LOGO-Net: Large-scale Deep Logo Detection and Brand Recognition with Deep Region-based Convolutional Networks

Steven C.H. Hoi⋆, Xiongwei Wu⋆, Hantang Liu⋆, Yue Wu⋆, Huiqiong Wang‡, Hui Xue‡, Qiang Wu‡
⋆School of Information Systems, Singapore Management University, Singapore
‡Alibaba Group, Hangzhou, China
{chhoi,yuewu,htliu,xwwu.2015@phdis}@smu.edu.sg
{huiqiong.whq,hui.xueh,qiangwu.wq}@alibaba-inc.com

Abstract

Logo detection from images has many applications, particularly for brand recognition and intellectual property protection. Most existing studies for logo recognition and detection are based on small-scale datasets which are not comprehensive enough when exploring emerging deep learning techniques. In this paper, we introduce "LOGO-Net"[1], a large-scale logo image database for logo detection and brand recognition from real-world product images. To facilitate research, LOGO-Net has two datasets: (i)"logos-18" consists of 18 logo classes, 10 brands, and 16,043 logo objects, and (ii) "logos-160" consists of 160 logo classes, 100 brands, and 130,608 logo objects. We describe the ideas and challenges for constructing such a large-scale database. Another key contribution of this work is to apply emerging deep learning techniques for logo detection and brand recognition tasks, and conduct extensive experiments by exploring several state-of-the-art deep region-based convolutional networks techniques for object detection tasks.

1. Introduction

Logo detection and recognition has been extensively studied in computer vision and pattern recognition literature [4, 1, 6, 2, 3, 26, 14, 12, 7, 15, 18, 20, 13]. From a computer vision perspective, *logo recognition*, which can be viewed as a special case of image recognition, aims to recognize the logo name of an input image, and *logo detection* is often more challenging in that it not only needs to recognize the logo name but also need to find the locations of logo objects in the input image. Logo detection and recognition found a wide range of applications in many domains, such as product brand recognition for intellectual property protection in e-commerce platforms, vehicle logo recognition for intelligent transportation [18], product brand management on social media [8], etc.

Logo objects typically consist of mixed text and graphic symbols. Although it may be viewed as a special type of object detection, logo detection from real-world images (e.g., product images) can be quite challenging since the same logo when appearing in different real scenarios can become very different due to the changes in sizes, rotations, lighting, occlusion, rigid and even non-rigid transformations. For example, a rigid logo object when appearing in a real clothing image often becomes non-rigid, making it difficult to be detected and recognized.

Although logo-related research has been explored for a long history of over two decades in literature [4, 1, 6], most existing studies use small datasets and very few large datasets are publicly available. For example, among the existing publicly available logo databases [12], one of the largest is "FlickrLogos-32", which only consists of 32 logo classes, 8240 images, and 5644 logo objects. Clearly, the existing datasets are not sufficient for conducting large-scale logo related research, particularly when exploring emerging data-intensive deep learning techniques.

To this end, we propose "LOGO-Net" — a large-scale logo image database to facilitate the research of logo detection and product brand recognition. The current LOGO-Net database consists of 160 logo classes, 100 brands, 73,414 images, and a total of 130,608 logo objects manually labeled with bounding boxes by human beings. Constructing such a large-scale database is challenging, time-consuming, and expensive. We discuss the details of how to construct the database and resolve the challenges in the project. In addition to the database, another key contribution of this work is to explore a family of emerging state-of-the-art deep learning techniques for generic object detection with application to large-scale logo detection and brand recognition tasks and conduct extensive empirical evaluations.

The rest of this paper is organized as follows. Section

[1]The LOGO-net will be released at http://logo-net.org/ This project was initialized in early of 2015, and the main tasks were completed in July 2015 when Prof Hoi visited Alibaba Group.

Figure 1: Examples of Brands and logo images from real-world product images

2 presents the problem formulation of logo detection and brand recognition tasks from real-world product images. Section 3 introduces the proposed "LOGO-Net" database. Section 4 presents the deep logo detection framework using the emerging deep region-based convolutional networks techniques. Section 4 presents our empirical studies. Section 5 concludes this work.

2. Problem Formulation

2.1. Logo Detection

In general, logo detection can be viewed as a special case of generic object detection in computer vision [17, 24, 21], which is often more challenging than generic object recognition/classificatin tasks. Logo detection aims to detect logo instances of some pre-defined logo classes in digital images or videos. Similar to a generic object detection task, given an input image, a logo detection method not only needs to indicate if a logo is found in the image, but also needs to report the locations of the detected logo object instances/regionds found in the image.

Despite being a special case of generic object detection, logo detection from real-world product images (e.g., online shopping portals like Taobao.com or Aliexpress.com) is very challenging for several reasons. First of all, a logo instance occurring in a real product image can be extremely small. Second, a "rigid" logo instance (in its original form) can become non-rigid when it is embedded in a natural product image, e.g., a logo occurring in a clothing image. Last but not least, logo instances of the same logo class occurring in real-world product images may differ very much due to a variety of changes, such as sizes, rotations, transformations, lighting, coloring, and occlusion, etc.

2.2. Brand Recognition

Brand recognition aims to recognize the brand names of products in a real-world product image. From a machine learning and pattern recognition perspective, brand recognition is essentially a multi-class image classification task, where an input product image is classified into one of multiple pre-defined brand categories. The techniques for brand recognition from product images have many important applications, such as intellectual property protection in e-commerce, tracking brand-specific products for business intelligence, and visual online advertising, etc.

