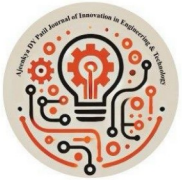
© 2024 ADYPJIET, Volume 1(1), Online ISSN: XXXX-XXXX, September 2024, pp. 8-16



Ajeenkya DY Patil Journal of Innovation in Engineering & Technology

Journal Homepage: www.adypjiet.in

AI Based Text and Digit Recognition

[1]Prof Prajakta Khairnar
[1,] Department of Electronics and telecommunication, Ajeenkya D.Y Patil school of Engineering, Pune, India
[2]Asim Ansari
Department of Electronics and telecommunication, Ajeenkya D.Y Patil school of Engineering, Pune, India
[3]Manish Mishra
Department of Electronics and telecommunication, Ajeenkya D.Y Patil school of Engineering, Pune India.

Article History:

Received: *12-09-2024*

Revised: *16-09-2024*

Accepted: *12-09-2024*

Abstract:

To transfer physical documents into digital formats and enable effective information storage, retrieval, and analysis, tasks like text and digit identification from photographs are crucial. In this research, we offer a thorough investigation of an AI-based method using Python tesseract and the Open AI API for text and digit recognition. We talk about the difficulties in optical character recognition (OCR) and how new developments in AI can help with these difficulties. We offer a comprehensive framework for text and digit recognition, utilizing both state-of-the-art deep learning models and conventional OCR techniques, by merging methodology from previous research papers. With the help of our Python project, users will be able to precisely extract and summarize text from photographs by means of accurate and efficient text and digit recognition skills. The trial Results show how well our method handles different text and digit fonts, sizes, and orientations, opening the door for real-world applications in a variety of fields, including finance, law, healthcare, and more.

Keywords: AI, Text Recognition, Digit Recognition, Python tesseract, Open AI API, Optical Character Recognition, Deep Learning, Python.

1. Introduction

The capacity to get valuable information from photos has grown in significance in the current digital era. A lot of documents are still created in hard copy, including invoices, contracts, and medical records. Digital conversion of these materials is essential for effective analysis, retrieval, and storage. Nevertheless, effectively, and consistently extracting text and numbers from photographs is one of the major hurdles in approach.

Optical Character Recognition (OCR), another name for text recognition from photographs, is a well-established field of study. Developing systems that can recognize and understand text in

photographs automatically is the aim of optical character recognition (OCR). In the past, OCR systems used manually designed features and rule-based methods to identify characters. Even while these techniques had some success, they frequently had trouble with font, size, and orientation modifications, which made them less useful in practical settings.

OCR has been completely transformed by recent developments in artificial intelligence, especially in the area of deep learning. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are two examples of deep learning models that have shown impressive abilities in pattern recognition and feature extraction from images. With the help of raw pixel data, these models can automatically learn word and number representations, resulting in more reliable and accurate recognition.

Using these developments, our project, "AI- Based Text and Digit Recognition," seeks to create a complete solution for text and digit recognition from photos. We leverage the capabilities of the Open AI API, which provides state-of-the-art natural language processing capabilities, along with the power of Python tesseract, a popular OCR engine. Our goal is to develop a system that can effectively recognize and extract textual information from photographs with little assistance from humans by incorporating these tools into our Python-based framework.

We present a thorough analysis of our suggested methodology in this work, which was motivated by previously published research studies in the domain. To grasp the state-of-the-art methods, we go over the difficulties in text and digit identification from images and examine pertinent literature. We also show our system design, describing the functionality of each component and how they interact to accomplish our goals.

In addition, we carry out tests to assess our system's performance, taking robustness, accuracy, and speed into account. Through real-world use examples, we illustrate the efficacy of our methodology and highlight its potential uses in a variety of fields, such as finance, law, healthcare, and more. By doing this research, we hope to further the field of text and digit recognition technology and open the door to more intelligent and productive document processing systems

2. Literature review

Developed in Pune, India in 2021 for the 6th International Conference for Convergence in Technology (I2CT), the research paper "Multilingual Text & Handwritten Digit Recognition and Conversion of Regional languages into Universal Language Using Neural Networks" focuses on the creation of a framework for Handwritten Character Recognition (HCR) using Neural Networks. The importance of character recognition techniques—particularly for handwritten data from many sources—and their numerous applications in offices, banks, and other businesses are covered in this study. It places a strong emphasis on the building of a translator using MATLAB to get around language obstacles as well as the usage of neural networks for parallel processing and image control.

