

HAND GESTURES RECOGNITION USING DEEP NEURAL NETWORK

LIM BOON CHEONG

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

HAND GESTURES RECOGNITION USING DEEP NEURAL NETWORK

LIM BOON CHEONG

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

HAND GESTURES RECOGNITION USING DEEP NEURAL NETWORK

LIM BOON CHEONG

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

BORANG PENGESAHAN STATUS LAPORAN
PROJEK SARJANA MUDA II

Tajuk Projek : HAND GESTURES RECOGNITION USING DEEP NEURAL NETWORK
Sesi Pengajian : 2018/2019

Saya LIM BOON CHEONG 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:

Jon
(TANDATANGAN PENULIS)


(COP DAN TANDATANGAN PENYELIA)

Alamat Tetap: 33, JALAN 4,
TAMAN
SEANGIN
45800 JERAM,
SELANGOR

PROF. DR. ZULKALAIN BIN MOHD YUSSOF
Penyelia
Fakulti Kejuruteraan Elektronik Dan Kejuruteraan Komputer
Universiti Teknikal Malaysia Melaka (UTeM)
Hala, Tanah Merah
76100 Durian Tunggal, Melaka

Tarikh : 31 May 2019

Tarikh : 31 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 "HAND GESTURES RECOGNITION USING DEEP NEURAL NETWORK " is the result of my own work except for quotes as cited in the references.

Signature : *Jm*

Author : *Lim Boon Cheong*

Date : *31 May 2019*

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

Prof. Dr. Zulkainain Mohd-Yusoff

Date

31/5/2019

DEDICATION

Dedicated to my beloved family and friends who supported and encouraged me to finish this project.

ABSTRACT

Gestures are another type of communication tool which can be used to express idea, thought and feeling. It is usually used by those who have hearing and speech disability as mother language. This project was about implementation of a deep learning-based recognitions system for 30 MSL static hand gestures and evaluation of the performance of the system designed in term of recognition accuracy. This project was designed for those users who using the recognition system in front of an USB camera and the system designed was capable to recognize 30 MSL hand gesture sonly. The recognition system designed was made up by two hierarchical CNN architectures, in which YOLOV2 model was used for hand detection while MobileNet pretrained model was used for gestures classification of this project. Throughout this project, the hand gestures recognition system designed achieved 98% average testing accuracy on self-generated testing dataset. Given the limitations of the datasets and the encouraging results achieved, a fully generalizable translator for all 30 MSL static hand gestures can be produced with further research and inclusion of more dataset for MobileNet classification training.

ABSTRAK

Gerak isyarat adalah satu alat komunikasi yang boleh digunakan untuk menyatakan idea, pemikiran dan perasaan. Ia biasanya digunakan oleh mereka yang mempunyai kecacatan pendengaran dan pertuturan sebagai bahasa ibu. Projek ini adalah mengenai pelaksanaan sistem pengiktirafan berasaskan pembelajaran mendalam untuk 30 tangan gerak isyarat statik MSL dan penilaian prestasi sistem yang direka dari segi ketepatan pengiktirafan. Projek ini direka untuk pengguna yang menggunakan sistem pengiktirafan di hadapan kamera USB dan sistem yang direka mampu mengiktiraf 30 tanda tangan MSL sahaja. Sistem pengiktirafan yang direka dalam project ini terdiri daripada dua seni bina CNN hierarki, di mana model YOLOV2 digunakan untuk pengesanan tangan manakala model MobileNet digunakan untuk pengkelasan gerak isyarat. Melalui projek ini, sistem pengenalan gerak isyarat tangan direka mencapai 98% ketepatan ujian purata atas data yang dikumpulkan sendiri. Memandangkan batasan data yang ada dan hasil dicapai yang menggalakkan, penterjemah yang sempurna dan umum untuk semua 30 gerak isyarat tangan statik MSL dapat dihasilkan dengan penyelidikan lanjut dan memasukkan lebih banyak data untuk latihan klasifikasi MobileNet.

