

# CS231N Stanford Lecture 1 Notes

Eric Xia

March 3, 2018

CS231-N is computer vision: the study of visual data. The amount of image data has drastically increased in the past, due to the increased number of sensors (think about smart phones). This produces a vast amount of image data. About 80% of the data on the internet today is visual data.

Visual data is very hard to understand. It can be likened to "dark matter", which consists of most of the matter in the universe but we don't really understand it. There exists a significant amount of visual data, but many algorithms don't do analysis of this data very well.

An important problem in computer vision is object recognition. One example is the PASCAL Visual Object Challenge, which was a dataset of images and 20 categories, and the goal was to develop algorithms that are capable of distinguishing categories.

One issue with machine learning algorithms for image data is that usually image data is very complex, so these algorithms have a tendency to overfit. Researchers organized a project called ImageNet, which is a data set of 14 million images organized in 22,000 categories. An international challenge was started, called the "Large Scale Visual Recognition Challenge", which was to test image classification results of a subset of the data. The challenge consisted of the algorithm outputting 5 classification labels, and it must contain the true category. The best algorithm from a few years ago had an accuracy that was on par with the performance of humans (the poor soul that had to go through 1.4 million images).

In this challenge, for the first two years the error rate hovered around 25%, but the next year it rapidly dropped, due to the advent of convolutional neural networks. This model became an important part of modern day computer vision.

CS231-N is focused on image classification. Other problems will also be considered, such as object detection, action classification (produce a sentence that describes the image), captioning, and so on.

For the past few years, the winner of ImageNet have always been deep and deeper convolutional neural networks. At a certain point, it gets too deep and the model becomes too large before it becomes computationally unreasonable.

Convolutional neural networks was established by a paper from Yann LeCun and others in 1998. It was originally developed to take in an image of a character and classify what digit or letter it was. The reason that even though this model has been around for 20 years, it has only gotten popular reasonable thanks to more powerful processors (and the use of GPUs which are specialized to do matrix computations). Another key reason is the abundance of data now: these models were very complex and would not work well on small data sets. ImageNet (and PASCAL) were large and well labeled data sets that allowed people to train very complex networks on.

The important part of computer vision is for a computer to be able to look at an image and understand it the way in which we humans can. Computers are still far from being able to see an image and "understand" it.