

REAL-TIME CONVERTING RECOGNIZED TEXT INTO VOICE

MOHAMAD DANIAL UZAEER BIN MOHD RAZEEF



اونيورسيتي تيكنيكل مليسيا ملاك

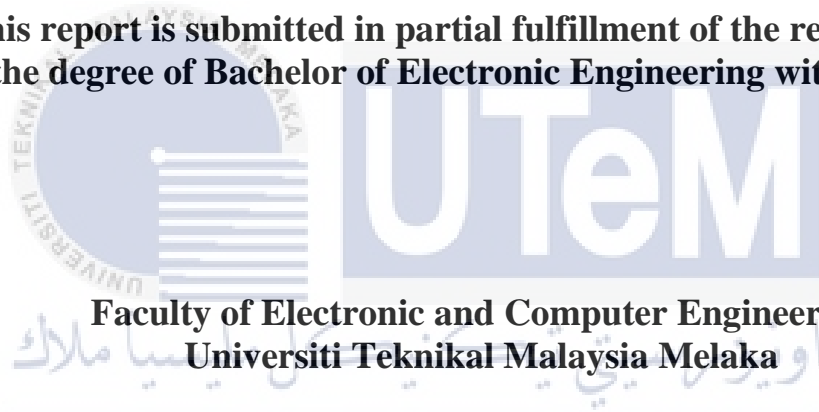
UNIVERSITI TEKNIKAL MALAYSIA MELAKA

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

REAL-TIME CONVERTING RECOGNIZED TEXT INTO VOICE

MOHAMAD DANIAL UZAEER BIN MOHD RAZEEF

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



**Faculty of Electronic and Computer Engineering
Universiti Teknikal Malaysia Melaka**

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2022

**BORANG PENGESAHAN STATUS LAPORAN
PROJEK SARJANA MUDA II**

Tajuk Projek : REAL-TIME CONVERTING RECOGNIZED
TEXT INTO VOICE
Sesi Pengajian : 2021/2022

Saya MOHAMAD DANIAL UZAEER BIN MOHD RAZEEF 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.)

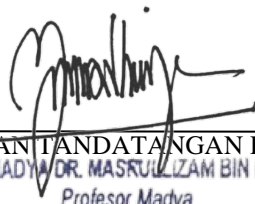
TIDAK TERHAD

Disahkan oleh:


(TANDATANGAN PENULIS)

Alamat Tetap: No 7, Jalan
Muhibbah, Taman
Muhibbah, 43000
Kajang, Selangor

Tarikh : 17 Jun 2022


(COP DAN TANDATANGAN PENYELIA)
PROFESOR MADYA DR. MASRULIZAM BIN MAT IBRAHIM
Profesor Madya
Fakulti Kejuruteraan Elektronik dan Kejuruteraan Komputer
Universiti Teknikal Malaysia Melaka (UTeM)
Hang Tuah Jaya
75100 Durian Tunggal, Melaka

Tarikh : 17 Jun 2022

DECLARATION

I declare that this report entitled “REAL-TIME CONVERTING RECOGNIZED TEXT INTO VOICE” is the result of my own work except for quotes as cited in the references.



Signature : 

Author : MOHAMAD DANIAL UZAEER BIN MOHD RAZEEF

Date : 17 JUNE 2022

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 : 

