DAYANANDA SAGAR UNIVERSITY

SCHOOL OF ENGINEERING

Bachelor of Technology in

COMPUTER SCIENCE AND ENGINEERING

DIGITAL IMAGE PROCESSING

MINI PROJECT REPORT

On

Pix2Text

SUBMITTED BY

Sameer S Katte (ENG22CS0148)

Sai Shravan V (ENG22CS0144)

Chhavi Sharma (ENG22CS0278)

Sriram Ravindra (ENG22CS0185)

UNDER THE SUPERVISION

Dr Rajesh T M, Associate Professor,
CSE, DSU

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
SCHOOL OF ENGINEERING DAYANANDA SAGAR UNIVERSITY

(2024)

School of Engineering
Department of Computer Science & Engineering
Harohalli, Ramanagara - 562112 Karnataka, India

CERTIFICATE

This is to certify that the DIP MINI PROJECT titled "**Pix2Text**" carried out by **Sameer S Katte (ENG22CS0148), Sai Shravan V (ENG22CS0144), Chhavi Sharma (ENG22CS0278), Sriram Ravindra (ENG22CS0185)** bonafide students of Bachelor of Technology in Computer Science and Engineering at the School of Engineering, Dayananda Sagar University, Bangalore in partial fulfillment for the award of degree in Bachelor of Technology in Computer Science and Engineering,
during the year 2023-2024.

Dr Rajesh T M
Associate Professor
Dept. of CSE
School of Engineering

Dr. Girisha G S
Chairman, CSE
of Engineering

Dayananda Sagar
University

Dr. Uday Kumar Reddy K R
School Dean

School of Engineering
Dayananda
Sagar University

ACKNOWLEDGEMENT

It is a great pleasure for us to acknowledge the assistance and support of many individuals who have been responsible for the successful completion of this DIP MINI PROJECT

First, we take this opportunity to express our sincere gratitude to the School of Engineering & Technology, Dayananda Sagar University for providing us with a great opportunity to pursue our bachelor's degree in this institution.

We would like to thank **Dr. Uday Kumar Reddy K R, Dean**, **School of Engineering & Technology**, **Dayananda Sagar University for** his constant encouragement and expert advice. It is immense pleasure to express our sincere thanks to **Dr. Girisha G S, Chairman**, **Department of Computer Science, and Engineering**, **Dayananda Sagar University,** for providing the right academic guidance that made our task possible.

We would like to thank our teacher **Dr Rajesh T M**, Associate Professor, **Department of Computer Science and Engineering**, **Dayananda Sagar University**, for sparing his valuable time to extend help in every step of our DIP MINI PROJECT, which paved the way for smooth progress and the fruitful culmination of the project.

We are also grateful to our family and friends who provided us with every requirement throughout the course. We would like to thank one and all who directly or indirectly helped us in the DBMS MINI PROJECT.

ABSTRACT

This software is an advanced image-to-text solution designed for a wide range of applications, from digitizing ancient manuscripts to automating modern document workflows. At its core, the tool employs sophisticated image processing techniques to optimize input images, addressing challenges such as noise, distortions, uneven lighting, faded text, and background interference. These enhancements ensure that even low-quality or degraded images are transformed into clear, high-quality visuals suitable for text extraction.

Powered by artificial intelligence, the software not only recognizes and converts text but also analyses its context for enhanced accuracy. Its AI models are trained to handle diverse scripts, fonts, and languages, making it suitable for use cases like preserving historical archives, automating data entry, extracting handwritten notes, and more. By prioritizing robust image optimization alongside intelligent text understanding, this software delivers a reliable and efficient solution for transforming visual data into accessible, editable formats, bridging the gap between analog and digital text processing.

TABLE OF CONTENTS

PAGE NO

Chapter 1
INTRODUCTION

The increasing need to digitize text from images has led to the development of versatile tools that bridge the gap between analog and digital content. This software is designed to process a wide range of text-based images, from ancient manuscripts to modern handwritten notes or printed documents. It combines cutting-edge image processing techniques with artificial intelligence to deliver a powerful solution for extracting and interpreting text. Whether preserving fragile historical documents or streamlining business workflows, this software provides a reliable, efficient, and highly accurate method for converting visual text into accessible, editable formats.

A standout feature of this tool is its robust image processing capability, which addresses common challenges found in text-based images. For historical documents, issues like faded ink, noise, and physical degradation often obscure the text. In modern contexts, uneven lighting, shadows, or textured surfaces can hinder clarity. The software uses advanced algorithms to optimize these images by enhancing contrast, reducing noise, and correcting distortions, ensuring the text is legible for further processing. These pre-processing steps are critical, enabling even poor-quality or damaged images to be transformed into high-quality visuals suitable for text recognition.