ACKNOWLEDGEMENTS

First of all, I would like to express sincere appreciate to FKEKK, Universiti Teknikal Malaysia Melaka for giving me an opportunity to conduct this project. In the process to complete this project, there were also some respectable persons, who have given me great support and guidance. I would like to manifest my gratitude to Prof. Dr. Zulkalnain Bin Mohd Yusof, my final year project supervisor in Universiti Teknikal Malaysia Melaka, who helps me by his effectively advices and guidelines.

Furthermore, I would like to express my special gratitude to Tan Kien Leong and Mohamed for sharing their experience and lending a helping hand whenever I faced a problem.

Last but not least, I would like to express my deeply appreciation to all participants who taking part in my dataset collection. Without their involvement, I am not able to complete this project.

TABLE OF CONTENTS

Declaration	
Approval	
Dedication	
Abstract	i
Abstrak	ii
Acknowledgements	iii
Table of Contents	iv
List of Figures	viii
List of Tables	xi
List of Symbols and Abbreviation	xii
List of Appendices	xiii
CHAPTER 1: INTRODUCTION	1
1.1 Project Overview	2
1.2 Problem Statement	2
1.3 Objectives	3
1.4 Scope of Project	3

1.5	Project Significant	4
1.6	Thesis Outline	5
CHAPTER 2: BACKGROUND STUDY		6
2.1	Hand Gesture	7
2.2	Sign language	7
2.3	Vision based Hand Gestures Recognition System	8
2.4	Neural Network	9
2.5	Deep Learning	10
2.5.1	Transfer learning and fine-tuning	11
2.6	Deep Neural Network (DNN)	13
2.6.1	Convolutional Neural Network (CNN)	13
2.6.1.1	Convolutional Layers	14
2.6.1.2	Activation Function Layer	15
2.6.1.3	Pooling Layer	16
2.6.1.4	Fully Connected Layer	17
2.7	Backpropagation learning algorithm	18
2.8	Review of project-related work	18
2.9	Summary	26
CHAPTER 3: METHODOLOGY		28
3.1	Block diagram for Hand Gestures recognition system	29

3.2	Project Methodology	29
3.3	Hardware specification for DNN training	30
3.4	Dataset Collection	31
3.5	YOLOV2 architecture	32
3.6	DNN hand detection training	33
	3.6.1 Preprocessing of raw images for YOLOV2 training	33
	3.6.2 Training setup for YOLOV2	34
3.7	MobileNet architecture	35
3.8	DNN hand gestures classification training	37
	3.8.1 Dataset arrangement and distribution	37
	3.8.2 Training set up for MobileNet training	38
	3.8.3 Training strategies	40
3.9	Performance test on test dataset and in real time	41
	CHAPTER 4: RESULTS AND DISCUSSION	44
4.1	Result for hand detection layer	45
	4.1.1 Summary	46
4.2	Result for hand gestures classification layer	47
	4.2.1 Hand Gestures classification result for MobileNet (Without using dataset generated by data augmentation, without fine-tuning)	47

4.2.2	Hand Gestures classification result for MobileNet (With data augmentation generated through data augmentation, without fine-tuning)	51
4.2.3	Hand Gestures classification result for MobileNet (With data augmentation generated through data augmentation, Fine-tuned last 3 layers of MobileNet)	55
4.2.4	Summary and Discussion	59
4.3	Result and discussion for hand gestures recognition in real time environment	62
CHAPTER 5: CONCLUSION AND FUTURE WORKS		66
5.1	Conclusion	66
5.2	Future works	67
REFERENCES		68
APPENDIX A		73
APPENDIX B		78