Although it is viewed as a multi-class image classification task, brand recognition cannot be solved directly by applying traditional image recognition techniques that simply do classification based on the visual contents of the whole product image. This is because the same brand can have multiple kinds of products (e.g., bags, shoes, or shirts, etc), and thus visual contents of product images from the same brand could be completely different.

To tackle the above challenge, we propose to explore logo detection techniques for brand recognition tasks. By detecting the appearance of logo objects related to a certain brand in a product image, one can solve the brand recognition task in an effective way. Therefore, the challenge of brand recognition can be reduced into solving a logo detection task from real product images. Finally, we note that a single brand can consist of multiple logo classes.

2.3. Deep Learning Framework

Deep learning has achieved promising results in varied object detection tasks recently [22, 10, 5]. One of the most successful deep learning paradigms for object detection is the series of region-based convolutional neural networks (R-CNN) [10, 11, 9, 16, 19], which had obtained state-of-the-art results in many object detection benchmarks [21]. Motivated by the successes of R-CNN related techniques for generic object detection, we propose a deep logo detection framework for brand recognition from real-world product images by exploring a family of state-of-the-art deep region-based convolutional networks (DRCN) techniques.

Figure 2 shows the proposed DeepLogo-DRCN framework for logo detection and brand recognition from product images. Specifically, given an input image, the region proposal step yields a set of region of interests (RoIs) using an efficient implementation of Selective Search (SS) [23]. The RoIs are then input into a fully convolutional neural network (CNN), in which each RoI is pooled into a fixed-size feature map and then mapped to a feature vector by fully connected

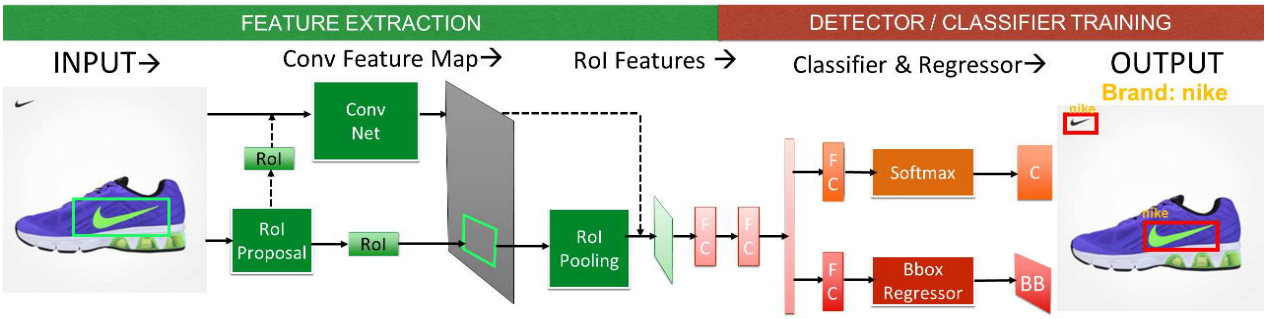


Figure 2: DeepLogo-DRCN: The proposed architecture of Deep Logo Detection using Deep Region-based Convolutional Networks (DRCN) techniques. Given an input image, the region proposal generation step yields a set of region of interests (RoIs) using an efficient implementation of Selective Search (SS) [23]. The RoIs are then input into a fully convolutional neural network, in which each RoI is pooled into a fixed-size feature map and then mapped to a feature vector by fully connected layers (FCs), which is followed by training the final object classifiers and bounding box regressors. For each RoI, the network yields two output vectors: softmax probabilities and per-class bounding-box regression offsets. The overall architecture typically can be trained in an end-to-end approach, where the DRCN technique can be any existing R-CNN variants for generic object detection. For example, the solid-line process includes several emerging fast variants of R-CNN, such as fast R-CNN (FRCN) [9] and SPPnet [11], while the dashline process represents the classical R-CNN [10].

layers (FCs), followed by training the object classifiers and bounding box regressors. For each RoI, the network yields two kinds of outputs: softmax probabilities and per-class bounding-box regression offsets. The overall architecture can be trained end-to-end. The DeepLogo-DRCN framework can take any recent DRCN algorithms for generic object detection, e.g., traditional R-CNN [10] (as shown by dashline), fast R-CNN (FRCN) [9] and SPPnet [11], etc.

3. LOGO-Net: large-scale logo image database

In this paper, we present "LOGO-Net" — a large-scale logo image database to facilitate the research of logo detection and brand recognition tasks with emerging deep learning techniques. Constructing such a large-scale real-world logo image database is very challenging, time-consuming, and costly. In the following, we discuss some major tasks and efforts in constructing our logo image database, especially data collection and data annotation tasks. We will then present two versions of datasets in our current LOGO-Net database, including a medium-scale dataset to ease the evaluations of varied settings, and a large-scale dataset. Finally, we compare our database against some of existing publicly available logo datasets.

3.1. Product Image Collection

The first task of LOGO-Net is to construct a list of brands and their associated popular logos according to the application needs. After building the list, for each brand and logo, we crawled their related product images from two online retail marketplaces: www.taobao.com — the world's largest marketplace for online shopping targetted at Chinese consumers, and www.aliexpress.com — a global retail marketplace targeted at consumers worldwide. All the product images crawled in our database were publicly available in the online marketplaces. In our database, each brand may have different number of images and each logo class may have different number of logo object instances. Our principle in the data collection is to ensure that each logo class at least has a minimal number of logo object instances for deep-learning research purposes.

