

# Billboard Content Recognition for Driver Assistance Systems

Punarva Katte  
Georgia Institute of Technology  
pkatte3@gatech.edu

Prabhudev Prakash  
Georgia Institute of Technology  
dev@gatech.edu

## Abstract

*In this project, we propose a Deep CNN based solution for logo recognition which can be used for Billboard Content Recognition in Driver Assistance Systems. The logo datasets available are not large enough to train a fresh Deep CNN. Hence we used a pretrained network and trained on the available logo datasets. The end system is a logo recognition system which can detect logos in iconic images. Hence we also integrated an Object proposal method which helps in getting the logos in test images as an icon and also localization of the detected logos.*

## 1. Introduction

One of the tasks as part of infotainment for Driver Assistance Systems is detection of sign boards and billboard advertisements while driving. Detection of sign boards, albeit very important for safe driving, is an easy problem to solve as there are specific rules for how a sign board should be and robust solutions already available. This is evident from German Traffic Sign Recognition Benchmark[10] results. Committee of CNNs, IDSIA who came first in the IJCNN 2011 competition were able to detect 99.46% of sign boards from test data.

A harder problem to solve would be to understand the surroundings of the car, such as terrain, weather, vehicles moving around, buildings, billboards, etc. This project aims at understanding the billboards encountered during the drive, specifically to recognize the company or brand advertised on the billboard. Exploiting the fact that all the billboards used for advertisements display their respective logos, we try to build a logo recognition system in order to recognize its company or brand.

In recent years, Deep Convolutional Neural Networks has been used extensively for computer vision applications, especially object recognition, because they can be directly applied to real life images rather than extracting features from input image and then use it to learn a network (like Viola Jones). They also perform better than the traditional methods, which is evident from the results of ImageNet

competitions[9] in recent years.

One of the papers which try to address a similar problem is DeepLogo [2]. This paper talks about applying Deep CNNs for logo detection and localization. They used FlickrLogos-32 dataset[8] for training Deep CNN. They propose use of improved GoogLeNet such as GoogLeNet-GP, GoogLeNet-FullClassify or Full-Inception.

Deep CNNs require a huge set of training data for optimal performance, which is not the case with any available logo datasets. The most sought out solution for this problem is pretraining a network on a huge dataset with similar application and using it to continue training on the small dataset. In this project, we use Alexnet[4] pretrained on Imagenet-2010[1] dataset, to train a logo recognizer. Apart from making up for smaller logo dataset, pretraining also allows faster convergence.

The training images in FlickrLogos 27 and 32 are iconic and as we discuss below, the images in BelgaLogos are also converted to be iconic. This cripples the logo detector to work only on iconic images. A simple solution would be use sliding window on test image, although it is computationally expensive. This leads us to use of Object proposal techniques, which are less expensive. EdgeBoxes[11] in particular is least expensive and provides object proposals in fraction of a second. After training the logo recognizer, while testing, we first use EdgeBoxes to propose objects in the image, consider some top ranked object boxes and test each of them for logos. This also solves the problem of localization of the logo and detecting multiple logos in an image.

## 2. Datasets

Some of the logo datasets available are listed below. The pretrained network was trained on these datasets without using the logo annotations.

### 2.1. Flickr Logos 27

Flickr Logos 27 dataset[3] is an annotated logo dataset downloaded from Flickr and contains 27 logo classes in total. The training set contains 810 annotated images, corresponding to 27 logo classes/brands (30 images for each

class). All images are annotated with bounding boxes of the logo instances in the image.

## 2.2. FlickrLogos-32

FlickrLogos-32[8] dataset 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. The dataset contains logos of 32 different logo brands downloaded from Flickr. The images comes with pixel-level annotations and there are also many images with no logos meant to be negative data.

## 2.3. BelgaLogos

BelgaLogos[6] dataset is composed of 10,000 images which includes images with single, multiple or no logos in them. The logos in the images, belonging to one of 37 classes, are manually annotated. Compared to Flickr Logos 27 and 32, BelgaLogos have very small logo instantiates in training images (Figure 1a,1b,1c) which makes it hard to train without annotations. Instead we extracted the logo from the image using the annotation with suitable background and jitter (Figure 2). Along with making the training data iconic, this process also allows to train on multiple logos in an image one by one.

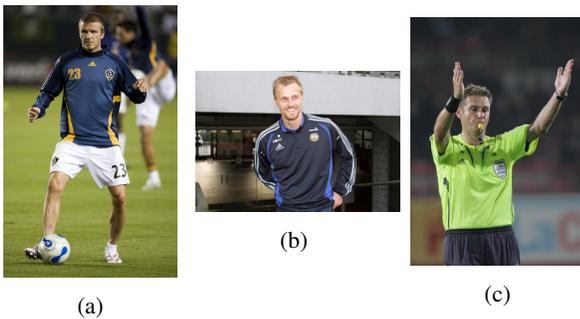


Figure 1: Example images from BelgaLogos showing very small logo instantiates

## 2.4. FlickrBelgaLogos

FlickrBelgaLogos[7] is a synthetic dataset created by cutting and pasting the cropped logos of BelgaLogos II into a dataset of 10K distractor images crawled from Flickr. This dataset is not used in this project.

