Machine Learning Engineer Nanodegree

Capstone Proposal

Khalid Aljaghthami
January 12th

1. Domain Background :

Detecting objects in images is an active research area by many researchers and companies. There are many innovative research in this area. Some of the research includes detecting the kind of objects which are in the pictures/videos such as boy or animal. The prominent Facebook tagging feature is another example which detects faces in an image and prompt the user to tag a friend.

One of the most interesting research areas I have found is logo recognition. This is a really interesting idea and comes along the research paradigm which eventually will benefit ends users by making information much easier than before.

2. Problem Statement:

The goal of this project to predict a logo name of an image given to the model. This addresses a real problem encountered many people in their daily life. Many times, we come across a problem where we want know to the brand of a product or service we liked or disliked. The situation get worse when the logo is an image not containing any character. Such model can help in recognizing logos or many taken a step further by incorporate it into a game. In summary, the problem is a classification problem where the input is an image and the output is a predicted logo.

3. Datasets and Input

As more broad and even greater size than the **FlickrLogos-32** proposed in the initial proposal, the **QMUL-OpenLogo** dataset will used a basis of our inputs. The dataset consists of real-world images depicting company logos in various circumstances. It is maintained by Queen Mary University of London. The datasets contains photos showing brand logos and is meant for the evaluation of logo retrieval and multi-class logo detection/recognition systems on real-world images.

QMUL-OpenLogo has 352 logo classes. The images are aggregated and refined from 7 existing datasets and establishing an open logo detection evaluation protocol. All images are colored, in JPEG format, containing different dimensions .

QMUL-OpenLogo is around supplying the model with varying data examples per class. So, their goal is to generate a model predict brand logo even if there is small training set. For the sake of simplicity, we will limit the evaluation to the number of classes to 10 classes contains the most training numbers including **Adidas, Aldi, Apple, BMW, Cannon, Caterpillar, Chimay, DHL, Google, Intel.**

4. Solution Statement

The solution will use Convolutional Neural Network for the classifying the images. The solution will use a pre-built and pre-trained model applying transfer learning. The data will be split according to the proposed split by the **QMUL-OpenLogo** for fully supervised Split. With every logo classes contains 70% of training split and 30% of evaluation split. The below table shows the total number of images per set. The second proposed split is Semi-Supervised. We will also try to implement in Semi-Supervised Split if time permits.

Fully Supervised Split

Classes	Train Images	Val Images	Test Images	Total Images
352	15,975	2,777	8,331	27,083

Semi-Supervised Split

Type	Classes	Train Images	Val Images	Test Images	Total Images
Supervised	32	1,280	1,019	9,168	11,467
Unsupervised	320	0	1,562	14,054	15,616
Total	352	1,280	2,581	23,222	27,083

5. Benchmark Model

As per the challenge leaderboard, there 6 methods applied to this problem. While 5 methods used Semi-Supervised, 1 method used Fully Supervised method. The table below shows the result.

Semi-Superivised Result

Method	All Class	Unsupervised Class	Supervised Class
YOLO9000 [1]	4.19	1.98	26.33
YOLOv2 [1]+SCL [2]	12.10	8.75	45.48
YOLOv2 [1]+CAL [4]	**13.14**	**9.55**	49.17
FR-CNN [3]+SCL [2]	12.35	8.51	50.74
FR-CNN [3]+CAL [4]	13.13	9.34	**51.03**

Fully Supervised Result

Method	mAP
FR-CNN [3]	**48.3**

[1]. Jpseph Redmon and Ali Farhadi. Yolo9000: better, faster, stronger. CVPR 2017.

[2]. Hang Su, Xiatian Zhu and Shaogang Gong. Deep learning logo detection with data expansion by synthesising context. WACV 2017.

[3]. Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real time object detection with region proposal networks. NIPS 2015.

[4]. Hang Su, Xiatian Zhu and Shaogang Gong. Open Logo Detection Challenge. BMVC 2018.

6. Evaluation Metrics

We will follow the same principle as the leaderboard show where for Fully Supervised Result, we will find mAP (Mean Average Precision) per class. We will also find the accuracy for the 10 selected classes.

7. Project Design

The project will consist of the following:

1. **Data Exploration:** In this section, we will explore data and try to understand it from different perspective.
2. **Pre-processing:** In this section, we will do all necessary setup to pre-process the data as needed to prepare for the neural network. Setup may include
    a. Image augmentation to double the dataset size
    b. Images resizing to 28x28
    c. Images re-scaling by 255
    d. We will use ResNet-50 as our pre-trained model
3. **Setup Conv Neural Network**: in this section, we will setup the natural by importing pre-trained models and set up the final layer. Setup may include
    a. Import RasNet-50 into the model. I am planning to use bottleneck features provided in the Deep learning project.
    b. Add the fully connected layer which output the desired classes.
4. **Training the model :** in this section. We will train the neural network to on the training data and on the validation set.
5. **Optimizing and Evaluating the model :** in this section will optimize and evaluate the result from the trained model. Steps may include
    a. Calculating the mean average precision
    b. Calculating the accuracy per class.

8. References

The data set is requested from the authors of the below article

Open Logo Detection Challenge.
Hang Su, Xiatian Zhu and Shaogang Gong.
In Proc. British Machine Vision Conference (BMVC), Newcastle, UK, September 2018.
BibTex arxiv