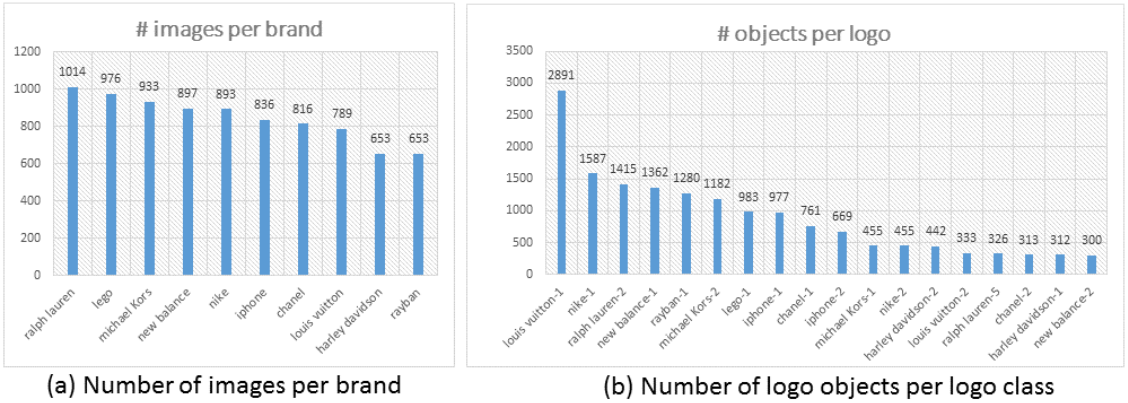
3.2. Logo Object Annotation

One of the most time-consuming and costly processes in constructing the LOGO-Net database is to annotate logo objects from the collected product images. For each product image, a human annotator needs to identify the logo objects, annotate the bounding box of each logo object, and then tag it with the corresponding logo class id. Figure 3 shows examples of logo object annotation on product images.

Figure 3: Instruction example of logo object annotation. The left-hand side is rejected due to too loose bounding box.

Statistics	Logos-18	Logos-160
# brands classes	10	100
# logo classes	18	160
# images	8460	73414
# logo objects	16043	130608
mean image width	564 pixels	687 pixels
mean image height	498 pixels	707 pixels

Table 1: Dataset summary of Logos-18 and Logos-160



(a) Number of images per brand (b) Number of logo objects per logo class

Figure 4: The statistics of numbers of images per brand and logo objects per logo class in the "logos-18" dataset

Figure 5: Examples of logo images from the LOGO-Net database

3.3. LOGO-Net Datasets: Logos-18 vs. Logos-160

Object detection is computationally very intensive for both training and test. To facilitate research, we design two datasets of different scales: Logos-18 versus Logos-160. Table 1 shows the dataset summary of LOGO-Net.

3.4. Comparison to Other Logo Data Sets

We compare our Logos-18 and Logos-160 datasets with some other publicly available logo datasets. Table 2 gives the summary of the dataset comparisons.

Dataset	#Image	#Logo	#Brand	#Logo Object
Logos-160	73414	160	100	130608
Logos-18	8460	18	10	16043
BelgaLogos	10000	37	37	2695
FlickrLogos-27	1080	27	27	4671
FlickrLogos-32	8240	32	32	5644

Table 2: Comparisons of existing logo datasets with Logos-18 and Logos-160. Note that BelgaLogos, FlickrLogos-27 and FlickrLogos-32 contain many non-logo images.

BelgaLogos Dataset[12] It contains 37 logo categories and a total of 10,000 images, in which the maximum height or width of each image has been re-sized to 800 pixels. However, the total of logo object instances is less than 3000.

FlickrLogos-27 Dataset[13] It has only 27 logo categories and a total of 1080 images, and each logo class has less than 50 images on average.

FlickrLogos-32 Dataset [20] It has 32 logo categories and a total of 8240 images. Similar to the FlickrLogos-27 Dataset, the number of images for each logo class is small, which is less than 70 images per logo class.

Unlike the existing datasets, the LOGO-Net database has a much larger scale in terms of both total number of logo objects and average number of logo objects per class, which is important and critical to explore any data-driven machine learning techniques for logo detection and recognition. Figure 4 shows some detailed statistics of numbers of images per brand and logo objects per logo class in the logos-18 dataset, and Figure 5 shows some examples of logos in our LOGO-Net database. More details about the database can be found in the supplemental materials.

4. Experiments

4.1. Experimental Setup

To enable the benchmark research, we follow standard competition setups for data partitions in our experiments. Specifically, for Logos-18, we randomly divide the dataset into three parts: 50% for training, 20% for validation, and

Algorithm(model)	mAP(%)	Accuracy (%)	AUC (%)	total train time	test time / image	GPU memory
RCNN(CaffeNet)	69.1	95.2	95.3	2444 (min)	20886(ms)	2.39 (GB)
RCNN(CaffeNet-w/o-ft)	55.1	86.5	86.4	1549 (min)	20881(ms)	2.39 (GB)
FRCN(CaffeNet)	58.8	93.2	92.0	147 (min)	448 (ms)	1.67 (GB)
FRCN(VGG1024)	59.8	94.8	93.6	253 (min)	529 (ms)	3.04 (GB)
FRCN(VGG16)	61.4	94.7	93.2	1312 (min)	836 (ms)	10.86 (GB)
SPPnet(ZF-w/o-bb)	54.5	92.5	92.3	707 (min)	968 (ms)	2.21 (GB)
SPPnet(ZF)	59.1	92.5	92.3	749 (min)	1199 (ms)	2.21 (GB)

Table 3: **Logos-18 test set** Logo Detection and Brand Recognition Results by DeepLogo-DRCN with different algorithms. Note that the test time cost includes the average region proposal time (310ms) by SS which is the same for all the algorithms.