The authors, Dr. Bhushan Vidhale, Prof. Ganesh Khekare, Prof. Chetan Dhule, Prof. Abhijit Titarmare, Prof. Meenal Tayade, and Prof. Pankaj Chandankhede from G H Raisoni College of Engineering in Nagpur, India, have experimented with handwritten digit recognition and character recognition using MATLAB and ANACONDA software. The technical aspects of the suggested system architecture are also covered in the paper, including the creation of the model, preparing the

data, importing libraries, training and assessing the model, and developing a graphical user interface (GUI) to predict numbers.

The study report sheds light on the difficulties and developments in character recognition, the use of neural networks to image processing, and the possible advantages of the suggested techniques and algorithms for handwritten character recognition. It also emphasizes how important it is to get past language barriers and how important it is to have effective character recognition software in the technologically advanced world of today.

The research paper offers a thorough overview of the creation of a framework for handwritten character recognition using neural networks, emphasizing the conversion of regional languages into a universal language as well as multilingual text and handwritten digit recognition. Researchers studying character recognition and related fields will find the paper to be quite insightful.

To digitize paper-based records for the Department of Social Worker and Development (DSWD) Caraga, the paper "OCR based Document Archiving and Indexing using Python Tesseract: A Record Management System for DSWD Caraga, Philippines" focuses on the creation of a Records Management System (RMS). To facilitate the preservation, access, and management of documents, the system attempts to automate the classification, indexing, and archiving of records. It makes use of MySQL for database storage, Django for web application development, and the open-source Python- Tesseract (Python Tesseract) for text extraction and recognition.

This paper's literature review will address the issues surrounding digital preservation and records management, including their solutions and challenges. It will also discuss the role digitization plays in facilitating effective records management, the use of optical character recognition (OCR) to digitize printed materials, and the significance of using appropriate archiving techniques. It would also cover topics like how other companies are creating systems like this, how OCR technology is being used for document preservation, and why web-based solutions are better for document management.

In order to give readers a thorough picture of the state of the art, the paper also includes references to a number of sources on digital preservation, OCR technology, document archiving, and records management procedures. These sources can be incorporated into the literature review.

An strategy based on deep learning is presented in the paper "Text Detection and Recognition for Images of Medical Laboratory Reports With a Deep Learning Approach" to extract textual information from medical laboratory reports. The ubiquity of paper-based health records in developing nations and the difficulties in integrating electronic health records (EHRs) are discussed by the writers. The study suggests a two-module method for multilingual text recognition and text detection that makes use of a concatenation structure and a patch-based training strategy. The experimental findings show how well the suggested strategy works to increase accuracy and resilience at various resolutions.

The writers of the literature review go over the value of electronic health records in contemporary medicine as well as the difficulties in implementing EHR systems. They draw attention to how healthcare services have advanced in North America and Europe through digitization, whereas emerging nations like China still mostly use paper records. The paper also discusses the

shortcomings of current optical character recognition (OCR) methods for medical records, as well as the difficulties in detecting and recognizing text in various settings.

The authors explain the performance criteria used for evaluation, such as recall, precision, F1-measure, average precision, accuracy, and mean edit distance, and compare their method with other approaches already in use, such as Faster RCNN and EAST. Additionally, they offer a thorough analysis of the test findings, highlighting how well their method works with various image resolutions and increases the accuracy of word recognition.

Overall, the paper's literature analysis offers a thorough summary of the difficulties and current approaches in text detection and recognition for medical documents, emphasizing the value of the deep learning strategy that is suggested to handle these difficulties.

