Enhancing Accuracy in Optical Character Recognition of Sensor Readings: A Comparative Study of Tesseract and CRNN Models with Emphasis on Image Preprocessing

by

Aidan Dennehy [R00145278]

For the module DATA9003 - Research Project as part of the
Master of Science in Data Science and Analytics, Department of Matematics

Supervisor: Dr Alex Vakaloudis

July 2023

Declaration of Authorship

I, Aidan Dennehy , declare that this thesis titled, "Enhancing Accuracy in Optical Character Recognition of Sensor Readings: A Comparative Study of Tesseract and CRNN Models with Emphasis on Image Preprocessing" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for an undergraduate degree at Munster Technological University.

- Where any part of this thesis has previously been submitted for a degree or any other qualification at Munster Technological University or any other institution, this has been clearly stated.

- Where I have consulted the published work of others, this is always clearly attributed.

- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this project report is entirely my own work.

- I have acknowledged all main sources of help.

- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

_____

Date:

_____

Munster Technological University

Abstract

Department of Matematics

Master of Science in Data Science and Analytics

by Aidan Dennehy [R00145278]

This research is primarily dedicated to the formulation of an innovative method for accurately interpreting sensor data obtained from digitized images. Confronting inherent challenges such as diminished contrast and subpar image quality, often associated with sensor readings, the study exploits Optical Character Recognition (OCR). This is accomplished employing two distinct techniques: Tesseract and Convolutional Recurrent Neural Network (CRNN) models.

An unique feature of the research lies in its novel image preprocessing steps, specifically the masking of red and green colors prior to conversion to grayscale. This process considerably augments the efficacy of OCR. Additionally, the study underlines the critical importance of correct font selection for each sensor to enhance reading accuracy.

The findings highlight the essential role of image quality and contrast in OCR, while presenting an innovative approach to image preprocessing for improved results. The potential implications of this research are extensive and could shape future undertakings in the fields of OCR and sensor digitization. The research underscores the vital aspects of image preprocessing and reveals how precise interventions can markedly improve sensor data interpretation from digitized images.

Acknowledgements

Acknowledgements here . . .

Contents

List of Figures

List of Tables

Abbreviations

LAH   **L**ist **A**bbreviations **H**ere

For/Dedicated to/To my. . .

Chapter 1

Introduction

1.1 Area of Interest

The area of interest for this literature review is the intersection of computer vision, optical character recognition (OCR), and deep learning, with particular emphasis on the Tesseract OCR engine and Convolutional Recurrent Neural Networks (CRNNs). These technological advancements have revolutionized the way machines recognize and understand visual information, especially digits. Given their diverse and significant applications, ranging from digitizing written documents to aiding autonomous vehicle navigation, they hold vast potential for transforming many sectors. This research focuses on exploring the principles that underlie these tools, their performance in real-world applications, and the possibilities they offer for future development. This involves assessing the strengths of these systems, identifying their limitations, and suggesting potential areas of improvement. Moreover, it considers how these technologies are pushing the boundaries of OCR, paving the way for more sophisticated and versatile tools that can better navigate the complexities and variations in text size, font, and orientation often encountered in different visual scenes.

Optical Character Recognition (OCR) technology has seen substantial advancements in recent years, transforming the process of data extraction from visual mediums to digital formats. This technology, crucial in numerous fields ranging from document digitization to automated data entry systems. OCR holds specific importance when it comes to interpreting sensor readings, a key aspect of data-driven industries. The necessity for accurate, efficient, and automated reading of sensor-generated data has led to the investigation of various techniques and models within the OCR domain.

