Computer Vision Final Report - Logo Classification

Animesh Sharma
Rutgers University
as3592

Kunj Mehta
Rutgers University
kcm161

Abstract

Logo detection and classification is a subset of the object classification domain in computer vision, concerning logos of popular brands. There is a huge real world use case of classifying logos effectively – it can be used for advertisement effectiveness tracking and measuring brand reach. We propose to implement the approach in (Tuzko et al., 2017) from scratch with a few modifications, and then extend it. Specifically, traditional approaches focus on seeing logo classification as a image problem but we also treat it as a separate "character" identification problem for textual logos. We discuss our approach, extensions, results and challenges faced.

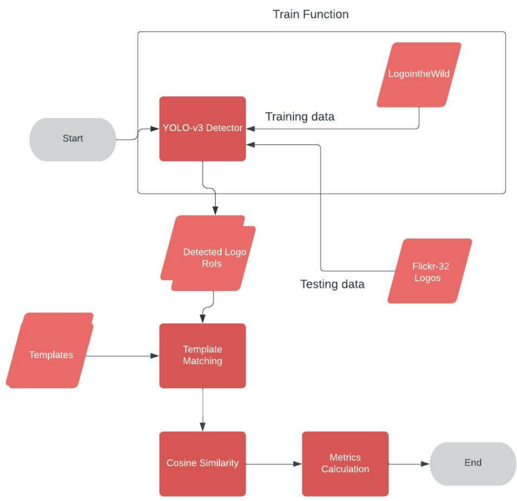

Figure 1: Architecture of Reproduced System (Tuzko et al., 2017)

1 Introduction

Logo detection and classification as a subset of object classification can be categorized into two: open set and closed set. Open set detection is when the dataset contains logos captured naturally in the real-world while closed set detection datasets have images which contain logos as the only thing in the image. Closed set detection is traditionally easier to do because of less variability in the data. For open set detection, two major problems arise: the size of the logo RoIs in the image which can be very small in comparison to image size, and the issue of logos being different for the same brand across different points in time. (Jin et al., 2020)

To help solve these problems, certain benchmark datasets exist. FlickrLogos class of datasets are closed set datasets, QMULOpenLogo is a compilation of many smaller logo datasets and LogosintheWild is a open set dataset. The Flickr-32 dataset contains 32 classes of logos extracted from Flickr. LogosintheWild is constructed using Google image search for certain brands and then cleaning the results obtained based on a strategy outlined in (Tuzko et al., 2017). The LogosintheWild dataset contains the "public" dataset images; these include images from Flickr-27, Flickr-32, FlickrBelgaLogos, TopLogo-10 and Logos-32plus datasets. Flickr-47 is the same dataset as Flickr-32 but with the textual type of brand logos categorized as different classes.

Because approaching the logo classification problem from an open set point of view reflects the real world better, we prefer that approach and try to reproduce the results in (Tuzko et al., 2017). We implement and compare closed set detection results using Faster-RCNN and open set detection results using YOLO-v3 instead of ResNet or VGGNet. As an enhancement to the above approach and to deal with the problem of duality in abstract and textual logos for the same brand, we also try to incorporate a textual character identification component to classify textual logos after detection.

2 Related Work

Much of the work around logo detection follows the closed set approach, but a few open set approaches are also present. Authors in (Tuzko et al., 2017), which we re-implement, introduce and use the LogointheWild dataset for object detection via a two-stage approach: use Faster-RCNN as a generic logo detector and then classify the features of the extracted logo RoIs using cosine similarity. 1

This is very advantageous for the open set scenario because this designs a system that can look at new logos in the real-world and classify them, something which is not possible with closed set systems and something we aim to achieve as well. The authors compare their system trained on the open set data by testing it on the Flickr-32 dataset.

Authors in (Fehervari and Appalaraju, 2018) also follow the same two-stage open set approach on a proprietary private dataset and test on the Flickr-32 dataset, with the major difference being that the classification uses Triplet Loss with a network similar to the Siamese network instead of cosine similarity. A two-stage approach is also followed in (Jin et al., 2020) but it incorporates attention mechanism during the classification stage for better performance.