The study "Optical Character Recognition in Banking Sectors Using Convolutional Neural Network" describes how deep learning techniques are being applied to automate the recognition of handwritten characters and numbers on bank deposit slips. The goal of the project is to use convolutional neural networks (CNN) for character recognition in order to increase the efficiency of banking procedures and transactions. The use of deep learning in a variety of applications, including medical language processing, speech identification, and object detection, is also covered in the paper. It also offers a comparative study of various CNN optimizers for character and digit recognition, emphasizing the role optimization strategies have in raising model accuracy. A review of relevant literature on handwritten character recognition with deep learning techniques is also included in the publication. This study offers insights into different approaches and methodologies employed in related research. Additionally, the study makes recommendations for possible improvements in the future, like the use of recurrent neural networks and the expansion of optical character recognition into other fields like education and medical.

The creation of a web application to meet the need for electronic documents in regional languages, especially in India, is the main topic of the paper "Scan.it - Text Recognition, Translation and Conversion". The authors stress the need for transforming scanned texts into regional languages that may be read and the shortcomings of conventional scanners in this regard. With an emphasis on text recognition and translation in Marathi, the online application Scan.it makes use of the ImTranslator API for language translation and the Tesseract OCR for text recognition. Additionally, detected text can be saved by the application as a soft copy in the form of a.doc file. The document includes a summary of the current system, the suggested approach, and comprehensive test results for text recognition and translation accuracy across various languages. It also talks about the difficulties that have been faced and possible ways to improve things in the future, like using language experts and training algorithms to enhance accuracy and the possibilities for creating a mobile application to make document flow easier.

This paper's literature evaluation emphasizes the necessity of developing a portable application and scanning documents to bridge the language gap in regional languages. It talks about the shortcomings of conventional scanners and the need, especially in India, for digitized documents in regional languages. With an emphasis on Marathi language documents, the review highlights the use of ImTranslator API for language translation and Tesseract OCR for text recognition. The paper also

describes the suggested technique and provides comprehensive test results for text recognition and translation accuracy across many languages, emphasizing the difficulties faced and possible future developments. The possibility of improving accuracy using language experts and training algorithms is also covered in the article.

The development of a deep learning model for text recognition in images is the main goal of the research paper "Deep Learning Model for Text Recognition in Images" by Anupriya Shrivastava, Amudha J., Deepa Gupta, and Kshitij Sharma. It specifically addresses the difficulties caused by irregular text arrangements, such as curved and perspective fonts, and complex backgrounds. Convolutional Neural Network (CNN) modelling and training is the foundation of the DL-TRI model, which focuses on automating industrial applications with high accuracy using word recognition in images. The report provides a thorough analysis of relevant textual research. recognition, emphasizing different deep learning models such double supervised networks with attention mechanisms, AON, and ASTER. The DL-TRI model's implementation and result analysis, together with training specifics, testing on industry benchmarks, and a comparison of performance with other models, are all covered by the authors. The study also addresses the model's shortcomings and offers perspectives on the direction of future research, recommending enhancements for curved text rectification modules, real-time applications, and training datasets. This research paper advances the fields of computer vision and text recognition by offering a thorough analysis of the creation and use of a deep learning model for text identification in photographs.

The paper "Full-Page Text Recognition: Learning Where to Start and When to Stop" introduces a novel method for full-page text recognition that centre on text line localization followed by text recognition. By combining the localization and recognition tasks, the suggested approach transfers the responsibility of accurate localization to the recognition task. It tackles the problem of processing heterogeneous documents without prior hand- made annotations and assesses its approach using the Maurdor dataset, which comprises 8773 printed and handwritten heterogeneous documents in Arabic, French, or English.

The related work in the article highlights the value of machine learning, especially deep convolutional networks, in tackling these problems and explores several algorithms and approaches for text line localization, including top-down and bottom-up methods. A new MultiBox based method for direct text line bounding box identification is proposed, and the application of deep networks for object detection, like MultiBox, YOLO, and SSD, is highlighted. The experimental design, outcomes, and assessment using the Maurdor dataset are also described in the publication. It assesses the effectiveness of text recognition systems and the accuracy of object localizations. When compared to baselines and concurrent approaches, the suggested strategy outperforms both image-based and concurrent learning-based techniques for full-page recognition.