Alg(model)	mAP(%)	Accuracy (%)	AUC(%)	total train time	test time / image	GPU memory
FRCN(CaffeNet)	61.0	81.6	82.6	169 (min)	685 (ms)	1.71 (GB)
FRCN(VGG1024)	60.3	81.3	82.3	283 (min)	752 (ms)	3.05 (GB)
FRCN(VGG16)	65.8	85.8	86.8	1362 (min)	1044 (ms)	10.89 (GB)
SPPnet(ZF-w/o-bb)	53.6	81.8	82.0	1360 (min)	1218 (ms)	3.63 (GB)
SPPnet(ZF)	58.1	81.8	82.0	1494 (min)	1639 (ms)	3.63 (GB)

Table 4: **Logos-160 test set** Logo Detection and Brand Recognition Results by DeepLogo-DRCN with different algorithms. Note that the test time cost includes the region proposal time (467ms) by SS which is the same for all the algorithms. We were not able to run RCNN on this large-scale dataset due to the huge computational cost.

30% for test. Similarly, we also divide the Logos-160 dataset into three parts: 20% for training, 20% for validation, and 60% for test. We purposely set less training data for the Logos-160 task to make it more challenging and realistic in real-world settings as requiring too much training data is not so impractical to scale.

We develop the proposed DeepLogo-DRCN scheme for logo detection and brand recognition by exploring several state-of-the-art Deep Region-based Convolutional Networks (DRCN) techniques for object detection, including

- RCNN [10][2] is the most popular and widely used deep learning framework for object detection by combining region proposal (e.g., selective search [23]) with convolutional neural networks (CNNs);
- FRCN [9][3] is a Fast R-CNN framework for object detection with deep region-based convolutional neural networks, which significantly improves computational efficiency of traditional R-CNN methods.
- SPPnet [11][4] is another variant of fast R-CNN by improving computational efficiency and exploring spatial pyramid pooling strategies.

Another important step in the DeepLogo-DRCN framework is the region proposal solution, which affects both logo detection quality and computational efficiency. In our approach, we employs the Selective Search (SS) method [23] which has been shown as the state-of-the-art region proposal that often achieves the best quality. However, the original SS implementation is notably slow particularly when dealing with a large image. Instead of exploring other fast region proposal techniques [27] which often sacrifice quality, we have done a fast implementation of the SS method, which yields almost the same quality as the original SS implementation but is much faster in an order of magnitude. In our experiments, we choose the number of RoIs to 2000 for all schemes according to the validation set in order to balance the trade-off between quality and efficiency.

For parameter settings, we choose the parameters of different algorithms using the same validation set. We set the number of fine-tuning iterations to 50,000 for all schemes whenever applicable, and the default threshold of Intersection over Union (IoU) to 0.5 for validating object bounding boxes. We will evaluate how different settings (e.g., the number of RoIs, the number of fine-tuning iterations, IoU, etc) affect the performance in parameter sensitivity section.

For performance evaluation metrics, we adopt the widely used mean Average Precision (mAP) for evaluating logo object detection tasks. For brand recognition tasks, we adopt the standard metrics for object recognition, including classification accuracy and Area under the ROC curve (AUC). The experiments were conducted in a GPU cluster with the NVIDIA high-end Tesla K80 GPU (2x Kepler GK210, 2496 cores per GPU, 12GB memory per GPU).

4.2. Main Results

Table 3 and Table 4 summarize the main results of logo object detection and brand recognition on the test sets by the proposed DeepLogo-DRCN with different algorithms and models. For each dataset (logos-18 or logos-160), all

[2] https://github.com/rbgirshick/rcnn
[3] https://github.com/rbgirshick/fast-rcnn
[4] https://github.com/ShaoqingRen/SPP_net

Algorithm(model)	mAP	cls1	cls2	cls3	cls4	cls5	cls6	cls7	cls8	cls9	cls10	cls11	cls12	cls13	cls14	cls15	cls16	cls17	cls18
RCNN(CaffeNet)	69.1	78.9	57.2	58.3	56.7	79.9	68.6	99.6	50.8	60.8	62.8	54.0	89.2	52.7	68.1	79.5	90.0	67.3	69.5
FRCN(VGG16)	61.4	75.2	50.8	57.0	56.2	69.4	67.2	99.5	42.5	47.0	46.1	28.2	89.1	44.6	58.6	77.2	82.2	48.7	64.9
FRCN(VGG1024)	59.8	74.6	46.1	58.1	53.8	74.4	73.2	99.6	43.1	39.2	47.9	20.4	88.8	37.0	55.4	74.6	82.0	46.1	62.1
FRCN(CaffeNet)	58.8	69.0	43.2	55.7	48.6	69.4	71.5	99.7	46.7	41.5	47.4	27.2	88.9	39.4	48.6	71.5	78.7	46.8	63.5
SPPNet(ZF Net)	59.2	69.6	49.3	54.4	50.9	70.2	73.8	90.9	41.3	38.6	48.4	48.5	84.3	37.7	49.8	67.2	79.3	50.6	61.5

Table 5: **Logos-18 test set logo detection** results of average precision (%). The notations from "cls 1" to "cls-18" denote chanel-1, chanel-2, harley davidson-1, harley davidson-2, iphone-1, iphone-2, lego-1, louis vuitton-1, louis vuitton-2, michael Kors-1, michael Kors-2, new balance-1, new balance-2, nike-1, nike-2, ralph lauren-2, ralph lauren-1, rayban-1, respectively.

