



THE LEARNING MACHINES

Using massive amounts of data to recognize photos and speech, deep-learning computers are taking a big step towards true artificial intelligence.

BY NICOLA JONES

Three years ago, researchers at the secretive Google X lab in Mountain View, California, extracted some 10 million still images from YouTube videos and fed them into Google Brain — a network of 1,000 computers programmed to soak up the world much as a human toddler does. After three days looking for recurring patterns, Google Brain decided, all on its own, that there were certain repeating categories it could identify: human faces, human bodies and ... cats¹.

Google Brain's discovery that the Internet is full of cat videos provoked a flurry of jokes from journalists. But it was also a landmark in the resurgence of deep learning: a three-decade-old technique in which massive amounts of data and processing power

help computers to crack messy problems that humans solve almost intuitively, from recognizing faces to understanding language.

Deep learning itself is a revival of an even older idea for computing: neural networks. These systems, loosely inspired by the densely interconnected neurons of the brain, mimic human learning by changing the strength of simulated neural connections on the basis of experience. Google Brain, with about 1 million simulated neurons and 1 billion simulated connections, was ten times larger than any deep neural network before it. Project founder Andrew Ng, now director of the Artificial Intelligence Laboratory at Stanford University in California, has gone on to make deep-learning systems ten times larger again.

Such advances make for exciting times in

BRUCE ROLFF/SHUTTERSTOCK

artificial intelligence (AI) — the often-frustrating attempt to get computers to think like humans. In the past few years, companies such as Google, Apple and IBM have been aggressively snapping up start-up companies and researchers with deep-learning expertise. For everyday consumers, the results include software better able to sort through photos, understand spoken commands and translate text from foreign languages. For scientists and industry, deep-learning computers can search for potential drug candidates, map real neural networks in the brain or predict the functions of proteins.

“AI has gone from failure to failure, with bits of progress. This could be another leapfrog,” says Yann LeCun, director of the Center for Data Science at New York University and a deep-learning pioneer.

“Over the next few years we’ll see a feeding frenzy. Lots of people will jump on the deep-learning bandwagon,” agrees Jitendra Malik, who studies computer image recognition at the University of California, Berkeley. But in the long term, deep learning may not win the day; some researchers are pursuing other techniques that show promise. “I’m agnostic,” says Malik. “Over time people will decide what works best in different domains.”

INSPIRED BY THE BRAIN

Back in the 1950s, when computers were new, the first generation of AI researchers eagerly predicted that fully fledged AI was right around the corner. But that optimism faded as researchers began to grasp the vast complexity of real-world knowledge — particularly when it came to perceptual problems such as what makes a face a human face, rather than a mask or a monkey face. Hundreds of researchers and graduate students spent decades hand-coding rules about all the different features that computers needed to identify objects. “Coming up with features is difficult, time consuming and requires expert knowledge,” says Ng. “You have to ask if there’s a better way.”

In the 1980s, one better way seemed to be deep learning in neural networks. These systems promised to learn their own rules from scratch, and offered the pleasing symmetry of using brain-inspired mechanics to achieve brain-like function. The strategy called for simulated neurons to be organized into several layers. Give such a system a picture and the first layer of learning will simply notice all the dark and light pixels. The next layer might realize that some of these pixels form edges; the next might distinguish between horizontal and vertical lines. Eventually, a layer might recognize eyes, and might realize that two eyes are usually present in a human face (see ‘Facial recognition’).

The first deep-learning programs did not perform any better than

NATURE.COM
Learn about another approach to brain-like computers:
go.nature.com/ftktnso

simpler systems, says Malik. Plus, they were tricky to work with. “Neural nets were always a delicate art to manage. There is some black magic involved,” he says. The networks needed a rich stream of examples to learn from — like a baby gathering information about the world. In the 1980s and 1990s, there was not much digital information available, and it took too long for computers to crunch through what did exist. Applications were rare. One of the few was a technique — developed by LeCun —

“OVER THE NEXT FEW YEARS WE’LL SEE A FEEDING FRENZY. LOTS OF PEOPLE WILL JUMP ON THE DEEP-LEARNING BANDWAGON.”

that is now used by banks to read handwritten cheques.

By the 2000s, however, advocates such as LeCun and his former supervisor, computer scientist Geoffrey Hinton of the University of Toronto in Canada, were convinced that increases in computing power and an explosion of digital data meant that it was time for a renewed push. “We wanted to show the world that these deep neural networks were really useful and could really help,” says George Dahl, a current student of Hinton’s.

As a start, Hinton, Dahl and several others tackled the difficult but commercially important task of speech recognition. In 2009, the researchers reported² that after training on a classic data set — three hours of taped and transcribed speech — their deep-learning neural network had broken the record for accuracy in turning the spoken word into typed text, a record that had not shifted much in a decade with the standard, rules-based approach. The achievement caught the attention of major players in the smartphone market, says Dahl, who took the technique to Microsoft during an internship. “In a couple of years they all switched to deep learning.” For example, the iPhone’s voice-activated digital assistant, Siri, relies on deep learning.

