Deep Learning for Computer Vision - A crash course

Shubham Arora

Contents

1 Standard tasks in Computer Vision

- Recognition
  - Image classification
  - Image captioning
  - Object localization
  - Segmentation
  - Object Detection
- Motion Analysis
  - Tracking
  - Optical flow
- Other
  - Image recolorization
  - Super-Resolution

2 Important Ideas of Deep Learning-Vision. A historical perpective

2.1 Convolutional Neural Networks (CNN)

LeNet 1998 - LeNet-5 was one of the first NN that utilized backpropagation using Supervised Learning. CNN's are by far the most used Neural layer architecture in any machine vision tasks. Yann LeCun was a co-recipent of the 2018 Turing Award for his work in Ai & vision, owning a lot to the success of CNN in practical applications.

How does it work

- ConvNet architecture is particularly optimized for images - vastly reduces the number of parameters neede to train the network.
- essentially a combination of matrix dot products and $max()$ operations
- Function from raw pixels $->$ n numbers (classs scores in the case of image classification)

2.2 Deep Convolutional Neural Networks (CNN)

AlexNet 2012 - Image classifier on the ImageNet database. It Built upon the 1998 work on CNN, it was scaled massively due to availability of exponentially more data and parallel compute using GPU's

Major Improvements

- Reduced code complexity as it utilized homogeneous architectures
- No need to do complex step of feature extraction
- Made it easier to do Transfer Learning

Industry applications

- Face Recognition
- Self driving cars
- Image captioning?
- Building block in Reinforcement Learning

2.3 Residual Networks (ResNet)

ResNet 2015 - Microsoft Research - Much better performance than plain deep neural networks. Utilized the idea of skip connections across non sequential layers. Won the ImageNet challenge in 2015.

3 Why vision is important?

- Vision and NLP have been the core of Machine Learning innovation
- Vision is a very important ability in robotics for perception, localization, mapping and motion planning tasks.
- Automate critical tasks that rely on human vision - fault detection, autonomous vehicles

4 Important Applications

4.1 Object Detection

It is the task of assiging a label to an image from a fixed set of categories.

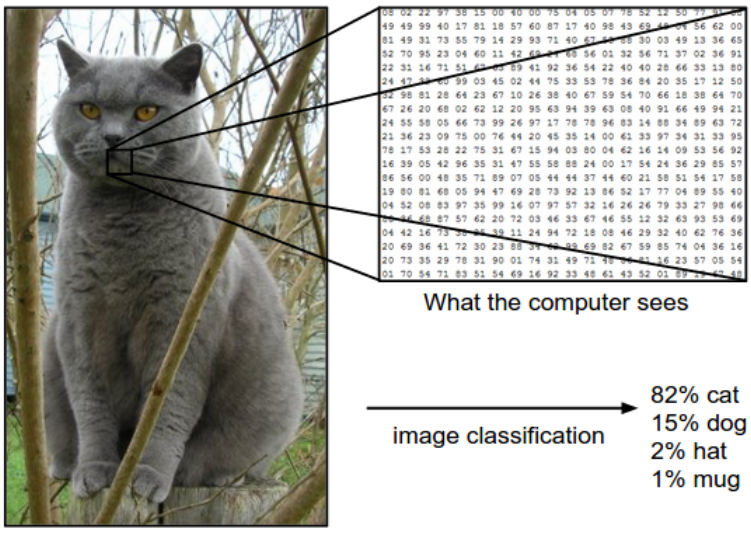

Figure 1: Example of classifying a cat image

As shown in Figure:1, thse system outputs various probabilities for different possible categories.

4.2 Human Activity Recognition

The task of identifying specific movement or action using data from visual sensors. These activities are like: walking, jumping, talking, standing, etc. The task can be online or offline depending upon the real time nature of the application. With the dawn of cheap consumer electronics like smartphones and cameras, lots of data is available now, which made Deep Learning viable and very effective.

5 Important Tools

- OpenCV
- Keras
- Tensorflow

6 Refrences

- Andrej Karpathy
- Wikipedia
- Stanford - cs231n

7 Peer Review

7.1 Thuy Pham

Topics covered

7.2 Bowen Qin

Topics covered Bowen does a deep analysis of the TensorFlow platform for use in creating Machine Learning models using Neural Networks. He discusses the following:

- How Tensorflow has tooling and support for multiple programming languages, neural network arcchitectures and deployment environments
- An analysis of common tasks performed using TF - Image classification and segmentation using CNN's
- How TF enables quick learning, prototyping and development

Learnings I learned how Tensorflow is much more than an ML library. The community and tooling support for TF alongwith backing by Google makes it a very lucrative ML tool.

7.3 Mengting Song

Topics covered Mengting elaborates on the various tasks in Object detection, classification and segmentation. He discusses the following:

- Different approaches in detection, namely CNN's, YOLO and SSD
- The different outputs that we expect in the 3 cases of detection, classification and segmentation.
- Detailed explanations and motivation of the different CNN architectures, and a comparison between them in object detection.

Learnings I learned about the various state of the art techniques in object detection. I got to know more about the evolution of CNN architectures, along with their strength and limitations.

7.4 Ningrong Chen

Topics covered Ningrong gives an overview of Object detection