Once the image is optimized, the software employs artificial intelligence to extract and understand the text. Unlike traditional OCR tools, which are often limited to pattern matching, this system is AI-driven and trained on diverse datasets, allowing it to recognize text in various languages, fonts, and even cursive handwriting. Additionally, the AI analyses the context of the text to improve accuracy, distinguishing similar characters, inferring missing content, and adapting to specialized or historical scripts. This makes the software not just a tool for transcription but a smart, context-aware system capable of producing reliable results across multiple use cases.

The applications for this technology are broad and impactful. It can be used to digitize historical records for preservation and research, automate data entry in business processes, or even enhance accessibility for visually impaired individuals by converting text into screen-reader-compatible formats.

Chapter 2
PROBLEM STATEMENT

In today's digital age, a significant amount of valuable information remains locked in physical or visual formats, such as historical manuscripts, handwritten notes, and printed documents. These materials are often difficult to access, search, or preserve, especially as they deteriorate over time. For instance, ancient scriptures and archival documents face challenges like fading ink, physical damage, and fragile materials that require careful handling. Similarly, modern documents like forms, handwritten notes, or scanned pages are often cluttered with noise, shadows, or uneven lighting, making manual transcription time-consuming and prone to errors. The inability to efficiently process and digitize such materials limits their usability, accessibility, and longevity.

Traditional optical character recognition (OCR) technologies, while useful, struggle with complex scenarios. They often fail to handle degraded images, non-standard fonts, or cursive handwriting effectively. Additionally, these systems typically lack the ability to understand the context of the text, resulting in errors when dealing with multi-language content, historical scripts, or specialized domains. Furthermore, manual efforts to clean, optimize, and extract text from images can be tedious and inefficient, requiring significant time and expertise, especially for large-scale projects.

These challenges underscore the need for an advanced solution that combines powerful image processing with intelligent text recognition. Such a tool must address both the technical difficulties of cleaning and optimizing complex images and the cognitive task of accurately interpreting the extracted text. By tackling these problems, the solution would enable users to unlock the value of text-based images, whether for preserving history, automating workflows, or enhancing accessibility in diverse applications.

Chapter 3
PROJECT DESCRIPTION

Overview

The project aims to develop a sophisticated image-to-text processing system that combines advanced image optimization, deep learning, and state-of-the-art AI-based OCR tools to extract, clean, and deliver readable text from images. Designed for versatility, the system is capable of handling various use cases, including traffic cameras and historical documents. By leveraging cutting-edge technologies in image processing and text recognition, the solution ensures high accuracy and usability across different scenarios. The system comprises a robust backend infrastructure powered by Python Flask and OpenCV, integrated with third-party APIs such as AWS Textract, Google Vision, and Gemini GenAI for text extraction and contextual cleaning. A user-friendly frontend built with HTML and CSS provides seamless interaction with the system.

Backend Architecture

The backend is the core of the system, developed using the Python Flask framework. It handles the image preprocessing, text recognition, and text cleaning processes through an organized pipeline. The main components of the backend include:

1. **Image Preprocessing with OpenCV and Deep Learning:** The input images are first processed using a combination of OpenCV and deep learning models. OpenCV is used to mask the text regions and enhance the image quality. The preprocessing steps include:
    - **Histogram Equalization:** This technique improves the contrast of the image, making faint or unclear text more visible.
    - **Binary Filtering:** This step converts the image into a binary format, reducing noise and highlighting the text for better readability.
    - **Deskewing:** For use cases such as traffic camera images, deskewing algorithms are applied to correct image orientation and ensure proper alignment of text.

These preprocessing steps significantly improve the quality of the image, laying the foundation for accurate text recognition.

2. **OCR Models for Text Extraction:**
   Two powerful OCR models, AWS Textract and Google Vision API, are integrated to extract text from the processed images:
   - **AWS Textract:** Known for its depth in recognizing text from complex documents, AWS Textract provides comprehensive text recognition and is optimized for a wide range of languages.
   - **Google Vision API:** Offering high-speed recognition with support for an extensive set of languages, this API complements AWS Textract, ensuring broader compatibility and robust performance.

   Both OCR models work in tandem to extract raw text from the image. Symbols and ambiguous characters are also captured during this stage.

3. **Text Cleaning with Gemini GenAI API:**
   The extracted text, often noisy and symbol-heavy, is passed through the Gemini GenAI API for contextual cleaning and correction. Gemini GenAI leverages generative AI to:
   - Understand the semantic context of the text.
   - Remove unwanted symbols, errors, and artifacts introduced during the OCR process.
   - Produce a clean, meaningful final text output.

   This step ensures that the extracted text is accurate, readable, and suitable for real-world applications.

