

AlphaGo, Deep Learning, and the Future of the Human Microscopist

Scott R. Granter, MD; Andrew H. Beck, MD, PhD; David J. Papke Jr, MD, PhD

In March of last year, Google's (Menlo Park, California) artificial intelligence (AI) computer program AlphaGo beat the best Go player in the world, 18-time champion Lee Se-dol, in a tournament, winning 4 of 5 games.¹ At first glance this news would seem of little interest to a pathologist, or to anyone else for that matter. After all, many will remember that IBM's (Armonk, New York) computer program Deep Blue beat Garry Kasparov—at the time the greatest chess player in the world—and that was 19 years ago. So, what's so significant about a computer winning another board game?

The rules of the several-thousand-year-old game of Go are extremely simple. The board consists of 19 horizontal and 19 vertical black lines. Players take turns placing either black or white stones on vacant intersections of the grid with the goal of surrounding the largest area and capturing their opponent's stones. Once placed, stones cannot be moved again. Despite the simplicity of its rules, Go is a mind-bogglingly complex game—far more complex than chess. A game of 150 moves (approximately average for a game of Go) can involve 10^{360} possible configurations, “more than there are atoms in the Universe.”² As complex as it is, chess is vastly less complex than Go, and chess is amenable to “brute force” algorithmic computer approaches for beating expert chess players like Kasparov. To beat Kasparov, Deep Blue analyzed possible moves and evaluated outcomes to decide the best move.

Go's much higher complexity and intuitive nature prevents computer scientists from using brute force algorithmic approaches for competing against humans. For this reason, Go is often referred to as the “holy grail of AI research.”² To beat Se-dol, Google's AlphaGo program used artificial neural networks that simulate mammalian neural architecture to study millions of game positions from expert human-played Go games. But this exercise would, at least theoretically, only teach the computer to be on par

with the best human players. To become better than the best humans, AlphaGo then played against itself millions of times, over and over again, learning and improving with each game—an exercise referred to as reinforcement learning.^{3,4} By playing itself and determining which moves lead to better outcomes, AlphaGo literally learns by teaching itself. And the unsettling thing is that we don't understand what AlphaGo is thinking. In an interview with FiveThirtyEight, one computer scientist commented, “It is a mystery to me why the program plays as well as it does.”⁵ In the same article, an expert Go player said, “It makes moves that no human, including the team who made it, understands,” and “AlphaGo is the creation of humans, but the way it plays is not.”⁵ It is easy to see how some viewed AlphaGo's victory over Se-dol as a turning point in the history of humanity—we have created machines that truly think and, at least in some areas like Go, they are smarter, much smarter, than we are.

Also see p. 622 and p. 624.

Importantly, whereas Big Blue was purpose built to play chess and only to play chess (and was mothballed after beating Kasparov), AlphaGo's technology is much more general and able to tackle diverse problems. It implements machine learning algorithms (including neural networks) that are effectively an extension of simple regression fitting. In a simple regression fit, we might determine a line that predicts an outcome y given an input x . With increased computational power, machine learning algorithms are able to fit a huge number of input variables (for example, moves in a game of Go) to determine a desired output (maximizing space gained on the Go board). It has been predicted that the algorithms used in AlphaGo “will be incredibly useful in medical research, diagnosis, [and] complex treatment planning”⁵—which is where the future of the microscopist comes in. Will AI programs like AlphaGo replace the microscopist?

There are many hurdles to replacing the human microscopist with computer algorithms—some practical and some theoretical. On a practical level, there are financial barriers to incorporating slide scanners and computers into pathology workflow,⁶ although presumably hospitals would undertake these steps if computers could improve diagnostic accuracy or increase the efficiency of pathologists. The more interesting question is, will computer algorithms surpass humans in diagnostic abilities?

Accepted for publication October 28, 2016.

From the Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts (Drs Granter and Papke); Harvard Medical School, Boston (Drs Granter and Beck); and the Department of Pathology, Beth Israel Deaconess Hospital, Boston (Dr Beck).

