WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Image to Text Conversion using Pre-trained Deep Learning Models

Tushar Aggarwal, Deepanshu, Mr. Varun Goel

Maharaja Agrasen Institute of Technology

aggarwaltushar36@gmail.com, raghavdwivedi6533@gmail.com, varungoel@mait.ac.in

ABSTRACT

The increasing digitization of information has emphasized the need for efficient Optical Character Recognition (OCR) mechanisms to convert textual data from images into editable formats. This research focuses on a comparative analysis of three widely used pre-trained Python libraries—keras_ocr, pytesseract, and easyocr—in real-world scenarios such as license plate recognition, medical record processing, and ID verification.

A manually curated dataset comprising 200+ images, each meticulously captioned through human intervention, forms the basis of our experimentation. The study is conducted on Google Colab, employing Python as the runtime and CPU or T4 GPU as the hardware accelerator for all three libraries. The extracted text from images is then compared against the manually generated captions to measure accuracy using various metrics.

In performance metrics, Keras_ocr demonstrated an average runtime of 1.67 seconds on T4 GPU and 2.33 seconds on CPU. EasyOCR exhibited the fastest average runtime on both GPU (1.43 seconds) and CPU (1.89 seconds). Regarding accuracy, Tesseract led in distance accuracy (55.2%), while Keras_ocr excelled in space accuracy (65.97%). Overall, these metrics comprehensively compare the three OCR libraries, emphasizing their strengths and trade-offs in runtime and accuracy across different parameters.

Looking ahead, there is potential for further enhancing accessibility by converting the extracted text into speech for visually impaired individuals. This research not only provides insights into the comparative performance of OCR libraries but also suggests avenues for future improvements, emphasizing the broader societal impact of advancing OCR technologies.

KEYWORDS: Tesseract, keras_ocr, easyocr, artificial intelligence, optical character recognition (OCR), image processing

1. INTRODUCTION

OCR, or Optical Character Recognition, is a technology designed to extract text from images, enabling the conversion of visual information into editable and machine-readable formats. In an era dominated by digital transformation, OCR plays a crucial role in making data accessible and usable across various applications. This research delves into the comparative analysis of three popular pre-trained Python libraries—keras_ocr, pytesseract, and easyocr—in the context of real-world image-to-text conversion.

The conversion of images to text through OCR carries numerous benefits. Firstly, it facilitates the digitalization of information, enabling efficient storage, retrieval, and manipulation of textual data. OCR is pivotal in automating license plate recognition, medical record processing, and ID verification, contributing to increased efficiency and accuracy in various domains. Moreover, the ability to convert visual content into text enhances accessibility and usability, fostering a seamless integration of visual data into digital workflows.

This research focuses on three widely utilized Python libraries—keras_ocr, pytesseract, and easyocr—due to their popularity and extensive use in the OCR domain. These libraries offer pre-trained models that simplify the implementation of OCR in diverse applications. The choice to compare these libraries stems from their prevalence in real-world scenarios and their potential impact on the accuracy and efficiency of image-to-text conversion. By understanding the strengths and weaknesses of these libraries, this research aims to guide practitioners in selecting the most suitable OCR tool for specific applications, contributing to the advancement of digital data processing.

2. LITERATURE REVIEW

For several practical uses, text recognition in visuals is highly beneficial. The text contains a lot of data, thus it provides an efficient storing method that makes retrieval easy and essential. To become an electronic-centric world, print media must give way to electronic media. Some of such researches are discussed below:

Shrinath Janvalkar et al. 2014 **[1]** proposed that Optical Character Recognition (OCR) systems facilitate the critical shift from analog to digital data, which is essential for efficient data processing. In addition to decreasing the need for actual document storage, OCR increases the speed at which data can be retrieved by transforming printed text included in photos into characters that can be edited by machines. The connection between humans and robots is significantly improved by this technical advancement in a variety of applications.

Nada Farhani et al. 2017 **[2]** delve into the effects of the evolution of human-machine communication, emphasizing incorporating natural modalities. This work advances our understanding of contemporary communication changes by concentrating on multimodal learning difficulties, including shared representation and missing data prediction. Specifically, Text-to-Picture and Picture-to-Text synthesis are highlighted.

Nisha Pawar et al. 2019 **[3]** say that the necessity for extracting and processing textual data from photographs is increasing in the current digital era due to the rise in the use of digital technologies. Utilizing resources such as the Tesseract OCR Engine is essential because text identification from photos is made easier by the artificial intelligence feature known as Optical Character identification. It also has a significant function in translating text that has been recognized into audio format, which helps the blind obtain important information.