LIST OF FIGURES

Figure 2.1: Examples of MSL static hand gestures. (Jooli Khoo, 2016)	8
Figure 2.2: A neural network architecture [6]. (Online source, 2017)	10
Figure 2.3: Illustration of transfer learning [8]. (Matlab Expo, 2017)	11
Figure 2.4: Performance of several pre-trained model against efficiency in ImageNet Large Scale Visual Recognition Challenge 2017 [9]. (Online Source)	12
Figure 2.5: Performance of several pre-trained model using Imagenet dataset [10]. (Online Source)	12
Figure 2.6: An example for CNN architecture [6]. (Online source, 2017)	14
Figure 2.7: Convolution in convolutional layer of CNN [8]. (Matlab Expo, 2017)	15
Figure 2.8: Types of popular activation functions [13]. (Online Source)	16
Figure 2.9: An example for max pooling and average pooling operation [8]. (Matlab Expo, 2017)	17
Figure 3.1: Illustration of project system.	29
Figure 3.2: Project flow chart.	30
Figure 3.3: Example of dataset collected.	32
Figure 3.4: YOLOV2 architecture details [27]. (Sik-Ho Tsang, 2018)	33
Figure 3.5: Content inside a xml file.	34
Figure 3.6: Configuration changed in yolov2-voc.cfg file.	35
Figure 3.7: MobileNet architecture [25]. (A. G.Howard <i>et al</i> , 2017)	36

Figure 3.8: Example of cropped raw input image by trained YOLOV2 model.	37
Figure 3.9: Illustration of MobileNet architecture with new top layer.	39
Figure 3.10: Formula of precision and recall [28]. (Ren Jie Tan, 2019)	41
Figure 3.11: Formula for AP [28]. (Ren Jie Tan, 2019)	42
Figure 3.12: Theoretical formula for f1 -score [28]. (Ren Jie Tan, 2019)	43
Figure 4.1: Training loss graph for YOLOV2 training.	45
Figure 4.2: mAP for trained YOLOV2 model.	46
Figure 4.3: Detection results for validation dataset.	46
Figure 4.4: Accuracy and loss of training and validation for model trained without data generated from data augmentation and without fine-tuning.	47
Figure 4.5: Normalized confusion matrix obtained from testing on test images.	48
Figure 4.6: Confusion matrix obtained from testing on test images.	48
Figure 4.7: F1-score obtained for each class/gesture.	50
Figure 4.8: Accuracy and loss of training and validation for trained model.	51
Figure 4.9: Normalized confusion matrix obtained from testing on test images.	52
Figure 4.10: Confusion matrix obtained from testing on test images.	52
Figure 4.11: F1-score obtained for each class/gesture.	54
Figure 4.12: Accuracy and loss of training and validation for trained model.	55
Figure 4.13: Confusion matrix obtained from testing on test images.	56
Figure 4.14: Normalized confusion matrix for testing on test images.	56
Figure 4.15: F1-score obtained for each class/gesture.	58
Figure 4.16: Accuracies obtained from three trained model.	59
Figure 4.17: Average f1-score from 3 trained models.	60
Figure 4.18: a) Correct label b) Wrong label	63

Figure 4.19: a) Correct label b) Wrong label

LIST OF TABLES

Table 2.1: Accuracy for each model test run.	21
Table 2.2: Average classification and training epoch duration on a GPU.	21
Table 2.3: Validation results.	22
Table 2.4: Summary of reviewed works.	24
Table 3.1: Hardware specification for this project.	30
Table 3.2: Deep learning Framework used in this project.	31
Table 3.3: Distribution of dataset for classification training.	38
Table 3.4: Dataset distribution for new generated dataset.	40
Table 4.1: Classification report for test performance of trained model.	49
Table 4.2: Classification report for test performance of trained model.	53
Table 4.3: Classification report for test performance of trained model.	57
Table 4.4: Result obtained from three trained models.	59
Table 4.5: Average F1-score for each trained model.	60

LIST OF SYMBOLS AND ABBREVIATION

CNN	:	Convolutional Neural Network
ANN	:	Artificial Neural Network
DNN	:	Deep Neural Network
RELU	:	Rectified Linear Unit
TP	:	True positive
TN	:	True Negative
FP	:	False positive
FN	:	False negative
MSL	:	Malaysia Sign Language
ASL	:	American Sign Language
PCA	:	Principal Component Analysis
LDA	:	Linear Discriminant Analysis
KNN	:	k-nearest neighbors
SVM	:	Support Vector Machine
IoU	:	Intersection over Union
mAP	:	Mean Average Precision
AP	:	Average Precision

LIST OF APPENDICES

Appendix A: Source code for classification training	73
Appendix B: Source code for classification evaluation	78

.....

