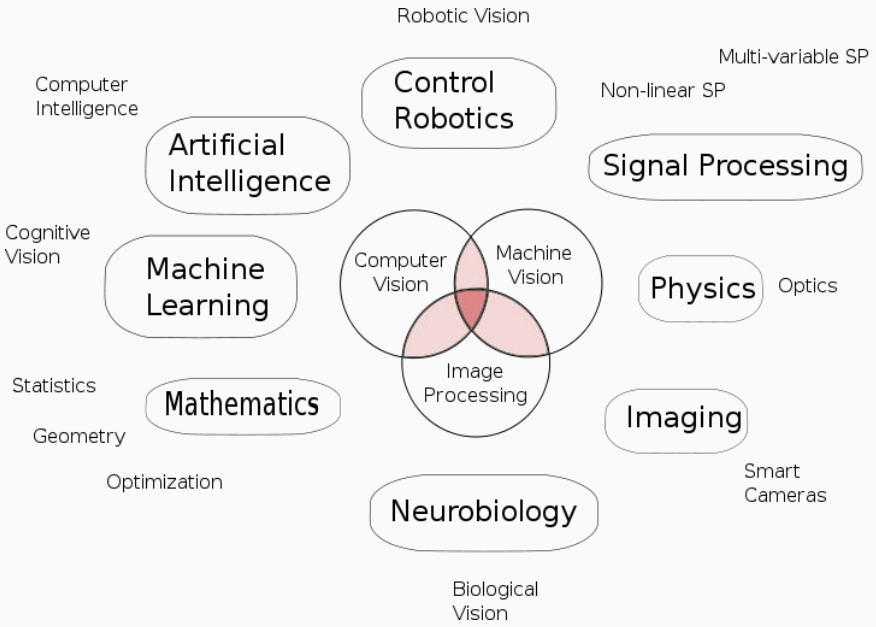
COMPUTER VISION

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

What is Computer Vision?

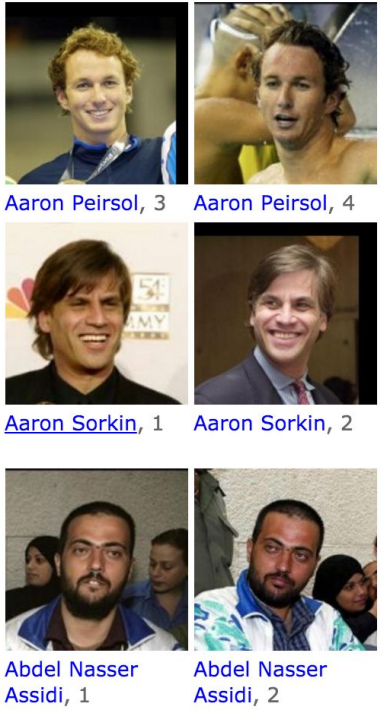
It is a field of Computer Science that deals with acquiring, processing and analyzing visual data from the real world.

Subfields: scene reconstruction, event detection, video tracking, object recognition, object pose estimation.

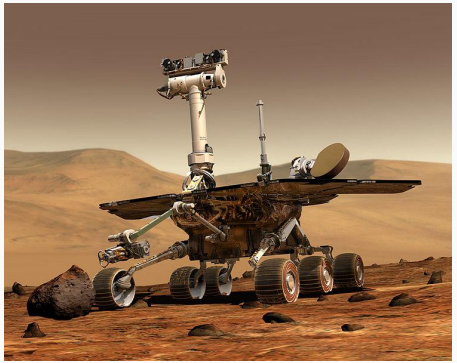


What can CV currently do?

- FaceNet - identifies human faces with over 99% accuracy. Was trained on a 260 Million image dataset.
- Navigation and scene reconstruction for autonomous vehicles, robots and drones.
- Tracking a set of moving points in an image sequence.



[6]



Google's Challenge

Spotting house numbers and street signs from Street View images.

There is a database (SVHN) of 200,000 images, many of which are blurry, for testing.

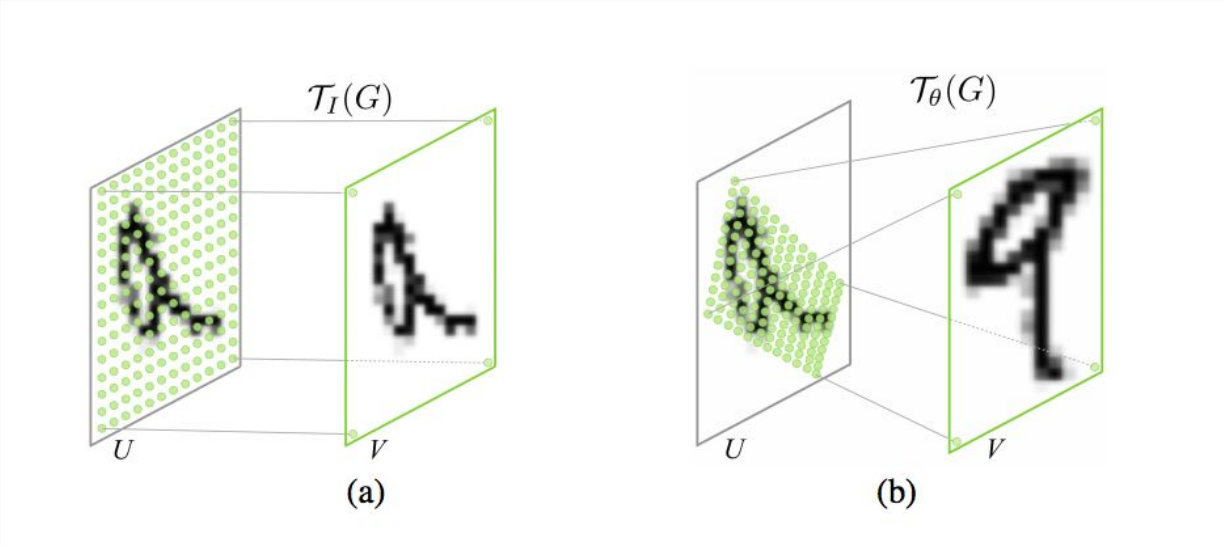
Technical Challenge:

The important information is small and can be present anywhere on the image. This information can be blurry, distorted and skewed.

