DEEP LEARNING BASED RACING BIB NUMBER DETECTION AND RECOGNITION

CHOI LI JIA

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

DEEP LEARNING BASED RACING BIB NUMBER DETECTION AND RECOGNIITION

CHOI LI JIA

This report is submitted in partial fulfilment of the requirements for the degree of Bachelor of Electronic Engineering with Honours

Faculty of Electronic and Computer Engineering
Universiti Teknikal Malaysia Melaka

2019

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

FAKULTI KEJUTERAAN ELEKTRONIK DAN KEJURUTERAAN KOMPUTER

BORANG PENGESAHAN STATUS LAPORAN

PROJEK SARJANA MUDA II

Tajuk Projek : Deep Learning Based Racing Bib Number Detection and Recognition

Sesi Pengajian : 2018/2019

Saya CHOI LI JIA mengaku membenarkan laporan Projek Sarjana Muda ini disimpan di Perpustakaan dengan syarat-syarat kegunaan seperti berikut:

1. Laporan adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan laporan ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. Sila tandakan (✓):

☐	**SULIT***	(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)
☐	**TERHAD***	(Mengandungi maklumat terhad yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan.
☐	**TIDAK TERHAD**	

Disahkan oleh:

(TANDATANGAN PENULIS) (COP DAN TANDATANGAN PENYELIA)

Alamat Tetap: 34, Jalan Harmoni 5/3, Taman Desa Harmoni 81100, Johor Bahru, Johor

Tarikh : 30 May 2019

Tarikh : 30 May 2019

*CATATAN: Jika laporan ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa/organisasi berkenaan dengan menyatakan sekali tempoh laporan ini perlu dikelaskan sebagai SULIT atau TERHAD.

DECLARATION

I declare that this report entitled "Deep Learning Based Racing Bib Number Detection And Recognition" is the result of my own work except for quotes as cited in the references.

Signature : ....

Author : ....

Date : ....

APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of Bachelor of Electronic Engineering with Honours.

Signature	:	…………………………………
Supervisor Name	:	…………………………………
Date	:	…………………………………

DEDICATION

I would like to dedicate my project report to my beloved family.

ABSTRACT

Nowadays, healthy lifestyle trends are more prominent globally around the world. There are numerous number of marathon running race events that have been held and inspired interest among peoples of different ages, genders and countries. Such diversified truths increase more difficulties to comprehending large number of marathon images since such process is often done manually. Therefore, a deep learning based racing bib number (RBN) localization and recognition for marathon running races is implemented. RBN is the identification tag number wore by each runner during marathon running races. The architecture pipeline is consists of two phases. Two different networks are utilized which are You Only Look Once version 3 (YOLOv3) and Convolutional Recurrent Neural Network (CRNN). During first phase, YOLOv3 consists of single convolutional network that used to predict RBN by multiple bounding boxes and class probabilities of boxes. Next the RBN detected can be parsed into CRNN to undergo RBN recognition. For second phase, CRNN is used to output a label sequence for each input image and then selecting the label sequence that has the highest probabilities. As a result, CRNN will output the contents of RBN detected. All of the experimental results including of mean average precision (mAP) and edit distance have been analysed and evaluated in the project thesis.

ABSTRAK

Pada masa kini, trend gaya hidup sihat semakin menonjol di seluruh dunia. Terdapat banyak acara perlumbaan dijalankan dan kejadian ini telah memberi inspirasi dalam kalangan orang berumur, jantina dan negara yang beza. Fenomena ini telah mengakibatkan banyak kesukaran untuk menguruskan imej and video dirakam daripada acara perlumbaan kerana proses ini dilakukan secara manual. Oleh sebab itu, penyetempatan dan pengenalian racing bib number (RBN) berasaskan deep learning telah dilaksanakan. Perancangan seni bina projek ini terdiri daripada dua peringkat. Dua rangkaian berbeza telah diguna iaitu You Only Look Once version 3 (YOLOv3) dan Convolutional Recurrent Neural Network (CRNN). Pada peringkat pertama, YOLOv3 mempunyai rangkaian konvolusi tungga yang digunakan untuk meramal RBN dengan beberapa peti bounding dan kebarangkalian kelas kotak. Selepas itu, RBN yang ditempatkan akan dihantar ke CRNN untuk menjalankan RBN pengenalian. Pada peringkat kedua, CRNN diguna untuk menghasilkan urutan label bagi setiap imej input dan kemudian memilih urutan label yang mempunyai kebarangkalian tertinggi. Akibatnya, CRNN akan menyembahkan kandungan RBN yang dikesan. Semua keputusan eksperimen termasuk ketepatan, min purata ketepatan (mAP) dan edit jarak telah dianalisis dan dinilai dalam projek tesis ini.

