IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, May 2022
Impact Factor: 6.252

Image-Based Text Recognition & Translation Using Machine Learning

Aishwarya Vijay Deshmukh[1], Sumit Rajendra Mali[2], Manish Kumar[3], Dhairya Deshmukh[4]
Department of Information Technology[1,2,3,4]
Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India
Savitribai Phule Pune University, Pune, Maharashtra, India

Abstract: *Text recognition in images is a research area that aims to create a computer system that can automatically read text from images. There is high demand these days for storing information available in paper document format on a computer storage disc and then reusing this information through the searching process. One simple method for transferring information from paper documents to a computer system is to scan the documents and then save them as images. However, it is extremely difficult to reuse this information by reading the individual contents and searching the contents of these documents line-by-line and word-by-word. The difficulties involved in this are the font characteristics of the characters in paper documents and the image quality. Because of These difficulties, the computer is unable to recognize the characters while reading them. As a result, character recognition mechanisms are required to perform In this paper, we will look at a method for recognizing text from images. The goal of this paper is to recognize text from images for better reader comprehension by employing a specific sequence of different processing modules. Many people in today's world face the problem of language translators. A technology known as machine translation can help to solve this type of problem. This paper proposed machine translation using a recurrent neural network with an attention mechanism, where a recurrent neural network (RNN) is a type of neural network designed to capture information from sequence and time-series data. As human language is one big complex pattern or a complicated pattern, RNN is useful for learning patterns in a given set of data. In machine translation, two recurrent neural networks collaborate to transform one sequence into the other. An encoder network modifies an input sequence in a vector, whereas a decoder network modifies the vector in a new sequence. To improve the previously stated RNN model, we will employ the attention mechanism, which aids in focusing on a specific range of input sequences.*

Keywords: Text Detection, Text Recognition, OCR, Machine Translation, Language Translator, RNN

I. INTRODUCTION

1.1 Machine Learning

Machine learning is a branch of artificial intelligence (AI), and computer science focuses on using data and algorithms to mimic human learning methods and gradually improve accuracy. Machine learning is an important part of the growing field of data science. By using statistical methods, algorithms are trained to classify or make predictions and gain important insights in data mining projects. These insights facilitate decision-making within the application or organization and ideally influence key growth indicators. As the large data grows and continues to grow, the market demand for data scientists grows, and you need to identify the most relevant business questions and identify the data to answer them. For machine learning, use PythonOpenCV, TesseractOCR4, and Langdetect.

1.2 Django

Django is a high-level Python web framework that enables the rapid development of secure and maintainable websites. Built by experienced developers, Django handles many of the tedious tasks of web development, so you can focus on creating your app without having to reinvent the wheel. It's free and open-source, with a lively and lively community, good documentation, and lots of free and paid support options.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, May 2022
Impact Factor: 6.252

1.3 Heroku

Heroku is a cloud platform (PAAS) as a container-based service. Developers use Heroku to deploy, manage, and scale their latest apps. Elegant, flexible, and easy to use, our platform provides developers with the easiest way to get their apps to market.

II. Introduction to Project Topic

The content of digital images is rapidly expanding due to the advancement of multimedia and network technology. According to 2010 [1] statistics, more than 30 million photos were uploaded to the social networking site Facebook every month, and more than 35 hours of video are uploaded to the video-sharing site YouTube every minute. As a result, the method for correctly understanding this information is critical; text information includes important semantic clues in images and videos, has the visual sense in images, and can be easily extracted when compared to other semantic information. As a result, it is widely used in video summarization, content-based image indexing, and video sequence retrieval. Character recognition research began in the 1950s, and many methods have been proposed since then, including pattern matching [2,3], blurring [4,], hierarchical classification [5,6], and multiple classifiers combination [7,8]. Character recognition has found widespread application in a wide range of fields, and several commercial optical character recognition (OCR) systems have been developed. Microsoft Corporation, for example, has created a Microsoft office document imaging components in the Microsoft office progprogramBYY Corporation created FineReader, which is widely used in digital libraries, embedded devices, and other applications.

