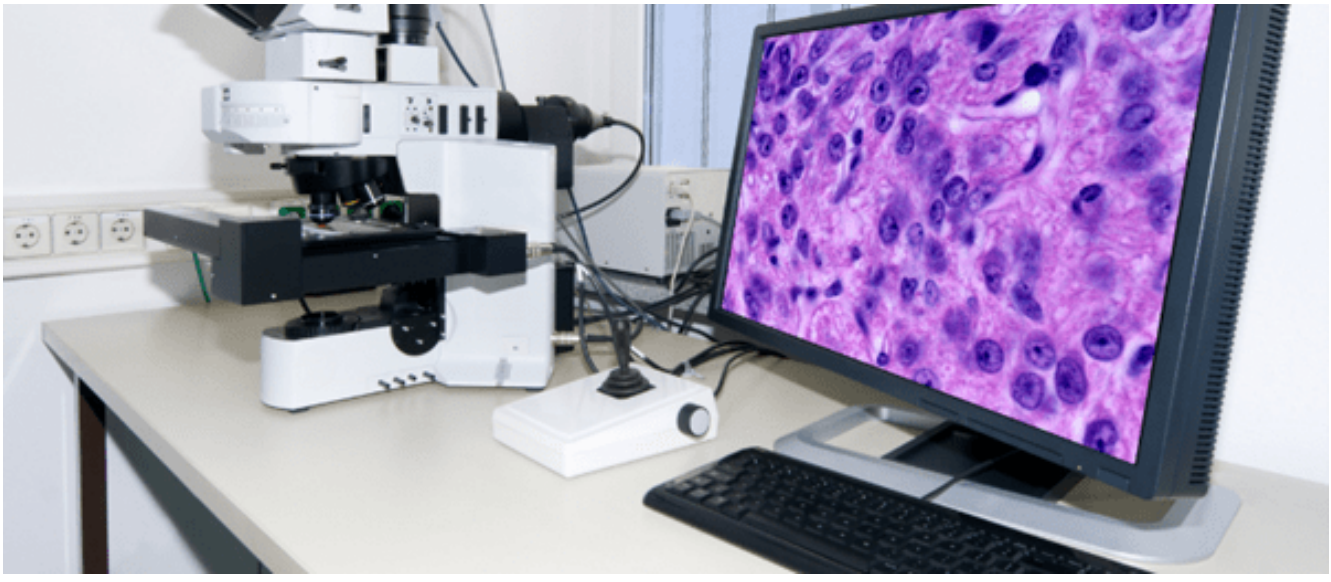
Bioimage analysis: has deep learning changed the game?

29 APR 2022 | WRITTEN BY GEORGIA BICKERTON (ASSISTANT EDITOR)

INTERVIEWS | LAB DESIGN AND MACHINERY



We recently spoke to Beth Cimini (left), an analyst at the Broad Ins (MA, USA), about her role at the institute and how bioimage analysis has evolved in recent years due to COVID-19 and the application of deep learning.

Bringing together researchers and labs from both Harvard and MIT, the Broad Institute is a not-for-profit research center, which was initially developed for genomic medicine and has since expanded across many disciplines of biomedical research. Nearly 20 years on from its inception, the Broad Institute focuses on doing biology at scale; for example, the institute currently performs around 1 in 30 COVID tests in the United States. Cimini is part of the imaging platform team, and develops open-access image analysis software using deep learning to streamline analysis leading to better and faster answers to biological questions.

The imaging platform contains two labs working on this mission. The first, led by Anne Carpenter and Shantanu Singh, is focused on turning the data gathered from images into answers to scientific questions using deep learning and informatics. The second lab, which Cimini leads, is focused on making image analysis easier, leading to quicker and more expansive image analysis. One function of the lab is to maintain the Cell Profiler and Cell Profiler Analyst tools, which were created by Carpenter and are freely available tools that streamline bioimage analysis.

“You shouldn’t need to know how to do [coding] in order to do good microscopy and good image analysis”

When a researcher has an image to analyze, the first step is to upload it onto an image analysis program such as ImageJ or Fiji and explore the thresholding and filtering options to produce the best quality images for analysis. But what happens when a researcher has 5000 images they’d like to analyze? Well, currently, they’d need to learn how to code. As Cimini pointed out, “you shouldn’t need to know how to do [coding] in order to do good microscopy and good image analysis”. Some people wouldn’t know where to begin, and many of us just don’t have the time to commit to mastering this skill, and that’s where Cell Profiler comes into play.

Cell Profiler allows you to string together a ‘pipeline’ of different image analysis steps called ‘modules’ without getting anywhere close to a line of code. This could mean taking measurements, finding objects, smoothing an image or highlighting the edge of the nucleus – or all the above. Once you have created a pipeline using Cell Profiler all you need to do is upload the whole set of images into the program and the rest is done for you.

