibson, S. F. and Lanni, F. (1989). Diffraction by a circular aperture as a model for three-dimensional optical microscopy. *J. Opt. Soc. Am. A* **6**, 1357. doi:10.1364/JOSAA.6.001357
- Gibson, S. F. and Lanni, F. (1992). Experimental test of an analytical model of aberration in an oil-immersion objective lens used in three-dimensional light microscopy. *J. Opt. Soc. Am. A* **9**, 154. doi:10.1364/JOSAA.9.000154
- Gilpin, W. (2019). Cellular automata as convolutional neural networks. *Phys. Rev. E* **100**, 032402. doi:10.1103/PhysRevE.100.032402
- Gómez-De-Mariscal, E., García-López-De-Haro, C., Ouyang, W., Donati, L., Lundberg, E., Unser, M., Muñoz-Barrutia, A. and Sage, D. (2019). DeepImageJ: a user-friendly environment to run deep learning models in ImageJ. *bioRxiv*. doi:10.1101/799270
- Goodfellow, I., Bengio, Y. and Courville, A. (2016). *Deep Learning*. Cambridge, USA: MIT Press.

- Graving, J. M., Chae, D., Naik, H., Li, L., Koger, B., Costelloe, B. R. and Couzin, I. D. (2019). DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. *eLife* **8**, e47994. doi:10.7554/eLife.47994
- Gundersen, G., Dumitrascu, B., Ash, J. T. and Engelhardt, B. E. (2019). End-to-end training of deep probabilistic CCA on paired biomedical observations. In *Proceedings of PMLR*, pp. 945-955.
- Gupta, A., Harrison, P. J., Wieslander, H., Pielawski, N., Kartasalo, K., Partel, G., Solorzano, L., Suveer, A., Klemm, A. H., Spjuth, O. et al. (2018). Deep learning in image cytometry: a review. *Cytometry Part A* **95**, 366-380. doi:10.1002/cyto.a.23701
- Hailstone, M., Waithe, D., Samuels, T. J., Yang, L., Costello, I., Arava, Y., Robertson, E., Parton, R. M. and Davis, I. (2020). CytoCensus, mapping cell identity and division in tissues and organs using machine learning. *eLife* **9**, e51085. doi:10.7554/eLife.51085
- Han, L. and Yin, Z. (2017). Transferring Microscopy Image Modalities with Conditional Generative Adversarial Networks. In *Proceedings of CVPRW 2017*, pp. 99-107. IEEE.
- He, K., Gkioxari, G., Dollr, P. and Girshick, R. B. (2017). Mask R-CNN. *Proceedings of ICCV 2017*, pp. 2980-2988. doi:10.1109/ICCV.2017.322
- Heinrich, L., Bennett, D., Ackerman, D., Park, W., Bogovic, J., Eckstein, N., Petrunco, A., Clements, J., Xu, C. S., Funke, J. et al. (2020). Automatic whole cell organelle segmentation in volumetric electron microscopy. *bioRxiv*. doi:10.1101/2020.11.14.382143
- Hoffman, D. P., Slavitt, I. and Fitzpatrick, C. A. (2021). The promise and peril of deep learning in microscopy. *Nat. Methods* **18**, 131-132. doi:10.1038/s41592-020-01035-w
- Henninger, J. E., Oksuz, O., Shrinivas, K., Sagi, I., Leroy, G., Zheng, M. M., Andrews, J. O., Zamudio, A. V., Lazaris, C., Hannett, N. M. et al. (2021). RNA-Mediated feedback control of transcriptional condensates. *Cell* **184**, 207-225.e24. doi:10.1016/j.cell.2020.11.030
- Hollandi, R., Diódsi, Á., Hollandi, G., Moshkov, N. and Horváth, P. (2020a). AnnotatorJ: an ImageJ plugin to ease hand annotation of cellular compartments. *Mol. Biol. Cell* **31**, 2179-2186. doi:10.1091/mbc.E20-02-0156
- Hollandi, R., Szkalitsy, A., Toth, T., Tasnadi, E., Molnar, C., Mathe, B., Grexa, I., Molnar, J., Balind, A., Gorbé, M. et al. (2020b). nucleAlzer: a parameter-free deep learning framework for nucleus segmentation using image style transfer. *Cell Syst.* **10**, 453-458.e6. doi:10.1016/j.cels.2020.04.003
- Jin, L., Liu, B., Zhao, F., Hahn, S., Dong, B., Song, R., Elston, T. C., Xu, Y. and Hahn, K. M. (2020). Deep learning enables structured illumination microscopy with low light levels and enhanced speed. *Nat. Commun.* **11**, 1934. doi:10.1038/s41467-020-15784-x
- Khosravi, P., Kazemi, E., Zhan, Q., Malmsten, J. E., Toschi, M., Zisimopoulos, P., Sigaras, A., Lavery, S., Cooper, L. A. D., Hickman, C. et al. (2019). Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization. *npj Digit. Med.* **2**, 21. doi:10.1038/s41746-019-0096-y
- Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Proceedings of NeurIPS 2012*, pp. 1097-1105.
- Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Commun. ACM* **60**, 84-90. doi:10.1145/3065386
- Krull, A., Buchholz, T.-O. and Jug, F. (2019). Noise2Void - learning denoising from single noisy images. In *Proceedings of CVPR 2019*, pp. 2124-2132.
- Lecun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. and Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Comput.* **1**, 541-551. doi:10.1162/neco.1989.1.4.541
- Lecun, Y., Bengio, Y. and Hinton, G. (2015). Deep learning. *Nature* **521**, 436-444. doi:10.1038/nature14539
- Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M. and Aila, T. (2018). Noise2Noise: Learning Image Restoration without Clean Data. In *Proceedings of ICML 2018*, pp. 2965-2974.
- Libbrecht, M. W. and Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nature Review Genetics* **16**, 321-332. doi:10.1038/nrg3920
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A. W. M., Van Ginneken, B. and Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, **42**, 60-88. doi:10.1016/j.media.2017.07.005
- Long, F. (2020). Microscopy cell nuclei segmentation with enhanced U-net. *BMC Bioinformatics* **21**, 8. doi:10.1186/s12859-019-3332-1
- Lopez, R., Nazaret, A., Langevin, M., Samaran, J., Regier, J., Jordan, M. I. and Yosef, N. (2019). A joint model of unpaired data from scRNA-seq and spatial transcriptomics for imputing missing gene expression measurements. In *Proceedings of ICML 2019 Workshop in Computational Biology*.
- Lucas, A. M., Ryder, P. V., Li, B., Cimini, B. A., Eliceiri, K. W. and Carpenter, A. E. (2021). Open-source deep-learning software for bioimage segmentation. *Mol. Biol. Cell* **32**, 823-829. doi:10.1091/mbc.E20-10-0660
- Lugagne, J. B., Lin, H. and Dunlop, M. J. (2020). DeLTA: Automated cell segmentation, tracking, and lineage reconstruction using deep learning. *PLOS Computational Biology* **16**, e1007673. doi:10.1371/journal.pcbi.1007673
- Mandal, S. and Uhlmann, V. (2021). SplineDist: automated cell segmentation with spline curves. In *Proceedings of ISBI 2021*.
- Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W. and Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **21**, 1281-1289. doi:10.1038/s41593-018-0209-y
- Mcculloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **5**, 115-133. doi:10.1007/BF02478259
- McDole, K., Guignard, L., Amat, F., Berger, A., Malandain, G., Royer, L. A., Turaga, S. C., Branson, K. and Keller, P. J. (2018). In toto imaging and reconstruction of post-implantation mouse development at the single-cell level. *Cell* **175**, 859-876.e33. doi:10.1016/j.cell.2018.09.031
- McGinn, J., Hallou, A., Han, S., Krizic, K., Ulyanchenko, S., Iglesias-Bartolome, R., England, F. J., Verstreken, C., Chalut, K. J., Jensen, K. B. et al. (2021). A biomechanical switch regulates the transition towards homeostasis in oesophageal epithelium. *Nat. Cell Biol.* **23**, 511-525. doi:10.1038/s41556-021-00679-w
- McQuin, C., Goodman, A., Chernyshev, V., Kamentsky, L., Cimini, B. A., Karhohs, K. W., Doan, M., Ding, L., Rafelski, S. M., Thirstrup, D. et al. (2018). CellProfiler 3.0: Next-generation image processing for biology. *PLoS Biol.* **16**, e2005970. doi:10.1371/journal.pbio.2005970
- Meijering, E. (2020). A bird's-eye view of deep learning in bioimage analysis. *Comput. Struct. Biotechnol. J.* **18**, 2312-2325. doi:10.1016/j.csbj.2020.08.003
- Moen, E., Bannon, D., Kudo, T., Graf, W., Covert, M. and Van Valen, D. (2019). Deep learning for cellular image analysis. *Nat. Methods* **16**, 1233-1246. doi:10.1038/s41592-019-0403-1
- Mohri, M., Rostamizadeh, A. and Talwalkar, A. (2018). *Foundations of Machine Learning*. Cambridge, USA: MIT Press.
