Insert Fancy Project Title

Joseph George, Tammer Haddad

Northeastern University, Boston, 02115

December 10, 2025

Abstract

This project explores machine learning approaches to handwritten text recognition and prediction. We developed a system that takes in digital text using a pipeline that includes Tesseract OCR, Convolutional Neural Network, and Markov Model. We source training data from EMNIST and MINST datasets. Our CNN model achieves 83.27% accuracy across 62 character classes.

Introduction

Even the rise of computers, handwritten notes are still a popular method to write down information for classes, meetings, etc. However, the organization and functionality aspect of written notes is inconvenient. For example, finding notes on a previous subject can be difficult manually but it would be easier if the notes were digitized and searchable. Functionally, digitized text allows for editing and sharing of notes. Traditionally, OCR (Optical Character Recognition) has been used to detect characters through rule-based algorithms. However, with the rise of AI approaches, deep learning and convolutional neural networks have improved the accuracy of OCR software. We took inspiration from this technological advancement to create an integrated pipeline that not only transcribes handwritten notes but also predicts subsequent text, making digitization more efficient and useful.

1 Example Section

This is an example of a numbered section. Sections typically seen in a research paper are Related Work, Preliminaries, Data, Methodology, Experiments, Results, and Discussion. The order in which these sections appear may depend on the venue.

1.1 Example Subsection

This is an example of a subsection. For example, Definitions could be a subsection of the Preliminaries section.

Figure 1: A cool graph. This image does not reflect the author's allegiance to the DC Comics universe.

2 Formulae, Tables

2.1 Typing Equations in LaTeX

This subsection demonstrates the use of inline formulae, like $e = mc^2$, or equations like the one below on a separate line.

$$\int_a^b x^2 \ \mathrm{d}x = \left. \tfrac{1}{3}x^3 \right|_a^b$$

For numbered equations, we use the *equation* environment:

$$\int_a^b x^2 \ \mathrm{d}x = \left. \tfrac{1}{3}x^3 \right|_a^b \tag{1}$$

2.2 Tables

This subsection demonstrates the use of tables in a LaTeX document. Resources like `tablesgenerator.com` can be used to generate code for tables like the one below.

Here, I refer to Table 1 without actually listing the actual Table number!

References to books [?] or articles [?] can be made like this, with the full citations defined in a *.bib* file.

Country List			
Country Name or Area Name	ISO ALPHA 2 Code	ISO ALPHA 3 Code	ISO numeric Code
Afghanistan	AF	AFG	004
Aland Islands	AX	ALA	248
Albania	AL	ALB	008
Algeria	DZ	DZA	012
American Samoa	AS	ASM	016
Andorra	AD	AND	020
Angola	AO	AGO	024

Table 1: A Random Table

3   Github Link & Team Contributions

Please include a link to your GitHub repo and a 65-word (total) statement summarizing each member's contributions.