Air Canvas: Real-Time Drawing and Text Recognition

Aamish Rafique, Fateh Aayan, Arooba Ali
November 25, 2023

Abstract

This paper presents a real-time hand gesture-based paint application using computer vision techniques. The system employs the *MediaPipe* hands module for hand tracking and recognition, allowing users to draw on a virtual canvas through hand movements. The application includes features such as color selection and clearing the canvas. A Transformer-Based Optical Character Recognition (TrOCR) is fine-tuned on custom handwritten equations for better interpretation of the drawn content. Moreover, the fine-tuned model is compared with state-of-the-art *Microsoft's TrOCR* on the same custom dataset.

Introduction

The development of gesture-controlled applications using hand tracking has gained prominence in human-computer interaction. This project integrates hand tracking and TrOCR to create a unique drawing application. The users can interact with a digital canvas using hand gestures to draw and/or write text. The application also displays the text drawn by the user using our fine-tuned TrOCR.

The primary objectives of this project include automatic data curation, implementing real-time hand tracking, integrating TrOCR to convert user drawings into text, and comparing our fine-tuned model's result with Microsoft's TrOCR.

Data Curation

A script was developed using Selenium in Python to automate retrieving images from *Calligrapher.ai*. It is a web-based tool that enables users to generate realistic computer-generated handwriting. The script was used to create a dataset of handwritten equations, such as *one plus two is three* and *three minus two is one*. The total number of images curated was 1500+. The images were of the same size and format to ensure smooth fine-tuning of the TrOCR model.

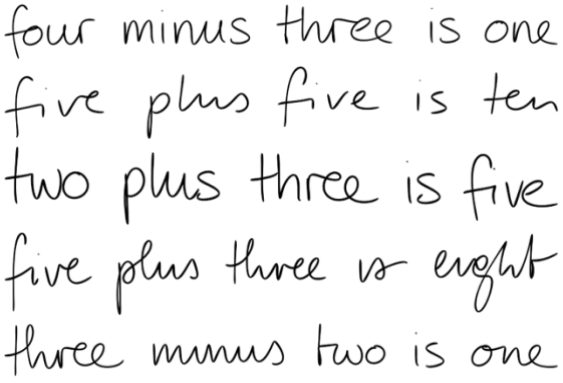
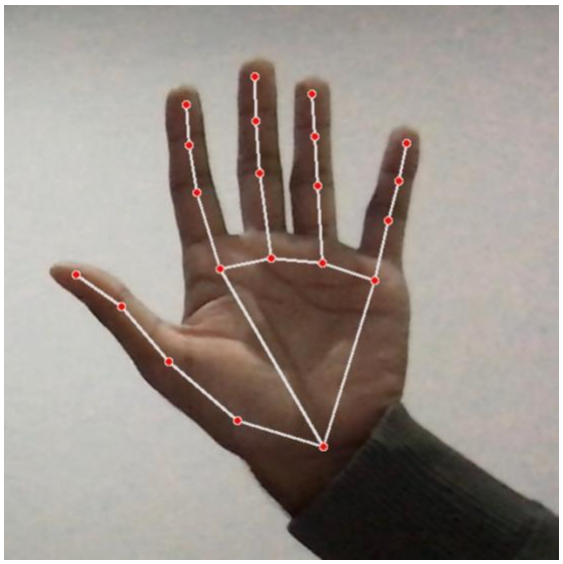


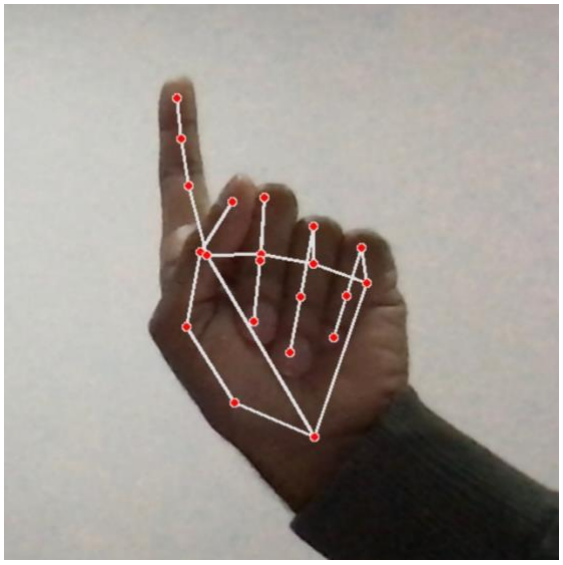
Figure 1: A Snapshot of the Custom Dataset.

