INFORMATICS INSTITUTE OF TECHNOLOGY

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

Handwritten source code recognition for python

Final Thesis

Mr. Rajeev Kodippily

W1761538 / 2018639

Supervised by

Mr. Saman Hettiarachchi

Submitted in partial fulfilment of the requirements for the BEng (Hons) Software
Engineering degree at the University of Westminster.

May 2021

Abstract

Modern programming Integrated Development Environments (IDE's) use keyboard and mouse as their main input method for text entry. However, some programmers suffer from disabilities such as Repeated Strain Injury (RSI) which cannot be productive using traditional text entry methods such as keyboard and mouse. This project focuses on creating a programming IDE in which handwriting would be the main input method for text entry. For this, the research involves training models using the open source OCR Tesseract and comparing the results to a state of the art English handwriting recognition engine in the Google ML Kit. The test results shows that the trained model performs better than base Tesseract on source code with 11.82% less Character Errors and 12.51% less Word Errors. Furthermore the ML Kit engine outperforms the trained model with 16.4 % less Character Errors and 26.5 % less Word Errors resulting in a 12.02 Character Error Rate and 41.65 Word Error Rate. Comparing this result to the benchmark presented turned out to be problematic due to the inaccuracy of the approach and is discussed in the results.

The project has the potential to be beneficial for software engineering professionals, educators and students who write computer programs in their day-to-day life and who are looking for alternate input methods to the traditional keyboard and mouse.