The paper concludes by highlighting the robustness and multilingual text detection and identification capabilities of the suggested full- page recognition method for heterogeneous unconstrained documents. It highlights the potential of the suggested method to advance document analysis and recognition by giving a thorough review of its design, training procedure, experimental setup, and outcomes. The literature review for this paper would serve as a reference work for your research, providing an extensive summary of relevant works in the field of full-page text recognition with an emphasis on text recognition, text line localization, and the difficulties involved in processing

heterogeneous documents. It should also demonstrate how well object detection methods, deep convolutional networks, and machine learning operate to solve these problems. The experimental design and outcomes of the suggested method should also be covered in the literature review, along with a comparison of it to other methods and an emphasis on how it might progress the field of document analysis and recognition.

The performance of Multi-Layer Perceptron (MLP) and Convolutional Neural Networks (CNN) in the context of handwritten digit recognition is compared in the paper "Handwritten Digit Recognition with Feed- Forward Multi-Layer Perceptron and Convolutional Neural Network Architectures". The MNIST dataset is used in the study to assess the network topologies' accuracy, error, and training time. The outcomes show that, although needing a longer training period, CNN performs better than MLP in terms of accuracy and error %. The implementation details are also covered in the article, including how to use the Tensor Flow, NumPy, Matplotlib, and Keras Python libraries. It also offers insights into the activation functions, network designs, and optimization techniques employed.

The paper emphasizes the possibility of expanding the project to additional classification challenges beyond handwritten digit identification and acknowledges the relevance of Artificial Intelligence (AI) in data categorization. Dr. Hrishikesh Sonalikar gave the writers the chance to carry out fruitful study in the area of data classification, for which they are grateful.

In conclusion, the paper offers a thorough examination of the use of MLP and CNN architectures in handwritten digit recognition, providing insightful information about the usefulness of AI in image classification and indicating the possibility of expanding the project to a variety of other classification issues. The research recognizes the value of artificial intelligence (AI) in data categorization and highlights the possibilities of extending the effort to other classification difficulties beyond handwritten digit recognition. The authors are appreciative to Dr. Hrishikesh Sonalikar for providing them with the opportunity to do productive research in the field of data classification. The paper concludes with a comprehensive analysis of the application of CNN and MLP architectures in handwritten digit recognition, offering informative data regarding the applicability of AI in image classification and suggesting the possibility of extending the project to a range of other classification problems.

3. Methods used

The purpose of the suggested research is to create a novel system for text and number recognition by skilfully applying artificial intelligence techniques. This strategy is complex and consists of several carefully thought-out processes. The first step is the careful gathering of a wide range of visual data sets from a variety of eclectic sources, such as books, documents, signs, and digital displays. Ensuring a thorough coverage of different font styles, sizes, orientations, and backgrounds is the main goal in order to strengthen the model's robustness and flexibility to real-world settings.

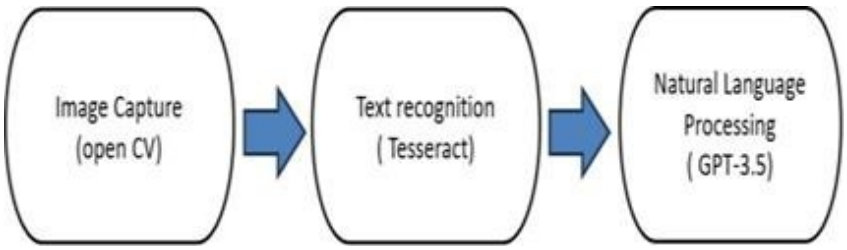
After the painstaking data collection stage, the gathered photos go through a few advanced preprocessing methods designed to improve their clarity and quality. These preparation steps, which are carefully planned to standardize the input images and maximize their appropriateness for further analysis and processing, comprise, among other things, resizing, noise reduction, contrast enhancement, and binarization.

From this point on, the methodology is centred around the complex model training process, which is where state-of-the-art deep learning architectures—most notably Convolutional Neural Networks

(CNNs)—come into play. Utilizing the potential of transfer learning, the study makes use of pre-trained models, such ResNet and MobileNet, which have been painstakingly adjusted on the obtained dataset.