GIANT LEAP

When Google adopted deep-learning-based speech recognition in its Android smartphone operating system, it achieved a 25% reduction in word errors. “That’s the kind of drop you expect to take ten years to achieve,” says Hinton — a reflection of just how difficult it has been to make progress in this area. “That’s like ten breakthroughs all together.”

Meanwhile, Ng had convinced Google to let him use its data and computers on what became

Google Brain. The project’s ability to spot cats was a compelling (but not, on its own, commercially viable) demonstration of unsupervised learning — the most difficult learning task, because the input comes without any explanatory information such as names, titles or categories. But Ng soon became troubled that few researchers outside Google had the tools to work on deep learning. “After many of my talks,” he says, “depressed graduate students would come up to me and say: ‘I don’t have 1,000 computers lying around, can I even research this?’”

So back at Stanford, Ng started developing bigger, cheaper deep-learning networks using graphics processing units (GPUs) — the super-fast chips developed for home-computer gaming³. Others were doing the same. “For about US\$100,000 in hardware, we can build an 11-billion-connection network, with 64 GPUs,” says Ng.

VICTORIOUS MACHINE

But winning over computer-vision scientists would take more: they wanted to see gains on standardized tests. Malik remembers that Hinton asked him: “You’re a sceptic. What would convince you?” Malik replied that a victory in the internationally renowned ImageNet competition might do the trick.

In that competition, teams train computer programs on a data set of about 1 million images that have each been manually labelled with a category. After training, the programs are tested by getting them to suggest labels for similar images that they have never seen before. They are given five guesses for each test image; if the right answer is not one of those five, the test counts as an error. Past winners had typically erred about 25% of the time. In 2012, Hinton’s lab entered the first ever competitor to use deep learning. It had an error rate of just 15% (ref. 4).

“Deep learning stomped on everything else,” says LeCun, who was not part of that team. The win landed Hinton a part-time job at Google, and the company used the program to update its Google+ photo-search software in May 2013.

Malik was won over. “In science you have to be swayed by empirical evidence, and this was clear evidence,” he says. Since then, he has adapted the technique to beat the record in another visual-recognition competition⁵. Many others have followed: in 2013, all entrants to the ImageNet competition used deep learning.

With triumphs in hand for image and speech recognition, there is now increasing interest in applying deep learning to natural-language understanding — comprehending human discourse well enough to rephrase or answer questions, for example — and to translation from one language to another. Again, these are currently done using hand-coded rules and statistical analysis of known text. The state-of-the-art of such techniques can be seen in software such as Google Translate, which can produce results that are comprehensible (if

sometimes comical) but nowhere near as good as a smooth human translation. “Deep learning will have a chance to do something much better than the current practice here,” says crowd-sourcing expert Luis von Ahn, whose company Duolingo, based in Pittsburgh, Pennsylvania, relies on humans, not computers, to translate text. “The one thing everyone agrees on is that it’s time to try something different.”

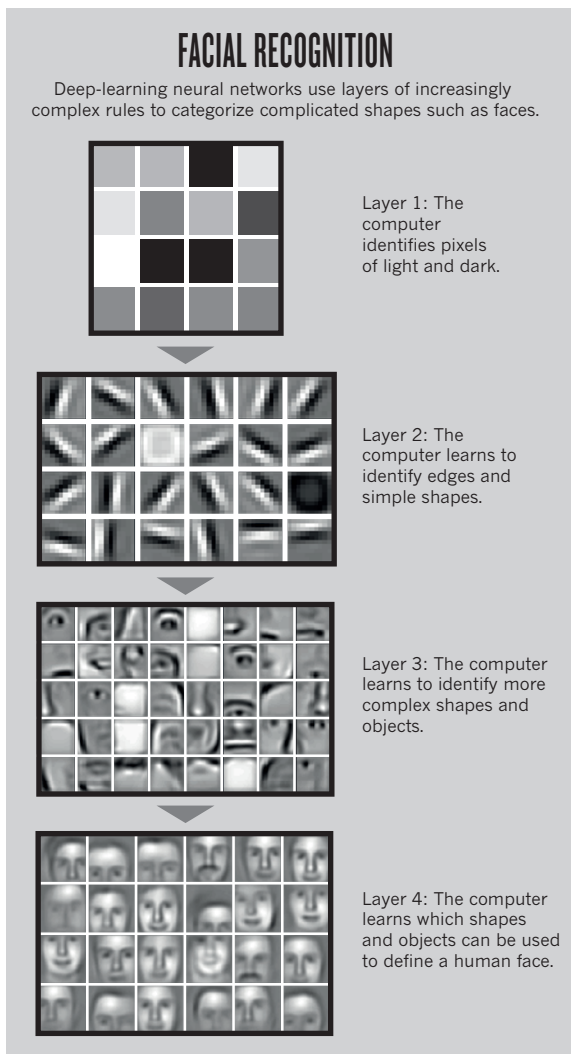
DEEP SCIENCE

In the meantime, deep learning has been proving useful for a variety of scientific tasks. “Deep nets are really good at finding patterns in data sets,” says Hinton. In 2012, the pharmaceutical company Merck offered a prize to whoever could beat its best programs for helping to predict useful drug candidates. The task was to trawl through database entries on more than 30,000 small molecules, each of which had thousands of numerical chemical-property descriptors, and to try to predict how each one acted on 15 different target molecules. Dahl and his colleagues won \$22,000 with a deep-learning system. “We improved on Merck’s baseline by about 15%,” he says.