Methodology

1. Hand Tracking: The hand tracking is implemented using the *MediaPipe* library, which provides a robust solution for detecting and tracking hand landmarks in real time. The application captures video frames from the webcam, converts them to

RGB format, and processes them using the *MediaPipe* hands module to detect and obtain the coordinates of hand landmarks. The landmark at index 8 (the tip of the index finger) is used as the marker position.





Figures 2 and 3: Hand Landmark Detection and Tracking in Real-Time Using MediaPipe.

2. Canvas and Color Selection: A digital canvas is created as an array of pixels and different color options are provided for drawing. The application allows users to select colors by using hand gestures, enhancing the interactive drawing experience. Similarly, the users can also choose to clear the canvas. To enhance the drawing experience, the application employs morphological operations with a predefined kernel. These operations help smooth the drawn lines on the canvas.

3. TrOCR Integration: The TrOCR is fine-tuned on handwritten equations for accurately interpreting the user drawings. The TrOCR model is integrated into the application to interpret user drawings. The canvas undergoes processing, transforming the visual content created by the user into a textual representation. The resulting text is promptly displayed to the user, offering a comprehensive overview of their creative expressions within the virtual canvas.

Model Evaluation and Comparison

In a comparative analysis between Microsoft's TrOCR model and our custom fine-tuned model, the Character Error Rate (CER) values reveal notable distinctions. Microsoft's TrOCR achieved a commendable CER of 0.0167, signifying a high level of accuracy in recognizing characters during the text recognition process. However, our custom fine-tuned model surpassed this performance, boasting an even lower CER of 0.0014 in only 10 epochs of training. This lower CER underscores the superior character recognition capabilities of the custom model, suggesting that it excels in accurately transcribing text. The results indicate that the fine-tuning process and adjustments made to the model have significantly enhanced its accuracy

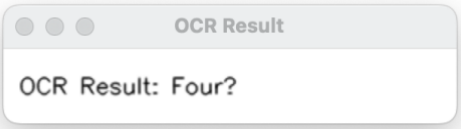
compared to the pre-trained Microsoft's TrOCR model.

Model	CER
Microsoft's TrOCR	0.0167
Custom Fine-Tuned	0.0014

Table 1: Comparison of Character Error Rate (CER) between Microsoft's TrOCR and Custom Fine-tuned Model.

Result and Output

The system successfully allows users to create drawings on a virtual canvas using hand gestures. The color selection mechanism and canvas clearing gestures are intuitive, providing a seamless user experience. The TrOCR feature enables the system to interpret the drawn content, adding a layer of functionality to the paint application.



Figures 4, 5, and 6: Implementation of the Application with Outputs.

Real-Life Applications

1. A notes-making application with the ability to identify the meaning of words.
2. An application that can solve mathematical equations.
3. A learning platform for individuals who cannot speak or use/navigate a computer.

References

Liuberskis, R. (2023, March 20). *Handwriting words recognition with PyTorch*. Retrieved from PyLessons: https://pylessons.com/handwriting-recognition-pytorch#

Hoffstaetter, S. (2022, August 17). *pytesseract 0.3.10*. Retrieved from The Python Package Index: https://pypi.org/project/pytesseract/

Rogge, N. (2023, June 16). *Transformers-Tutorials*. Retrieved from GitHub: https://github.com/NielsRogge/Transformers-Tutorials/tree/master/TrOCR