

Faculty of Applied Sciences  
Bachelor of Science in Computing

COMP490 Final Year Project  
Project Proposal  
  
Academic Year 2022/23

|  |  |
| --- | --- |
| Text Detection and Recognition in Image-Based on Deep Learning | |
|  |  |
| Project number: | Your project number |
| Student ID: | P1908370 |
| Student Name: | Eden |
|  |  |
| Supervisor: | Jacky Tang |
| Assessor: | Your assessor |
|  |  |
| Submission Date: | 08/09/22 |

Table of Contents

[1 Project Description 2](#_Toc113375424)

[1.1.1 Background 2](#_Toc113375425)

[1.1.2 Motivation 2](#_Toc113375426)

[1.1.3 Main objective 2](#_Toc113375427)

[1.1.4 Difficulties 3](#_Toc113375428)

[2 Summary of Related Work and Key References 3](#_Toc113375429)

[3 Project Work Plan 3](#_Toc113375430)

[3.1.1 Dataset preparation 3](#_Toc113375431)

[3.1.2 Model design 3](#_Toc113375432)

[3.1.3 Train and improve model 3](#_Toc113375433)

[4 Risk Assessment 4](#_Toc113375434)

[4.1.1 Ethics 4](#_Toc113375435)

[4.1.2 Hardware 4](#_Toc113375436)

[5 References 4](#_Toc113375437)

# Project Description

### Background

Reading text in realistic life scenarios becomes a hot topic in computer vision area. There are two main reasons. Firstly because of the complexity of the real-world environment and the artistic design of a store sign, for example, the text may not be arranged from left to right as it is known. Performing text recognition under this environment is very challenging. The second is the application of text recognition is very wide. A well-known application of text recognition is Optical character recognition (OCR) [1]. It can be found in many places, such as PDF conversion, printers and cell phone photo translation.

(diagram of real-world environment)

This project mainly includes three sub-tasks: dataset preparation, text detection and text recognition.

### Motivation

The current OCR technology for recognizing handwritten Simplified Chinese and Traditional Chinese handwriting is not very good.

(diagram)

### Main objective

This project should

In this project, I am planning to develop a Chinese OCR program using neural network. This program contains two parts, detection, and recognition. For this program, the user inputs the image containing Chinese characters and returns computer-encoded text.

1. The model can process images
2. The model can detect the text and do the segmentation
3. The model can recognize the text and output the computer-encoded text

(diagram)

### Difficulties

The core difficulty of this project is to ensure the accuracy of detection and recognition. This problem may need to be solved in many aspects.

# Summary of Related Work and Key References

In text detection, Traditional approaches for text detection mostly employ a bottom-up pipeline. But for this project, I decide to use Connectionist Text Proposal Network (CTPN) [2].

竞品分析

使用场景

# Project Work Plan

### Dataset preparation

Duration: Week 4, 5

Description: Collect handwriting picture and label them

### Model design

Duration: Week 8 – Week 9

Description: Choose proper deep learning model and redesign it if it needs. (e.g., reinforcement learning)

### Train and improve model

Duration: Week 11 – Week 14

Description: Feed the model the feature we extract

# Risk Assessment

### Ethics

Most of the training data can be found on the internet, but this project may still need human participants and collect their handwriting. Some ethical problems may be involved in this project. For participants, I will avoid collecting data from vulnerable groups and protect the rights of participants. For data collected, all data stored should be anonymized and should not contain sensitive data about living individuals.

### Hardware

In theory, the deep learning’s computational requirement is hard to estimate [3], but there are important reasons to believe that the computational requirement is related to model complexity. My hardware may not meet this requirement. A substitute solution is rent a high compute capability server and train model on it.

# References

|  |  |
| --- | --- |
| [1] | “Optical character recognition - Wikipedia,” 30 08 2022. [Online]. Available: https://en.wikipedia.org/wiki/Optical\_character\_recognition. |
| [2] | L. N. J. M. Michal Busta, “FASText: Efficient Unconstrained Scene Text Detector,” IEEE International Conference on Computer Vision (ICCV), 2015. |
| [3] | K. G. K. L. G. F. M. Neil C. Thompson, The Computational Limits of Deep Learning, 2022. |