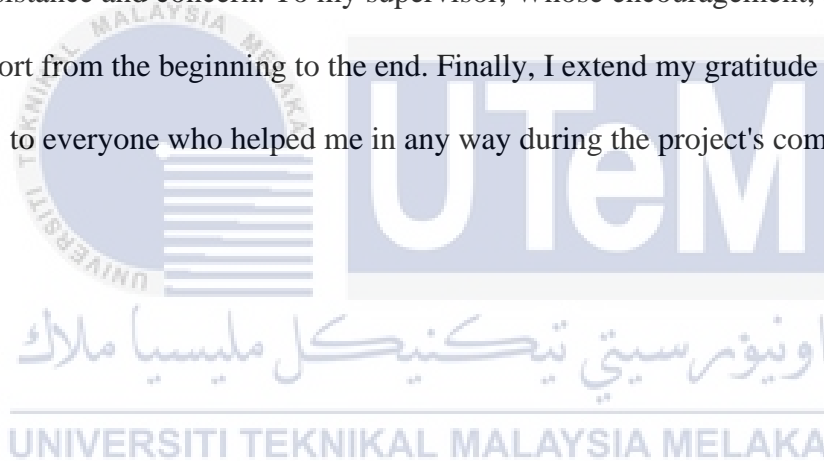
UNIVERSITI TEKNIKAL MALAYSIA MELAKA

Supervisor Name : PM DR. MASRULLIZAM BIN MAT IBRAHIM

Date : 17 JUNE 2022

DEDICATION

To my parents. The reason for what I become today. Thank you for all your assistance and concern. To my supervisor, Whose encouragement, advice, and support from the beginning to the end. Finally, I extend my gratitude and blessings to everyone who helped me in any way during the project's completion.



ABSTRACT

The goal of this project is to create an algorithm that converts text into speech. Most of the work involves coding and software. The algorithm built should be able to convert recognized text from an image captured into speech. According to studies, text-to-speech conversion is extremely beneficial for persons who have difficulty reading. Various experiments are being carried out, as well as a few techniques gained from the research. The research problem statement as well as the project's goals have been determined. The study's scope has been discussed. In the literature review chapter, every aspect of the project and study has been thoroughly explained. This project combines a few approaches, and it is divided into four sections: camera captured image, optical character recognition, language translation, and text to speech. Each component uses a different technique. In the methodology chapter, the development of real-time conversation recognized text to speech is covered. The system was then put to the test, with accuracy and code execution time being measured. This project is running smoothly and achieving all its goals

ABSTRAK

Matlamat projek ini adalah untuk mencipta algoritma yang menukar teks kepada ucapan. Kebanyakan kerja melibatkan pengkodan dan perisian. Algoritma yang dibina seharusnya dapat menukar teks yang dikesan daripada imej yang ditangkap kepada pertuturan. Menurut kajian, penukaran teks ke pertuturan amat berfaedah bagi mereka yang mengalami kesukaran membaca. Pelbagai eksperimen telah dijalankan, serta beberapa teknik yang diperolehi daripada penyelidikan. Penyataan masalah kajian serta matlamat projek telah ditentukan. Skop kajian telah dibincangkan. Dalam bab kajian literatur, setiap aspek projek dan kajian telah dijelaskan dengan teliti. Projek ini menggabungkan beberapa pendekatan, dan ia dibahagikan kepada empat bahagian: imej yang ditangkap kamera, pengesanan aksara optik, terjemahan bahasa dan teks ke pertuturan. Setiap komponen menggunakan teknik yang berbeza. Dalam bab metodologi, pembangunan masa nyata menukar teks yang diiktiraf kepada pertuturan diliputi. Sistem itu kemudiannya diuji, dengan ketepatan dan masa pelaksanaan kod diukur. Projek ini berjalan lancar dan mencapai semua matlamatnya

ACKNOWLEDGEMENTS

First and foremost, praise and thank God, the Almighty, for His blessings on my Final Year Project work, which enabled me to complete it successfully. This job would not have been finished on time without the guidance of our Almighty.

I'd like to convey my heartfelt appreciation to my supervisor, Prof. Madya Dr. Masrullizam Bin Mat Ibrahim, for his patient instruction, enthusiastic support, and for providing me with the opportunity to study with him about an exciting case study for my Final Year Project. His encouragement and constructive comments were the most valuable tools in bringing this project in on time. Finally, I'd want to thank my parents and all of my friends for their invaluable assistance and consistent encouragement throughout 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.	x
LIST OF SYMBOLS AND ABBREVIATIONS	xi
LIST OF APPENDICES	xii
CHAPTER 1 INTRODUCTION	1
1.1 Project Background	1
1.2 Problem Statement	3
1.3 Objectives	4
1.4 Project Scopes	4