3 Approach and Improvements

As described above, our approach for logo detection and classification follows closely with (Tuzko et al., 2017) and we reproduce their results before enhancing it with using YOLO-v5 as the detector, and using techniques to try and separately identify textual logos

We refer to a third party implementation of (Tuzko et al., 2017). We download the LogosintheWild dataset and train a YOLO-v3 model on the same until convergence. Due to the dataset being a bit old, there are some image links that are broken due to which the distribution of the data is changed in comparison to that of the original. Nonetheless, we follow the experiment for open set detection as described in (Tuzko et al., 2017) as closely as possible: we use the (public + LogosIntheWild dataset) for training; we keep aside 10% of the dataset for validation and use the Flickr32 dataset for testing. We ensure that the brands present in Flickr32 are not present in the training dataset.

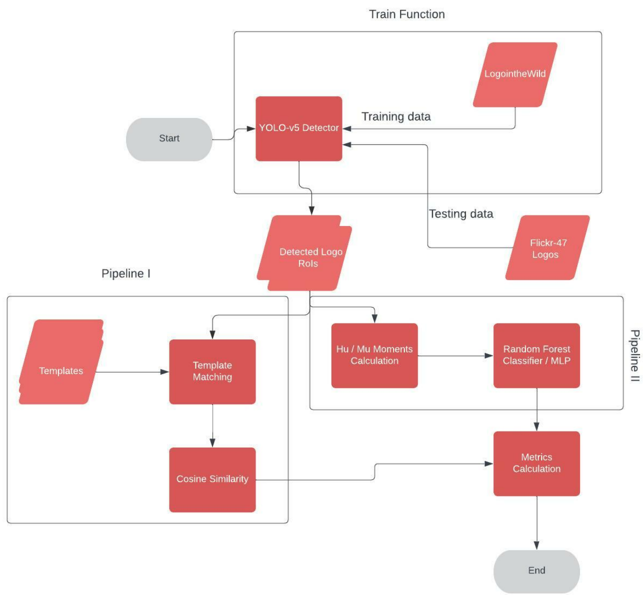

Figure 2: Architecture of Proposed System (Two Pipelines Tried)

For the closed set retrieval with Faster-RCNN, we deviate from the approach in (Su et al., 2018a), cited in (Tuzko et al., 2017) train and test the model on Flickr-32. This is because in closed set retrieval, the only data we work with is contained in the dataset itself. We do this training in accordance with the suggested train-test split of the dataset.

As part of the enhancement, we use the LogointheWild and Flickr-47 datasets. We train an improved version of YOLO, YOLO-v5 on the LogointheWild dataset for detection in the same fashion as before. However, instead of using Flickr-32 as the test dataset, we use Flickr-47 dataset. This is the same dataset but contains classes for the textual logo separately. We use this because separately classifying textual logos is the focus of our enhancement. After training the detector, for the part of the architecture not specifc to textual logos, we follow the same cosine similarity approach. For the part of textual logos, we tried an approach of feature extraction using Hu and Mu features because they have been found to classify characters better, followed by classification. 2

4 Results

The results that we obtain for logo detection while reproducing the approach in (Tuzko et al., 2017) are very close to what the authors obtained. We

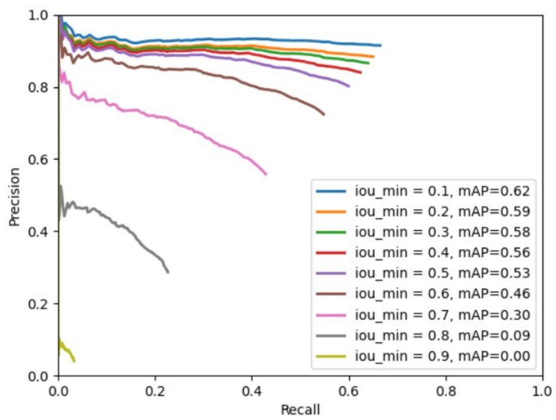

Figure 3: Reproduced open set logo detection results following approach in (Tuzko et al., 2017)

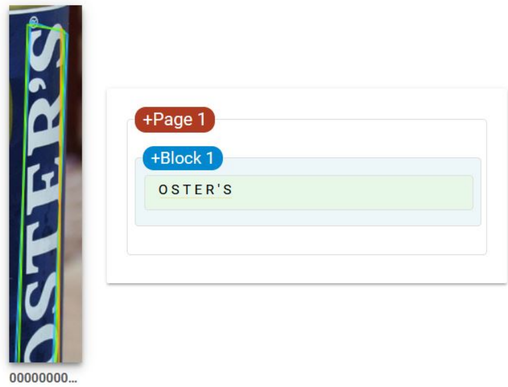

Figure 4: Reproduced open set logo detection results following approach in (Tuzko et al., 2017)

obtain a mAP = 0.53 against an IoU of 0.5 in comparison to the authors mAP of 0.464 while using YOLO-v3 detector for open set detection. Results for other IoUs are shown in 3. For the closed set detection, the Faster-RCNN that we trained obtained a mAP of 0.74 compared to the original mAP of 0.81. We attribute this difference in results to the fact that the downloaded dataset has a changed distribution from the original due to broken image links.

