ImageNet Object Detection Challenge

Machine Learning Project Status Report

By

HARSH BHALANI(hpb170030)
MAITRI SHAH (mxs172030)
NIDHI VAISHNAV(ntv170030)
SHIVANI THAKKAR (sdt170030)

Eric Jonsson School Of Engineering and Computer Science
UNIVERSITY OF TEXAS AT DALLAS

NOVEMBER 2017

1 Introduction and Problem Description

ImageNet Object Detection Challenge is an official Kaggle Competition problem. It is an estimate that by the end of 2017, there will be 1.2 trillion images available online. If someone takes around 1 second to annotate, then also it will take around 38000 years to classify them all. Researchers have been working effectively to design efficient algorithms for automatically classification of images. But due to the limitation of good quality of datasets, many researchers are not successful to accomplish this task.

ImageNet dataset is provided which contains large number of human annotated images. This massive dataset gives all researchers an opportunity to analyse different techniques for automatic image classification. Moreover, ImageNet dataset contains variety of images containing a great amount of distinct objects.

A good commercial use of image classification can be found in the field of stock photography and video. Stock websites gives platforms to sell photos and videos. Contributors require a method to tag many visual materials, which consumes large amount of time and is even tedious.

If any visual database contains metadata about the images, a huge tedious task of categorizing them can be avoided. Classification of images with the help of machine learning is a prime solution for this. With image classification, companies can easily classify and categorize their database. This helps them to manage their visual database without investing large amount of hours for manual sorting and tagging.

2 Dataset Details

The dataset contains around 200 categories of objects. These categories are selected based on factors like average number of objects contained in any image, scene-based and many others.

2.1 Training dataset

The training set contains around 475,000 objects for classification from around 450,000 images. Training dataset contains three folders

- Annotations
  - This folder contains different folders for years 2013 and 2014. In these folders, there are xml formatted files for each particular image.
  - This xml file contains general information like image filename, folder name and size of image.

- It also contains specific information for each object contained in that image. Corresponding to each image object, its name, xmin (minimum x co-ordinate), xmax (maximum x co-ordinate), ymin (minimum y co-ordinate) and ymax (maximum y co-ordinate)

- Data
  - This folder contains several folders for years 2013 and 2014.
  - Each of these folders contains images.

- ImageSets
  - This folder contains 200 text files corresponding to 200 categories.
  - This test file constitutes paths of images belonging to that category.

2.2 Testing data

Testing dataset contains two folders: •

Data

- This folder contains 65,500 images to perform testing.

• ImageSets

- This folder contains text file containing names of 65,500 images spanning across 200 categories.

2.3 Features and Instances

In this dataset, the features used for the task of image classification is the image itself. Instances are also images.

3 Techniques

Images in the training dataset constitutes multiple images. So, to classify different images, we first planned to extract individual objects from each image. For extraction of individual objects, we thought of utilizing the annotated file given in xml format. We planned to train our model using images containing single objects.

4 Experimental Methodology

4.1 Pre-processing Methodology

For pre-processing, we used Annotation folder of the training dataset. For each xml formatted file, we extracted image file database, folder name, filename and features of each object like xmin (minimum x co-ordinate), xmax (maximum x co-ordinate), ymin (minimum y co-ordinate), ymax (maximum y co-ordinate) and its unique name.

For each individual object in each image, we created a separate image and stored it in a folder. For training purpose, these images containing a single object are used.

Figure 1: Original Image containing multiple objects.

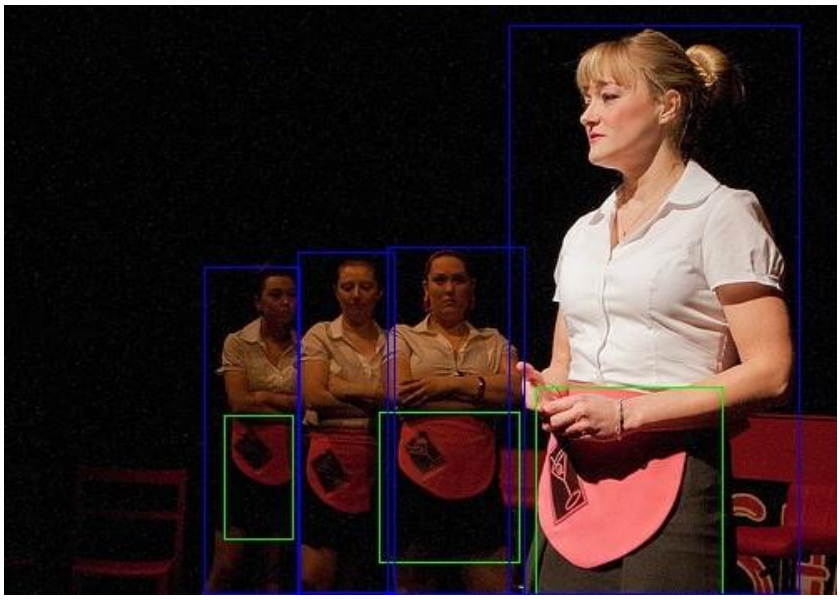

Figure 2: Segmented Image according to the co-ordinates given in xml format image containing multiple objects

(a) (b) (c)

(d) (e) (f) (g)

Above 7 images are the cropped images containing a single object each which can be given as an input to the training model.

4.2 Dataset Splitting

For evaluation of any training model, it should be first modelled on training data and tested on testing data. In our dataset, we have been provided three different datasets- training, validation and testing. For our convenience, we first used training dataset only. So, we used hold-out validation approach to split training dataset into training and testing.

5 Coding Language/technique to be used

We planned to use Python language for our problem. We explored some of the machine learning libraries and planned to use mainly keras, tensorflow, scikit-learn for our project.

We are planning to use Convolutional Neural Networks to train our model. We want our model to be able to distinguish between all images and figure out unique features to identify objects. The general view of a Convolutional Neural Network is that a series of convolutional layers performs image classification by looking for low level features like curves and edges.

6 Preliminary Results

We are currently working on training our model using Convolutional Neural Networks yet not received appropriate results.