1.5	Significant of Study	5
CHAPTER 2 LITERATURE REVIEW		6
2.1	Deep Learning for Computer Vision	6
2.2	Text Classification	7
2.2.1	Convolutional Neural Network (CNN)	8
2.2.2	Recurrent Neural Network (RNN)	9
2.2.3	Long Short-Term Memory (LSTM)	10
2.3	Google Colaboratory	11
2.4	Optical Character Recognition	12
2.5	EasyOCR	13
2.5.1	Convolutional Recurrent Neural Network (CRNN)	13
2.5.2	Character Region Awareness for Text detection (CRAFT)	14
2.6	Speech Synthesis	15
2.7	Previous Work	16
2.7.1	Reading Device for Blind People Using Python, OCR and GTTS	16
2.7.2	Raspberry Pi Based Reader for Blind Peoples	17
2.7.3	A Smart Reader for Blind People	17
2.7.4	Optical Character Recognition from Text Image	18
2.7.5	Automatic number plate recognition	18
CHAPTER 3 METHODOLOGY		20

3.1	Project Introduction	20
3.2	Work Procedure	22
3.3	Gantt Chart	23
3.4	Working Principle	23
3.5	System Overview	24
3.6	Main Tools	24
3.6.1	Google Colab	24
3.6.2	Python Language	25
3.6.3	Camera Sensor	26
3.6.3.1	Specification	27
3.6.4	EasyOCR	28
3.6.5	Google Translator	30
3.6.6	Google Text-to-Speech	30
3.7	Writing Coding	31
3.8	Validation of Project	31
CHAPTER 4 RESULTS AND DISCUSSION		33
4.1	Code Implementation	33
4.2	Dataset	34
4.3	Code Execution Time	36
4.3.1	Result for Code Execution Time	36

4.4	Project Accuracy	38
4.5	Confusion Matrix	39
4.5.1	Accuracy	40
4.5.2	Result of Confusion Matrix	40
4.6	Project Functionality	42
4.7	Environment and Sustainability	43
CHAPTER 5 CONCLUSION AND FUTURE WORKS		44
5.1	Conclusion	44
5.2	Recommendation	45
REFERENCES		46
APPENDIX A		51



اونيورسيتي تيكنيكل مليسيا ملاك

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

LIST OF FIGURES

Figure 2.1 English billboard text recognition using deep learning [8]	8
Figure 2.2 LSTM gate [6]	10
Figure 2.3 Network architecture of CRNN [8]	14
Figure 2.4 Example of CRAFT [19]	15
Figure 3.1: Work procedure	22
Figure 3.2: Working principle	23
Figure 3.3: System overview	24
Figure 3.4: Google Colaboratory	25
Figure 3.5: Python	25
Figure 3.6: ANBIUX 1080p webcam	26
Figure 3.7: Easy OCR framework	28
Figure 3.8: CRNN Layer	29
Figure 3.9: Text to speech	31
Figure 4.1: Samples of Large Word Size Images	34
Figure 4.2: Samples of Medium Word Size Images	35
Figure 4.3: Samples of Small Word Size Images	35
Figure 4.4: Samples of Typeface Word Images	35
Figure 4.5: Load Auto-time coding	36

Figure 4.6: Images that have been tested	36
Figure 4.7: Example of failed OCR	39
Figure 4.8: Confusion Matrix for Binary Classification	39
Figure 4.9: Confusion Matrix for Large and Medium Word Size	41
Figure 4.10: Confusion Matrix for Small Word Size	41
Figure 4.11: Confusion Matrix for Typeface Word	41

