Master's Thesis

Authors:
Christoffer Bøgelund Rasmussen

Supervisors:
Kamal Nasrollahi

Aalborg University
VGIS
10th semester
Group 17gr1041

TBA

AALBORG UNIVERSITY

Title:
Master's Thesis
Subject:
Vision, Graphics and Interactive Systems
Project period:
1/2-2017 to TBA
Project group:
17gr1041
Participants:
Christoffer Bøgelund Rasmussen

TBA

Supervisor:
Kamal Nasrollahi
Printed copies:
TBA
Number of pages:
TBA
Appendix media:
AAU digital exam zip file
Finished:
TBA

Preface

Reading Guide

Tables, code listings and figures are numbered sequentially within each chapter. Citations are written as [x] where x denotes the reference number used in the bibliography. Code classes and functions are written as `class` and `function()`, respectively. Additional files have been uploaded to the AAU Digital Exam.

Christoffer Bøgelund Rasmussen

Contents

1 Introduction

Object detection is a fundamental area of computer vision that has had a great amount of research over the past decades. The general goal of object detection is to find a specific object in an image. The specific object is typically from within a pre-defined list of categories that are of interest for a given use case. Object detection generally consists of two larger tasks; localisation and classification. It is assumed that the objects of interest are not already located in the image and as objects can vary in number of pixels depending on factors such as distance and scale, objects must be both localised in an image and classified accurately. Localisation is typically done by with a bounding-box indicating where a given object is in the image. However, other methods such as objects' centres and closed boundaries can also be used. Not only is object detection an important task in localising and classifying, it is also a necessary earlier step in larger computer vision pipelines. For example, object detection is needed within the tasks such as activity and event recognition, scene understanding, and robotic picking.

Object detection is a challenging problem due to both some large scale issues and minute differences. Firstly, there is the challenge of differentiating objects between classes. Depending on the problem at hand the sheer number of potential categories present can be into the thousands or tens of thousand. On top of this separate object categories can be both very different in appearance, for example an apple and an aeroplane, but separate categories can also be similar in appearance, such as dogs and wolves.

Current state-of-the-art within object detection is also within the realm of deep learning with Convolutional Neural Networks (CNNs). This is exemplified with almost all entries in benchmark challenges such as Pattern Analysis, Statistical Modelling and Computational Learning Visual Object Classes (PASCAL VOC) [1], ImageNet [2], and Microsoft Common Objects in Context (MS COCO) [3] consisting of CNN based approaches. However, improvements are still needed before object detection can be used in real-world scenarios that require a high level of precision, accuracy, and performance.

1.1 Initial Problem Statement

An initial problem statement can be formed as follows:

> *What are the specific challenges within object detection?*

Based upon this, Chapter *2 Problem Analysis* will outline these challenges. On top of this, related work into object detection, both current state-of-the-art and also classic methods, will be researched.

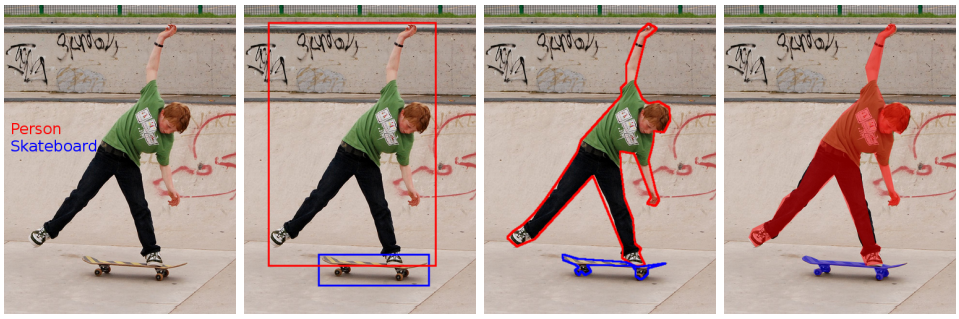
2 Problem Analysis

This chapter will outline object detection and it's key challenges. This includes aspects within robustness, computational-complexity and scalability. Once completed the key works within object detection will be analysed, both current state-of-the-art and notable older methods.