When the text recognition task is extended from clearly scanned documents to complex scene images, the background becomes complex and diverse. As a result, existing commercial OCR systems struggle to produce good results. To address these issues, ICDAR [9–11] and TRECVID [12] organized a series of competitions on text recognition in natural scene images, and many effective methods have been developed. Many people in today's world are facing language translation issues, such as conversing with someone who only speaks a language you don't understand or having information in a language other than English, such as French, German, etc. Machine translation is a technique that can solve this type of difficulty. This paper proposes machine translation using a recurrent neural network with an attention mechanism, where recurrent neural networks (RNN) are types of neural networks designed for capturing information from sequence and time-series data. As human language is one big complex pattern or a complicated pattern, RNN is useful for learning patterns in a given set of data. In machine translation, two recurrent neural networks collaborate to transform one sequence into the other. An encoder network modifies an input sequence in a vector, whereas a decoder network modifies the vector in a new sequence.

III. Tools and Technology Used

3.1 Python

Python is an open, interpreted, high-level language that provides a great approach to object-oriented programming. It is one of the best languages used by data scientists for a variety of data science projects/applications. Python provides great features for processing math, statistics, and scientific functions. It provides an excellent library for data science applications. One of the main reasons Python is so widely used in the science and research community is that its ease of use and simple syntax make it easy to adapt to non-technical people. It is also suitable for rapid prototyping.

3.2 Jupyter Notebook

Jupyter Notebook is an open-source web application that allows you to create and share documents containing live code, equations, visualizations, and text. Jupyter Notebook is managed by Project Jupyter staff. Jupyter Notebooks is a spin-off project from the IPython project that had the IPython Notebook project itself. The name Jupyter comes from the main supported programming languages, Julia, Python, and R. Jupyter comes with an IPython kernel that allows you to program in Python.

3.2 Pytesseract

Python-tesseract is an Optical Character Recognition (OCR) tool for Python. That is, it recognizes and "reads" the text embedded in the image. Pythontesseract is a wrapper for Google's Tesseract OCR engine. It can also read all image types

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, May 2022
Impact Factor: 6.252

supported by Pillow and Leptonica imaging libraries, including jpeg, png, gif, BMP, tiff, etc., making it useful as a stand-alone calling script for Tesseract. In addition, when used as a script, Pythontesseract prints the recognized text instead of writing it to a file.

3.3 Google Colab

Collaboratory, or Colab for short, is a product of Google Research. With Colab, anyone can write and execute Python code from a browser, making it especially suitable for machine learning, data analysis, and education. Technically, Colab is a hosted Jupyter notebook service that provides free access to computing resources, including the GPU. Most importantly, no setup is required, and the notebooks you create can be edited by team members at the same time, much like editing a document in Google Docs. Colab supports many popular machine learning libraries that you can easily load into your notebook.

3.4 OpenCV

OpenCV is the huge opensource library for computer vision, machine learning, and image processing and now it plays a major role in real-time operation which is very important in today`s systems. By using it, one can process images and videos to identify objects, faces, or even the handwriting of a human. When integrated with various libraries, such as NumPy, python is capable of processing the OpenCV array structure for analysis. To Identify image patterns and their various features we use vector space and perform mathematical operations on these features.

3.5 Matplotlib

Matplotlib is a great Python visualization library for the 2D plotting of arrays. Matplotlib is a cross-platform data visualization library based on the NumPy array, designed to work with the broader SciPy stack. Introduced by John Hunter in 2002. One of the greatest benefits of visualization is the visual access to the vast amount of data in an easily digestible image. Matplotlib consists of multiple graphs such as lines, bars, scatter plots, and histograms. Matplotlib comes with various plots. Charts help you understand trends and patterns and establish correlations. These are usually tools for thinking about quantitative information.