Biologists and computational researchers including Sebastian Seung of the Massachusetts Institute of Technology in Cambridge are using deep learning to help them to analyse three-dimensional images of brain slices. Such images contain a tangle of lines that represent the connections between neurons; these need to be identified so they can be mapped and counted. In the past, undergraduates have been enlisted to trace out the lines, but automating the process is the only way to deal with the billions of connections that are expected to turn up as such projects continue. Deep learning seems to be the best way to automate. Seung is currently using a deep-learning program to map neurons in a large chunk of the retina, then forwarding the results to be proof-read by volunteers in a crowd-sourced online game called EyeWire.

William Stafford Noble, a computer scientist at the University of Washington in Seattle, has used deep learning to teach a program to look at a string of amino acids and predict the structure of the resulting protein — whether various portions will form a helix or a loop, for example, or how easy it will be for a solvent to sneak into gaps in the structure. Noble has so far trained his program on one small data set, and over the coming months he will move on to the Protein Data Bank: a global repository that currently contains nearly 100,000 structures.

For computer scientists, deep learning could earn big profits: Dahl is thinking about start-up opportunities, and LeCun was hired



last month to head a new AI department at Facebook. The technique holds the promise of practical success for AI. “Deep learning happens to have the property that if you feed it more data it gets better and better,” notes Ng. “Deep-learning algorithms aren’t the only ones like that, but they’re arguably the best — cer-

“DEEP LEARNING HAS THE PROPERTY THAT IF YOU FEED IT MORE DATA, IT GETS BETTER AND BETTER.”

tainly the easiest. That’s why it has huge promise for the future.”

Not all researchers are so committed to the idea. Oren Etzioni, director of the Allen Institute for Artificial Intelligence in Seattle, which launched last September with the aim of developing AI, says he will not be using the brain for inspiration. “It’s like when we invented flight,” he says; the most successful designs for aeroplanes

were not modelled on bird biology. Etzioni’s specific goal is to invent a computer that, when given a stack of scanned textbooks, can pass standardized elementary-school science tests (ramping up eventually to pre-university exams). To pass the tests, a computer must be able to read and understand diagrams and text. How the Allen Institute will make that happen is undecided as yet — but for Etzioni, neural networks and deep learning are not at the top of the list.

One competing idea is to rely on a computer that can reason on the basis of inputted facts, rather than trying to learn its own facts from scratch. So it might be programmed with assertions such as ‘all girls are people’. Then, when it is presented with a text that mentions a girl, the computer could deduce that the girl in question is a person. Thousands, if not millions, of such facts are required to cover even ordinary, common-sense knowledge about the world. But it is roughly what went into IBM’s Watson computer, which famously won a match of the television game show *Jeopardy* against top human competitors in 2011. Even so, IBM’s Watson Solutions has an experimental interest in deep learning for improving pattern recognition, says Rob High, chief technology officer for the company, which is based in Austin, Texas.

Google, too, is hedging its bets. Although its latest advances in picture tagging are based on Hinton’s deep-learning networks, it has other departments with a wider remit. In December 2012, it hired futurist Ray Kurzweil to pursue various ways for computers to learn from experience — using techniques including but not limited to deep learning. Last May, Google acquired a quantum computer made by D-Wave in Burnaby, Canada (see *Nature* **498**, 286–288; 2013). This computer holds promise for non-AI tasks such as difficult mathematical computations — although it could, theoretically, be applied to deep learning.

Despite its successes, deep learning is still in its infancy. “It’s part of the future,” says Dahl. “In a way it’s amazing we’ve done so much with so little.” And, he adds, “we’ve barely begun.” ■

Nicola Jones is a freelance reporter based near Vancouver, Canada.

1. Le, Q. V. et al. Preprint at <http://arxiv.org/abs/1112.6209> (2011).
2. Mohamed, A. et al. 2011 *IEEE Int. Conf. Acoustics Speech Signal Process.* <http://dx.doi.org/10.1109/ICASSP.2011.5947494> (2011).
3. Coates, A. et al. *J. Machine Learn. Res. Workshop Conf. Proc.* **28**, 1337–1345 (2013).
4. Krizhevsky, A., Sutskever, I. & Hinton, G. E. In *Advances in Neural Information Processing Systems 25*; available at go.nature.com/ibace6.
5. Girshick, R., Donahue, J., Darrell, T. & Malik, J. Preprint at <http://arxiv.org/abs/1311.2524> (2013).