# 3. Experiments and Results

## 3.1. Experiment 1

In this experiment, we combined BelgaLogos and FlickrLogos-27 datasets and trained them on pretrained Alexnet. We then observed qualitative results.



Figure 2: Logos extracted from the image using their annotations with suitable background and jitter

1. Extract each logo from BelgaLogos using their annotations
2. Add background and jitter
3. Append FlickrLogos 27 images

This creates approximately 31000 training images

4. Divide the new dataset into training (85%) and validation (15%) sets
5. Create a .imdb file from this new dataset
6. Load the pretrained Alexnet on Caffe
7. Use the .imdb file to train the network for 10 epochs

Figure 3 shows the evolution of Validation accuracy over training epochs.

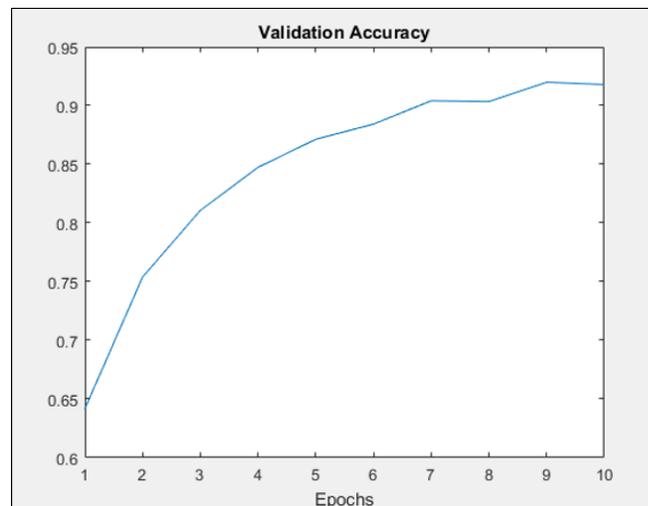
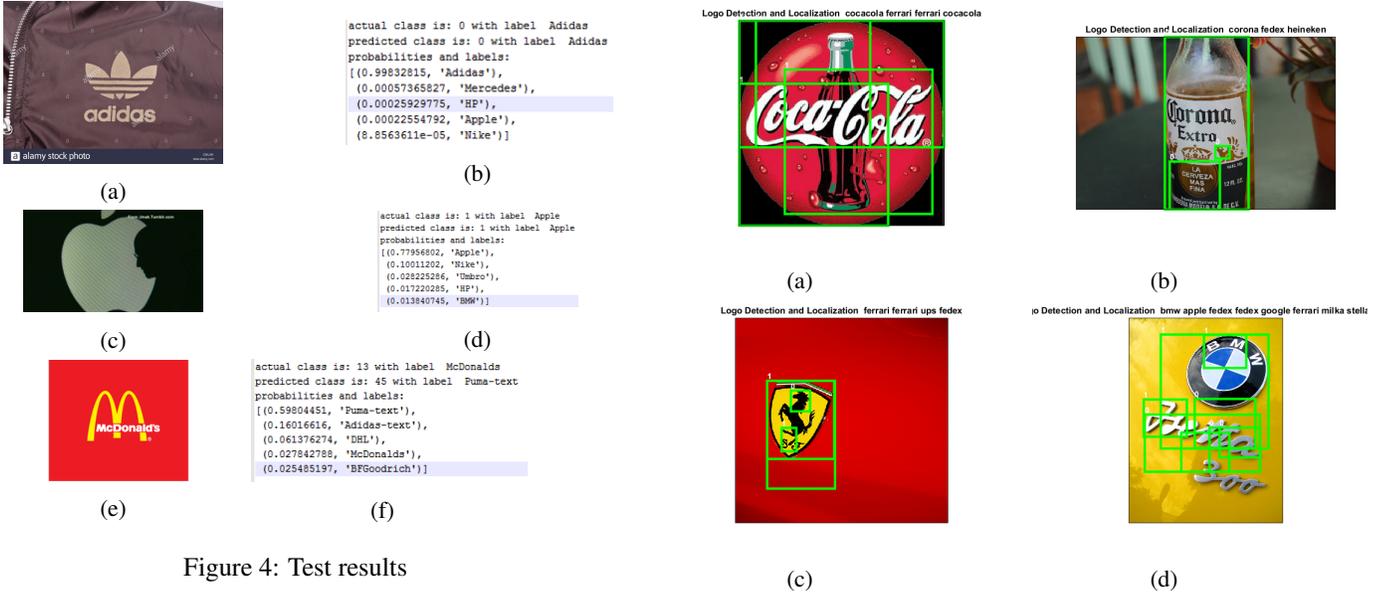


Figure 3: Validation accuracy while training

The logo detector was then tested on some of the images as shown in Figure 4a,4b,4c,4d,4e,4f,



### 3.2. Experiment 2

In this experiment, we used FlickrLogos-32 dataset and trained it on pretrained Alexnet using Matconvnet. We then observed qualitative results. FlickrLogos-32 has 70 images per class and the logos are iconic. Hence we can directly use them for training without pre-processing.