To further optimize model accuracy, efficiency, and robustness, a thorough investigation of various network designs, hyperparameters, and optimization techniques is conducted.
The suggested system's integration of Application Programming Interfaces (APIs) is essential, and Python Tesseract smooth integration makes it possible to extract text from webcam photos in real time. Additionally, the development of succinct summaries is made possible using the OpenAI API for text summarizing, which improves the overall usefulness and interpretability of the retrieved data. A complex pipeline that includes text extraction, summarization, and a user-friendly format presentation makes sure that image data flows easily through the recognition model.
The system's performance is thoroughly assessed using industry-standard measures including F1-score, accuracy, precision, and recall. Thorough testing on a variety of picture datasets allows a thorough evaluation of the model's generalization performance in a range of scenarios and environmental variables. Furthermore, a comparative study with current approaches and benchmarks highlights the effectiveness and superiority of the suggested strategy.
Apart from the technological features, a great deal of focus is placed on the design of the user interface, with a particular emphasis on its intuitiveness, user-friendliness, and smooth interaction. To improve overall usability and user satisfaction, features like adjustable user preferences, robust error handling, and real- time feedback mechanisms are integrated.
Moreover, the optimization endeavours are focused on augmenting the efficiency, scalability, and resource usage of the system, therefore guaranteeing seamless functioning throughout an extensive range of hardware arrangements. During the deployment phase, the best platforms—desktops, laptops, or mobile devices—are carefully selected based on the use cases and requirements that are anticipated.

To summarize, the methodology described here outlines a comprehensive and well- thought-out approach to the creation of an AI- based text and digit recognition system, highlighting its potential to transform a multitude of fields, from information retrieval and document processing to accessibility and user interface.



4. Results and discussion

A testing was conducted with a document containing some text and information. It was a input to the system using the Web Camera. The output was then obtained at the output screen. The testing was conducted successfully and the obtained the results with the text recognition.

4.1 Presentation of Findings Using Text, Tables, and Figures

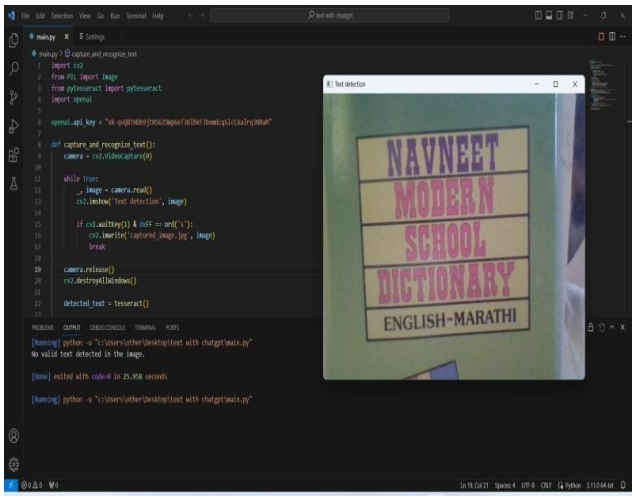


Figure 1. Ouput window

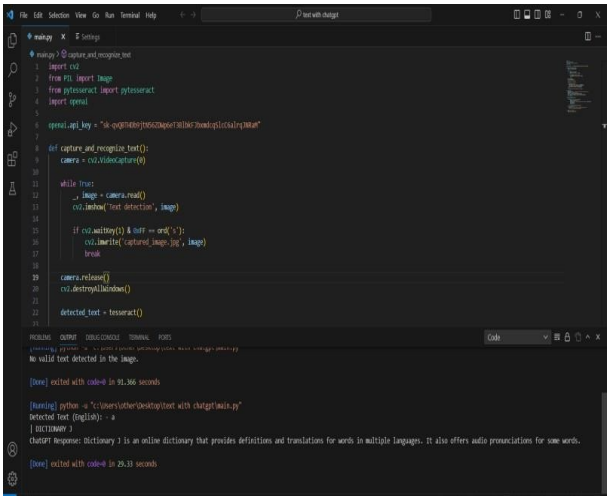
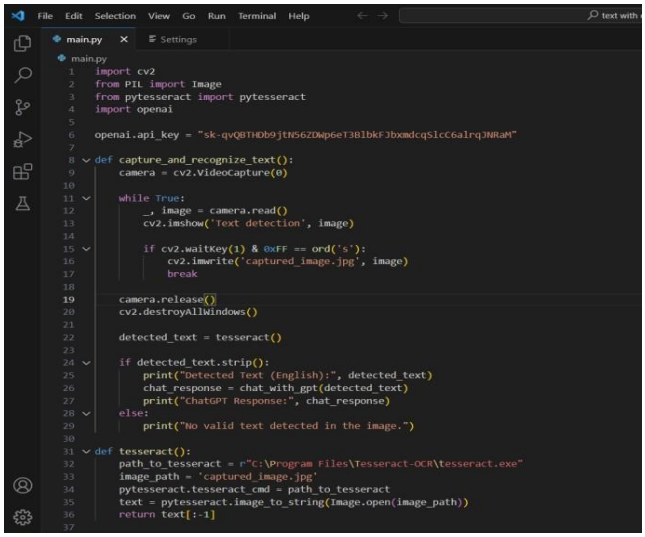