ACKNOWLEDGEMENTS

This acknowledgement was written after I finish the paramount important task that is final year project in Universiti Teknikal Malaysia Melaka (UTeM). Throughout proceeding this project, I have received countless help from supervisor, lecturer and friends for their support. Without their contribution, this project will never be accomplished. First and foremost, I would like to appreciate to my parents, Tan Ling Ling and Choi Mun Hin for their unlimited encouragement and attention anytime.

Moreover, I would like to express my sincere thankfulness to my supervisor, Dr Wong Yan Chiew for providing me with all the knowledge and information about this project. She has given many valuable guidance and encouragement extended to me all the times.

Furthermore, I would also like to take this opportunity to express my gratitude to all the lecturers of Faculty of Electronic and Computer Engineering who had provide me comments and advices. Last but not least, I would like to express my gratitude to one and all who directly or indirectly lend their hand for completing this project and thesis.

TABLE OF CONTENTS

Table 1: Results of the proposed approach for RBN detection and recognition

Methods	Training Images	Technique	Time (s)	P	R	F
P. Shivakumara [1]	212	Torso + HOG	N/A	**0.60**	**0.74**	**0.66**
Ami *et al* [2]	217	Face + SWT	N/A	0.53	0.40	0.45
N. Boonsim [3]	400	Face + Torso + Edge	~ 0.75	0.59	0.69	0.64

According to Table 1, "P" represents for Precision, "R" represents for recall and "F" represents for F-score that are used to indicate the performance of the proposed approaches. F-score is calculated by using equation (1) where α = 0.5:

$$F = \frac{P \,.\, R}{\alpha \,.\, R + (1-\alpha).P} \text{ ------------------- } (1)$$

It can be seen that P. Shivakumara [1] outperforms the other approaches by achieved the highest accuracy, precision and F-score. They proposed a multi-modal approach that includes of three steps which are torso localization, foreground segmentation and torso extraction regarding to appearance models and image parsing. This method combines torso and text detection method that is not related to runner movement directions and do not requires the face information as method proposed by Ami et al. [2] who used stroke width transform (SWT) [3], [4] method to extract characters in input image and the characters with similar stroke width are grouped together for producing text region. Then face detection method is applied in order to detect the face of runner. At the end the detected torso is used to indicate and recognize the true RBN by using Tesseract OCR [5] engine. The limitation of such method is there will be a tremendous large quantity of runners run with different speeds during running races such as Marathon and hence it is very difficult to localize and detect each faces of them. Moreover, N. Boonsim [6] applied edge-based technique [7], [8] to extract edges of an input image captured during marathon races and local contrast improvement method [7] is employed to enhance the contrast of image.

In this project, artificial intelligent cascade network which can automatically detect and recognize RBN during marathon race will be implemented on GPU by using Deep Learning. It follows with analyzing and enhancing the performance of the system that able to achieve high accuracy and precision. This project is predominantly to develop a RBN detection and recognition cascade network in marathon running races based on deep learning. The text language is limited to representative tag number of runner only, where the font size and type of the text dataset are typically depends on the resources of such dataset. Name of logo and name of runners are not concerned in this project.

II. Methods

Two different networks are utilized which are You Only Look Once version 3 (YOLOv3) and Convolutional Recurrent Neural Network (CRNN) to localize and recognize RBN. These two networks is chosen due to their state-of-the-art performance with high accuracy and precision. The priority benefit of the proposed cascade network is that it has the capability to automatically record down RBN from each runners at finish line without human assistant and recognize the RBN from images or videos captured during marathon events.

A. Phase 1- YOLO v3

In this project, first dataset used to train for YOLOv3 has the total amount of 4096 images which consists of 1070 pixels as width and 1600 pixels as height in png format and it is created by collecting images one by one from Internet [9] and such images are captured at the places near to finish line during Delaware marathon running race year 2018. It is then annotated by a graphical image annotation tool which named as Labelimg. Every image is provided with ground truth values which indicate the detail annotation and temporal localization of each target objects that are runner, bib and number.

