Bradley Zhu

304627529

Lab 5 Fan, R

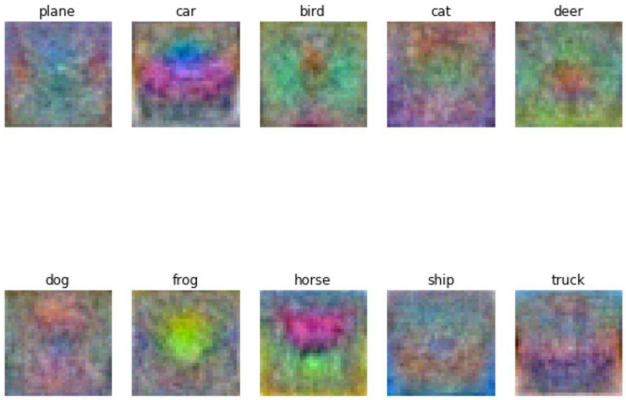
5 March 2016

Computers That See Like Humans

When asked to name the most interesting and upcoming field inside the broad category of computer science, both ordinary people and computer scientists will bring up Artificial Intelligence over and over again. However, Artificial Intelligence is a broad field that can be divided into many specific focuses, one of them being computer vision. As human beings, vision is crucial to our logic and understanding of the world, so it unsurprisingly follows that vision is one of the priorities of researchers studying Artificial Intelligence. The article *Digital Baby Project's Aim: Computers That See Like Humans* published on IEEE spectrum addresses current methods for computer vision, the problems associated with them, as well as a possible solution.

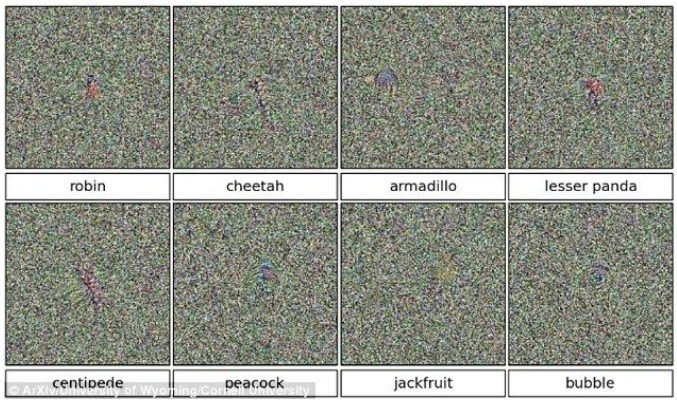
All digital images are consisted of many pixels that each have a color. Seeming like a simple task on the surface, programming a computer to be able to process and understand an image is actually a huge problem. Right now the most commonly used method for image processing is Deep Learning, specifically Deep Neural Networks (commonly referred to as DNN's). They usually involve several layers of electronic neurons that allows for higher and higher levels of abstraction. These layers of neurons combine in order to find patterns that might allow the network to distinguish between different images. A simple layer might just measure color differences in the upper left

hand corner, but a more complex layer might recognize vertical pillars of solid colors (which might be found in images of, for example, trees). Through the combined processing of all these neurons, DNN's can successfully categorize images almost to the same ability of humans when given a large enough training set of images and labels.



One algorithm to do this is the Softmax Regression algorithm which basically creates "average pictures". As seen in the diagram to the right supplied by a fellow student of the class, Ray Zhang, these pictures produced by Softmax Regression, though not accurate to a human, contain characteristics commonly found in pictures of the said objects. For example, you can almost see the four legs of the horse. A DNN will take an image and compare it with its "average pictures", and find the object that the given image is closest to, hopefully successfully categorizing the image.