Frontend Architecture

The frontend of the system is built using HTML and CSS to provide an intuitive, user-friendly interface. It enables users to:

- Upload images directly for processing.
- Visualize the intermediate steps, such as the processed image and raw OCR output.
- View and download the final cleaned text output.

The frontend is designed to be simple yet effective, catering to both technical and non-technical users. Its responsiveness ensures accessibility across devices.

Use Case: Traffic Camera Images

One of the primary use cases for this system is processing images captured by traffic cameras. These images often suffer from issues such as skewed alignment, noise, and poor lighting conditions, which make text recognition challenging. The system addresses these challenges as follows:

- **Preprocessing for Deskewing and Noise Reduction:** The deskewing functionality corrects misaligned text in traffic signs or license plates, while binary filtering and histogram equalization enhance image quality.
- **OCR Recognition:** AWS Textract and Google Vision API extract text from traffic images, ensuring high accuracy even under adverse conditions.
- **Contextual Cleaning:** Gemini GenAI cleans and contextualizes the extracted text, producing error-free results suitable for automated systems, such as issuing tickets or traffic analysis.

Technologies Used

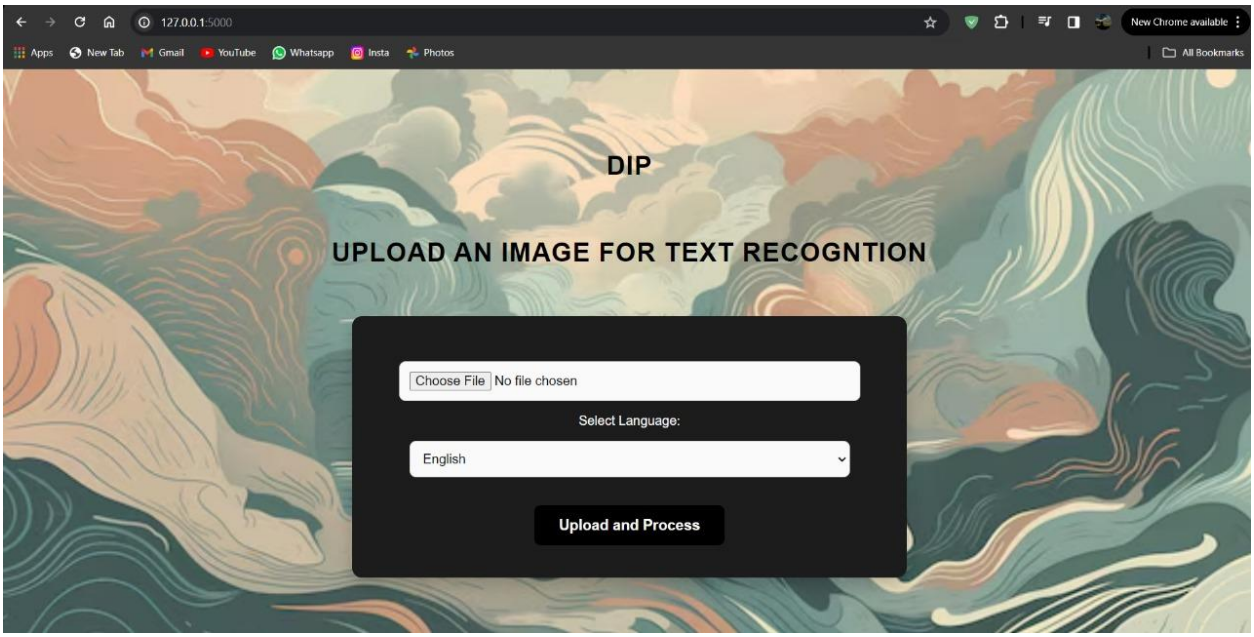
- **Backend Framework:** Python Flask
- **Image Processing:** OpenCV and deep learning models
- **OCR APIs:** AWS Textract, Google Vision API
- **Text Cleaning API:** Gemini GenAI API
- **Frontend Development:** HTML and CSS