YOLO [10] is a single convolutional network for object detection. The concept is you only look once (YOLO) at an image to predict what are the objects presented and its location. It uses single regression to detect target object directly from image pixels by predicting multiple bounding boxes and class probabilities for those boxes. YOLO trains on full images and straight away optimizes detection performance. YOLOv3 [11] is the third object localization algorithm in family. The accuracy by using it to detect small object has been refined compared to other members.

In Phase 1, runner and racing bib and number are detected by using YOLOv3. The overall process of fetching the labeled images to YOLOv3 model is associated by using OpenCV library [12]. By applying the OpenCV library, all the input images are scaled to the same size. After resizing the input images then such images will be fed to YOLOv3. The network architecture of YOLOv3 is separated into several modules. First layer is the total of 32 convolutional layer where the input images have been supplied to. These convolutional layers is implemented to extract features from input images.

Apart from that, second layer is the residual layers. Such layers are proposed to vary the training process of the deep neural network from layer-by-layer into phase-by-stage training. Therefore the deep neural network is separated into few segments in order to realize the problem of gradient explosion or gradient dispersion of the network. Third layer is darknet-53 which range from 0th layer until 74th layer. There are total of 53 convolutional layers and the remaining are the residual layers. Darknet-53 acts as a key components of the network architecture to extract feature and it implements a series of 3 x 3 and 1 x 1 convolutional layers as shown in Figure 2.

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Figure 1: Network Structure of YOLO v3 [11]

Last layer is the feature interaction layer of the YOLO network. It is separated into three scales that are region 82, 94 and 106 in the feature pyramid network. Local feature interaction is implemented by convolutional kernel which known as a fully connected layer in each regions. It is used to obtain local feature interaction among feature maps and hence accomplish regression and classification.

B. Phase 2- CRNN

Second dataset used to train for CRNN is composed of 100,000 images that contains digits numbers only in grey scale as shown in Figure 3. These images are generated by a text generator python script. Types of font that are similar to the images of first dataset are searched and downloaded from Internet and the path of font type are specified in the script. Type of fonts used to generate the second dataset includes of Identikal Sans Bold, Contax Bold, Roadway and Sofia Pro Semibold and so on. Only one type of font is utilized to generate texts at one time and then keep switching another font types to generate dataset until it accumulates to a total of 100,000 images. In the text generator script, only digit numbers are specified and thus it will generate images which composed of only digit numbers in random arrangement.

Figure 2: Second Dataset Generated

RBN text is recognized by using CRNN model. CRNN [13] is the combination network of DCNN and RNN. For sequence-like objects, CRNN composed of several unique advantages in six aspects. Firstly, CRNN can be straight leaned from sequence labels such as characters. Secondly it has the same characteristic as DCNN on learning informative representations straightly from image data which do not need hand-craft features and preprocessing steps which includes of binarization, segmentation or component localization. Next, it possess the same characteristic of RNN that able to generate a sequence of labels.

According to Figure 4, the network architecture of CRNN composed of three components which from bottom to top that are convolutional layers, recurrent layers and a transcription layer. At the most bottom of CRNN, the convolutional layers automatically extract a sequential feature from each input image. While on top of the convolutional network, a recurrent network is constructed for making prediction for each frame of the feature sequence which is outputted from the convolutional layers. The transcription layer at the top of CRNN is utilized to translate the per-frame predictions by the recurrent layers into a label sequence.

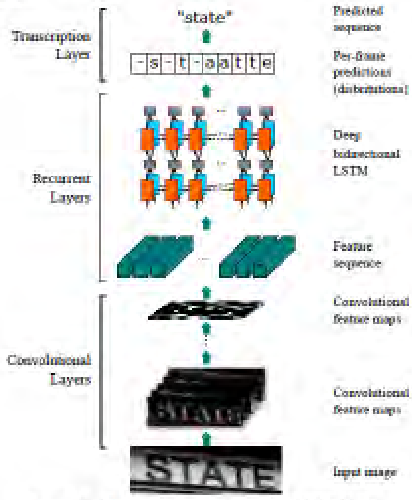

Figure 3: Network Structure of CRNN [12]

Before start to do training for CRNN model, it is essential to convert the second dataset that composed of 100,000 images to Lightning Memory-Mapped Database (LMDB) format. LMDB is a software library that supplies an embedded transactional database in key-value store format. Key-value store is considered as data storage built for saving, extracting and manipulating associative arrays. Since LMDB employs memory-mapped files and hence it enhances the performance of input and output.