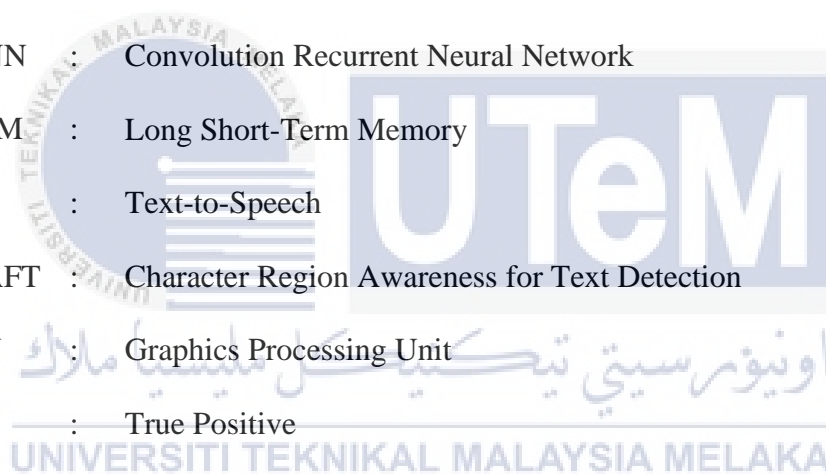


LIST OF TABLES.

Table 1: Camera Specification.....	27
Table 2: Hardware and Software Specification	34
Table 3: Time Taken for the Process	37
Table 4: Accuracy of the Process.....	38
Table 5: Result of Confusion Matrix	42



LIST OF SYMBOLS AND ABBREVIATIONS



OCR	:	Optical Character Recognition
gTTS	:	Google Text-to-Speech
CNN	:	Convolution Neural Network
RNN	:	Recurrent Neural Network
CRNN	:	Convolution Recurrent Neural Network
LSTM	:	Long Short-Term Memory
TTS	:	Text-to-Speech
CRAFT	:	Character Region Awareness for Text Detection
GPU	:	Graphics Processing Unit
TP	:	True Positive
TN	:	True Negative
FP	:	False Positive
FN	:	False Negative
SDG	:	Sustainable Development Goal

LIST OF APPENDICES

Appendix A: Gantt Chart

51



CHAPTER 1

INTRODUCTION



This chapter describes the introduction of the project. This chapter discusses the study's background, objectives, problem statement, and project scopes.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

1.1 Project Background

Reading words may look straightforward, but if you have ever listened to a little kid or the elderly read a book that was simply too tough for them. The main problem is that written content is ambiguous: the same written information can frequently mean more than one thing, and reading it properly usually necessitates knowing the meaning or making an educated guess. As a result, the initial stage of voice synthesis, known as pre-processing or normalization, is all about reducing ambiguity: narrowing down the many ways you may read a piece of text into the best-suited one.

Optical character recognition (OCR) is the technique of converting scanned images of typewritten text into machine-editable data[1]. This technique enables a machine to detect letters automatically via an optical mechanism. Many things are recognized by humans in this way; our eyes are the optical mechanism. However, while the brain interprets the data, each person's capacity to grasp these signals differs depending on a variety of conditions such as language and the similarity between numerical and alphabetical symbol shapes.

Text-to-Speech (TTS) technology converts any text into a speech signal. It can be used for a variety of purposes, including car navigation, announcements in train stations, telecommunications response services, and e-mail reading. TTS is a common assistive technology in which a computer or tablet reads out loud to the user the text on the screen. This technology is popular among people who have reading issues, particularly those who struggle with decoding. By delivering the words orally, the learner may concentrate on the meaning of the words rather than using all their brain ability to sound them out. While this can assist folks in overcoming their reading issues and gaining access to the reading content [2].

1.2 Problem Statement

People who have illiteracy issues have a difficult time comprehending the information they are presented with. Illiteracy is defined as a person's inability to read and write, preventing them from entering the workforce or working as unskilled labor, as well as a lack of awareness to make informed decisions that affect them and their community. People with inadequate literacy skills may find it difficult to read a book or newspaper, understand traffic signs or pricing labels, understand a bus or train schedule, complete paperwork, read medicine instructions, or use the internet. Illiterate people will face difficulties in their daily activities.