- Mordvintsev, A., Randazzo, E., Niklasson, E. and Levin, M. (2020). Growing neural cellular automata. *Distill* **5**, e23.
- Murphy, K. P. (2012). *Probabilistic Machine Learning: An Introduction*. Cambridge, USA: MIT Press.
- napari contributors (2019). napari: a multi-dimensional image viewer for python. doi:10.5281/zenodo.3555620
- Nehme, F., Freedman, D., Gordon, R., Ferdman, B., Weiss, L. E., Alalouf, O., Naor, T., Orange, R., Michaeli, T. and Shechtman, Y. (2020). DeepSTORM3D: dense 3D localization microscopy and PSF design by deep learning. *Nat. Methods* **17**, 734-740. doi:10.1038/s41592-020-0853-5
- Niklasson, E., Mordvintsev, A., Randazzo, E. and Levin, M. (2021). Self-organising textures. *Distill* **6**, e00027-e00003. doi:10.23915/distill.00027.003
- Ouyang, W., Mueller, F., Hjelmar, M., Lundberg, E. and Zimmer, C. (2019). ImJoy: an open-source computational platform for the deep learning era. *Nat. Methods* **16**, 1199-1200. doi:10.1038/s41592-019-0627-0
- Paternot, G., Devroe, J., Debrock, S., D'hooghe, T. M. and Spiessens, C. (2009). Intra- and inter-observer analysis in the morphological assessment of early-stage embryos. *Reprod. Biol. Endocrinol.* **7**, 105. doi:10.1186/1477-7827-7-105
- Pathak, D., Lu, C., Darrell, T., Isola, P. and Efros, A. A. (2019). Learning to control self-assembling morphologies: a study of generalization via modularity. In *Proceedings of NeurIPS 2019*.
- Pereira, T. D., Tabris, N., Li, J., Ravindranath, S., Papadoyannis, E. S., Wang, Z. Y., Turner, D. M., McKenzie-Smith, G., Kocher, S. D., Falkner, A. L. et al. (2020). SLEAP: Multi-animal pose tracking. *bioRxiv*, 2020.2008.2031.276246.
- Pratapa, A., Doron, M. and Caicedo, J. C. (2021). Image-based cell phenotyping with deep learning. *Curr. Opin. Chem. Biol.* **65**, 9-17. doi:10.1016/j.cbpa.2021.04.001
- Quiñero-Candela, J., Sugiyama, M., Schwaighofer, A. and Lawrence, N. D. (2008). *Dataset Shift in Machine Learning*. The MIT Press.
- Reinke, C., Etcheverry, M. and Oudeyer, P.-Y. (2020). Intrinsically Motivated Discovery of Diverse Patterns in Self-Organizing Systems. In *Proceedings of ICLR 2020*.
- Ricci, M., Jung, M., Zhang, Y., Chalvidal, M., Soni, A. and Serre, T. (2021). KuraNet: systems of coupled oscillators that learn to synchronize. *arXiv:2105.02838*.
