Machine Learning Engineer Nanodegree

Capstone Proposal

Benjamin Rouillé d'Orfeuil

December 27, 2017

Domain Background

Convolutional Neural Networks (CNNs) have successfully been applied in the field of image recognition. These algorithms have proven to be incredibly efficient at classifying images and often outperforms other machine learning algorithms at this task. To illustrate, very accurate predictions can be achieved on the well known MNIST database of handwritten digits[1] and the CIFAR-10 dataset[2] using a very simple network architecture. These datasets, however, are quite simple. Not only the number of classes is very low (10 digits for the MNIST database and 10 object categories for the CIFAR-10 dataset) but each class is vey different from one another.

Image recognition on real world images requires the development of deep CNNs. To illustrate, the ImageNet Large Scale Visual Recognition Challenge[3] (ILSRVC) have led various teams to build very complex models to accurately classify a large number of photographs collected from search engines. Unlike the MNIST database, the ILSRVC dataset has a large number of classes and the difference between the object categories can be very tenuous. Among the 1000 classes in this dataset, 118 are dog breeds. It is hence not surprising that the ResNet50 model developed by Microsoft and that won the 2015 ILSRVC edition has 50 convolutional layers and a total of 168 layers.

Problem Statement

Datasets and Inputs

Solution Statement

Benchmark Model

Evaluation Metrics

Project Design

[1] The MNIST database is available at `http://yann.lecun.com/exdb/mnist/`. A quick analysis of this dataset can be found here.

[2] The CIFAR-10 dataset can be found at the following url: `https://www.cs.toronto.edu/~kriz/cifar.html`. Predictions on this dataset are presented here.

[3] The 2017 challenge is described on the ImageNet website: `http://image-net.org/challenges/LSVRC/2017/index` and on Kaggle: `https://www.kaggle.com/c/imagenet-object-localization-challenge`.