Logo Recognition

Naveen Arivazhagan
Department of Computer Science
Stanford University
naveen67@stanford.edu

1. Introduction

In this paper we will work on effective and scalable approaches to extracting logos from images of natural scenes. This could be very useful for contextual advertisement placement, which is about placing relevant ads on website, images, and videos. It is also useful for brand managers as it could help them understand the context that their product appear in in the images that users upload to the Internet.

It is important that the system we build can do two things. First it should be able to scale well so as to be able to process the huge number of images and videos that are being uploaded every minute. Secondly, it should be easily adaptable to recognize new logos that are made everyday. In this paper we will apply deep learning methods to achieve both the above goals.

2. Related Work

[1] propose an effective and scalable framework for recognizing logos in images which is based on a method for encoding and indexing the relative spatial layout of local features detected in the images with logos. Based on the analysis of the local features and other spatial structures they use an automatic method for constructing a model for each logo-class out of multiple training images. This allows them to detect logos under varying conditions such as perspective tilt.

[2] view the problem of logo recognition as an instantiation of the broader problem of object recognition. They divide the overall problem into four sub-problems: logo classification, logo localization, logo detection without localization, and logo detection with localization. They then use evaluate their approach using modification of GoogLeNet. Using this they are able to achieve greater than 90% accuracy and establish the state of the art in logo recognition.

This problem is also very similar to reading text in the wild like house numbers, for example. Thus we can expect that approaches from that domain with a little adaptation would work very well for the task of logo recognition. [3] use an approach as follows: : word bounding box proposal generation, proposal filtering and adjustments, text recognition and final merging. The detection stage is done using the R-CNN object detection framework in which a region proposal is converted into a fixed size to which we can apply a CNN. This helps avoid the computationally expensive task of scanning the full image for text. Before applying the CNN they use various classifiers to filter out the large number of false-positive region proposals. Once they have filtered out some of the region proposals and only retain ones that are highly likely to contain text, they apply whole-word approach for recognition by providing the entire region proposal to a deep convolutional neural network.

3. Data

We will be using the FlickrLogos-32 dataset [4] for this paper. The dataset contains photos showing the brand logos of 32 different brands. The images were obtained from Flickr and all the logos have an approximately planar surface.

The images are labeled with the brand they contain. There is also a pixel level annotation available for each of the images to indicate the location of the logos. These pixel level annotations consist of binary masks and bounding boxes.

The dataset is already partitioned into a training set, validation set, and a test set each of which contains an equal number of images from each of the brands. The validation and test sets may also contain images with multiple instances of a logo. For example, many cans of Fosters on a shelf. The validation and test sets also contain images that do not have a logo at all.

4. Evaluation

We evaluate using the standard metrics of precision, recall, and F1. Here recall is the fraction of images that have a logo that are recognized to have some logo. Precision is the fraction of images that are recognized to have the correct logo over the over the number of images recognized to have a logo.

Table 1. Data Summary

Partition	Description	Images	Number of Images
P_1 (training set)	hand picked	10 per class	320 images
P_2 (validation set)	at least one logo present no logos in the image	30 per class 3000	3960 images
P_2 (test set)	at least one logo present no logos in the image	30 per class 3000	3960
Total			8240

5. System Setup

I am using Theano for my project. I am currently set up on AFS but I'm considering switching to AWS when I start working with larger models.

6. Preliminary Results

I trained a 2 layer fully connected network with a hidden dimension size of 100 with no regularization and a learning rate of 0.1. With that I got a precision of only 1.2% and a recall of 100%. While the parameters have not yet been tuned at all, the main reason for this bad performance is due to the system being unable to handle the 3000 images in the test set that have no label; it always predicts a label even on these images. This is why the recall is 100%.

7. Future Work

I plan on following the same approach that Jaderberg et al applied for localising and recognizing text in natural scenes. Basically dividing the problem into two parts: 1) locating the logo and 2) recognizing it. This will allow us to get better recall and avoid all the false positives. Additionally, it is likely that the recognition system could highly benefit the localization module. Considering that, we would have to apply a joint inference kind of approach to get better overall results.

Another thing I would like to look into is logo class expansion. In the current set up of Iandola et al they train on the images of a given set of logos and predict on only those. Similarly Jaderberg et al are only doing text recognition which also draws form a closed set of classes. However, in a real work application of logo recognition, we can expect the the set of logos can grow very frequently. This could be because we are interested in detecting the logos of new brands, a brand changes its logo, or adds a new logo. In these cases it would be very expensive to constantly train a new system every time our class of logos changes. It would be very useful to see if something better can be done here. One approach would be to somehow divide the system into multiple components so that only part of the system needs to be retrained. Another approach could be to simply have a second logo classifier in addition the the original one. The second classifier would only be trained on the new logos. During prediction we could compare the scores of the first and second classifier and predict accordingly. Then at a particular frequency the fist classifier can be retrained on all the data. This could significantly reduce training costs

References

[1] S. C. H. Hoi, X. Wu, H. Liu, Y. Wu, H. Wang, H. Xue, and Q. Wu. Logo-net: Large-scale deep logo detection and brand recognition with deep region-based convolutional networks. *CoRR*, abs/1511.02462, 2015.

[2] F. N. Iandola, A. Shen, P. Gao, and K. Keutzer. Deeplogo: Hitting logo recognition with the deep neural network hammer. *CoRR*, abs/1510.02131, 2015.

[3] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman. Reading text in the wild with convolutional neural networks. *International Journal of Computer Vision*, 116(1):1–20, 2016.

[4] 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.