From the
RETAIL TECH BULLETIN
Second Quarter 2019

Computer Vision Is Here. Now What?

By George Shaw, CEO/CTO, Brain of the Store

By now, you have probably heard about computer vision and recent advances in the field. Compared to even just 10 years ago when computer vision was in its infancy, deep learning software and algorithms now make it possible to accurately recognize objects, faces, and even emotions from common video streams. Additionally, the computer hardware that performs these computer vision operations is advancing at an unusually fast pace. Where do these developments lead? And what happens then? More importantly, what does this all mean for brick and mortar retail?

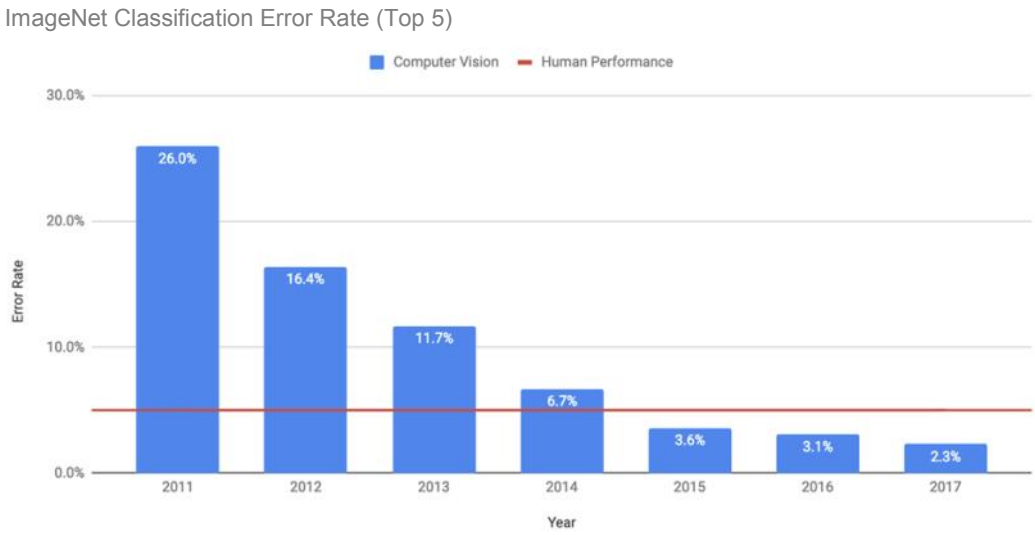
Computer vision was mentioned in 49 percent of all AI-related patents and grew by 24 percent from 2013 to 2016. Deep learning is the fastest-growing technique in AI, with a 175 percent increase between 2013 and 2016.[1]

It's useful to consider how the human brain processes vision, in order to understand current vision technology, as well as the likely roadmap for the years ahead. Even a small child can point out everyday objects in images, such as a dog, cat, car, or bus. If these objects are shown in different colors, from new angles, and in various lighting and photographic conditions, they are still recognizable. A child also has no trouble following people or objects as they move through a scene in a video sequence, or picking out details about those people or objects, such as the gender and approximate age of a person, or the color or size of an object.

Similarly, computer vision algorithms, fueled in large part by advances in deep learning and related technologies, have made incredible strides in recent years and are now more accurate than humans in performing many of these tasks. In one of the leading benchmarks, the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), deep learning algorithms first surpassed a 5 percent error rate, which is classified as human performance, in 2015, and as of 2017, was down to a 2.3 percent error rate. For this vision task, this means that algorithms are now more than twice as accurate as humans.[2] Results for other vision tasks across many different areas are similar, with algorithms gaining ground and surpassing humans.

[1] https://www.wipo.int/edocs/pubdocs/en/wipo_pub_1055.pdf.

[2] https://www.researchgate.net/figure/Winner-results-of-the-ImageNet-large-scale-visual-recognition-challenge-LSVRC-of-the_fig7_324476862.



Computers are now better than humans at visual tasks like tracking people, identifying products, and discerning human features such as age, gender, and even mood. As new products become available, teaching the machine to recognize and label those products will be faster and more straightforward.

Companies like Intel, Qualcomm, and Nvidia are producing increasingly powerful chips to perform deep learning analyses more efficiently, everywhere -- from cameras at the edge to servers in the cloud. Devices built specifically for retail from RetailNext, ShopperTrak, Xovis, and others are deployed to thousands of brick and mortar retail stores worldwide to detect the locations of shoppers. Dor, MotionLoft, and others can efficiently count how many people pass by a particular location or through the doorway of a store. Computer vision hardware is becoming increasingly sophisticated, specialized, and ubiquitous, especially in retail.

If computer vision is about to be ubiquitous and more powerful than human vision, what is next?

To start, let's revisit the human analogy. Once we've processed what our eyes take in, we begin to **think** about the results. Imagine our eyes, along with the visual cortex portion of our brains, as sensors that produce data about what's happening in the physical world. Similarly, systems consisting of video cameras and computers running deep learning algorithms can also be thought of as sensors that produce data from a visual representation of the world. Value can be unlocked when higher level reasoning is applied to raw sensor data.

Potential applications of higher reasoning to retail are plentiful. Given locations of customers and store associates, for example, machine learning might be used to alert a nearby sales associate when a sensor determines that a customer needs help. Given thousands of examples of "normal" movement patterns through a physical store, similar machine learning might be able to pick out "abnormal" behaviors and patterns, identifying customers who are acting suspiciously. With facial recognition, tracking, and emotion analysis (all derived from computer vision), we can easily imagine coaching store staff in real time as to who has walked into the store and for what purpose, when and how to approach this person, and then providing company management with metrics about the effectiveness of such interactions.

Why should brick and mortar stores care about computer vision? And what should those stores be thinking in five years?

In retail, there is a growing stockpile of raw data. More sensors are being added to brick and mortar stores every day, from purpose-built stereoscopic tracking cameras to RFID, UWB, iBeacon, and others. Existing surveillance cameras are also beginning to be utilized more readily to generate new data streams about where customers, associates, and products are and what they're doing in the complex dance that takes place in a store. It's now up to the technology community to build on this dataset, standing on the shoulders of giants to revolutionize the way stores operate.