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

Todo list

1

Introduction

Advertising through static logos is one of the most important marketing method. A very effective possible way to reach a lot of people with these static logos, to sponsor sport teams or to buy advertising spaces in sport events. However the prices of these surfaces defined by the broadcasting companies mean huge expenses for the companies. The calculation of the costs is based firstly on the proportional size of the logo to the screen, and secondly on the time one particular logo is visible on the screen. 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 these data to pay only for the provided service.

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

The challenge of this task is manifold. 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 of the ambient illuminance variation just as for other computer vision tasks. Second challenge is that however the logo detector is trained to recognise only one class, different brand's logos have a huge appearance variation. 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 last, hand-crafted feature extraction was prevalent in computer vision tasks. It needed an expert and often did not yield so good results. Deep learning methods for computer vision problems are dominant since the success of convolutional neural networks in 2012 [VGG]. This development is mainly powered by the annually organised ImageNet classification challenge (ILSVRC). 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 does not play a specific role. To be able to classify and recognise objects, which have a much smaller size relative to the size of the whole image, region based classification can be utilised.

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 I will introduce the proposal based object detection with convolutional neural networks. In Section 4 I describe the logo retrieval system. Afterwards in Section 5 I evaluate and compare the system with another logo retrieval method. Finally I conclude my work and give prospects on future work in Section 6.

2

Related Work

2.1 Image Retrieval

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 . In that paper they introduce how to transform a neural network with classificator layers at the end to a fully convolutional network. The fully connected layers should be removed from the end of the network. Fully connected layers need a fix sized input. As the 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 As we saw in the chapter two, the multiscale sliding window method induces a lot of computational costs. In order to reduce this burden region proposal systems can be used. Earlier solutions used external proposal systems, e.g. selective search. 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 network consists of a fully connected neural network, also called as base network [OverFeat], 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 much lower computational costs. This is 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

4

Logo Retrieval System

4.1 Logo Datasets

TABLE

4.1.1 FlickrLogos-32

4.1.2 Flickr Logos 27

4.1.3 TopLogo10

4.1.4 BelgaLogos

4.1.5 Logos-32plus

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

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