2.1 Object Detection

As mentioned in Section *1.1 Initial Problem Statement*, object detection consists of two larger tasks; classification and localisation. Depending on the problem at hand, object detection can be split into two categories. If only a single class is of interest, such as detecting a specific traffic sign, the object detection task is denoted as class-specific detection. Whereas, the more general case when multiple classes are of interest in an image is denoted as multi-class detection [4]. Key challenges such as PASCAL VOC, ImageNet, and MS COCO are of the latter task. This thesis will be within the multi-class detection domain and take these challenges into account when analysing related works in Section **?? ??** and determining the algorithm to be implemented and evaluated in Section **?? ??**. An analysis of these key challenges is done in Section **?? ??**. The goal of a detector is to output a list of labels from a predefined list of categories indicating which objects are present and where they are located in an image. Object detection has a number of related fields which share to common goal of categories relevant objects. This can be seen in Figure 2.1. In all four instances the goal is to categorise the two objects person and skateboard, however, the difference lies in the level of localisation precision. In Figure 2.1a, object categorisation aims to only classify the objects in the image without providing any indication as to where the objects are located. Object class detection in Figure 2.1b, localises the classified objects with the use of bounding-boxes, where ideally the bounding-boxes are placed as tightly around the given object as possible. Figure-ground segmentation in Figure 2.1c, indicates localisation with a lasso outline around the objects. Finally, in Figure 2.1d, semantic-segmentation localises objects at a pixel-level classifying each pixel that is related to the given object.

correct section refs to above



(a) **(b)** **(c)** **(d)**

Figure 2.1: Example of vision tasks related to object detection. All tasks have the common goal of categorising predefined objects. Methods are: object categorisation (a), object class detection (b), figure-ground segmentation (c), semantic Segmentation (d). Image and class labels taken from MS COCO [3].

A more recent example of segmentation is that of instance segmentation. Instance

segmentation varies to semantic segmentation in that individual instances of objects are classified as such. If multiple instances of the same object is present, such as an image of a crowd with many people, in semantic segmentation all people will be given the same label as one large group. However, in instance segmentation the people are still given the same label but individual instances of a person is also found. This area of research within segmentation is relatively new, however, is beginning to become more popular in comparison to semantic segmentation. For example, the MS COCO segmentation challenge which has been held in 2015 and 2016 only accepts instance segmentation entries.

2.2 Main Challenges

The challenges of object detection can be split into two groups as per [4]:

- Robustness-related.
- Computational-complexity and scalability-related.

Robustness-related refers to the challenges in appearance within the both of intra-class and inter-class. Intra-class is the differences in appearance of objects which are of the same class. For example as seen in Figure 2.2, all of the images belong to the superclass chair from the ImageNet training set [2], however, vary greatly in their overall appearance.

(a) **(b)** **(c)** **(d)**

Figure 2.2: Examples of intra-class appearance variation. All images have the label *chair* in the ImageNet training set [2].

An object detection system must be able to learn

explanation that typically obj detection is supervised learning

the appearance variations that can occur intra-class. These variations can be categorised into two types as per [5]:

- Object variations.
- Image variations.

2.3 Implementation Outline

As per [4] the steps in the general pipeline for an object detection system is as follows:

1. Find all possible object regions in the image.
2. Determine if the regions correspond to any of the predefined categories.
3. Evaluate all responses from step 2 to determine final detections.

2.4 Related Work

3 Technical Analysis

4 Discussion

5 Conclusion

Bibliography

[1] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results," http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html.

[2] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[3] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, *Microsoft COCO: Common Objects in Context.* Cham: Springer International Publishing, 2014, pp. 740–755. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-10602-1_48

[4] X. Zhang, Y.-H. Yang, Z. Han, H. Wang, and C. Gao, "Object class detection: A survey," *ACM Comput. Surv.*, vol. 46, no. 1, pp. 10:1–10:53, Jul. 2013. [Online]. Available: http://doi.acm.org/10.1145/2522968.2522978

[5] F. Schroff, *Semantic Image Segmentation and Web-supervised Visual Learning.* University of Oxford, 2009. [Online]. Available: https://books.google.dk/books?id=4EqZYgEACAAJ

Appendices

Notes