- Romero-Ferrero, F., Bergomi, M. G., Hinz, R. C., Heras, F. J. H. and De Polavieja, G. G. (2019). idtracker.ai: tracking all individuals in small or large collectives of unmarked animals. *Nat. Methods* **16**, 179-182. doi:10.1038/s41592-018-0295-5
- Ronneberger, O., Fischer, P. and Brox, T. (2015). U-net: Convolutional Networks for Biomedical Image Segmentation. In *Proceedings of MICCAI 2015*, pp. 234-241. Springer International Publishing.
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **65**, 386-408. doi:10.1037/h0042519
- Schäfer, J. and Strimmer, K. (2005). An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics* **21**, 754-764. doi:10.1093/bioinformatics/bti062
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B. et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676-682. doi:10.1038/nmeth.2019

- Schmidt, U., Weigert, M., Broaddus, C. and Myers, G. (2018). Cell Detection with Star-Convex Polygons. In *Proceedings of MICCAI 2018*, pp. 265-273. Springer International Publishing.
- Schneider, C. A., Rasband, W. S. and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671-675. doi:10.1038/nmeth.2089
- Speiser, A., Müller, L.-R., Matti, U., Obara, C. J., Legant, W. R., Kreshuk, A., Macke, J. H., Ries, J. and Turaga, S. C. (2020). Deep learning enables fast and dense single-molecule localization with high accuracy. *bioRxiv*, 2020.2010.2026.355164.
- Stephens, M. (2017). False discovery rates: a new deal. *Biostatistics* **18**, 275-294.
- Storkey, A. J. (2008). When Training and Test Sets are Different: Characterising Learning Transfer. In *Dataset Shift in Machine Learning*, pp. 2-28. The MIT Press.
- Stringer, C., Wang, T., Michaelos, M. and Pachitariu, M. (2021). Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100-106. doi:10.1038/s41592-020-01018-x
- Sudhakaran, S., Grbic, D., Li, S., Katona, A., Najarro, E., Glanois, C. and Risi, S. (2021). Growing 3D artefacts and functional machines with neural cellular automata. In *Proceedings of ALIFE 2021*.
- Sugawara, K., Cevrim, C. and Averof, M. (2021). Tracking cell lineages in 3D by incremental deep learning. *bioRxiv*, 2021.2002.2026.432552.
- Sui, L. and Dahmann, C. (2020). Increased lateral tension is sufficient for epithelial folding in *Drosophila*. *Development* **147**, dev194316. doi:10.1242/dev.194316
- Sui, L., Alt, S., Weigert, M., Dye, N., Eaton, S., Jug, F., Myers, E. W., Jülicher, F., Salbreux, G. and Dahmann, C. (2018). Differential lateral and basal tension drive folding of *Drosophila* wing discs through two distinct mechanisms. *Nat. Commun.* **9**, 4620. doi:10.1038/s41467-018-06497-3
- Toda, S., Mckeithan, W. L., Hakkinen, T. J., Lopez, P., Klein, O. D. and Lim, W. A. (2020). Engineering synthetic morphogen systems that can program multicellular patterning. *Science* **370**, 327-331. doi:10.1126/science.abc0033
- Tokuoka, Y., Yamada, T. G., Mashiko, D., Ikeda, Z., Hiroi, N. F., Kobayashi, T. J., Yamagata, K. and Funahashi, A. (2020). 3D convolutional neural networks-based segmentation to acquire quantitative criteria of the nucleus during mouse embryogenesis. *NPJ Syst. Biol. Appl.* **6**, 32. doi:10.1038/s41540-020-00152-8
- Vergara, H. M., Pape, C., Meechan, K. I., Zinchenko, V., Genoud, C., Wanner, A. A., Titze, B., Templin, R. M., Bertucci, P. Y., Simakov, O. et al. (2021). Whole-body integration of gene expression and single-cell morphology. *Cell* (in press). doi:10.1016/j.cell.2021.07.017