Alg(model)	Acc	bnd1	bnd2	bnd3	bnd4	bnd5	bnd6	bnd7	bnd8	bnd9	bnd10
RCNN(CaffeNet)	95.2	91.1	88.1	96.8	100.0	93.7	91.0	99.3	98.2	97.7	92.3
FRCN(VGG16)	94.7	91.9	88.1	95.3	97.6	93.3	89.6	99.3	97.8	97.0	94.9
FRCN(VGG1024)	94.8	92.7	87.6	91.7	99.7	93.3	91.4	98.9	97.0	98.7	93.3
FRCN(CaffeNet)	93.2	90.3	85.5	88.5	99.7	90.4	86.7	98.9	97.4	97.4	92.8
SPPNet(ZF Net)	92.5	81.9	88.6	91.3	99.7	93.7	84.2	97.8	94.8	98.7	91.3

Table 6: **Logos-18 test set brand recognition accuracy** results(%). "bnd 1" to "bnd-10" denote "chanel", "harley davidson", "iphone", "lego", "louis vuitton", "michael Kors", "new balance", "nike", "ralph lauren", "rayban", respectively.

the models were trained on the same training data set, and tested on the same test set. The validation set was only used for choosing key parameters of each scheme. Several observations can be drawn from the main experimental results.

First of all, among all the methods, RCNN(CaffeNet with fine-tuning) obtained the best logo detection and brand recognition results on the logos-18 dataset, while FRCN(VGG16) obtained the second best results on the logos-18 dataset. This seems a bit surprising as both FRCN and SPPnet have been reported with state-of-the-art results, if not better than, at least comparable to RCNN for generic object detection on PASCAL VOC object detection benchmarks. This is mainly because unlike PASCAL VOC datasets, our LOGO-Net database has many small logo objects for which FRCN and SPPnet might fail to detect if the convolutional feature map is not large enough. In contract, RCNN suffers less from this issue since RCNN first takes an RoI (a small region) and then resize it to a fixed size (essentially enlarged) before it is passed to the convolutional network.

Second, in terms of training time cost, we found that RCNN is the most computationally expensive among all the solutions. This is because RCNN has to repeatedly perform convolutional operations for each RoI with the original image, which can be somewhat redundant and thus very computationally expensive. By comparing FRCN and SPPnet, when using a simpler network (e.g., CaffeNet), FRCN can be trained several times faster than SPPnet (based on the Zeiler-Fergus's ZF-net) [25]. However, when using a very deep network (VGG16), FRCN is computationally more expensive than SPPnet with ZF-net. Moreover, by examining GPU memory cost during training, we found that FRCN consumes a large amount of GPU memory (e.g., more than 10GB) when training the very deep VGG16 network. This poses a very high requirement for the GPU hardware (e.g., requiring very high-end GPU, e.g., K80 in our cluster).

Moreover, by examining the test time cost which is critical when being deployed in real-world applications, we found that R-CNN takes more than 20 seconds for processing an image, which is an order of magnitude slower than the others. The poor prediction efficiency makes RCNN infeasible to be deployed in a real-world application which may need to deal with millions of product images daily.

Finally, Table 5 and Table 6 give the detailed results of specific logo detection and specific brand recognition accuracy, respectively. Similar observations can be found.

4.3. Parameter Sensitivity

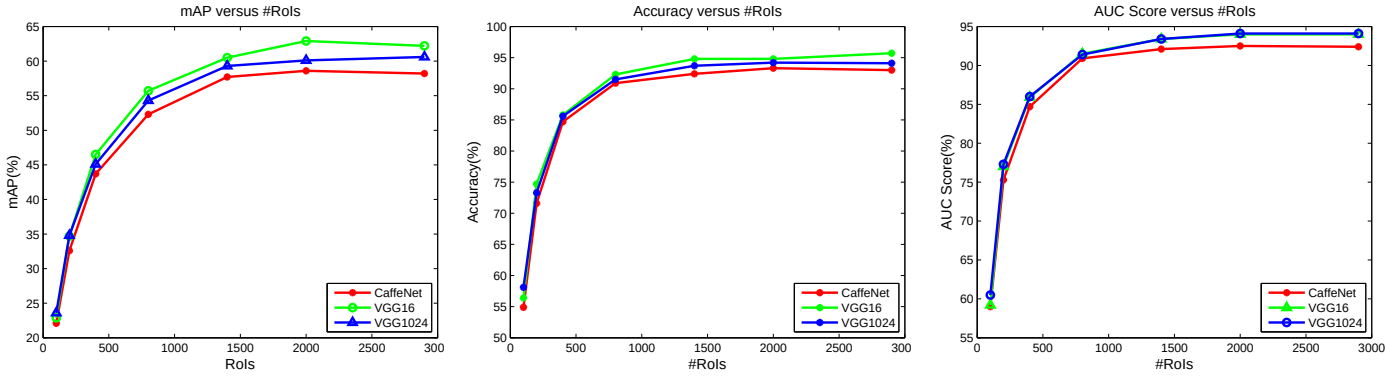
4.3.1 Overview

For the proposed DeepLogo-DRCN, there are some key parameters that may significantly affect logo detection and brand recognition performance, including the number of regions of interests (RoIs), the number of fine-tuning (ft) iterations, the amount of training data, the IoU setting, extra acceleration using SVD, etc. In this section, we evaluate how the performance is sensitive to each of these factors.