Dr Beck is a cofounder of PathAI, Inc (Brookline, Massachusetts). The other authors have no relevant financial interest in the products or companies described in this article.

doi: 10.5858/arpa.2016-0471-ED

Reprints: Scott R. Granter, MD, Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, 75 Francis St, Boston, MA 02115 (email: sgranter@partners.org).

In contrast to AlphaGo's success at surpassing human experts, computer vision algorithms fall short of matching some basic human vision capabilities. When humans look at pictures, they can interpret scenes and predict within seconds what is likely to happen after the picture is taken ("object dynamic prediction" in computer terminology). However, although algorithms can be trained to make similar predictions in abstract cartoon scenes, they cannot accurately predict what will happen next in real-world photographic scenes.⁷ And algorithms are notoriously bad at some aspects of image analysis. For example, algorithms cannot accurately predict whether a human will find a photograph funny or not,⁸ to the point where humor detection is considered an "AI-complete problem."⁹ But is some intangible quality that only a human can provide necessary for accurate pathology interpretation?

A recent and somewhat unsettling study answers this question. Levenson et al¹⁰ showed that pigeons can be trained to distinguish malignant from benign breast tissue with 85% accuracy for individual birds and with an impressive 99% accuracy for a flock consensus. Like pigeons, computer algorithms can be trained to excel at image classification. Training involves giving algorithms a set of labeled example images (for example, pictures of houses or cars labeled "house" and "car," respectively), and then allowing the program to determine how to use the information in the training set to correctly label unknown images. Given an adequate number of training examples, computers in principle can learn to classify images into correct categories.

One major hurdle in this training process is obtaining a sufficiently large number of training examples. Currently, development of state-of-the-art computer vision algorithms requires millions of training images.¹¹ For AlphaGo, it was feasible to generate millions of example games, but establishing a whole-slide image database of millions of images is currently not practical. A second problem is the size of each data point: it is currently impossible to directly feed whole-slide images into algorithms because each one contains on the order of 10 GB of data.¹² However, there are workarounds, such as extracting information from images and then feeding this information into the algorithms.¹³

A few recent studies have demonstrated a promising approach to circumvent this problem: whole slide images are divided into smaller "patches," and then an algorithm is trained to classify these patches into different categories.^{14,15} Statistical summaries of patch diagnoses are then fed into a machine learning algorithm to classify the entire image into a single diagnosis. Impressively, algorithms were able to distinguish subtypes of non-small cell lung carcinoma with an accuracy similar to that of expert pulmonary pathologists.¹⁴ And combining the predictions of the algorithms with those of humans led to an 85% decrease in human error in detecting metastatic breast cancer in lymph nodes.¹⁵

We believe these studies point to the future of computational pathology: computers will increasingly become integrated into the pathology workflow when they can improve accuracy in answering questions that are difficult for pathologists. Programs could conceivably count mitoses or quantitatively grade immunohistochemistry stains more accurately than humans, and they could identify regions of interest in a slide to reduce the time a pathologist spends screening, as is done in cytopathology. We predict that, over time, as computers gain more and more discriminatory abilities, they will reduce the amount of time it takes for a

pathologist to render diagnoses and, in the process, reduce the demand for pathologists as microscopists, potentially enabling pathologists to focus more cognitive resources on higher-level diagnostic and consultative tasks (eg, integrating molecular, morphologic, and clinical information to assist in treatment and clinical management decisions for individual patients).

Readers of our (S.R.G.'s)¹⁶ recent essay in this journal will recall that we predicted that the microscope (we made no mention of the microscopist) will have continued utility in diagnosis and medical research. In that essay we argued that inside-view arguments are often unreliable in making predictions such as the future longevity of the microscope. Instead, we presented an outside-view perspective to predict the microscope will likely be around for a long time. Readers will also note that in this essay our prediction that AI technology will likely play an increasingly important role in diagnostic microscopy was based on what we believe are less-reliable inside-view perspectives. In fact, an outside-view perspective would emphasize that there have been several "AI winters" since the 1960s—periods of deep pessimism that followed initial excitement in the field of AI. But the potential of "deep learning" is already being demonstrated and the trajectory of accomplishments is astonishing. For example, when the Defense Advanced Research Projects Agency held a self-driving car challenge in the Mojave Desert in 2004, the most successful car was able to complete only 7.3 miles of the 150-mile course before driving into a rock. Only 3 years later, in 2007, 6 teams were able to complete the 60-mile Defense Advanced Research Projects Agency-sponsored Urban Challenge that required stopping at signs, traffic lights, and generally following traffic laws, as well as responding to the presence and actions of other vehicles in a much more complex and challenging urban environment. Today, Uber (San Francisco, California) is offering rides to customers with self-driving cabs in Pittsburgh.¹⁷ If winter is coming, we expect it will be gentle and mild.