The results obtained as part of the enhancement where we train a YOLO-v5 model for open set logo detection are a mAP of 0.72 on the 10% stratified validation data handcrafted out of the Flickr-47 dataset. Following the same approach of cosine similarity + template matching on the Flickr-47 dataset, we obtain an accuracy of 22.57%. For the pipeline involving Hu and Mu feature extraction we obtained an accuracy of 3.53% when using RF and 4.126% when using MLP as a classifier.

5 Discussion

We must note that the results that we obtain and describe above are in all probability skewed from the original due to broken links in the LogointheWild dataset that we downloaded. This was the only link that we found for the dataset and there was no other alternative.

For the enhancement concerning the part of the textual logos, we note that we tried applying OCR to the separated textual logo RoIs using the tesseract engine but found that the performance for vertically aligned text is not good. We experimented with Google Vision API for the same and found that it performs satisfactorily well only for the topmost tier of the service, which is paid. An example of the result of the Google Vision API is shown in 4.

Further, we would like to list down and discuss other papers, approaches and datasets we explored and experimented with in Phase 3 of the project before settling down and implementing the approach in (Tuzko et al., 2017), and the challenges we faced while doing so.

We explored (Bhunia et al., 2018) but found that while the training data was publicly available, the test data was not. In addition, the authors were using a custom pipeline to preprocess the data, the code for which is not public. A similar problem was faced while looking at (Zhang et al., 2021b) where the authors pickled their pre-trained models in a specific format; the code for unpickling this was not available publicly at that time.

There are two papers (Zhang et al., 2021a) and (Jia et al., 2021) whose pre-trained models are hosted on Baidu. We downloaded these models but found that due to the dependencies and size of the datasets used, it was not feasible to run the models on the available storage and compute resources available to us.

Problems regarding the dataset being too large or the official or third-party code and pre-trained models not being maintained properly were also faced when we tried to reproduce results in (Su et al., 2018b) and (Wang et al., 2019)

6 Conclusion

The problem of logo detection and classification is a very specific one with real-world use cases. It can be solved using two approaches: closed set and open set, but for the proposed approach to mirror the real world the best, it has to be an open set system. In this vein, we tried and successfully reproduced the open set detection results in (Tuzko et al., 2017). We also recognized that logos can be both abstract and textual and tried to augment our detection and classification pipeline with a better detector and OCR.

References

A. K. Bhunia, A. K. Bhunia, S. Ghose, A. Das, and P. P. Roy. 2018. A deep one shot network for query-based logo retrieval.

I. Fehervari and S. Appalaraju. 2018. Scalable logo recognition using proxies.

X. Jia, H. Yan, Y. Wu, X. Wei, X. Cao, and Y. Zhang. 2021. An effective and robust detector for logo detection.

X Jin, W. Su, R. Zhang, Y. He, and H. Xue. 2020. Scalable logo recognition using proxies.

H. Su, X. Zhu, and S. Gong. 2018a. Deep learning logo detection with data expansion by synthesising context.

H. Su, X. Zhu, and S. Gong. 2018b. Open set logo detection.

A. Tuzko, C. Herrmann, D. Manger, and J. Beyerer. 2017. Open set logo detection and retrieval.

J. Wang, W. Min, S. Hou, S. Ma, Y. Zheng, H. Wang, and S. Jiang. 2019. Logo-2k+: A large-scale logo dataset for scalable logo classification.

B. Zhang, W. Min, J. Wang, S. Hou, Q. Hou, Y. Zheng, and S. Jiang. 2021a. Discriminative semantic feature pyramid network with guided anchoring for logo detection.

H. Zhang, Z. Cao, Z. Yan, and C. Zhang. 2021b. Sill-net: Feature augmentation with separated illumination representation.