Two models which feature prominently emerged as potential solutions, namely Tesseract, an open-source OCR engine sponsored by Google, and Convolutional Recurrent
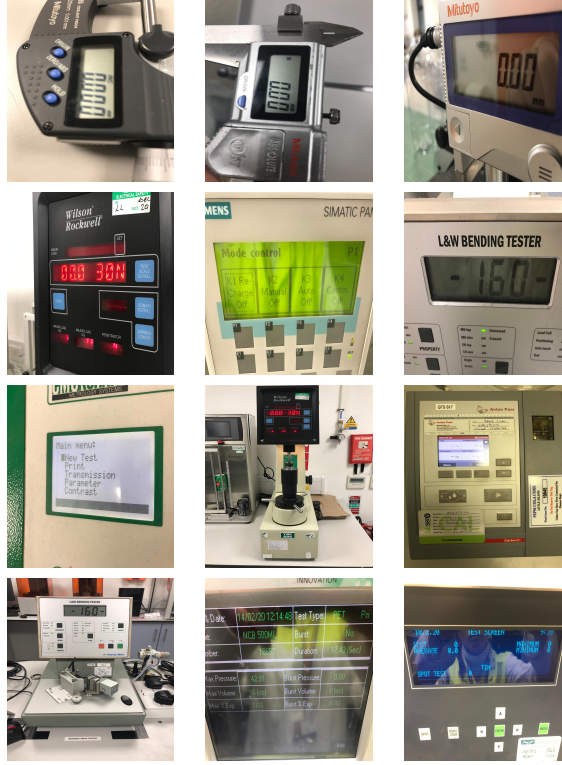
TABLE 1.1: Nimbus Sensor Images

Neural Network (CRNN), a combination of CNN, RNN, and Connectionist Temporal Classification that offers promising results in scene text recognition tasks.

In OCR applications, image preprocessing has a pivotal role. It prepares an image for further processing by reducing noise and unnecessary details and enhancing features that are important for later stages, thereby directly influencing the accuracy of the final output. Among various preprocessing techniques, the novel approach of red and green color masking, followed by conversion to grayscale, has shown to significantly improve the accuracy of digit recognition.

1.2 Motivation

The motivation behind this research stems from the challenges encountered in the manual and infrequent readings of environmental sensors in various operational settings such as factories. These sensors, while accurate and essential, lack a means for continuous data capture. Typically, an individual manually reads the sensor outputs at fixed intervals, which could range from hourly to daily. This method, while necessary, is prone to human error, potentially leading to inaccuracies in the recorded data and subsequent analysis reports. Furthermore, the infrequency of readings may result in delays in responding to critical sensor data, which could precipitate further issues. These complications could be

mitigated with the implementation of Optical Character Recognition (OCR) technology. By enabling continuous, automated readings of these sensors, OCR has the potential to not only reduce errors but also ensure timely reaction to important sensor changes, optimizing the overall operation and efficiency of the systems.

1.3 Aims and Objectives

1.4 Structure of the Thesis

Chapter 2

Literature Review

2.1 Introduction

In the dynamic and continuously evolving field of computer vision and optical character recognition (OCR), two concepts have emerged as among the significant game-changers: the Tesseract OCR engine and Convolutional Recurrent Neural Networks (CRNNs). Tesseract, initially developed by Hewlett-Packard and later adopted by Google, is a pioneering engine that converts images of text into machine-encoded text, offering groundbreaking utilities across numerous applications. On the other hand, CRNNs, a deep learning-based approach, combine the spatial feature extraction capabilities of Convolutional Neural Networks (CNNs) with the sequential data processing capacity of Recurrent Neural Networks (RNNs). These networks have set new benchmarks in the realm of scene text recognition, overcoming the challenges posed by variations in text sizes, fonts, and orientations. This literature review delves into the intricacies of these advanced tools, shedding light on their principles, applications, strengths, and potential areas for improvement, thereby enriching our understanding of current trends in OCR technology and pointing to the future possibilities.

In addition to these techniques, the selection of the correct font for each sensor is another critical element that affects the accuracy of the OCR system. Despite its importance, this aspect has been less emphasized in existing literature, thereby forming a crucial area of exploration in this study.

This literature review explores the current state of OCR technologies, with a particular focus on Tesseract and CRNN models. It delves into various image preprocessing techniques, emphasizing the unique method of red and green color masking before conversion to grayscale. Lastly, it investigates the role of font selection in enhancing OCR accuracy, thereby setting the context for the subsequent research.