1. Divide the dataset into training ( $\frac{6}{7}$ )<sup>th</sup> and validation ( $\frac{1}{7}$ )<sup>th</sup> sets
2. Create a .imdb file for the dataset
3. Load the pretrained Alexnet on Matconvnet
4. use the .imdb file to train the network for 13 epochs

Top 10 Object proposals from each test image were then taken by EdgeBoxes and the logo detector was then tested on them. The results are as shown in Figure 5a,5b,5c,5d,5e,5f,5g,5h. In the figure, for each proposal if the top detection is a logo, it is listed at the top and the proposed box is marked in green.

### 4. Conclusion

Looking at the results, we see that Deep CNNs can be used for logo detection on available datasets, but they require much more tuning for better accuracy. Since the logo datasets are small we also observe a tendency of overfitting. Having a bigger logo dataset will certainly improve the results.

EdgeBoxes are also a culprit in bad results, because they are not tuned to extract logos in an image. One solution would be to use DeepBox [5] as object proposer which can be trained on ground truth of logos from dataset.

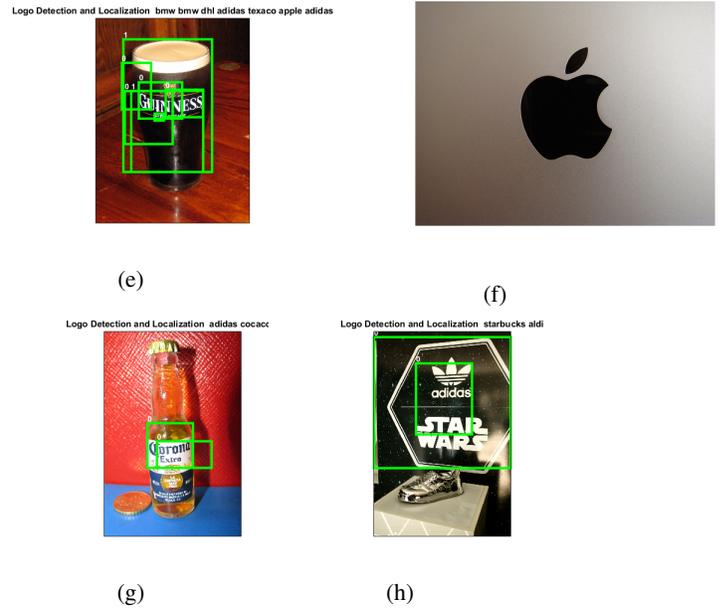


Figure 5: Test results

Implementing the whole system on a GPU might be the next step towards having a real time Billboard Ad recognition for Driver Assistance Systems.

### References

- [1] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.
- [2] F. N. Iandola, A. Shen, P. Gao, and K. Keutzer. DeepLogo: Hitting Logo Recognition with the Deep Neural Network Hammer. *ArXiv e-prints*, Oct. 2015.

- [3] Y. Kalantidis, L. Pueyo, M. Trevisiol, R. van Zwol, and Y. Avrithis. Scalable triangulation-based logo recognition. In *Proceedings of ACM International Conference on Multimedia Retrieval (ICMR 2011)*, Trento, Italy, April 2011.
- [4] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.
- [5] W. Kuo, B. Hariharan, and J. Malik. DeepBox: Learning objectness with convolutional networks. In *International Conference on Computer Vision (ICCV)*, 2015.
- [6] P. Letessier, O. Buisson, and A. Joly. Scalable mining of small visual objects. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 599–608. ACM, 2012.
- [7] P. Letessier, A. Joly, and O. Buisson. Scalable mining of small visual objects. In *MM '12: Proceedings of the 20th ACM international conference on Multimedia*, 2012.
- [8] S. Romberg, L. G. Pueyo, R. Lienhart, and R. van Zwol. Scalable logo recognition in real-world images. In *Proceedings of the 1st ACM International Conference on Multimedia Retrieval, ICMR '11*, pages 25:1–25:8, New York, NY, USA, 2011. ACM.
- [9] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.
- [10] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel. Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. *Neural Networks*, (0):–, 2012.
- [11] C. L. Zitnick and P. Dollár. Edge boxes: Locating object proposals from edges. In *ECCV. European Conference on Computer Vision*, September 2014.