C. Cascade Both Models

Two pre-trained models that trained from two different datasets are obtained as shown in Table 2. Such models are YOLOv3 for runners and RBN detection and CRNN for RBN recognition are combined together and form a cascade network that aimed for RBN detection and recognition in deep learning based as shown in Figure 5.

Table 2: Training and Testing Subset

Dataset	Training Subset	Testing Subset	Total
1 (RBN)	2868	1228	4,096
2 (Numbers)	700,000	300,000	100,000

LIST OF FIGURES

(Please delete this part): This list contains the titles of figures, together with their page numbers, which are listed in the text. For e.g., figures in Chapter 3 are numbered sequentially: Figure 3.1, Figure 3.2.

For title of list tables use *other title* and *TOC1* for style.

LIST OF TABLES

(Please delete this part): This list contains the titles of tables, together with their page numbers, which are listed in the text. The numbering system is according to chapter, for e.g.: tables in Chapter 3 are numbered sequentially: Table 3.1, Table 3.2.

LIST OF SYMBOLS AND ABBREVIATIONS

For examples:

RBN	:	Racing bib number
CNN	:	Convolutional Neural Network
RNN	:	Recurrent Neural Network
CRNN	:	Convolutional Recurrent Neural Network
DNN	:	Deep Neural Network
YOLO	:	You Only Look Once

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

This thesis proposes the implementation of deep learning-based racing bib number (RBN) detection and recognition for marathon running races. This chapter will present about the project background, problem statement, objectives, the scope of works, project significant and chapter review.

1.2 Project Background

Over the past decades, the great popularization of intelligence devices and rapid development of technology have brought forth enormous of new outcomes and services that have prompted the huge demand of practical computer vision technologies. As opposed to the scanning of document, natural scene text localization and recognition contributes a method to straightly access and utilize the textual data in the wild which apparently is the most pressing technologies. Subsequently, text localization and recognition in natural scene have enticed many significant attention from the communities of computer vision and document analysis.

Nowadays, healthy lifestyle trends are more prominent globally around the world. Such new trend is to organize such running activities in order to inspire the awareness of the public about the importance of health. There are numerous number of marathon or distance running race events that raises in different formats for different situations have been held and inspired interest among peoples of different ages, genders and countries. Such diversified truths increase more difficulties to comprehending marathon video or images. The runners or participants have a single racing bib number (RBN) to indicate themselves and this bib number presents on the heterogeneous backgrounds and different materials. Numerous number of images captured by organizers and photographers attending the marathon running races are significantly increased.

As a result, a deep learning based racing bib number (RBN) detection and recognition system is proposed. Two distinct neural network frameworks are utilized in this system and therefore combine into one cascade networks that purposed for RBN detection and recognition. During Phase 1 which is RBN detection, You Only Look Once version 3 (YOLOv3) system will be applied while phase 2 will be using Convolutional Recurrent Neural Network (CRNN) to recognize RBN detected.

1.3 Problem Statement

Scene text are usually short snippets written in different languages and fonts and text arrangement normally does not obey strict rules of printed documents. More broadly, natural scene text localization and recognition is a problem of interest to the optical character recognition community. Today, RBN identification is often done manually which a process made difficult by the sheer number of available photos. Tag number localization and detection is commonly used in various traffic and security applications, for instance parking, border control and analysis of traffic flow. There are multiple approaches for natural scene text localization and recognition in images and video form and also in high speed and high accuracy. However, these approaches are not better enough to obtain accurate and precise results for RBN localization in marathon images because they usually relies on characteristics of rich texts but not numerals. In running races, each competitors has an identification tag number named Racing Bib Number (RBN). Such tag number is generally printed on a paper of cardboard tag and pinned onto the T-shirt of different color and at different parts of each competitor's body during marathon race. There are some problems faced for application of RBN detection and recognition. Firstly, competitors run together with different speeds and background complexity varies continuously with moving objects, sky, buildings, tree and others. Secondly, input images captured in natural scene usually influenced by occlusion and the loss of information or quality. Thirdly, inconsistent light intensities can influence the images captured and it results in unbalanced illumination throughout the images. Therefore, achieving accurate performance for localizing and recognizing RBN in marathon races is a challenging and difficult task.

1.4 Objectives

- To implement an artificial intelligent cascade network which can automatically detect and recognize racing bib number during marathon race on GPU by using Deep Learning.
- To analyze and enhance the performance of the system that able to achieve high accuracy and precision.