Daniel O’Laughlin
CS 35L
Project 10

The Search for a Thinking Machine Research Report

Ideas concerning artificial intelligence date back 1950s. However, since then, AI has not merely remained an idea, but has become a reality. The latter part of the 20th century and the beginnings of the 21st century have bred incredibly advanced artificially intelligent machines such as IBMs DeepBlue, Sony’s AIBO pet robots, Siri, Google Now, and Cortana. However, a new type of artificially intelligent machine is on the horizon. Researchers are now attempting to allow machines to see.

Fei-Fei Li is a researcher at Stanford University who has spent the last 15 years of her career pioneering efforts to give computers vision. Li likens visual processing to a child learning how to see, maintaining that a child is not told how to see, rather he/she learns by exposure to real world examples. Thus, her approach to visual processing places more emphasis on providing existing algorithms with vast amounts of training data rather than improvement of the algorithms themselves. However, if we consider the analog of a three-year-old child, we see that this is no small task. We can think of a child’s eyes as a pair of cameras that take a picture of his/her environment every time an eye movement is made. On average this happens every 200 milliseconds. Thus, by the time a child is three years old, he/she would have cataloged hundreds of millions of pictures. This is a lot of training data (Wakefield, 2015).

Li sent out to conquer this colossal task by assembling a team of 50,000 workers from 167 countries for the purpose of cataloging millions of indiscriminate images. The images were compiled and used to create ImageNet, a database containing 15 million images across 22,000 categories. ImageNet has been employed by many research teams worldwide in attempts to develop seeing machines and has had tremendous success. Machines trained using ImageNet have become increasingly able to correctly interpret images, yielding as low as a 5% error rate (Wakefield, 2015).

These machines are able to interpret images via expansive extensions of neural networks. Neural networks are systems of data structures and programs that approximate the functionality of the human brain by running several parallel processes, each operating with its own small field of knowledge and given access to data in its local memory. These neural networks are trained using large data sets (Rouse). The determinations made by neural networks can be made based on different principles such as genetic algorithms or fuzzy logic. The artificial neurons that compose neural networks can be on the scale of millions and are arranged in layers. In the case of “seeing” machines, each layer is responsible for interpreting different elements of an image. For example one layer may be responsible for determining color, while another may determine shape. By the time each layer has carried out its assignment, this culmination of knowledge helps to demystify what the pixels that make up the image actually represent. Modern neural networks can contain as many as 30 layers (Wakefield, 2015).

An experiment carried out by three researchers at the University of Toronto helped to elucidate the utility value of deep convolution neural networks. The network used contained 8 learned layers- five convolutional and three fully connected. Using a subset of ImageNet coupled with supervised training, the researchers were able to achieve great success. Figure 1 displays test images and the five choices considered most probable by the machine.

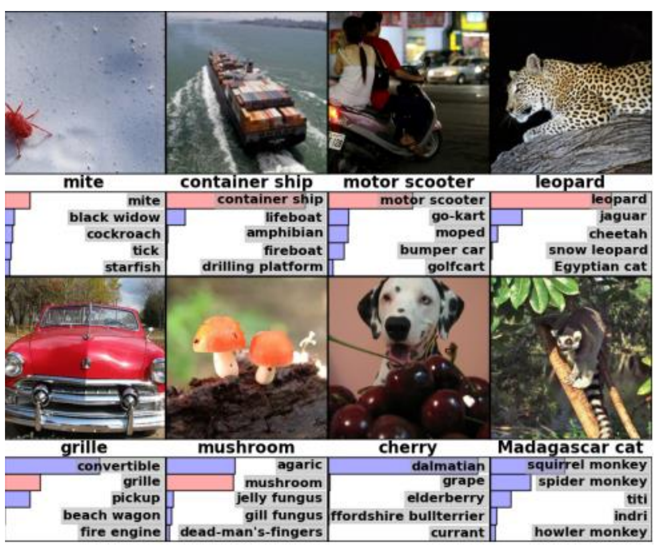

Figure 1: Eight test images are show. The top five most probable labels are displayed under each image with the correct label written under each image. The orange bar is the probability of the correct label that was assigned by the computer.

The top-5 error rate achieved in this experiment 18.2%. These are notably effective results considering the highly challenging nature of the data set chosen. Outside of the obtained results, a dependency between layers and correctness was realized. The program's performance was observed to degrade markedly if even a single convolution layer was removed, affirming that the depth of the neural network was crucial to achieving such successful results (Krizhevsky, Sutskever, Hinton, 2012).

Despite notable progress, much work still has to be done in order to improve computer vision. As of now, Stanford's machine only has the equivalent visual intelligence of a three-year-old. Also, the machine does not understand context. Li provides evidence of this reality by offering an example of machine interpretation without contextual understanding. When the computer was asked to interpret a picture of her son at a family party, the computer labeled it, "Boy standing next to cake." However, "What the computer doesn't see is that this is a special Italian cake that's only served during Easter time" (Wakefield, 2015). Thus, much work still needs to be done in order to allow machines to more accurately interpret images based on the relationship between objects within the image.

When perfected, one possible implementation of a seeing machine would be in the medical field, particularly in surgical applications. If done correctly, this could help increase, precision, safety, and efficiency of operations. Furthermore, Li speculates about the possibility of using such machines as a means to rescue people from disaster situations and generally improve people's lives for the better (Wakefield, 2015). Whether or not this is a realistic goal remains to be seen.

Artificial intelligence has progressed a great deal since the 1950s, but there is still a lot of work to be done. Modern machines are still not capable of thinking, right now they are just learning. There is debate about whether or not a thinking machine can be realized, as the nature of human thought has eluded scholars for centuries. However, one thing is sure, in this golden age of Artificial Intelligence, machines are only getting smarter.

References

Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems 25 (NIPS 2012)*. Retrieved November 24, 2015, from NIPS Proceedings, http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

Rouse, M. (n.d.). Neural network definition. Retrieved November 24, 2015, from http://searchnetworking.techtarget.com/definition/neural-network

Wakefield, J. (2015, September 17). The search for a thinking machine. Retrieved November 24, 2015, from http://www.bbc.com/news/technology-32334573