MSC Project Report

Sinhala Handwriting Character Recognition System Via a Deep Convolutional Neural Network

A Dissertation by

Sanjeewa Mangala Weerasinghe

Supervised by

Dr. Ruvan Weerasinghe

IIT STUDENT ID: 2017027

RGU STUDENT ID: 1616193

2019

A report submitted as part of the requirements for the degree of MSc Big Data Analytics at Robert Gordon University, Aberdeen, Scotland

Abstract

Sinhalese is a language mostly used in the country of Sri Lanka and it consists of a spoken and written form. As many Sri Lankans write in Sinhala, digitization of handwritten text will add value to many industries and domains that use the Sinhala language. Many attempts at Sinhala handwriting recognition has yielded average results and therefore the project scope is aimed at finding a solution.

Handwritten digit and character recognition for other languages such as English has seen a marked improvement in recent years since the modern advent of deep learning artificial neural networks such as Convolutional Neural Networks. The research aims to combine and apply the technological advancements in the deep learning field of Convolutional Neural Networks to the problem domain in order to explore the possibility of increasing the recognition accuracy for Sinhala handwriting recognition.

The research was able to create a deep learning model and a Sinhala handwritten character dataset that performs well at an 83% accuracy. The applications and implications of the convolutional neural network used is discussed and evaluated in the project report.

Keywords:

Convolutional Neural Network, Deep Learning, Artificial Neural Network, Sinhala, Handwritten character recognition, Open CV, Keras, TensorFlow, Machine learning, Object Detection, Image Processing