Published in final edited form as:

Nat Methods. 2018 November ; 15(11): 868–870. doi:10.1038/s41592-018-0194-9.

Deep learning to predict microscope images

Roger Brent and

Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, Washington 98109

Laura Boucheron

Klipsch School of Electrical and Computer Engineering, New Mexico State University, Las Cruces, New Mexico 88003

Abstract

A species of neural network first described in 2015 can be trained to translate between images of the same field of view acquired by different modalities. Trained networks can use information inherent in grayscale images of cells to predict fluorescent signals.

Unlabeled micrographic images of cells are full of blobs, whorls, patterns of different shades of grey. Those images seem as if they might well contain information about cell substructures and cell state, but, if so, that information seems to lie just beyond the powers of human observation. Two recent studies show how the breathtakingly rapid evolution of methods in Machine Learning (ML) might change that situation. Work by Christiansen et al¹, and Ounkomol et al²., exploits not-yet-mainstream developments in ML to train neural networks (NNs) to harvest this information and make it available to humans.

The key idea is that these researchers trained NNs to “translate” between “label-free” (brightfield, phase, DIC, and transmission electron microscope) and “labeled” (here, fluorescence) images of the same cells. Once trained to translate, the networks can convert hitherto-unseen label-free images into predicted fluorescence images (Figure 1a). The ability to predict fluorescence images from unpromising gray scaled data promises real advantages, including but not limited to increased speed of imaging and better time- lapse imaging free of artifacts induced by fluorophore photobleaching and phototoxicity, ultimately increasing microscopic seeing power. Moreover, because the computing power to run trained NNs is cheap and accessible, as close as the smartphones in the pockets and purses of contemporary researchers, this ability may help democratize research by diminishing the need for equipment, cell engineering, and training fluorescence approaches now require.

Given the short history of computer science and engineering, the foundations of the NNs in the current studies are positively ancient. 70 years ago, in work funded by the US Navy, Rosenblatt³ described “Perceptrons”, devices in which inputs from light sensors fed into a layer of McCulloch-Pitts neurons, each of which fired an output impulse when the sum of its multiple inputs exceeded a threshold⁴. Rosenblatt showed that such networks could be trained to perform classifications of input patterns in a finite number of cycles of pattern presentation, observation of network output, and adjustment of the weights that each neuron gives each input. Early networks were composed of actual analog “neurons”, physical

elements connected to a sensor array of photoelectric elements, by input wires with inline variable resistors, so that researchers could adjust with a dial the weight of each neuronal input. Work on NNs slowed during the 1960s, after circulation of a manuscript by Minsky and Papert (published 1969)⁵ convinced much of the computer science establishment and DoD funders that NNs were not able to perform key computational tasks⁶. Despite this prevailing view, some researchers continued to study ever-more-complex networks and so kept the approach alive⁷.

The current explosion of interest in NNs dates from this decade. One key development was the description of a distant descendant of the Perceptron, a particular multilayer (and thus “deep”) “convolutional” neural network (CNN), AlexNet⁸, in 2012. AlexNet decisively outperformed other methods in an open competition to classify thousands of different kinds of objects depicted in images. The decisive win greatly stimulated research in CNNs and other “deep learning” architectures. Six years later, progress in building, training and deploying NNs across different areas of human activity has been breathtaking, even terrifying. Today, trained NNs suggest books to read, authorize or deny access to credit, guide driverless automobiles, fly drones that track objects moving on the ground, and identify individuals captured in images from security cameras and social media posts. Rapid progress is fueled by widespread availability of image data for training, and two other key factors. One is the continuing increase in power of the GPUs used to train and run complex networks, itself spurred by the a market that demands ever-better performance in video games. The other is a now established convention in academic ML, in which actively competing research groups can build immediately on one another’s work, using work described in open preprints on arXiv and open source code deposited in repositories such as Github.

Today, the vast majority of tasks performed by CNNs on images involve *classification*. Figure 1b shows a simple CNN that decides whether an object in an image is a cat. One can imagine this type of CNN being used to help guide a self-driving car once trained on human-labeled images of pedestrians, bicycles and trucks, so that, given the image from a forward facing camera, it can identify image objects and trigger an appropriate response in the vehicle.

To appreciate the recent work, it’s helpful to consider the degree to which the rapid evolution of ML methods has resulted in means of processing information that seem non-intuitive, non-biological, alien. In the CNNs of 2018, recognizable simulations of McCulloch-Pitt “neurons” have mostly vanished. Replacing them are vast arrays of pure numbers, stored in dozens of different high dimensional matrices, those connected by permissible operations on those numbers. Input pixel intensities are operated on by matrix operations, in particular convolution, between for example pixel values and filter matrices. During training, the numbers in the filters come to represent operations necessary to extract abstractions, such as shapes and edges from images. Each convolution operation operates on multiple inputs, and each resulting number is operated on, not by “biological” neurons, but by a nonlinear “activation function” that gives finite slopes for every input. Output winds up in matrices called “feature maps”. The numbers in feature maps represent abstractions, again such as shapes and edges, the network recognizes in a particular image. Operations on

feature map numbers generate pooled maps. On the numbers in pooled maps, the networks repeat the convolution-filter-nonlinearize-pool steps, at least twice. In those downstream repetitions, the general information about images the network stores in its filter maps, and the specific information about particular images in its feature maps, becomes more successively abstract. Further downstream, linear/ affine operations on inputs from pooled maps followed by activation functions in “fully connected” layers output the classification (Figure 1a). To borrow a phrase⁹, image processing by today’s CNNs is no more like biological image processing that a Boeing 737 is like a seagull.

