

# CV proposal

Yuting Ji

Yifan Zhou

Jeremy Chan

## 1. Introduction

Text recognition and Natural Language Processing(NLP) are two of the greatest applications of machine learning at present. Nevertheless, prevalent algorithms of these two algorithms are limited to their own scope: Text recognition only deals with seeing words from pictures but doesn't interpret the meaning of words and NLP only deals with interpreting existing words in the forms of text or audio but doesn't see the words. We are aiming to design a Machine-Learning model that is capable of both vision and interpretation.

For the CV part of the project, we essentially are doing a segmentation of the text. In other words, we want to extract the words from given images. This can be achieved by a simple CNN model.

For the NLP part of the project, we are considering implementing a few features to make our project have multi-functionality. As a group, we want to conduct sentiment analysis, with the end goal of determining whether text is positive or negative. The models and NLP algorithms we are planning to use are tf-idf, word senses, lexical semantics, logistic regression, decision trees, and SVM. We will also look into newer techniques such as transformers. An example of this would be using the language model BERT. After the proposal, we will look more into the different algorithms we want to do and then work from there, these models and algorithms are up for discussion.

## 2. Related Works

These reviews summarize conventional approaches to image text recognition or OCR (Optical Character recognition) [2]. Typical classification in images are based on forms (printed/handwritten), language, and generation (Graphic text/ scene text). Most of the time each form requires a dedicated pre-processing module. We've also reviewed approaches such as super-resolution that creates a robust image text recognition model. [1]

The papers bring up the problem of different characteristics and fonts of text and images, and suggest different approaches to conduct Document Image Analysis (DIA). They propose an algorithm to bring paper documents to

electronic format, and then they use Artificial Neural Networks such as Kohonen Neural Network to conduct classification. For the problem of OCR, one of the papers suggests using a probabilistic model, with the ultimate goal using OCR to help acquire resources for new languages from printed text. [3]

## 3. Plan

1. Data Gathering
2. Pre-Process data, Build the CV model
3. Train and Test the vision model
4. Gather the info, establish the NLP model for interpretation
5. Train and Test the NLP Model
6. Post-Process data
7. Whole Process Pipeline establishment

## 4. Labor Division

1. Pre-Process Vision: All
2. Vision Model: All
3. Train and Test Vision model: All
4. Post-Process Vision to text: Yuting, Jeremy
5. NLP Model: Yifan, Jeremy
6. Train and Test NLP model: All
7. Verify the result: Jeremy, Yifan
8. Reference: All
9. Latex Formatting: Yuting
10. Final Report: All

## References

- [1] Xiaoxue Chen, Lianwen Jin, Yuanzhi Zhu, Canjie Luo, and Tianwei Wang. Text recognition in the wild: A survey, March 2021. URL: <https://dl.acm.org/doi/abs/10.1145/3440756>. 1
- [2] Okan Kolak, William Byrne, and Philip Resnik. A generative probabilistic ocr model for nlp applications. URL: <https://aclanthology.org/N03-1018.pdf>. 1
- [3] Yiwei Zhu, Shilin Wang, Zheng Huang, and Kai Chen. Text recognition in images based on transformer with hierarchical attention, 2019. URL: <https://ieeexplore.ieee.org/document/8803203>. 1