

# A Review of Name Board Text Recognition Using Python's Optical Character Reader Tools

S. Lathagini<sup>1\*</sup> and N. Sulaxan<sup>2</sup>

<sup>1</sup>Computer Unit, Sir Ponnambalam Ramanathan Faculty of Performing and Visual Arts, University of Jaffna, Sri Lanka.

<sup>2</sup>ICT Unit, Technical College, Vavuniya, Sri Lanka.

\*[lathagini@univ.jfn.ac.lk](mailto:lathagini@univ.jfn.ac.lk)

Name boards of places that are the most frequent visual aids in roadways, help to identify the areas by guiding the traveller. Their most significant utility is in aiding drivers who are inexperienced with the region and attempting to verify road names and locations as they follow a map or route. Although Optical Character Recognition (OCR) software is commonly available, OCR is still challenging in uncontrolled environments, such as natural scenery, because of geometrical distortions, complicated backgrounds, and various fonts. This study investigates the performance of some state-of-the-art models on OCR introduced by Python on scene text images: Keras-OCR, Pytesseract, and Easy-OCR. Besides traditional metrics such as Character Recognition Rate (CRR), minimum edit distance distribution is included to reflect more on the obtained results from each Python OCR tool. According to the minimum edit distance distribution, Easy OCR produces somewhat better transcription results than Pytesseract, while Keras-OCR produces noticeably better scene text transcription overall. CRR assessment measure is employed to provide transcription outcomes for various category name board pictures. Overall recognition rates at English character level for Pytesseract, Keras-OCR, and Easy OCR are 34.87%, 94.19%, and 83.09% respectively. Recognition rates at multi language character level for Pytesseract, Keras-OCR and Easy OCR are 66.66%, 35.80%, and 65.43%. The best Python tools for identifying name boards with improperly aligned text, incomplete letters, and background noise are Keras-OCR and Easy OCR. When detecting multilingual name boards with Tamil and English characters, Easy OCR and Pytesseract outperform Keras-OCR. The quality of the input data affects the Pytesseract output. Better outcomes come from precise text segmentation and a backdrop free of background noise. Thus, to improve Pytesseract results, various pre-processing approaches are needed.

Keywords: Character Recognition Rate, Easy-OCR, Edit Distance, Keras-OCR, Optical Character Recognition, Pytesseract