References

1. Choe S. Google's computer program beats Lee Se-dol in Go tournament. *New York Times*. March 3, 2016. http://www.nytimes.com/2016/03/16/world/asia/korea-alpha-go-vs-lee-sedol-go.html?_r=0. Accessed August 9, 2016.
2. Koch C. How the computer beat the Go master. *Sci Am*. <http://www.scientificamerican.com/article/how-the-computer-beat-the-go-master/>. Published March 19, 2016. Accessed August 8, 2016.
3. Gibney E. Google masters Go. *Nature*. 2016;529(7587):445–446.
4. Silver D, Huang A, Maddison CJ, et al. Mastering the game of Go with deep neural networks and tree search. *Nature*. 2016;529(7587):484–489.
5. Roeder O. Don't forget humans created the computer program that can beat humans at Go. FiveThirtyEight Web site. <http://fivethirtyeight.com/features/dont-forget-humans-created-the-computer-program-that-can-beat-humans-at-go/>. Published March 16, 2016. Accessed August 9, 2016.
6. Ghaznavi F, Evans A, Madabhushi A, Feldman M. Digital imaging in pathology: whole-slide imaging and beyond. *Ann Rev Pathol*. 2013;8:331–359.
7. Fouhey DF, Zitnick CL. Predicting object dynamics in scenes. Poster presented at: IEEE Conference on Computer Vision and Pattern Recognition; June 24–27, 2014; Columbus, Ohio.
8. Chandrasekaran A, Vijayakumar AK, Antol S, et al. We are humor beings: understanding and predicting visual humor. Paper presented at: IEEE Conference on Computer Vision and Pattern Recognition; June 26–July 1, 2016; Las Vegas, Nevada.
9. Hurley MM, Dennett DC, Adams RB. *Inside Jokes: Using Humor to Reverse-Engineer the Mind*. Cambridge, MA: MIT Press; 2011.
10. Levenson RM, Krupinski EA, Navarro VM, Wasserman EA. Pigeons (*Columba livia*) as trainable observers of pathology and radiology breast cancer images. *PLoS One*. 2015;10(11):e0141357.
11. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Cornell University Library Web site. <http://arXiv.org/abs/1512.03385>. Published December 10, 2015. Accessed August 25, 2016.
12. Chlipala E, Elin J, Eichhorn O, et al. Archival and retrieval in digital pathology systems. Digital Pathology Association Web site. <https://>

digitalpathologyassociation.org/_data/files/Archival_and_Retrieval_in_Digital_Pathology_Systems.pdf. Accessed August 25, 2016.

13. Beck AH, Sangoi AR, Leung S, et al. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Sci Transl Med*. 2011;3(108):108ra113.

14. Hou L, Samaras D, Kurc TM, Gao Y, Saltz JH. Patch-based convolutional neural network for whole slide tissue image classification. Cornell University Library Web site. <http://arXiv.org/abs/1504.07947v5>. Published March 9, 2016. Accessed October 28, 2016.

15. Wang D, Khosla A, Gargeya R, Irshad H, Beck AH. Deep learning for identifying metastatic breast cancer. Cornell University Library Web site. <http://arXiv.org/abs/1606.05718>. Published June 18, 2016. Accessed September 9, 2016.

16. Granter SR. Reports of the death of the microscope have been greatly exaggerated. *Arch Pathol Lab Med*. 2016;140(8):744–745.

17. Maselli J. Uber's "self-driving" test cars to be overseen by driver and engineer. *Sci Am*. <http://www.scientificamerican.com/article/uber-s-self-driving-test-cars-to-be-overseen-by-driver-and-engineer/>. Published August 25, 2016. Accessed August 28, 2016.