CHAPTER 1:

INTRODUCTION

This chapter consists of five sections. The overview of the project is firstly introduced in this chapter. The problem statement regarding to the study are described in section 1.2. The objectives for this study are stated in section 1.3. Section 1.4 discusses the scope and limitations of work while section 1.5 describes the project significant. The thesis outline is presented in the final section of this chapter.

1.1 Project Overview

Gestures are another type of communication tool which can be used to express idea, thought, feeling and even for devices controlling. Meanwhile, hand gestures or gestures performed by one or two hands is the largest category of gestures due to large number of signs can be performed by varying the fingers or hands orientation or position.

From the past, the researches on hand gestures recognition for both static and hand gestures had been applied in many fields for example entertainment, automotive device controlling, communication and etc. Hand gestures recognition system can be implemented using various approaches in which deep learning approach is currently the most popular. Authors in studies [1] and [2] were conducted their researches on hand gestures recognition using different deep learning approaches.

Although there are lot of research had been conducted on hand gestures recognition for sign language users, there is still difficult to construct a robust sign language recognition system due to the numerous sign languages that are used in the world and for a sign language, there are a lot of different hand gestures. Communication between those who have speech or hearing disability and normal people is still becoming an issue.

In light of this, a deep learning-based hand gestures recognition system had been proposed to solve the communication issue between disabled and normal people to certain level.

1.2 Problem Statement

According to World Federation of the Deaf, 70 million deaf people are using sign language as their first language or mother tongue. Sign language is one of the effective communication tools for the people who are not able to speak and hear anything. It is

also useful for the person who are able to speak but not able to hear or vice versa[3]. However, due to the disabilities, deaf and dumb people are having difficulty to communicate with the normal people. This cause a communication gap or ideas sharing obstacle and collaboration problem for the deaf -mute community and people who can speak and hear since not everyone learns sign language.

In addition, from previous studies, most of Malaysian Sign Language (MSL) static hand gestures recognition systems were implemented using conventional approach. Deep learning is a trending technology in image recognition which had been proven its successful in several studies. Thus, this project is aims to implement a hand gestures recognition system based on deep neural network (DNN) to recognize 30 Malaysian Sign Language (MSL) static hand gestures.

1.3 Objectives

The objectives of this project are:

- 1) To construct a deep learning-based recognition system for 30 static hand gestures.
- 2) To evaluate and analyze the performance of system designed in term of accuracy.

1.4 Scope of Project

In this project, a hand gesture recognition system is designed to recognize 30 static hand gestures only. This is due to dynamic hand gestures recognition system requires longer period of time, more complex network architecture and higher computation power to train. Static hand gestures are considered less useful than dynamic hand gestures in daily communication since most of static hand gestures are alphabets and

numbers. However, static hand gestures are effective in expressing certain word such as a person's name, place name and others. In addition, this system is designed for those users who sat in front of a camera. The hand gestures look different when a user shows the gestures in front of a camera in sitting or standing pose. The angle and orientation of hand gestures shown in front of camera will influence the recognition result of the system designed.

Besides, the system is designed to recognize Malaysia Sign language (MSL) hand gestures only since this is the most commonly used in Malaysia. In addition, the software for the system will be developed by using KERAS in Python language. KERAS is a high-level interface for deep learning and Python is a high-level programming language that easily to understand. The syntax in python helps the programmers to design coding in fewer steps as compared to Java and C++. Moreover, an USB camera is used to capture raw images for hand gestures due to it is cheap and within the budget of project.

1.5 Project Significant

This system can be used during communication between deaf dumb and normal people. With this system, the translation of gestures or sign language to text form will be carried out immediately after the system detected a hand gesture. This can help in reducing the communication gap between these disabled people and normal people. Through this system, these disabled people can adapt to normal people lifestyle and not just live in their small group.

Apart from that, hand gestures or sign language not only used by people who are unable to speak. It is also used by deaf, people who have trouble with spoken language due to disability, people who have problem in hearing and those with deaf family