Next, people with learning disabilities who, due to illness or other issues, find it difficult to read large amounts of text. People who suffer from illnesses like dyslexia have difficulty processing written materials, making it more difficult to recognize, spell, and interpret words. Elderly people who are experiencing vision-reducing eye disease will face the same issue. Over time, our eyes deteriorate, causing vision problems that make it difficult to read for lengthy periods.

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

- 1) The employment of OCR presents a serious challenge in terms of accuracy.
To ensure that it will help individuals be more efficient, it can be improved by training the OCR for better results.
- 2) The time it took for the system to perform was rather considerable. It is hoped that by leveraging a GPU and the internet, the coding will be executed more quickly and with better results.

1.3 Objectives

The objectives are as follows:

- I. To construct an algorithm to access the camera sensor and process the captured image [RO1].
- II. To develop an algorithm to recognize the text embedded in the image [RO2].
- III. To develop an algorithm to convert text into speech [RO3].

1.4 Project Scopes

The purpose of this project is to develop an algorithm capable of converting recognized text from an image to voice. This paper details the procedures involved in developing and designing an algorithm for converting detected text to voice. To improve understanding of the deep learning process for computer vision in OCR, the theories of the deep learning process for computer vision in OCR will be applied.

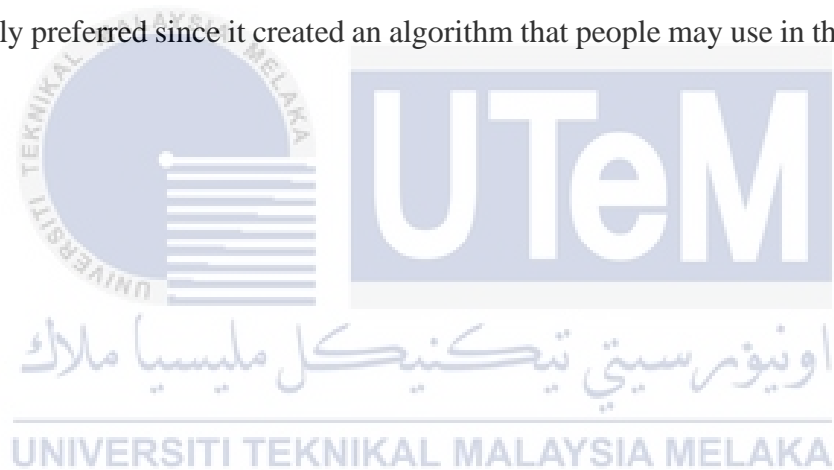
An image will be captured using a camera that functions as a sensor. The image is then subjected to an OCR procedure to extract the text embedded in it. The result of OCR will then be transformed into speech.

The project's limitation is that the text recognition pipeline fails when the text is skewed/rotated, even as the text's font is not like the model trained on. For EasyOCR, low-quality scans may result in low-quality OCR. If the image contains languages other than the ones specified by the language argument, the result may be inadequate. In this project, the OCR cannot recognize numerous words in a single image since the OCR accuracy falls as the number of words in the image increases, influencing all

other processes. There are a few challenges using OCR in this project such as the image containing complicated backgrounds, noise, lightning, a unique font, and geometrical distortions.

1.5 Significant of Study

The project's findings will benefit children and the elderly who have difficulty reading. Because digital text can be used with a variety of reading software. It will also allow individuals to concentrate on the content rather than the act of reading, which will result in a greater understanding of the information. This initiative is socially preferred since it created an algorithm that people may use in their daily lives.



CHAPTER 2

LITERATURE REVIEW



This chapter will explain the literature review from the research regarding deep learning of computer vision, text classification, Google Colaboratory, optical character recognition, EasyOCR, GoogleTrans and gTTS.

2.1 Deep Learning for Computer Vision

Deep learning evolved from artificial neural networks to become a dominating area of machine learning, seeking to extract high-level abstraction from data via hierarchical techniques. It is a rising approach that has found widespread use in domains such as pattern recognition, semantic parsing, audio recognition, computer vision, and natural language processing. Deep-learning models often use hierarchical architecture to connect various levels. Through simple linear or nonlinear connections,