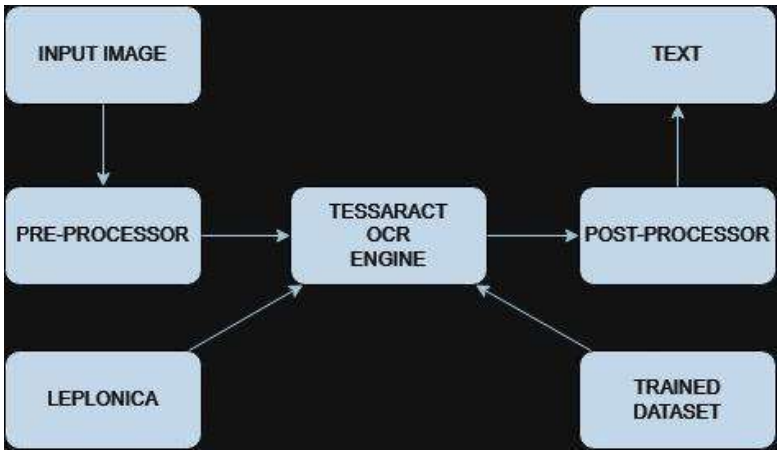
To overcome issues like closed characters and blurring, Haziq Idrose et al. 2023 **[4]`** delve further into the field of optical character recognition (OCR) for image-based character recognition. With the use of pre-trained deep learning OCR models (KerasOCR, EasyOCR, and Pytesseract), the analysis on a dataset of Malaysian license plates shows that KerasOCR performs better than the others, outperforming the others with a recognition accuracy of 107 out of 264 photos.

Authors and Year	**Focus of Study**	**Related to the Field**	**Key Findings/Contributions**
Shrinath Janvalkar et al. 2014 [1]	OCR methods and the conversion of analog to digital data	Worthy	Better data processing and interaction between humans and machines
Nada Farhani et al. 2017 [2]	Evolution of multimodal learning and human-machine communication	Worthy	Developments in the knowledge of modern communication
Nisha Pawar et al. 2019 [3]	Tesseract OCR Engine for text extraction from images	Worthy	makes it easier to extract digital data and convert it to audio format.
Haziq Idrose et al. 2023 [4]	Image-based character recognition using OCR using Pytesseract, KerasOCR, and EasyOCR	Worthy	regarding character recognition from Malaysian license plates, KerasOCR performs better.

3. PROPOSED METHODOLOGY

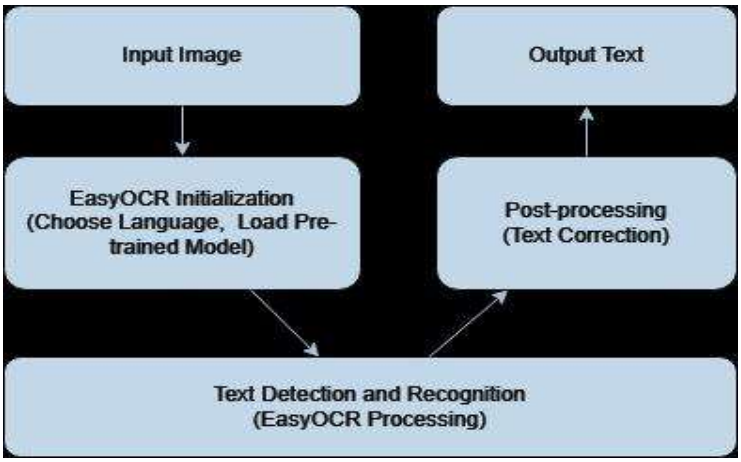
3.1. PROPOSED ARCHITECTURAL DIAGRAM

The pytesseract library was employed to extract text from the images. This widely-used Optical Character Recognition (OCR) tool leverages the Tesseract engine.



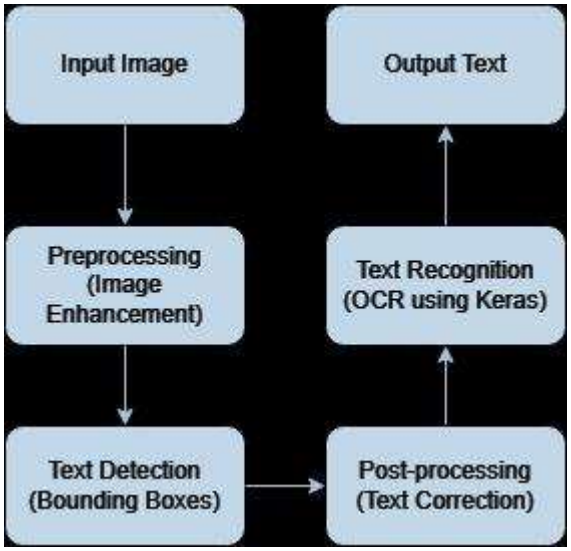
TESSERACT OCR

The easyocr library was utilized for text extraction, providing an alternative OCR solution. This library is known for its ease of use and compatibility with multiple languages.