While this review focuses on the promising capabilities of Tesseract OCR and Convolutional Recurrent Neural Networks (CRNNs) in the OCR domain, it's crucial to acknowledge that the OCR landscape is not limited to these technologies. Many other methods play equally significant roles in expanding the OCR frontiers and opening up new avenues for research and application. Long Short-Term Memory Networks (LSTMs), Transformers, attention-based OCR models, rule-based systems, Support Vector Machines (SVMs), Hidden Markov Models (HMMs), K-Nearest Neighbors (KNN), and template matching are some of these diverse methodologies that provide unique perspectives and solutions in the OCR realm. Each of these methods has its distinctive advantages, making them optimal for certain types of tasks, as well as its limitations, requiring continuous research and development for enhancement. However, the scope of this review will mainly revolve around Tesseract and CRNNs, while the mentioned methods provide an essential context for understanding the broader OCR ecosystem.

2.2 Tesseract OCR

2.3 Convolutional Recurrent Neural Networks (CRNNs)

2.4 Other OCR Methods

2.4.1 Long Short-Term Memory Networks (LSTMs)

Long Short-Term Memory Networks (LSTMs) are a special kind of recurrent neural network capable of learning long-term dependencies, which makes them highly suitable for OCR tasks. They've been used successfully to decode sequences of characters from images.[1]

2.4.2 Transformers

Originally developed for natural language processing tasks, Transformer models have been adapted for OCR. They treat the OCR problem as a sequence-to-sequence translation task, translating the input image into a sequence of characters.

2.4.3   Attention-based OCR models

Attention mechanisms allow models to focus on different parts of the input image while predicting each character in the output sequence, similar to how humans read. This can improve accuracy, especially on more complex images.

2.4.4   Rule-based systems

These were some of the earliest methods for OCR and use specific rules for identifying characters based on their shape, size, and relative position. They are now less commonly used due to their limitations with complex and diverse inputs.

2.4.5   Support Vector Machines (SVMs)

Support Vector Machines (SVMs) are used for character recognition in OCR due to their effective high-dimensional mapping and classification abilities. They work best when text is clearly segmented. In the their paper, Joshua et al. used SVMs to classify characters in a license plate image and achieved an accuracy of 98.3% using a local dataset of 10,000 images.[2]

FIGURE 2.1: Greyscale image with background cleaning
[2]

2.4.6   Hidden Markov Models (HMMs)

HMMs have been used in OCR for recognizing sequential data. HMMs are statistical models that assume an underlying process to be a Markov process with hidden states.

2.4.7   K-Nearest Neighbors (KNN)

KNN is a simple, instance-based learning algorithm used for OCR, particularly for isolated character recognition. [3]

2.4.8   Template Matching

Template Matching is a technique used to locate small-parts of the bigger image which match a template image. This can be useful in OCR when the set of possible characters is known and limited.

Chapter 3

Methodology

3.1

Chapter 4

Results

4.1 Introduction

Chapter 5

Discussion and Conclusion

5.1   Discussion

Bibliography

[1] T. M. Breuel, A. Ul-Hasan, M. A. Al-Azawi, and F. Shafait, "High-Performance OCR for Printed English and Fraktur Using LSTM Networks," in *2013 12th International Conference on Document Analysis and Recognition*. IEEE, pp. 683–687. [Online]. Available: http://ieeexplore.ieee.org/document/6628705/

[2] I. O. Joshua, M. O. Arowolo, M. O. Adebiyi, O. R. Oluwaseun, and K. A. Gbolagade, "Development of an Image Processing Techniques for Vehicle Classification Using OCR and SVM," in *2023 International Conference on Science, Engineering and Business for Sustainable Development Goals (SEB-SDG)*. IEEE, pp. 1–9. [Online]. Available: https://ieeexplore.ieee.org/document/10124622/

[3] T. K. Hazra, D. P. Singh, and N. Daga, "Optical character recognition using KNN on custom image dataset," in *2017 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON)*. IEEE, pp. 110–114. [Online]. Available: http://ieeexplore.ieee.org/document/8079572/

Appendix A

Code Snippets

Put appendix material in this section e.g. code snippets

USE THE APPENDICES