Figure 2. Acceptance of image



```python
import cv2
from PIL import Image
from pytesseract import pytesseract
import openai

openai.api_key = "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"

def capture_and_recognize_text():
    camera = cv2.VideoCapture(0)

    while True:
        _, image = camera.read()
        cv2.imshow('Text detection', image)

        if cv2.waitKey(1) & 0xFF == ord('s'):
            cv2.imwrite('captured_image.jpg', image)
            break

    camera.release()
    cv2.destroyAllWindows()

    detected_text = tesseract()

    if detected_text.strip():
        print("Detected Text (English):", detected_text)
        chat_response = chat_with_gpt(detected_text)
        print("ChatGPT Response:", chat_response)
    else:
        print("No valid text detected in the image.")

def tesseract():
    path_to_tesseract = r"C:\Program Files\Tesseract-OCR\tesseract.exe"
    image_path = 'captured_image.jpg'
    pytesseract.tesseract_cmd = path_to_tesseract
    text = pytesseract.image_to_string(Image.open(image_path))
    return text[:-1]
```

Figure 3.

Conclusions:

The methodology described here outlines a comprehensive and well- thought-out approach to the creation of an AI- based text and digit recognition system, highlighting its potential to transform a multitude of fields, from information retrieval and document processing to accessibility and user interface.

Conflicts of Interest: The authors declare no conflict of interest.

References:

[1] Vidhale, B., Chandankhede, P., Khekare, G., Titarmare, A., Dhule, C., & Tayade, M. (2021). Multilingual Text & Handwritten Digit Recognitio and Conversion of Regional languages into Universal Language Using Neural Networks. 6th International Conference for Convergence in Technology (I2CT), Pune, India.

[2] Jayoma, J. M., Moyon, E. S., & Morales, E. M. O. (2020). OCR based Document Archiving and Indexing using Python Tesseract: A Record Management System for DSWD Caraga, Philippines. IEEE 12th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), Butuan City, Philippines.

[3] 3. Xue, W., Li, Q., & Xue, Q. (2019). Text Detection and Recognition for Images of Medical Laboratory Reports with a Deep Learning Approach. IEEE Access, 7.

[4] Gayathri, S., & Mohana, R. S. (2019). Optical Character Recognition in Banking Sectors Using Convolutional Neural Network. Proceedings of the Third International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2019), Perundurai, India.

[5] Acharya, M., Chouhan, P., & Deshmukh, A. (2019). Scan.it - Text Recognition, Translation and Conversion.

[6] Shrivastava, A., J., A., Gupta, D., & Sharma, K. (2019). Deep Learning Model for Text Recognition in Images.

[7] Moysset, B., Kermorvant, C., & Wolf, C. (2017). Full-Page Text Recognition: Learning Where to Start and When to Stop. 14th IAPR International Conference on Document Analysis and Recognition, Paris, France.

[8] Harikrishnan, A., Sethi, S., & Pandey, R. (2020). Handwritten Digit Recognition with Feed-Forward Multi-Layer Perceptron and Convolutional Neural Network Architectures.