



INFORMATICS
INSTITUTE OF
TECHNOLOGY

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

Handwritten Text Detection System in Sinhala

A dissertation by

Mr. Nadin Kodithuwakku

Supervised by

Mr. Torin Wirasingha

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

July 2022

Abstract

Computer vision has been expanded to identify both handwritten documents and printed documents to improve human-computer interactions. However, in the case of Asian languages, this is still a major controversy. Sinhala language characters recognition is still a work in progress among Asian languages, as Sri Lanka is the only nation that utilizes Sinhala as its official language.

In comparison to other commonly spoken languages, Sinhala characters have circular, intricate forms with connections between characters. There are 60 essential characters in the Sinhala language (non-cursive). Modifiers which can be applied to primary characters further enhance the Sinhala character set. As a result, recognizing Sinhala handwritten characters is a complicated process.

Pattern matching and image processing techniques are used in most of the character segmentation and recognition research projects. These methods, on the other hand, are unable to adjust to the changes.

This research is aimed at a Convolutional Neural Network-based handwritten Sinhala character recognition system (CNN). CNN differs from other traditional methods to Sinhala handwritten character recognition in that it extracts features automatically throughout the training phase. The objective of this research is to use deep learning approach to develop a process for Sinhala handwritten character recognition.

This research project will be able to segment characters in handwritten character and train a CNN model to accurately recognise Sinhala handwritten characters.

Key Words – Image Processing, OCR, Convolutional Neural Network, Sinhala Handwritten Character Recognition, Deep Learning