

ABSTRACT

Receipts or documents such as printed receipts are still widely used by industry or companies when making buying and selling transactions. Receipts can be filled in with human handwriting or typed and then printed, therefore a tool is needed that can scan and recognize patterns of human handwriting as well as computer printed letter patterns. Optical Character Recognition (OCR) provides the ability to detect text in documents to convert it to digital form, but OCR in handwriting has a little special treatment.

This study uses OCR to create a portable scan-like tool that recognizes handwriting and print. In handwriting recognition using a combination of deep learning algorithms Convolution Neural Network (CNN) and Recurrent Neural Network (RNN). CNN and RNN are highly used machine learning architectures inspired by the human brain's natural visual perception techniques. In the introduction of print writing using Tesseract OCR in carrying out its feature extraction. Tesseract OCR is an open source project, developed by Google for the most accurate free OCR software engine on print with limited fonts.

The dataset used in the handwriting training model comes from the IAM Database. The test was carried out with the handwriting of 10 different people and obtained an accuracy of 83.33% on the amount of money stated on the receipt. For print writing, the dataset used is 400000 lines of text covering about 4500 fonts. Testing of printed writing was carried out with a font size of 11 with an accuracy range of 89.63% - 96.62%. For a font size of 12 the accuracy range of the tool is 96.65% - 97.75%. For a font size of 14 the accuracy range of the tool is 99.43% - 100%.

Keywords: Portable Scanner, Optical Character Recognition, Convolution Neural Networks, Tesseract.