4.3.2 Evaluation of Number of Regions of Interests

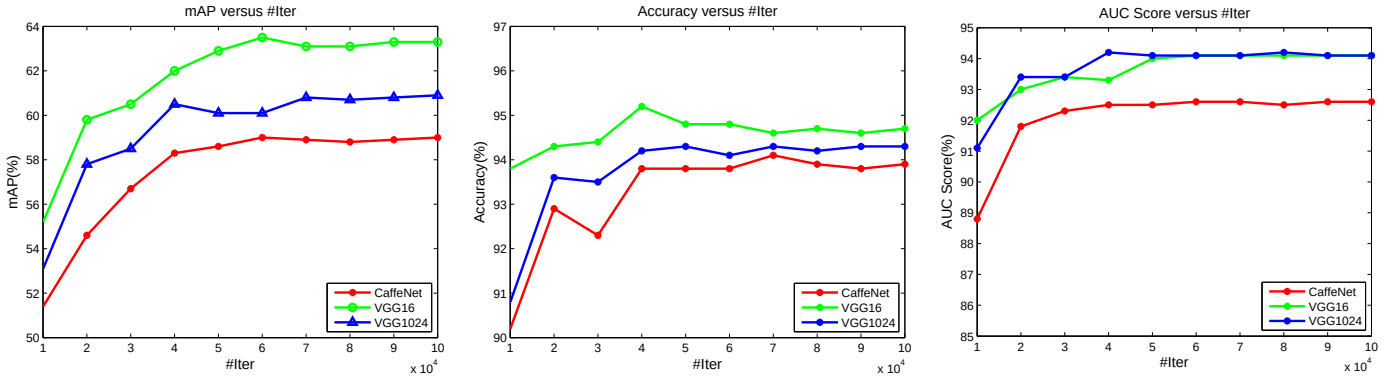
The number of RoIs (i.e., the number of bounding boxes) yielded by SS [23] plays a critical role in DeepLogo-DRCN, which affects both detection quality and recognition accuracy as well as computational efficiency. Computational cost is generally proportion to the number of RoIs. The higher the number of RoIs, the more computational costs for both training and test. However, increasing the RoI size may not always guarantee a significant improvement of mAP.

Figure 6 shows how the mAP performance on the validation set is changed with respect to different numbers of RoIs in the proposed DeepLogo-DRCN with FRCN using different CNN models. From the results, we can see that when the number of RoIs is small (e.g., less than 1000), increasing the number of RoIs always leads to a considerable improvement of overall mAP. However, when it is large enough (e.g., 2000), increasing it may only make a marginal



(a) mAP versus #RoIs (b) Accuracy versus #RoIs (c) AUC versus #RoIs

Figure 6: Evaluation of parameter sensitivity of #RoIs (bounding boxes) for object detection (mAP) and brand recognition performance (accuracy and AUC). The logo object detection algorithm is based on DeepLogo-FRCN.



(a) mAP versus #Iter (b) Accuracy versus #Iter (c) AUC versus #Iter

Figure 7: Evaluation of parameter sensitivity of #Iterations (fine-tuning) for object detection (mAP) and brand recognition performance (accuracy and AUC). The logo object detection algorithm is based on DeepLogo-FRCN.

improvement or even degrade the performance when it is too large (perhaps due to noise reasons). Thus, we found that setting the number of RoIs to 2000 is able to make a good tradeoff between logo detection efficacy and computational efficiency. Finally, by examining the brand recognition results, we found that as compared with mAP, both accuracy and AUC are less sensitive to the number of RoIs. When the number of RoIs is large than 1500, both accuracy and AUC results are almost saturated.

4.3.3 Evaluation of Fine-Tuning (FT) Iterations

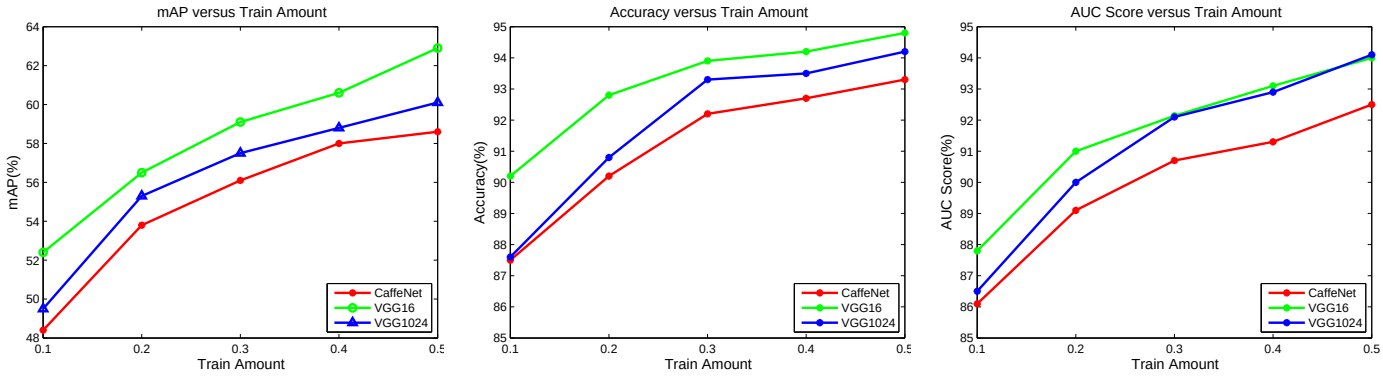
The fine-tuning procedure is one of critical steps for ensuring the proposed DeepLogo-DRCN scheme can adapt the existing pre-trained CNN models on the logo image dataset domain. In general, we need a significantly large number of fine-tuning iterations to ensure the proposed DeepLogo-DRCN is converged on the logo training data set. However, the training time cost grows linearly with the number of fine-tuning iterations. For a large-scale experiment, we need to set a proper number of fine-tuning iterations to balance the trade-off between efficacy and efficiency.

