Machine Learning Engineer Nanodegree

Capstone Proposal

Benjamin Rouillé d'Orfeuil

January 2, 2018

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 very simple network architectures. These datasets, however, are relatively 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, on the other hand, requires the building of complex models. Deep CNNs developed for the ImageNet Large Scale Visual Recognition Challenge[3] (ILSRVC) is a perfect illustration. Unlike the MNIST and the CIFAR-10 databases, the ILSRVC dataset has a large number of classes and the difference between some of the object categories can be very tenuous. The algorithm needs to be able to discriminate between different breeds of dog or types of snake. It is hence not surprising that the ResNet50 model[4] developed by Microsoft and that won the 2015 ILSRVC edition has 50 convolutional layers and a total of 168 layers.

It is common practice today to use pre-trained state of the art models to classify photographs. CNNs that have been pre-trained on a large and diverse dataset like ImageNet captures universal features in its early layers that are relevant and useful to most classification problems. The weights of the pre-trained CNNs can then be fine-tuned by continuing training it on the dataset under study.

Problem Statement

Yelp is a social networking site that publishes crowd-sourced reviews about local businesses. About two years ago, Kaggle hosted the Yelp Restaurant Photo Classification challenge (see `https://www.kaggle.com/c/yelp-restaurant-photo-classification`). Since labels are optional during the review submission process, some restaurants can be left uncategorized. For this reason, Yelp asked competitors to build a model that automatically predict attribute labels for restaurants using their user-submitted photographs. The goal of this project is to build such a model.

Datasets and Inputs

The photographs and attributes for this project can be found in the data section of the competition webpage on Kaggle: `https://www.kaggle.com/c/yelp-restaurant-photo-classification/data`. Yelp provides for this competition a training (234,842 photographs) and a test (1,190,225 photographs) datasets. Each image is

[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`.

[4] The Microsoft team presents their model in the following paper: `https://arxiv.org/pdf/1512.03385.pdf`.

mapped to a business identification number. There is a total of 2000 businesses (restaurants) that can be tagged with 9 different attributes. The labels are given below:

0. good_for_lunch
1. good_for_dinner
2. takes_reservations
3. outdoor_seating
4. restaurant_is_expensive
5. has_alcohol
6. has_table_service
7. ambience_is_classy
8. good_for_kids

Solution Statement

A convolutional deep learning network will be built using transfer learning to tag the restaurants. For this purpose, we will use the keras high-level neural networks Python library with TensorFlow as a backend.

Benchmark Model

Evaluation Metrics

Project Design