Cell Profiler is a powerful program with between 900 and 950 different image analysis settings; however, a powerful program like this doesn't come without a trade-off, as is often the case. Cimini recognizes that the program could initially come across as overwhelming and appreciates that this may be one of the largest challenges faced by users of Cell Profiler. "I was a Cell Profiler before I worked on the program, so I know how challenging it can be," Cimini explained. However, to combat this the team has developed a thorough guide along with video tutorials to help researchers use this tool.

"I was a Cell Profiler before I worked on the program, so I know how challenging it can be."

While Cell Profiler allows faster analysis of cells and turning images into data, Cell Profiler Analyst enables researchers to take this data set and turn it into tangible answers to questions. For instance, if a researcher measured the cell area in 100 images using Cell Profiler, they would then be able to use Cell Profiler Analyst to visualize this data. Cimini put it nicely, "Cell Profiler gets you the important numbers and then Cell Profiler Analyst allows you to explore that data to answer your questions."

These are powerful tools that are freely available to all researchers, something the Broad Institute feels is essential for a number of reasons. The first is that sometimes there are cases when a researcher wishes to use the Cell Profiler platform to do something that is not currently available on the program. As the code is open-source, updates can be suggested to the team at the Broad, which means the tool doesn't remain stagnant and the technology continues to develop. The second is ensuring that researchers in countries with fewer resources don't have to pay for expensive software licensing, which is not only good for accessibility but will also accelerate the advancement of scientific research.

"We think it makes science move a lot faster and it is a lot fairer"

Open-source image analysis and tools such as Cell Profiler and Cell Profiler Analyst have exploded recently, which Cimini agreed could be down to COVID. She observed that when researchers were forced out of the labs and into their homes, they were finding the time to dig out the data they had been meaning to analyze, and the Broad was able to put time and energy into teaching people how to use their tools through webinars and office hours.

But, Cimini thinks the recent advancements in image analysis software itself were down to something else entirely.

“Deep learning changed the game”

When Cimini first joined the Broad Institute in 2016, the idea that deep learning could be used in bioimaging was one the team believed possible; however, the sticking point was gathering enough data to train the models.

Neural networks are trained by data sets that contain labels. For example, if you are training a neural network to recognize a bus or a plane in an image, you would train it with a data set of pictures labeled either as ’bus’ or ’plane’. Each picture would then run through the neural network, which would decide if the picture was more likely to contain a bus or a plane. At this point, the label on the picture would tell the model if it was right or wrong. The model gets smarter by prioritizing the route through the neural network that results in the best predictions of what an image contains. So, if the goal is to train a neural network to isolate more specific features such as nuclei in squamous endothelial cells, a data set with these features and corresponding labels are required.

Industry leaders in deep learning and artificial intelligence, such as Facebook, had enormous data sets created by millions of people tagging their friend’s faces for years that could be used to train neural networks. So, for deep learning to be applied to bioimaging, relevant data sets need to be created to train the computer models.

Currently, Cell Profiler doesn’t have its own component of deep learning but can incorporate deep learning networks and data sets such as Cellpose and StarDist. As data sets improve, Cimini and the team at Broad Institute hope their tool can come with a neural network that works straight out of the box rather than with the complicated settings and extra plug-ins required now. Currently, Cimini spends time helping researchers understand how to pick the best thresholding algorithm for their work, which she thinks will eventually become redundant, as the tool will be able to do this itself. This will allow more time for the group to work with biologists to help them make Cell Profiler pipelines or create more open-source image analysis tools and workflows.

While deep learning is, as Cimini describes, “undoubtedly valuable” it does not come without limitations. There is no understanding of how a neural network makes certain decisions, which Cimini explained using the example of a deep learning algorithm trained to classify handwritten numbers. Once sorted and classified, Cimini filtered the results to show only the elements of the dataset that the algorithm had assigned as a three. While most of the results were a three, there were a few twos and an eight. “When mistakes are made it’s hard to know why and it is hard to know how to fix them,” says Cimini, and often, the only option is to continue training the

network. For this reason, Cimini doesn't believe that deep learning tools will become the be-all and end-all of bioimage analysis but there is no doubt it will continue to develop further.

When asked about what makes working on the Cell Profiler and Cell Profiler Analyst technology so rewarding, Cimini revealed that Cell Profiler is cited in over a thousand papers a year, meaning that there are "at least a thousand people a year who might not have been able to get an answer before." Cimini adds that, "there is a little piece of science that couldn't be done and now can". For example, Cell Profiler has been used to pick which therapeutic drugs would be best suited to patients with leukemia or lymphoma. The interface was able to predict which drugs would work best for a patient and made better judgments about treatment plans than a doctor alone, increasing the life expectancy of patients with cancer. "When you have a huge impact on somebody's life, that is great, but having thousands of tiny impacts on people's lives is also really rewarding."

As deep learning becomes more incorporated into bioimage analysis programs, like Cell Profiler and Cell Profiler Analyst, the speed at which data can be sorted and analyzed will lead to faster science dissemination. While these tools may look daunting at first, those developing and maintaining them have created a range of informative materials and enjoy assisting researchers to find answers to a scientific questions within their images.