Machine Learning Engineer Nanodegree

Capstone Proposal

Rachit Kumar Agrawal 1st Sept 2017

Proposal

(approx. 2-3 pages)

Domain Background

(approx. 1-2 paragraphs)

I have picked up the domain of Computer Vision for my capstone project. In last decade, the amount of data has exploded exponentially. These data are in different forms - text, image and video. With each of these forms of data exploiting it becomes extremely important to categorize them or classify them and classification of these images is not humanly possible. So, it become important to let machine do it for us. I am personally motivated for this problem because, I feel this is first step to my ultimate goal to understand the Real Time Object Detection for Autonomous Driving.

Problem Statement

(approx. 1 paragraph)

The problem I am trying to solve is - "Image Classification on CIFAR-10 dataset". The goal is to recognize objects in the image and classify each of this image into 10 object classes.

Datasets and Inputs

(approx. 2-3 paragraphs)

For this problem, I am using CIFAR-10 image dataset. CIFAR-10 is an established computer-vision dataset used for object recognition. It is a subset of the 80 million tiny images dataset and consists of 60,000 32x32 color images containing one of 10 object classes, with 6000 images per class. It was collected by Alex Krizhevsky, Vinod Nair, and Geoffery Hinton.

The dataset is divided into five training batches and one test bench, each with 10000 images. The test bench contains exactly 1000 randomly-selected images from each class.

Here are the classes in the dataset, as well as 10 random images from each:
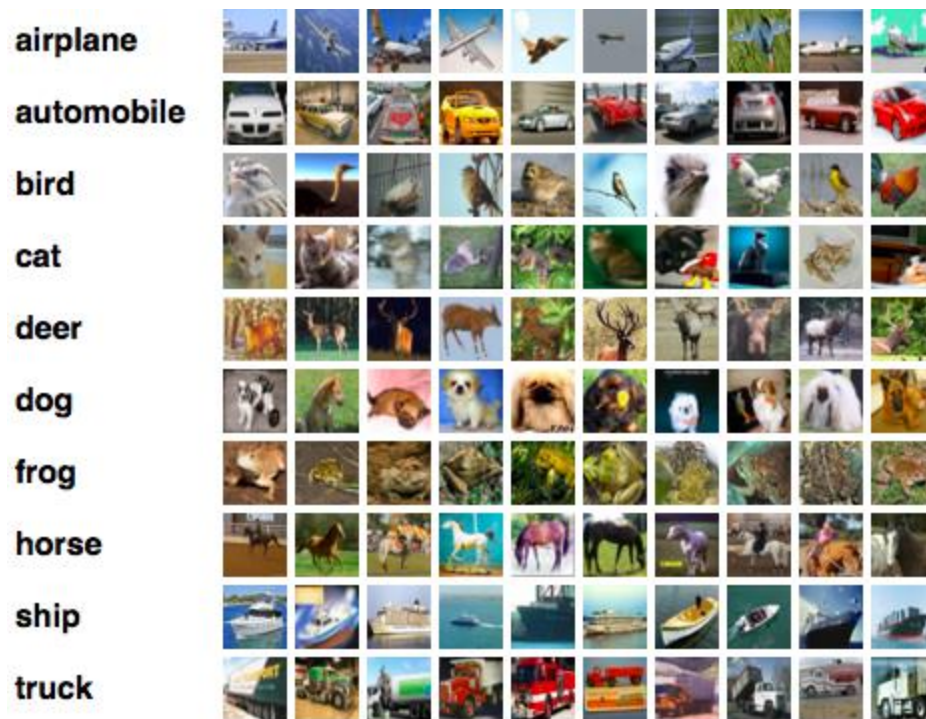
The classes are completely mutually exclusive. There is no overlap between automobiles and trucks. "Automobile" includes sedans, SUVs, things of that sort. "Truck" includes only big trucks. Neither includes pickup trucks.

Solution Statement

(approx. 1 paragraph)

I plan to use Convolution Neural Network (CNN) for this project because it's known to be best suited for any image dataset with the Network architecture proposed by Alex Krizhevsky, with a few differences in the top few layers.

Benchmark Model

(approximately 1-2 paragraphs)

CIFAR-10 is one of the popular Object Classification benchmark and there are many solutions proposed on the same. The best performing method is Fractional Max Pooling by Benjamin Graham [1]. This model by Benjamin achieves an accuracy of 96.53%.
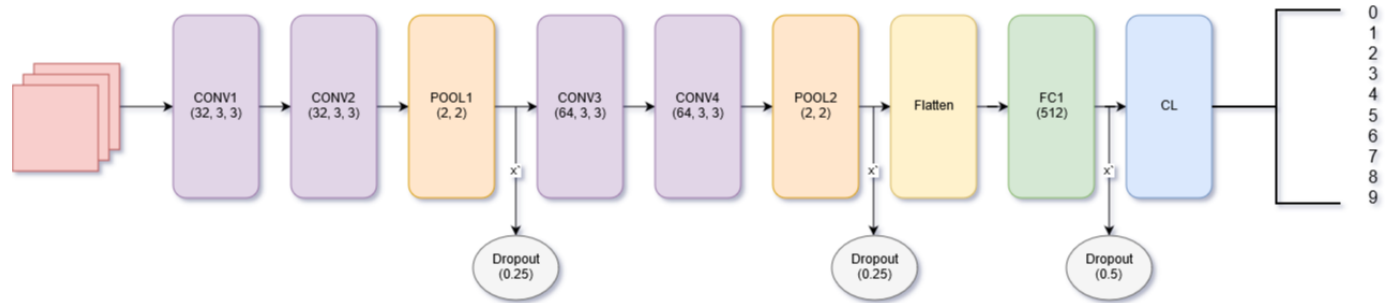
Evaluation Metrics

(approx. 1-2 paragraphs)

Model is evaluated on the 10,000 images in the evaluation set of CIFAR-10. It calculates the accuracy based on – how often the top prediction matches the true label of the image.

Project Design

(approx. 1 page)

The CNN architecture for this work consists of two Convolution layer with a ReLU activation function followed by a MaxPooling Layer. This then passed through another 2 CNV layers followed by a MaxPooling layer. Towards the end it's passed through 1 fully connected layer and at the end a softmax transformation to obtain logits.

References

1. Graham, Benjamin. *Fractional Max-Pooling*. In *arxiv:cs/arXiv:1412.6071*, 2015.