Recognition and Prominence Ranking of Alphanumeric Number Sequences in Images

By Alex Cummaudo

BSc *Swinburne*

Supervised by Prof. Rajesh Vasa, Assoc. Prof. Andrew Cain

A thesis submitted in partial fulfilment of the requirements for the
Bachelor of Information Technology (Honours)

Deakin Software and Technology Innovation Laboratory
School of Information Technology
Deakin University, Australia

27 October 2017

Abstract

Text detection in natural images is a growing area with increasing applications, including traffic sign and license plate recognition, and text-based image search. Robustly detecting and recognising text is especially challenging when text is deformed, such as the photometric and geometric distortions of text worn by a moving subject in unstructured scenes. Existing methods of text detection in such cases are classified as learning-based or connected component (CC)-based, applying a mix of enhanced detection techniques—such as stroke width transformation (SWT), canny-edge detection and maximally stable extremal regions (MSERs)—and feeding candidates into optical character recognition (OCR) engines or neural networks to recognise the text. This study proposes applying a learning-based approach using deep-learning strategies to automate the recognition of racing bib numbers (RBNs) in a natural image dataset of various marathons, and then ranking detected subject's photos in order of prominence. Experimental results showed that these deep-learning strategies performed favourably against other methods using a consistent dataset, prompting further investigation in the generality of the technique developed to other similar subject material.

Declarations

I certify that the the thesis entitled "Recognition and Prominence Ranking of Alphanumeric Number Sequences in Images" submitted for the degree of Bachelor of Information Technology (Honours) is the result of my own work and that where reference is made to the work of others, due acknowledgement is given. I also certify that any material in the thesis which has been accepted for a degree or diploma by any university of institution is identified in the text.

Alex Cummaudo, BSc *Swinburne*
27 October 2017

We certify that the thesis prepared by Alex Cummaudo entitled "Recognition and Prominence Ranking of Alphanumeric Number Sequences in Images" is prepared according to our expectations and that the honours coordinator can proceed to accept this submission for examination.

Prof. Rajesh Vasa
27 October 2017

Assoc. Prof. Andrew Cain
27 October 2017

Acknowledgements

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

I would also like to thank Andrew Cain for his extraordinary efforts over many years to teach hundreds of students (myself included) and who has developed a valued mentorship with me in guiding me throughout my academic life.

Contents

List of Figures

List of Tables

List of Abbreviations

CNN Convolutional Neural Network.

OCR optical character recognition.

RBN Racing Bib Number.

Chapter 1

Introduction

Ever since the development of low-cost cameraphones, we have seen an increasing interest for image processing analysis, but text recognition still faces challenges within images of unstructured scenes. While successes in character recognition have long been developed and improved upon with Optical Character Recognition (OCR) engines (Smith, 1987), these are typically applied under strict conditions. Once applied within the context of a natural scene, real-world discrepancies pose serious shortcomings, such as illumination and viewpoint conditions, blur and glare variations, geometric and photometric distortion, and differences in font size and style. Realising potential applications to overcome similar issues has motivated a variety of different proposed techniques.

Chen et al. (2011) develop a robust text detection on the fundamental principle that the text itself must be robustly located in the image. Similarly

1.1 Background

1.2 Research Goals

Goal 3: ***Detect RBNs using a CNN***

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet

aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Goal 3: ***Design a CNN that recognises RBN sequences without character segmentation***

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Goal 3: ***Rank prominence of alphanumeric sequences***

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

1.3 Thesis Organisation

This thesis is organised into the chapters as outlined below. An appendix follows with additional supplementary material.

Chapter 2 - Background Work Provides an overview of prior studies broadly around the areas of number recognition in image processing and artificial NNs.

Chapter 3 - Related Work Documents a number case studies within the literature directly or closely related to the aims of this research.

Chapter 4 - Research Methodology Describes possible techniques in closer depth to develop a number recognition pipeline, and explores ways to develop prominence ranking techniques.

Chapter 5 - Benchmarking Collates results of a series of experiments using our dataset amongst other tools and pipelines currently developed.

Chapter 6 - Processing Pipeline Discusses the proposed processing pipeline developed that satisfies the aims of this study.

Chapter 7 - Deep Learning Comparison Compares our deep-learning approach with those benchmarked.

Chapter 8 - Validation of Results Highlights a number of validation techniques used to ensure results found in the comparison are correct.

Chapter 9 - Discussion and Limitations Presents implications that were found from the results of our findings and possible limitations.

Chapter 10 - Conclusions and Future Work Draws a number of conclusions and alleviates gaps in the findings of this work by presenting future studies.

Chapter 2

Background Work

In this chapter, we survey a range of literature

2.1 Applications of Image Recognition

2.2 Detection Strategies

2.3 Recognition Strategies

2.4 Prominence Strategies

Chapter 3

Related Work

3.1 RBN Recognition

Fu et al. (2015)

3.2 Speed Limit Sign Recognition

3.3 License Plate Recognition

Chapter 4

Research Methodology

4.1 Overview

4.2 Prominence Ranking Survey

This section encapsulates an experiment to capture prominence rankings of a given sample of the dataset. In this context, prominence is defined as the prominence of a particular marathon runner is within a photo, as identified by the runner's RBN. Results gathered from this experiment will assist in developing a quantitative measure of humans identify prominence within our context. We present participants with a number of subjects and ask to rank them by a prominence Likert scale. The aggregated results of the findings are used as a prominence training dataset fed into a deep-learning neural network.

4.2.1 Survey Design

The survey published for the experiment was collected online via Google Forms[1]. The collection period was for ⟨ *number of months* ⟩ months between ⟨ *survey start date* ⟩ and ⟨ *survey end date* ⟩.

Previous chapters indicated that

Images

4.2.2 Ethics Approval

4.2.3 Demographics

[1] `http://forms.google.com` last accessed 8 May 2017.

Chapter 5

Benchmarking

5.1 Open Source Tools

5.2 Existing Pipelines From Literature

5.3 Hermes Approach

Chapter 6

Processing Pipeline

Chapter 7

Deep Learning Comparison

Chapter 8

Validation of Results

Chapter 9

Discussion and Limitations

Chapter 10

Conclusions and Future Work

References

Chen, H., S. S. Tsai, G. Schroth, D. M. Chen, R. Grzeszczuk, and B. Girod (2011). Robust text detection in natural images with edge-enhanced Maximally Stable Extremal Regions. *ICIP*.

Fu, C., C.-W. Cheng, W.-H. Shen, Y.-L. Wei, and H.-M. Tsai (2015). LightBib: Marathoner Recognition System with Visible Light Communications. In *2015 IEEE International Conference on Data Science and Data Intensive Systems (DSDIS)*, pp. 572–578. IEEE.

Smith, R. W. (1987). *The Extraction and Recognition of Text from Multimedia Document Images*. Ph. D. thesis, University of Bristol.

Appendix A

Ethics Clearance

Appendix B

Prominence Ranking Survey Results