Figure 7 shows how the overall object detection (mAP) and brand recognition (accuracy and AUC) performances on the validation set are changed when increasing the number of fine-tuning iterations. We found that when setting it to about 50,000 iterations, the mAP performance will converge for most settings for FRCN. Finally, by examining the accuracy and AUC results of brand recognition, we found that the performances are almost saturated after 40,000 fine-tuning iterations.

4.3.4 Evaluation of Training Data Sizes

For deep learning methodology, the amount of training data can affect considerably the resulting performance. In this experiment, we aim to examine how the logo detection and brand recognition results were sensitive to different amounts of training data used for training the DeepLogo-DRCN scheme.

Figure 8 shows the evaluation results of object detection and brand recognition performances with respect to differ-



(a) mAP versus training data size (b) Accuracy versus training data size (c) AUC versus training data size

Figure 8: Evaluation of different amounts of training data for object detection (mAP) and brand recognition performance (accuracy and AUC). The logo object detection algorithm is based on Fast R-CNN (FRCN).

ent amounts of training data. From the results, we can see that increasing the amount of training data generally gives a consistent improvement of both logo object detection and brand recognition performances. The improvement of mAP is particularly more significant while the improvement of brand recognition accuracy is relatively less obvious. This is primarily because classification accuracy is already very good and thus making a further improvement would be more difficult. This result also indicate classification accuracy may not be an ideal performance metric as compared with either mAP or AUC metrics.

4.3.5 Evaluation of Intersection over Union (IoU)

The IoU threshold is a parameter to decide if a detected bounding box is overlapping enough with a target object. Setting a high IoU threshold requires a more precise localization of the detected object which is often done by the bounding box regression step. For brand recognition, it is less important for detecting precise bounding boxes of target logos. Figure 9 shows an evaluation of mAP with respect to different settings of IoU. We can see that when IoU is larger than 0.5, decreasing IoU only leads to a marginal improvement of mAP. However, when IoU is less than 0.5, increasing IoU can result in a significant drop of mAP.

4.3.6 Evaluation of Acceleration by Truncated SVD

We realize the prediction time is crucial when applying DeepLogo-DRCN for real-world applications. To speed up the prediction of DeepLogo-DRCN, we explore an SVD truncation based acceleration technique from [9]. The basic idea is to simplify the most intensive and repeatedly computation parts, i.e., the fully connected layers of DRCN, using SVD-truncation based approximation. More details about SVD-approximation can be found in [9].

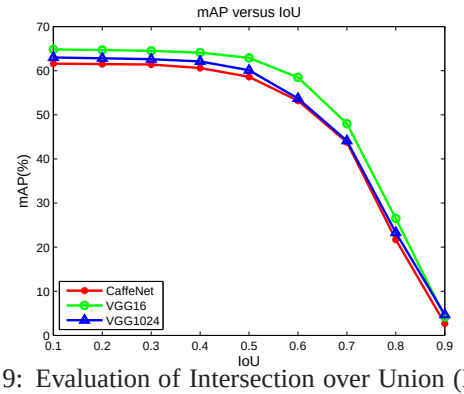


Figure 9: Evaluation of Intersection over Union (IoU) parameter settings for object detection performance (mAP).

Table 7 gives the speedup gains using the SVD-based acceleration. We can obtain about 30% speedup while sacrificing only a minor drop in predictive performance. Specifically the gains obtained for deeper models are especially impressive. For example, for VGG16, we can obtain 28% speedup in test time while only suffering no more than 0.2% drop in both accuracy and AUC.

5. Conclusions

This paper presented "LOGO-Net" — a large-scale logo image database to facilitate large-scale deep logo detection and brand recognition from real-world product images. LOGO-Net consists of two datasets: (i)"logos-18": 18 logo classes, 10 brands, and 16,043 logo objects, and (ii) "logos-160": 160 logo classes, 100 brands, and 130,608 logo objects. We discussed the challenges and solutions for constructing such a large-scale database, tackled the deep logo recognition and brand recognition tasks by exploring a family of emerging Deep Region-based Convolutional Networks (DRCN) techniques, and finally conducted an extensive set of benchmark evaluations.

metric	SVD(Y/N)	CaffeNet	VGG16	VGG1024
mAP	w/o SVD	58.8%	61.4%	59.8%
	w/ SVD	57.3%	60.8%	59.6%
	drop in mAP	1.5%	0.6%	0.2%
Acc.	w/o SVD	93.2%	94.7%	94.8%
	w/ SVD	92.4%	94.6%	94.5%
	drop in acc.	0.8%	0.1%	0.3%
AUC	w/o SVD	92.0%	93.2%	93.6%
	w/ SVD	91.5%	93.0%	93.5%
	drop in AUC	0.5%	0.2%	0.1%
Test Time	w/o SVD	0.137	0.542	0.230
	w/ SVD	0.091	0.391	0.149
	speedup	33.6%	27.9%	35.2%

Table 7: Evaluation of speedup gains obtained by SVD-based acceleration. Note that the above test time cost excludes the region proposal time cost by SS.

