Karlsruhe Institute of Technology

Fraunhofer IOSB

Logo retrieval in mass data using deep learning

Master thesis

KIT - Karlsruhe Institute of Technology
Fraunhofer IOSB - Fraunhofer Institute of Optronics, System Technologies and Image Exploitation

Andras Tüzkö

June 14, 2017

Main Advisor:	Dipl.-Inform. Christian Herrmann
Co-Advisor:	Dipl.-Inform. Daniel Manger

Interaktive Echtzeitsysteme

Statement of authorship

I hereby declare that I have produced this work by myself except the utilities known to the supervisor, that I have labeled all used utilities completely and detailed and that I have labeled all material that has been taken with or without modification from the work of others.

Karlsruhe, June 14, 2017 Andras Tüzkö Andras Tüzkö

Contents

1 Introduction

Advertising with static logos is one of the most important marketing methods. A very effective way to reach a lot of people with these static logos is, to sponsor sport teams or to buy advertising spaces in sport events broadcasted on the TV. However, the prices of these surfaces mean huge expenses for the advertiser. This is the reason why the need for logo appearance statistics of sport videos arises. In particular there is a desire for quantitative measurement of the proportional size of the logo to the screen, and of the time one particular logo is visible on the screen. This data is then used to judge the cost efficiency for the specific logo placement, i.e. to be able to decide on which sport event to advertise, with which size of logo, and where to place it.

In this work a system for logo retrieval with proposal based object detection and classification will be presented. The system consists of a logo detector, and a classifier used for feature extraction. The logo detector is a faster region based convolutional neural network [Ren15] trained to recognize logos on images. The features of the proposed regions are extracted with a ResNet neural network [He15]. To recognize logos in videos, the videos will be cut into frames, and then the system will be run on every image.

The challenge of this task is manifold. The first problem is that the logos in these videos are far from being perfectly clear. They can be partially occluded, blurred - if the camera is moving fast, perspectively transformed, rotated and can have various coloring, suiting well to the design of the shirt or the arena. In addition, there is a problem with the ambient illumination variation just as for other computer vision tasks. Second challenge is the large variety of different brand's logos. This makes the detection of logos very challenging. Furthermore, there are only a few smaller publicly available datasets, with bounding box annotated logos.

In the decade before, hand-crafted feature extraction was prevalent in computer vision tasks. It needed an expert to create such a system, and it yielded often only results. Deep learning methods for computer vision problems are dominant since the success of convolutional neural networks in 2012 [Kri12]. This development is mainly powered by the annually organized ImageNet classification challenge [Rus15]. Since the aim of this contest to classify an object, which is filling out the majority of an image, the location of the particular object is irrelevant. To be able to classify and recognize objects which have a much smaller size relative to the size of the whole image, region based classification can be utilized.

The rest of this thesis is organized as follows. Section 2 reviews the related work within image retrieval, object detection and logo retrieval. In Section 3 the proposal based object detection with convolutional neural networks will be introduced. Section 4 describes the logo retrieval system. Afterwards, Section 5 includes evaluation and comparison of the system with another logo retrieval method. Finally, the last section concludes the work and gives prospects on future work.

2

Related Work

2.1 Image Retrieval

Many technics outside the scope of deep learning exists for retrieving images from videos. SIFT features [] with bag-of-visual-words were used to efficiently get translation invariant descriptors around keypoints by Zisserman [].

2.2 Object Detection

Viola Jones

2.3 Logo Retrieval

3

Proposal Based Object Detection and Classification

3.1 Fully Convolutional Neural Networks

Fully Convolutional Neural Networks were proposed by Long et. al. [Lon14]. It is proposed how to transform a deep neural network with fully connected classifier layers at the end to a fully convolutional network. For this purpose the fully connected layers should be removed from the end of the network. Fully connected layers need a fix sized input. Since a fully convolutional network does not have fully connected layer anymore, it has the advantage of being able to train and test with images of arbitrary sizes. The output of such a network is usually a 2 dimensional feature map, which can be used as a per class heatmap. This conv map can also be used directly for semantic segmentation, where each pixel of an image should be assigned to a class.

3.2 Region proposal systems

Region proposals are possible object locations on the image. As we saw in the chapter two, the multiscale sliding window method induces a lot of computational costs. In order to reduce this computational burden, region proposal systems can be used. Earlier solutions used external proposal systems, e.g. Selective Search [Uij13]. Today the proposal system is already part of the neural network.

3.3 Region Based Convolutional Neural Network

Region based convolutional neural networks use external region proposals. The network is used to run the complete inference on every region proposal bounding boxes.

3.4 Fast Region Based Convolutional Neural Network

This type of network [Gir16] consists of a fully connected neural network, also called as base network in [Ser13], and a fully connected classifiation network. This version of region based convolutional neural network also uses external region proposals. First the image is completely pushed through the fully convolutional network. The output is a downscaled feature map, which is fed to the so called ROI Pooling layer. This layer crops regions from the map according to the appropriate downscaled region proposals, and executes pooling on each regions. After the pooling there are some fully connected layers, which make the final classification of the regions of interest. The advantage of this method is the much shorter inference time, achieved by the much lower computational redundancy of running convolutional layers on the whole image only once.