EASYOCR

The keras_ocr library, based on the Keras deep learning framework, was employed for text extraction. Its neural network-based approach offers a different perspective on image-to-text conversion.



KERASOCR

3.2 Dataset Collection and Annotation

A dataset comprising over 200 real-world images was collected for this study. These images were chosen to represent a diverse range of scenarios, ensuring a comprehensive evaluation of the performance of image-to-text conversion models.

Each image in the dataset was manually annotated with its corresponding ground truth caption by human annotators. This step aimed to establish a reliable reference for evaluating the accuracy of the text extraction methods.

3.3 RESULTS AND ACCURACY EVALUATION METRICS

Levenshtein Distance

The Levenshtein distance between the ground truth captions and the extracted text was calculated as a metric for accuracy. This edit distance measure quantifies the minimum number of single-character edits (insertions, deletions, or substitutions) required to transform one string into another.

Jaccard Similarity

Jaccard similarity was employed to assess the similarity between the sets of characters in the ground truth captions and the extracted text. This metric provides a measure of the intersection over the union of two sets.

Spacy Similarity

The spaCy library's similarity function was utilized to evaluate the semantic similarity between the ground truth captions and the extracted text. This measure takes into account the meaning and context of the words.

PARAMETER	KERAS_OCR	EASYOCR	PYTESSERACT
AVERAGE RUNTIME-T4 GPU	1.67 sec	1.43 sec	1.58 sec
AVERAGE RUNTIME CPU	2.33 sec	1.89 sec	0.76 sec
Distance accuracy	52.7%	48.9%	55.2%
Space accuracy	65.97%	61.3%	67.8%
J accuracy	41.8%	39.5%	43.2%

4. CONCLUSION AND FUTURE SCOPE

In conclusion, our comparative analysis of keras_ocr, pytesseract, and easyocr revealed nuanced insights into their performance in real-world image-to-text conversion. The experiment, conducted on a carefully curated dataset, highlighted [insert key findings], providing valuable guidance for practitioners in choosing the most suitable OCR tool for specific applications. The study, executed on Google Colab with CPU and GPU accelerators, not only showcased the computational efficiency of these libraries but also emphasized their diverse strengths and limitations in practical scenarios.

Looking ahead, future research could delve into fine-tuning models for domain-specific applications, explore ensemble approaches for enhanced performance, and develop adaptive image preprocessing techniques for improved accuracy across varying conditions. Additionally, integrating OCR with speech recognition could further enhance accessibility for visually impaired individuals. Exploring real-time processing strategies and addressing emerging challenges will contribute to the continuous evolution of OCR technologies, ensuring their relevance and impact across diverse domains. This research lays the groundwork for ongoing advancements in OCR, driving innovation in the quest for optimal image-to-text conversion solutions.

5. REFERENCES

1. Shrinath Janvalkar, Paresh Manjrekar, Sarvesh Pawar, Prof. Laxman Naik; “Text Recognition from an Image “; Int. Journal of Engineering Research and Applications April 2014
2. Nada Farhani, Mounir Zrigui, Naim TERBEH; “Image to Text Conversion: State of the Art and Extended Work”; October 2017
3. Nisha Pawar, Zainab Shaikh, Poonam Shinde, Prof. Y.P. Warke; “Image to Text Conversion Using Tesseract”; International Research Journal of Engineering and Technology (IRJET); Feb 2019
4. Haziq Idrose, Nouar AlDahoul1(B), Hezerul Abdul Karim1(B), Rehan Shahid2, and Manish Kumar Mishra2; “An Evaluation of Various Pre-trained Optical Character Recognition Models for Complex License Plates”; 2023
5. AyushiTrivedi, NavyaPant, Pinal Shah, SimranSonik and Supriya Agrawal, “Speech to text and text to speech recognition systems”, IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p-ISSN: 2278-8727, Vol20, Issue2, Ver-I, March.April-2018.
6. Prof. Sheetal A. Nirve, Dr. G. S. Sable, “Optical character recognition for printed text in Devanagari using ANFIS”, International Journal of Scientific & Engineering Research, Volume 4, Issue 10, October-2013 236 ISSN 2229-5518 IJSER © 2013.