Acknowledgements

The authors would like to acknowledge the support from Dr Rong Jin at Alibaba Group to facilitate the research collaboration between Dr Hoi's team and Alibaba Group. The authors would like to acknowledge the fruitful discussions and help from researchers and engineers from Alibaba Group, especially for the intern student Mr Jiewei Luo.

References

[1] F. Cesarini, E. Francesconi, M. Gori, S. Marinai, J. Sheng, and G. Soda. A neural-based architecture for spot-noisy logo recognition. In *Document Analysis and Recognition, 1997., Proceedings of the Fourth International Conference on*, volume 1, pages 175–179. IEEE, 1997. 1

[2] J. Chen, M. K. Leung, and Y. Gao. Noisy logo recognition using line segment hausdorff distance. *Pattern recognition*, 36(4):943–955, 2003. 1

[3] R. J. Den Hollander and A. Hanjalic. Logo recognition in video stills by string matching. In *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, volume 3, pages III–517. IEEE, 2003. 1

[4] D. S. Doermann, E. Rivlin, and I. Weiss. Logo recognition using geometric invariants. In *Document Analysis and Recognition, 1993., Proceedings of the Second International Conference on*, pages 894–897. IEEE, 1993. 1

[5] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov. Scalable object detection using deep neural networks. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 2155–2162. IEEE, 2014. 2

[6] E. Francesconi, P. Frasconi, M. Gori, S. Marinai, J. Sheng, G. Soda, and A. Sperduti. Logo recognition by recursive neural networks. In *Graphics Recognition Algorithms and Systems*, pages 104–117. Springer, 1998. 1

[7] K. Gao, S. Lin, Y. Zhang, S. Tang, and D. Zhang. Logo detection based on spatial-spectral saliency and partial spatial context. In *Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on*, pages 322–329. IEEE, 2009. 1

[8] Y. Gao, F. Wang, H. Luan, and T.-S. Chua. Brand data gathering from live social media streams. In *Proceedings of International Conference on Multimedia Retrieval*, page 169. ACM, 2014. 1

[9] R. Girshick. Fast R-CNN. In *Proceedings of the International Conference on Computer Vision (ICCV)*, 2015. 2, 3, 5, 8

[10] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 580–587. IEEE, 2014. 2, 3, 5

[11] K. He, X. Zhang, S. Ren, and J. Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. In *Computer Vision–ECCV 2014*, pages 346–361. Springer, 2014. 2, 3, 5

[12] A. Joly and O. Buisson. Logo retrieval with a contrario visual query expansion. In *Proceedings of the seventeen ACM international conference on Multimedia (MM'09)*, pages 581–584, 2009. 1, 4

[13] Y. Kalantidis, L. Pueyo, M. Trevisiol, R. van Zwol, and Y. Avrithis. Scalable triangulation-based logo recognition. In *in Proceedings of ACM International Conference on Multimedia Retrieval (ICMR 2011)*, Trento, Italy, April 2011. 1, 4

[14] J. Kleban, X. Xie, and W.-Y. Ma. Spatial pyramid mining for logo detection in natural scenes. In *Multimedia and Expo, 2008 IEEE International Conference on*, pages 1077–1080. IEEE, 2008. 1

[15] Z. Li, M. Schulte-Austum, and M. Neschen. Fast logo detection and recognition in document images. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 2716–2719. IEEE, 2010. 1

[16] W. Ouyang, X. Wang, X. Zeng, S. Qiu, P. Luo, Y. Tian, H. Li, S. Yang, Z. Wang, C. C. Loy, and X. Tang. Deepid-net: Deformable deep convolutional neural networks for object detection. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, pages 2403–2412, 2015. 2

[17] C. P. Papageorgiou, M. Oren, and T. Poggio. A general framework for object detection. In *Computer vision, 1998. sixth international conference on*, pages 555–562. IEEE, 1998. 2

[18] A. P. Psyllos, C.-N. E. Anagnostopoulos, and E. Kayafas. Vehicle logo recognition using a sift-based enhanced matching scheme. *Intelligent Transportation Systems, IEEE Transactions on*, 11(2):322–328, 2010. 1

[19] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *Neural Information Processing Systems (NIPS)*, 2015. 2

[20] 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. 1, 4

[21] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, pages 1–42, 2014. 2

[22] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229*, 2013. 2

[23] J. R. Uijlings, K. E. van de Sande, T. Gevers, and A. W. Smeulders. Selective search for object recognition. *International journal of computer vision*, 104(2):154–171, 2013. 2, 3, 5, 6

[24] P. Viola and M. Jones. Robust real-time object detection. *International Journal of Computer Vision*, 4:51–52, 2001. 2

[25] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In *Computer Vision–ECCV 2014*, pages 818–833. Springer, 2014. 6

[26] G. Zhu and D. Doermann. Automatic document logo detection. In *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on*, volume 2, pages 864–868. IEEE, 2007. 1

[27] C. L. Zitnick and P. Dollár. Edge boxes: Locating object proposals from edges. In *Computer Vision–ECCV 2014*, pages 391–405. Springer, 2014. 5