Mihir Mathur
TA Tomer Weiss
CS35L
08 March 2016

Computer Vision

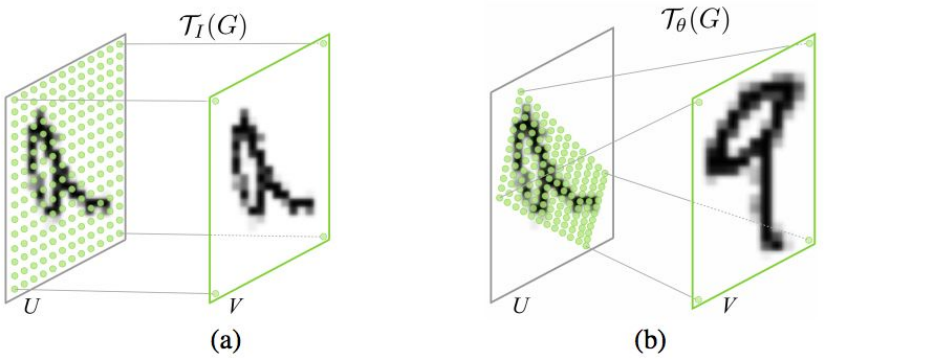
Google AI gets better at 'seeing' the world by learning what to focus on[1].

Computer Vision (CV) is a field of Computer Science that deals with acquiring, processing and analyzing visual data from the real world. It is closely connected to the fields of imaging, machine learning and signal processing and uses mathematics and physics extensively [2]. In recent years, Computer Vision has been used in an array of applications like scene reconstruction, event detection, video tracking, object recognition and object pose estimation. One example of an application of CV is Facenet—a system that identifies human faces with over 99% accuracy. It was trained on a 260 million image dataset. Another example is NASA's Mars Rover that uses CV for scene reconstruction and navigation. Due to large-scale research and development in CV, new technologies are becoming possible.

A challenge for Computer Vision is to learn 'what to focus on' in an image. For instance, Google had a technical challenge to detect house numbers and street signs from Google Street View Images. Google Deepmind was testing its algorithms on the publically available Street View House Numbers dataset. This task was challenging because the important information is small and can be present anywhere on a large image. This information can be blurry, distorted and skewed. To overcome this challenge, Google uses a Machine Learning approach called Convolutional Neural Networks. This approach uses several small neuron collections that independently work on parts of a photo. The neurons are like biological neurons that process

[1] "Google AI Gets Better at 'seeing' the World by Learning What to Focus on." *TechRepublic*. Web. 15 Feb. 2016.
[2] "Computer Vision." Wikimedia Foundation. Web. 16 Feb.2016<https://en.wikipedia.org/wiki/Computer_vision>

information, then send it to the next artificial neuron. The outputs are then tiled together to give a better representation of the image. Using Convolutional Neural Networks allows translation of parts of a photo since units are processed separately. Google created a spatial transformer module for their neural network [3] which significantly decreased the error rate in identifying 5 digit house numbers and increased the accuracy of detecting bird species by enabling their bodies and heads to be singled out and processed.



[3]

a) Sampling grid (U) to output (V)
b) Sampling grid (U) to transformer to output (V)

The next thing for the field of Computer Vision is to make computers see and process what humans cannot. For instance, UCLA researchers recently released an open sourced image detection algorithm called the "The Phase Stretch Transform algorithm"[4]. It is based on a research about 'photonic time stretch' which is used for detecting cancer cells in blood. This new technology has lots of potential applications in the field of medicine and for detecting internal structures of things. Thus the possibilities in the field of Computer Vision are limitless.

[3] Jaderberg, Max, and Karen Simonyan. "Spatial Transformer Networks." *ArXiv* (2016): 1-15. *ArXiv*. Web. <http://arxiv.org/pdf/1506.02025.pdf>.
[4] "UCLA Just Open-sourced a Powerful New Image-detection Algorithm." *InfoWorld*. Web. 15 Feb. 2016.