IV. PROPOSED ARCHITECTURE

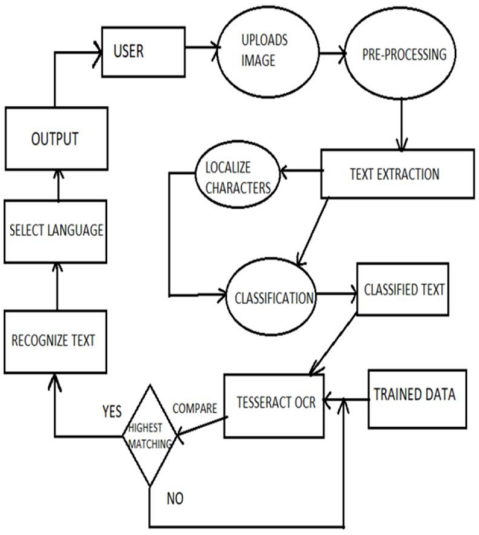

Figure: Proposed Architecture

The user uploads an image which first of all, undergoes preprocessing before going to the next stage of Text extraction. The extracted text is then classified and sent as an input to the Tesseract OCR engine with the trained data. The output is then compared and if the matching is highest then it moves to the next stages. And if not, the Tesseract engine is once again

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 4, May 2022
Impact Factor: 6.252

sent the input and the result is compared again. The next steps after highest matching involve text recognization, and language selection, after which finally the output is produced based on the selected language.

V. CONCLUSION

- This seminar report is based on how to recognize text from an image and translate the detected text from one language to another. Specifically, you learned:
- Detect Text from an image using OCR and OpenCV
- Recognise Text using Tesseract OCR 4
- Language Detection using NLP
- Language Translation using RNN and NLP
- If a document contains languages outside of those given in the -l LANG arguments, results may be poor.
- Doesn't do well with images affected by artifacts including partial occlusion, distorted perspective, and complex background.
- Poor quality scans may produce poor quality OCR. Perhaps it leads to processing an Image at a deployed website domain and gives an Application error in which the domain has 500MB capacity of data handling.
- It does not expose information about what font family text belongs to.

REFERENCES

[1]. Bal, M.; Bal, Y.; Ustundag, A. Knowledge Describes, and Discovery Using Formal Concept Analysis: An HRM Application. World Congress Eng. 2021, 2, 1068–1073.

[2]. Chien, C.-F.; Chen, L.-F. Data mining to advanced personnel selection and enhance human capital: A case study in the high-technology industry. Expert Syst. Appl. 2018, 2, 280–290.

[3]. R. Deepa, Associate Professor, Department of Information Technology, Loyola-ICAM College of Engineering and

[4]. Ganesh G. Patil, Department of Computer Science and Engineering SVERI's College of Engineering, Pandharpur, India.

[5]. A. Sharma (B) · P. S. Banerjee · A. Sharma · A. Yadav Department of Computer Science and Engineering, Jaypee University of Engineering and Technology, Guna, India

[6]. Zelenika, D., Povh, J., & Ženko, B. (2019). Text Detection in Document Images by Machine Learning Algorithms. Proceedings of the 9th International Conference on Computer Recognition Systems CORES 2015, 169–179.

[7]. Nicola Bertoldi and Marcello Federico. 2016. A new decoder for spoken translation based on confusion networks. In Automatic Speech Recognition and Understanding, 2005 IEEE Workshop on, pages 86–91. IEEE

[8]. Sharma, R., Kumar, A., Meena, D., & Pushp, S. (2020). Employing Differentiable Neural Computers for Image Captioning and Neural Machine Procedia Computer Science, 173, 234–244. [9] Klein, G., Kim, Y., Deng, Y., Senellart, J., & Rush, A. M. (2017). Openness: Open-source toolkit for neural machine translation.

[9]. Kanchan M. Tarwani, Swathi Edem "Survey on Recurrent Neural Network in Natural Language Processing", International Journal of Engineering Trends and Technology (IJETT), V48(6),301-304 June 2017.

[10]. D. Bahdanau, K. Cho and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate", Published as a conference paper at ICLR 2018

[11]. L. Jain and P. Agrawal, "English to Sanskrit Transliteration: an effective approach to design Natural Language Translation Tool", International Journal of Advanced Research in Computer Science, Volume 8, No. 1, Jan-Feb 20