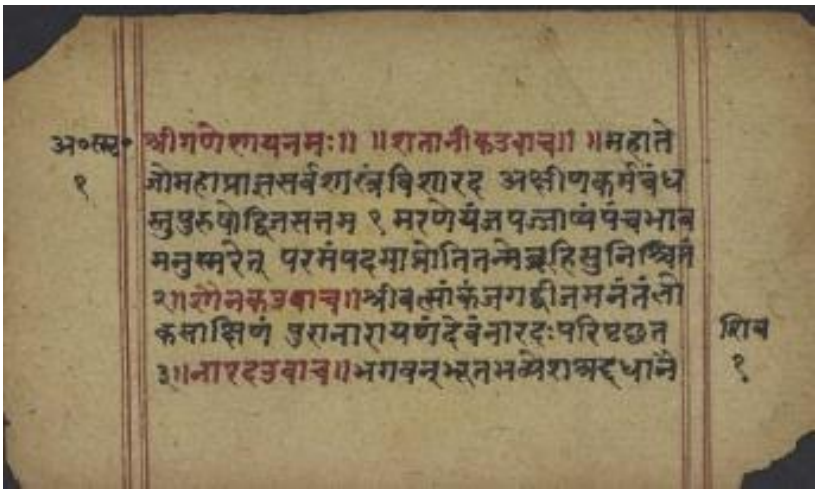
Key Features and Benefits

- **Advanced Image Preprocessing:** The system can handle low-quality images and optimize them for text extraction, ensuring high accuracy across use cases.
- **Multi-OCR Model Integration:** The combination of AWS Textract and Google Vision API ensures broader language support and accurate recognition.

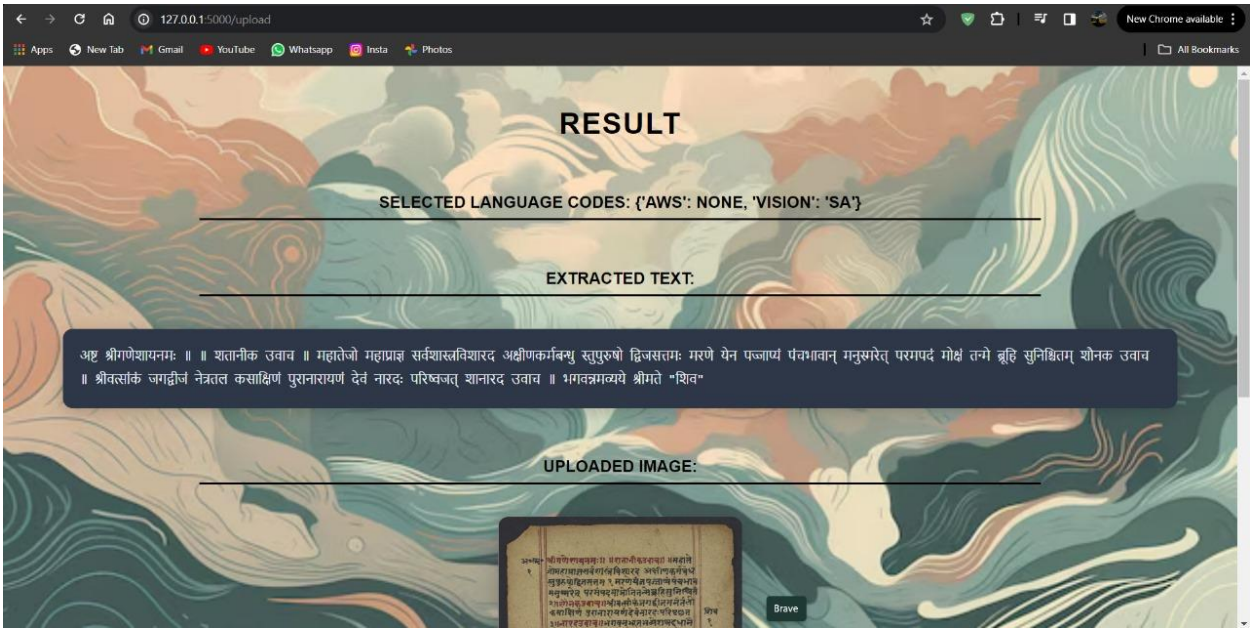
- **Contextual Text Cleaning:** Gemini GenAI enhances the extracted text by removing noise and understanding its context, delivering clean, meaningful output.
- **Scalability and Versatility:** The modular architecture can be adapted for various applications, such as digitizing documents, traffic monitoring, or archiving historical records.
- **User-Friendly Interface:** The frontend provides a seamless user experience for both technical and non-technical audiences.

Chapter 4
DESIGN





The above 2 images show the Front page of our application and the image we select for image processing.



These images depict how the application displays the image processing and the extracted text.

Chapter 5
CONCLUSION

In conclusion, this project successfully integrates advanced image processing techniques, AI-powered OCR models, and contextual text cleaning to create a robust and versatile image-to-text conversion system. By leveraging tools like OpenCV, AWS Textract, Google Vision API, and Gemini GenAI, the solution addresses challenges in handling complex and degraded images while ensuring high accuracy and readability of extracted text. Its modular architecture, combined with a user-friendly interface, makes it adaptable for a wide range of real-world applications, from traffic monitoring to document digitization.

This system not only bridges the gap between analog and digital text but also enhances the usability of previously inaccessible information. By automating labor-intensive processes and improving text accuracy, it stands as an innovative tool for both researchers and industries. With its scalability and efficiency, this solution demonstrates significant potential for streamlining workflows, improving accessibility, and preserving valuable textual content for future use.