



FINDING ONE FACE IN A MILLION

A new benchmark test shows that even Google's facial recognition algorithm is far from perfect

Helen of Troy may have had the face that launched a thousand ships, but even the best facial recognition algorithms might have had trouble finding her in a crowd of a million strangers. The first public benchmark test based on 1 million faces has shown how facial recognition algorithms from Google and other research groups around the world still fall well short of perfection.

Facial recognition algorithms that had previously performed with more than 95 percent accuracy on a popular benchmark test involving 13,000 faces saw significant drops in accuracy when taking on the new MegaFace Challenge. The best performer, Google's FaceNet algorithm, dropped from near-perfect accuracy on the five-figure data set to 75 percent on the million-face test. Other top algorithms dropped from above 90 percent to below 60 percent. Some algorithms made the proper identification as seldom as 35 percent of the time.

"MegaFace's key idea is that algorithms should be evaluated at large scale," says Ira Kemelmacher-Shlizerman, an assistant professor of computer science at the University of Washington, in Seattle, and

the project's principal investigator. "And we make a number of discoveries that are only possible when evaluating at scale."

The huge drops in accuracy when scanning a million faces matter because facial recognition algorithms inevitably face such challenges in the real world. People increasingly trust these algorithms to correctly identify them in security verification scenarios, and law enforcement may also rely on facial recognition to pick suspects out of the hundreds of thousands of faces captured on surveillance cameras.

The most popular benchmark until now has been the Labeled Faces in the Wild (LFW) test created in 2007. LFW includes 13,000 images of just 5,000 people. Many facial recognition algorithms have been fine-tuned to the point that they scored near-perfect accuracy when picking through the LFW images. Most researchers say that new benchmark challenges have been long overdue.

"The big disadvantage is that [the field] is saturated—that is, there are many, many algorithms that perform above 95 percent on LFW," Kemelmacher-Shlizerman says. "This gives the impression that face recognition is solved and working perfectly."

With that in mind, University of Washington researchers raised the bar by creating the MegaFace Challenge using 1 million Flickr images of 690,000 unique faces that are publicly available under a Creative Commons license.

The MegaFace Challenge forces facial recognition algorithms to do verification and identification, two separate but related tasks. Verification involves trying to correctly determine whether two faces presented to the facial recognition algorithm belong to the same person. Identification involves trying to find a matching photo of the same person among a million "distractor" faces. Initial results on algorithms developed by Google and four other research groups were presented at the IEEE Conference on Computer Vision and Pattern Recognition on 30 June. (One of MegaFace's developers also heads a computer vision team at Google's Seattle office.)

The results presented were a mix of the intriguing and the expected. Nobody was surprised that the algorithms' performances suffered as the number of distractor faces increased. And the fact that algorithms had trouble identifying the same person at different ages was a known problem. However, the results also showed that algorithms trained on relatively small data sets can compete with those trained on very large ones, such as Google's FaceNet, which was trained on more than 500 million photos of 10 million people.

For example, the FaceNet algorithm from Russia's N-TechLab performed well on certain tasks in comparison with FaceNet, despite having trained on 18 million photos of 200,000 people. The SIAT MMLab algorithm, created by a Chinese team under the leadership of Yu Qiao, a professor with Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, also performed well on certain tasks.

Nevertheless, FaceNet has so far performed the best overall. It delivered the most consistent performance across all testing.

Just seeing how Google's algorithm stacks up against those of competitors may be the challenge's most valuable result, says Stefanos Zafeiriou, a computer vision expert at Imperial College London. He and other researchers not involved with the MegaFace Challenge were impressed by the fairly consistent performance of FaceNet. On the other hand, its accuracy of 75 percent shows that even the best facial recognition algorithms may have trouble identifying faces on a "world scale" with millions or billions of distractor faces.

And, MegaFace could provide a focal point for future research efforts. Until now, most academic research groups have focused on improving their algorithms by using larger and larger training data sets rather than challenging them with larger benchmark data sets, says Jonathon Phillips, an engineer at the National Institute of Standards and Technology, in Washington, D.C.

The University of Washington researchers plan to release a training data set based on MegaFace photos for use by any researcher. That would help even the smallest academic teams marshal some of the resources already available to Silicon Valley giants.

"The more we can look at how these algorithms perform on giant data sets that are more characteristic of the images people have in their phones, the better," says Ross Beveridge, a computer scientist at Colorado State University, in Fort Collins. —JEREMY HSU

FOR SALE: DEEP LEARNING

Cloud-computing services
deliver AI to the rest of us



Facebook's deep-learning artificial intelligence systems have learned to recognize your friends in your photos, and Google's AI has learned to anticipate what you'll be searching for. But there's no need to feel left out, even if your company's computers haven't learned much lately.

A growing number of tech giants and startups have begun offering machine learning as a cloud service. That means other companies and startups do not need to develop their own specialized hardware or software to apply deep learning—the high-powered version du jour of machine learning—to their specific business needs.

"Deep-learning algorithms dominate other machine-learning methods when data sets are large," says Zachary Chase Lipton, a deep-learning researcher in the Artificial Intelligence Group at the University of California, San Diego, who has examined cloud AI services from companies such as Amazon and IBM. "Thus any company or application that has well-formed prediction problems—such as forecasting demand or translating between languages—could benefit from deep learning."

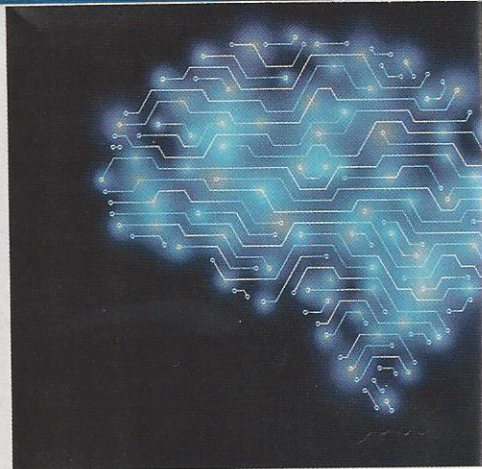
With cloud-based deep learning, companies can simply select a cloud service and browse its online offerings of application programming interfaces for software tasks such as recognizing images of corgi dogs or automatically translating a restaurant menu. Some services will even tailor their machine-learning

tools to the data and needs of individual companies.

According to Lipton, the rise of cloud services for machine learning hinges on at least two factors: first, a continued rise in the demand for machine learning as the technology has matured in its ability to solve a wide variety of problems with economic value; and second, the relative scarcity of machine-learning talent, which makes it tough for every company to build its own machine-learning team. Competition for talent has become even tougher with startups trying to compete with tech giants like Microsoft and IBM, which can afford to vacuum up the best and brightest.

Most commercial applications of machine learning rely on supervised learning. This involves algorithms that can observe correctly labeled examples and learn to perform certain tasks through imitation. Artificial neural networks are currently the most popular and successful algorithms for supervised machine learning on large data sets. They learn by passing information through an interconnected network of multiple nodes (also known as neurons). The connections between these nodes each have adjustable weights that influence the flow of information through the graph. Nodes are generally arranged in layers. But historically it was feasible to train networks with only one hidden layer of neurons in addition to the input and output layers.

Deep learning takes these methods to the next level by filtering the data through multiple layers of neurons,



ISTOCKPHOTO