The obvious flaw that appears in how DNN's work then is that their average pictures don't resemble the objects that they represent. While the current algorithms are very very good at distinguishing between different objects and even different locations, they unfortunately can label images as objects even when the images don't bear any resemblance to humans at all. According to the article *Images That Fool Computer Vision Raise Security Concerns* a group in Cornell reverse engineered images



specifically to fool visual networks. Starting from static, they slightly modified the image file every time the DNN labeled the image as an object, eventually ending up with the images shown above. Obviously none of these images are actually what the DNN is 99% confident in categorizing them as. This proves that DNN's can be tricked, and the more we rely on them the worse the potential risks are. For example, if in the future we rely on surveillance combined with DNN's to identify and track criminals, if they 3D print a fake nose and wear it, the DNN could possibly be unable to identify them correctly. A more extreme example could be a mask of abstract shapes that might even allow for complete invisibility to DNN tracking. Because DNN's are constantly learning, if they incorrectly identify an image but think they did it correctly, then they'll be learning in a wrong direction. If a group maliciously feeds search engines that rely on DNN's incorrectly labelled images, it would be possible to direct the DNN in specific ways to incorrectly label specific images, allowing images of the actual objects to potentially slip through filters and avoid identification. These risks are produced due to the inherent flaws in our current algorithms. Right now we train them with a focus on differentiating between objects, rather than with a focus on being able to produce an image of a specific object.

Shimon Ullman conducted a study of 14,000 participants with the aim of comparing the performance of DNN's with that of humans for identifying pictures. His results, not unsurprisingly, showed that humans are superior at correctly labeling images in general, but the difference in performance becomes more pronounced the lower resolution the image is. With his study, Ullman concluded that the reason for this increase in difference of performance for low quality images is due to the differences in

how humans process images and how computers do it. He believes that humans process images from “top-down”, comparing a standard model of certain objects with the particular object they’re trying to identify, while computers process images from “bottom-up”, filtering images based on the simplest features before using more complex ones. Ullman wants to create new computer models and algorithms capable of developing a complex understanding of the world they see based on the human “top-down” approach. He’s also now received funding for his ultimate goal of reverse engineering the infant mind. In his own words, “As a baby, you open your eyes, see flickering pixels, and somehow it all comes together and you know something about the world. You’re starting from nothing, absorbing information and getting a rich view of the world. We were thinking about what would it take to get a computer program where you put in the minimal structures you need and let it view videos or the world for six months. If you do it right, you can get an interesting system.” (Ullman).

The possible effects of this project are immense and broad. For starters, computer vision will be stronger and more accurate. This applies to many applications including everything from Facebook’s facial recognition, to driverless cars stopping suddenly if they recognize that a person is in front of the car. More complex consequences include producing a more complex and important Artificial Intelligence in general. This would be a significant advancement in the eventual development of Strong AI. Currently, we are very good at programming Weak AI, meaning Artificial Intelligence designed to do one specific task well. Everything from Deep Blue to Google are examples of Weak AI. Strong AI is Artificial Intelligence that can perform any task that a

human can do, as well or better than a human can do it. This is the ultimate goal of Artificial Intelligence to many researchers, and once we reach that point, Artificial Intelligence will be able to start improving itself to the point of what is commonly referred to as the “technological singularity”. In order to produce Strong AI however, the program needs to be able to learn and to understand, and understanding images would be a significant checkpoint in achieving that goal. Ullman’s project has the potential to permanently change the world.

Works Cited

Aron, Jacob. "Optical Illusions Fool Computers into Seeing Things." New Scientist. New Scientist, 11 Dec. 2014. Web. 02 Mar. 2016.

Basulto, Dominic. "Humans Are the World's Best Pattern-Recognition Machines, But for How Long?" Big Think. Big Think, 24 July 2013. Web. 02 Mar. 2016.

Hsu, Jeremy. "Digital Baby Project's Aim: Computers That See Like Humans." Computers That See Like Humans. IEEE Spectrum, 15 Feb. 2016. Web. 01 Mar. 2016.

Steele, Bill. "Images That Fool Computer Vision Raise Security Concerns." Cornell University. Cornell Chronicle, 20 Mar. 2015. Web. 02 Mar. 2016.

Vanhemert, Kyle. "Simple Pictures That State-of-the-Art AI Still Can't Recognize." Wired.com. Conde Nast Digital, 5 Jan. 2015. Web. 02 Mar. 2016.