Christiansen et al¹. and Ounkomol et al² extend this splendid artificiality to carry out a task quite different from classification. They use “fully convolutional” NNs introduced in 2015¹⁰. These (Figure 1c) replace the operations performed by the fully connected layers of the CNNs in Figure 1b with successive operations that reconstruct data from the downstreammost filter maps and, at each step, fuse that information with intermediately abstracted data from the original feature maps. Such networks can be trained to “translate”, pixel by pixel, between two kinds of images, for example raw images and “segmented” images in which boundaries had been drawn around different objects¹¹. And here, to use information present in unlabeled images to predict fluorescence images.

Use of NNs to predict fluorescence may turn out to be a case in which recondite computational methods, first developed for military purposes by a nation state, extended by large corporations¹ and institutions² in coastal enclaves, ultimately results in worldwide democratization of technical capability and a burst of new knowledge. That outcome isn’t preordained. The work in Christiansen et al. is covered by an issued US Patent¹² which seemingly governs all use of trained neural networks to label two types of unlabeled images. It’s too early to guess whether intellectual property concerns might hinder academic investigators and smaller companies from building on this work.

But for now, one consequence of the ability to use NNs to translate rather than classify may be take decisions about what’s worth noticing in images away from arrays of 64 bit floating point numbers, and deliver that agency back to the human investigator. We can hope that cheap computational methods that put empowered human observers back into the loop may enable discoveries not obvious to algorithms. Whether or not it delivers new discoveries, this application of NNs to augment human observation already counts as a non-terrifying use of ML.

Acknowledgements.

We are grateful to John Markoff, Jason Yosinski, Jeff Clune, Greg Johnson, Molly Maleckar, Bill Peria, and other colleagues for useful communications, and to R21 CA223901 to RB and U54 CA132831 (NMSU/ FHCRC Partnership for the Advancement of Cancer Research) for partial support.

References

- 1). Christiansen EM, Yang SJ, Ando DM, Javaherian A, Skibinski G, Lipnick S, Mount E, O’Neil A, Shah K, Lee AK, Goyal P, Fedus W, Poplin R, Esteva A, Berndl M, Rubin LL, Nelson P, and Finkbeiner S (2018). *In silico* labeling: predicting fluorescent labels in unlabeled images. *Cell*. 173(3), 792–803 [PubMed: 29656897]

- 2). Ounkomol C, Seshamani S, Maleckar MM, Collman F, and Johnson GR (2018) Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy. *Nature Methods*. 2018 9 17 doi: 10.1038/s41592-018-0111-2. [Epub ahead of print]
- 3). Rosenblatt F (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*. 65, 386–408. [PubMed: 13602029]
- 4). McCulloch WS, and Pitts WH (1943) A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115–133.
- 5). Minsky M, and Papert S (1969) *Perceptrons: An Introduction to Computational Geometry*, MIT Press, Cambridge Massachusetts.
- 6). Olazaran M (1996) A sociological study of the official history of the perceptrons controversy. *Social Studies of Science*. 26, 611–659
- 7). Markoff J (2015) *Machines of loving grace*. Harper Collins, New York
- 8). Krizhevsky A, Sutskever I, and Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* (pp. 1097–1105)
- 9). Stross C (2017). Dude, you broke the future! Keynote address, 34th Chaos Communications Congress, Leipzig, <http://www.antipope.org/charlie/blog-static/2018/01/dude-you-broke-the-future.html>, retrieved 11 September 2018.
- 10). Long J. Shelhamer E. Darrel T. Fully convolutional neural networks for semantic segmentation; Conference on Computer Vision and Pattern Recognition; Boston, Massachusetts. 2015.
- 11). Ronneberger O, Fischer P, and Brox T (2015) U-net: Convolutional networks for biomedical image segmentation In *International Conference on Medical Image Computing and Computer-Assisted Inter-Image Computing and Computer-Assisted Intervention*, pp. 234–241. Springer, New York.
- 12). Nelson P, Christiansen EM, Berndl M, and Frumkin M (2018) Processing cell images using neural networks, US Patent 9,971,996 B. Issued May 15, 2018, assigned to Google LLC.

Fig 1.

a) Translating to predict. Top. Network in Panel c is trained on pairs of images of cells acquired under Modality M1 (eg. brightfield) and Modality M2 (here, fluorescence images of signal from a protein the nuclear envelope). Bottom. Given a new brightfield image, the trained network predicts the nuclear envelope signal from information in an image of a field of unlabelled cells.

b) A network that classifies. A simple NN similar to Alexnet⁸ that identifies images of cats. Its operation shown in captions. In an untrained network, values of the filter matrices (which are convolved with upstream information to create feature maps) and of the weights of the individual inputs to the downstream "fully connected" "neurons" (affine operations and activation functions) are set randomly, and values of the filter matrices and input weights move toward optimum values during training. The network is trained by exposure to images, some of which are labeled "Cat" and some of which are labeled "Not cat". During training, feature maps at the top level come to recognize human-intelligible image elements such as edges, while those in deeper levels come to recognize more abstract aspects of images less easily described by the human observers. In these steps, information *processing* takes place via matrix operations and nonlinearization, while information *storage*, retention of knowledge acquired during training, takes place as changes in values stored in the filter matrices and weights of inputs to the downstream "neurons".

c) A network that translates. A "simple" NN descended from that in references^{10, 11}. This network translates between black and white images of cats and otherwise identical images

that reveal the cats' third or inner eyes. During training, pixelated M1 images are subjected to the series of the convolution, nonlinearization, and pooling steps as in panel b. Here, however, entries in downstreammost pooled maps are successively "upsampled" and combined with information coming from intermediate level feature maps on the left. The result for each pixel is compared with pixel intensities for the training M2 images. Training continues by adjustment of filter values until the network learns relationship between M1 and M2 images. The trained network can then operate on a new M1 image to produce a predicted M2 image.