Solution: Learning what to focus on

- Google used a Machine Learning approach called **convolutional neural network (CNN)**
- This approach uses several small neuron collections that independently work on parts of a photo.
- The outputs are then tiled together to give a better representation of the image.
- CNN allows for translation of parts of a photo.

Google created a **Spatial Transformer** module for neural network[1]



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

Success after adding Spatial Transformer module

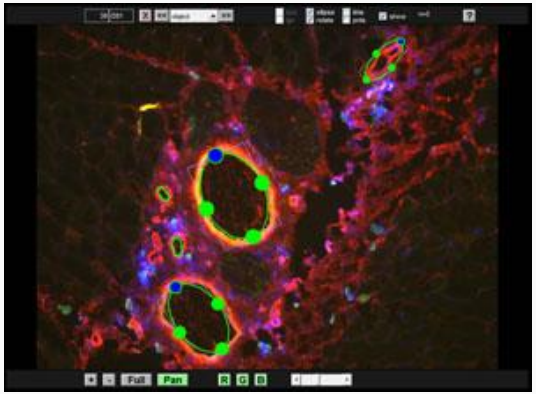
- Error rate in identifying upto 5 digit house numbers was significantly reduced
- Traffic signs were easily focussed on and background was removed.

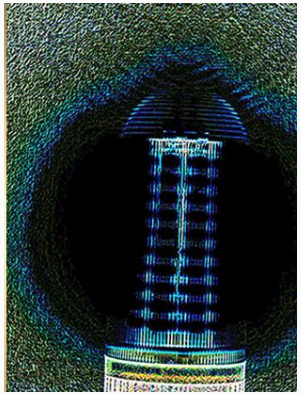
- Species of birds could be detected with 84% accuracy by singling out their heads and bodies.

[5]

What’s Next?

MAKING COMPUTERS SEE AND PROCESS WHAT HUMANS CANNOT



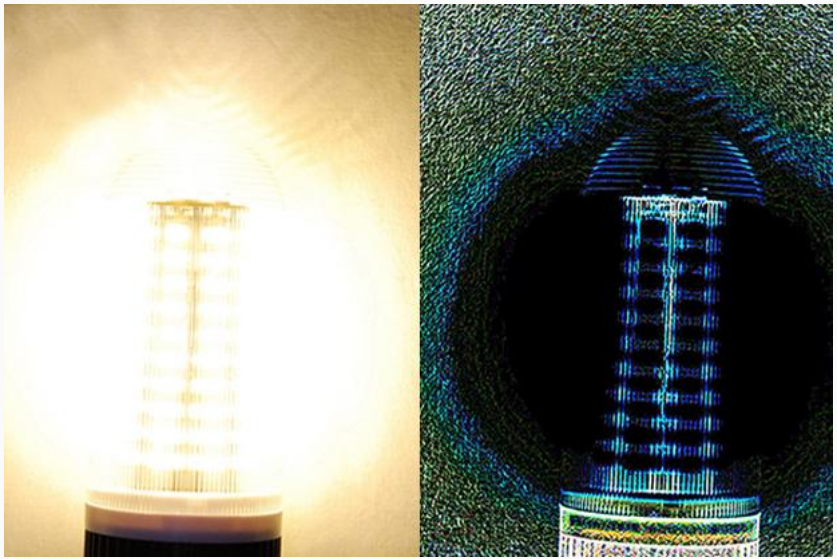


UCLA’s new open-sourced image-detection algorithm

“**The Phase Stretch Transform algorithm**” is based on a UCLA research about ‘photonic time stretch’ which is used for detecting cancer cells in blood.

Performs mathematical operations to detect edges and other features which may be obscured to the eye.

Example: Detecting the internal structure of a light bulb.



This algorithm is available on Github:
https://github.com/JalaliLabUCLA/Image-feature-detection-using-Phase-Stretch-Transform

Citations

[1] Jaderberg, Max, and Karen Simonyan. "Spatial Transformer Networks." *ArXiv* (2016): 1-15. *ArXiv*. Web. <http://arxiv.org/pdf/1506.02025.pdf>.

[2] "Google AI Gets Better at 'seeing' the World by Learning What to Focus on." *TechRepublic*. Web. 15 Feb. 2016.

[3] "UCLA Just Open-sourced a Powerful New Image-detection Algorithm." *InfoWorld*. Web. 15 Feb. 2016.

[4] “Convolutional Neural Network.” *Wikimedia Foundation*. Web. 15 Feb. 2016. <https://en.wikipedia.org/wiki/Convolutional_neural_network>.

[5] "Caltech-UCSD Birds-200-2011." *Caltech-UCSD Birds-200-2011*. Web. 16 Feb. 2016. <http://www.vision.caltech.edu/visipedia/CUB-200-2011.html>.

[6] "Labeled Faces in the Wild." *LFW Face Database*. UMass, Computer Science. Web. 16 Feb. 2016. <http://vis-www.cs.umass.edu/lfw/>.

[7] “Computer Vision.” Wikimedia Foundation. Web. 16 Feb. 2016 <https://en.wikipedia.org/wiki/Computer_vision>.