- Villoutreix, P. (2021). What machine learning can do for developmental biology. *Development* **148**, dev188474. doi:10.1242/dev.188474
- Von Chamier, L., Laine, R. F., Jukkala, J., Spahn, C., Krentzel, D., Nehme, E., Lerche, M., Hernández-Pérez, S., Mattila, P. K., Karinou, E. et al. (2021). Democratizing deep learning for microscopy with ZeroCostDL4Mic. *Nat. Commun.* **12**, 2276. doi:10.1038/s41467-021-22518-0
- Wathe, D., Brown, J. M., Reglinski, K., Diez-Sevilla, I., Roberts, D. and Eggeling, C. (2020). Object detection networks and augmented reality for cellular detection in fluorescence microscopy. *J. Cell Biol.* **219**, e201903166. doi:10.1083/jcb.201903166
- Wang, S., Yang, D. M., Rong, R., Zhan, X. and Xiao, G. (2019). Pathology image analysis using segmentation deep learning algorithms. *Am. J. Pathol.* **189**, 1686-1698. doi:10.1016/j.ajpath.2019.05.007
- Weigert, M., Schmidt, U., Boothe, T., Müller, A., Dibrov, A., Jain, A., Wilhelm, B., Schmidt, D., Broaddus, C., Culley, S. et al. (2018). Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nat. Methods* **15**, 1090-1097. doi:10.1038/s41592-018-0216-7
- Weigert, M., Schmidt, U., Haase, R., Sugawara, K. and Myers, E. (2020). Star-convex Polyhedra for 3D Object Detection and Segmentation in Microscopy. 2020 IEEE Winter Conference on Applications of Computer Vision (WACV) 3655-3662.
- Wen, C., Miura, T., Voleti, V., Yamaguchi, K., Tsutsumi, M., Yamamoto, K., Otomo, K., Fujie, Y., Teramoto, T., Ishihara, T. et al. (2021). 3DeeCellTracker, a deep learning-based pipeline for segmenting and tracking cells in 3D time lapse images. *eLife* **10**, e59187. doi:10.7554/eLife.59187
- Wolf, S., Pape, C., Bailoni, A., Rahaman, N., Kreshuk, A., Köthe, U. and Hamprecht, F. A. (2018). The Mutex Watershed: Efficient, Parameter-Free Image Partitioning. In *Proceedings of ECCV 2018*, pp. 571-587.
- Wolf, S., Bailoni, A., Pape, C., Rahaman, N., Kreshuk, A., Kothe, U. and Hamprecht, F. A. (2020). The mutex watershed and its objective: efficient, parameter-free graph partitioning. *IEEE Trans. Pattern Anal. Mach. Intell.* [Epub ahead of print]. doi:10.1109/TPAMI.2020.2980827
- Wollmann, T. and Rohr, K. (2021). Deep consensus network: aggregating predictions to improve object detection in microscopy images. *Med. Image Anal.* **70**, 102019. doi:10.1016/j.media.2021.102019
- Yang, K. D., Belyaeva, A., Venkatachalapathy, S., Damodaran, K., Katcoff, A., Radhakrishnan, A., Shivashankar, G. V. and Uhler, C. (2021). Multi-domain translation between single-cell imaging and sequencing data using autoencoders. *Nat. Commun.* **12**, 31. doi:10.1038/s41467-020-20249-2
- Yang, S. J., Berndl, M., Michael Ando, D., Barch, M., Narayanaswamy, A., Christiansen, E., Hoyer, S., Roat, C., Hung, J., Rueden, C. T. et al. (2018). Assessing microscope image focus quality with deep learning. *BMC Bioinformatics* **19**, 77. doi:10.1186/s12859-018-2087-4
- Young, H., Belbut, B., Baeta, M. and Petreanu, L. (2021). Laminar-specific cortico-cortical loops in mouse visual cortex. *eLife* **10**, e59551. doi:10.7554/eLife.59551
- Zhang, Z.-B., Wang, Q.-Y., Ke, Y.-X., Liu, S.-Y., Ju, J.-Q., Lim, W. A., Tang, C. and Wei, P. (2017). Design of tunable oscillatory dynamics in a synthetic NF- $\kappa$ B signaling circuit. *Cell Syst.* **5**, 460-470.e5. doi:10.1016/j.cels.2017.09.016
- Zhao, H., Braatz, R. D. and Bazant, M. Z. (2021). Image inversion and uncertainty quantification for constitutive laws of pattern formation. *J. Comput. Phys.* **436**, 110279. doi:10.1016/j.jcp.2021.110279