3.5 Faster Region Based Convolutional Neural Network

[Ren15]

4

Logo Retrieval System

4.1 Logo Datasets

The hunger of deep learning method for training data is well-known. As the publicly available logo datasets are quite small, a better training result can be achieved if the datasets are merged. The different datasets with the number of brands, images and bounding box rois can be seen in table 4.1. The total number of brands means the number of different brands altogether.

There are also trademark datasets available having a much greater cardinality [Tur17]. The images of this dataset contain however only the logo of a company, without any context of the logo. This dataset turned out to have no use for region based deep learning methods, since this approach needs to learn to distinguish between objects to be learned and the background. The network was trained with the fusion of FlickrLogos-32 and the trademark dataset, and tested with the evaluation method of FlickrLogos-32.

Table 4.1: Publicly available logo datasets with with bounding box annotations

	Number of brands	**Number of logo images**	**Number of ROIs**
BelgaLogos	37	1321	2697
Flickr Logos 27	27	810	1261
FlickrLogos-32	32	$70 \cdot 32 = 2240$	3404
Logos-32plus	32	7830	12300
TopLogo10	10	$10 \cdot 70 = 700$	863
Total (union)	**80**	12 901	20 525

4.2 Logo Detection

4.3 Logo Comparison

5 Evaluation

5.1 Logo Detection

5.2 Logo Retrieval

6

Conclusion

6.1 Summary

6.2 Future Work

Bibliography

[Gir16] GIRSHICK, Ross; DONAHUE, Jeff; DARRELL, Trevor und MALIK, Jitendra: Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* (2016), Bd. 38(1): S. 142–158, URL `http://dx.doi.org/10.1109/TPAMI.2015.2437384`

[He15] HE, Kaiming; ZHANG, Xiangyu; REN, Shaoqing und SUN, Jian: Deep Residual Learning for Image Recognition. *CoRR* (2015), Bd. abs/1512.03385, URL `http://arxiv.org/abs/1512.03385`

[Kri12] KRIZHEVSKY, Alex; SUTSKEVER, Ilya und HINTON, Geoffrey E: ImageNet Classification with Deep Convolutional Neural Networks, in: F. Pereira; C. J. C. Burges; L. Bottou und K. Q. Weinberger (Herausgeber) *Advances in Neural Information Processing Systems 25*, Curran Associates, Inc. (2012), S. 1097–1105, URL `http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf`

[Lon14] LONG, Jonathan; SHELHAMER, Evan und DARRELL, Trevor: Fully Convolutional Networks for Semantic Segmentation. *CoRR* (2014), Bd. abs/1411.4038, URL `http://arxiv.org/abs/1411.4038`

[Ren15] REN, Shaoqing; HE, Kaiming; GIRSHICK, Ross und SUN, Jian: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, in: C. Cortes; N. D. Lawrence; D. D. Lee; M. Sugiyama und R. Garnett (Herausgeber) *Advances in Neural Information Processing Systems 28*, Curran Associates, Inc. (2015), S. 91–99, URL `http://papers.nips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks.pdf`

[Rus15] RUSSAKOVSKY, Olga; DENG, Jia; SU, Hao; KRAUSE, Jonathan; SATHEESH, Sanjeev; MA, Sean; HUANG, Zhiheng; KARPATHY, Andrej; KHOSLA, Aditya; BERNSTEIN, Michael; BERG, Alexander C. und FEI-FEI, Li: ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)* (2015), Bd. 115(3): S. 211–252

[Ser13] SERMANET, Pierre; EIGEN, David; ZHANG, Xiang; MATHIEU, MichaÃ«l; FERGUS, Rob und LECUN, Yann: OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks. *CoRR* (2013), Bd. abs/1312.6229, URL `http://dblp.uni-trier.de/db/journals/corr/corr1312.html#SermanetEZMFL13`

[Tur17] TURSUN, Osman; AKER, Cemal und KALKAN, Sinan: A Large-scale Dataset and Benchmark for Similar Trademark Retrieval. *CoRR* (2017), Bd. abs/1701.05766, URL `http://arxiv.org/abs/1701.05766`

[Uij13] UIJLINGS, J.R.R.; VAN DE SANDE, K.E.A.; GEVERS, T. und SMEULDERS, A.W.M.: Selective Search for Object Recognition. *International Journal of Computer Vision* (2013), URL http://www.huppelen.nl/publications/selectiveSearchDraft.pdf

List of Figures

List of Tables

Acknowledgment

»Physics is to mathematics as sex is to masturbation«

R.P. Feynman

»In der Informatik geht es genauso wenig um Computer wie in der Astronomie um Teleskope.«

Dijkstra