

Media Practices: AI American Sign Language Translator

Sejal Nimbargi
MS Media Management
Email: nimbs720@newschool.edu

Concept:

This is a two-way communication model that translates Text-to-ASL and ASL-to-Text. This project aims to enhance user interaction with media content, making it more accessible and inclusive. I hoped to achieve a basic understanding of building a model from scratch, from research to implementation, and understanding the fundamentals of integrating coding with an AI tool.

Process:

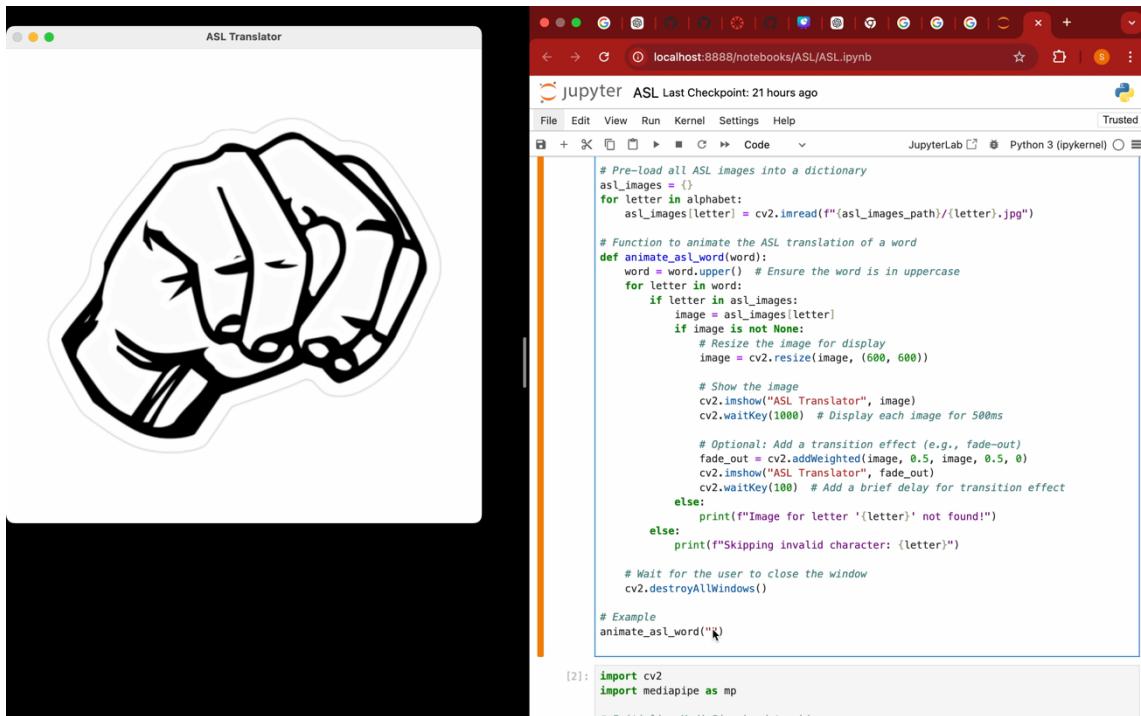
Part 1: Creating a Text-to-ASL model.

Since ASL is such a vast language that not only has alphabets but also gestures and expressions, I have only implemented this model with ASL Alphabets for now, considering my limited coding background and time constraints.

Starting with static sketches of ASL letters, I recognised each image of the letter based on its respective alphabet, e.g., A.jpg, B.jpg, C.jpg, etc.



After that, I coded the Text-to-ASL by using the ASL letters, which translates text by sequentially outputting each letter to form the word. I added slight animations to create smooth transitions from one letter to the next.



Part 2: ASL-to-Text model

Initial Plan: I tried using existing models such as and But these are fairly huge datasets and sadly, these were far more complicated for me to integrate into this model.

Initial Plan: I initially tried using existing models such as https://huggingface.co/Niharmahesh/Sign_language_recognition_v1 and <https://github.com/harshbg/Sign-Language-Interpreter-using-Deep-Learning>, but these datasets were fairly large and far too complicated for me to integrate into this model.

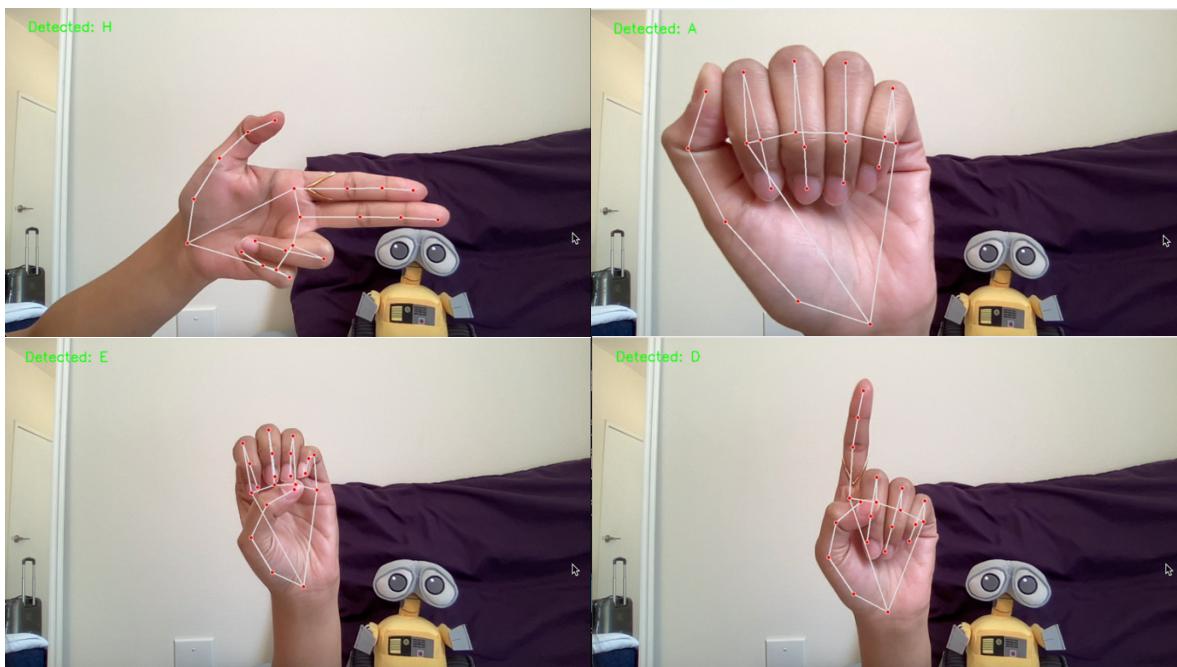
New Approach

I decided to apply the same approach as I did with the Text-to-ASL model. This is where I used MediaPipe, an AI tool used for hand tracking. MediaPipe creates 21 nodes for each hand gesture, and I decided to use this to train my own model.

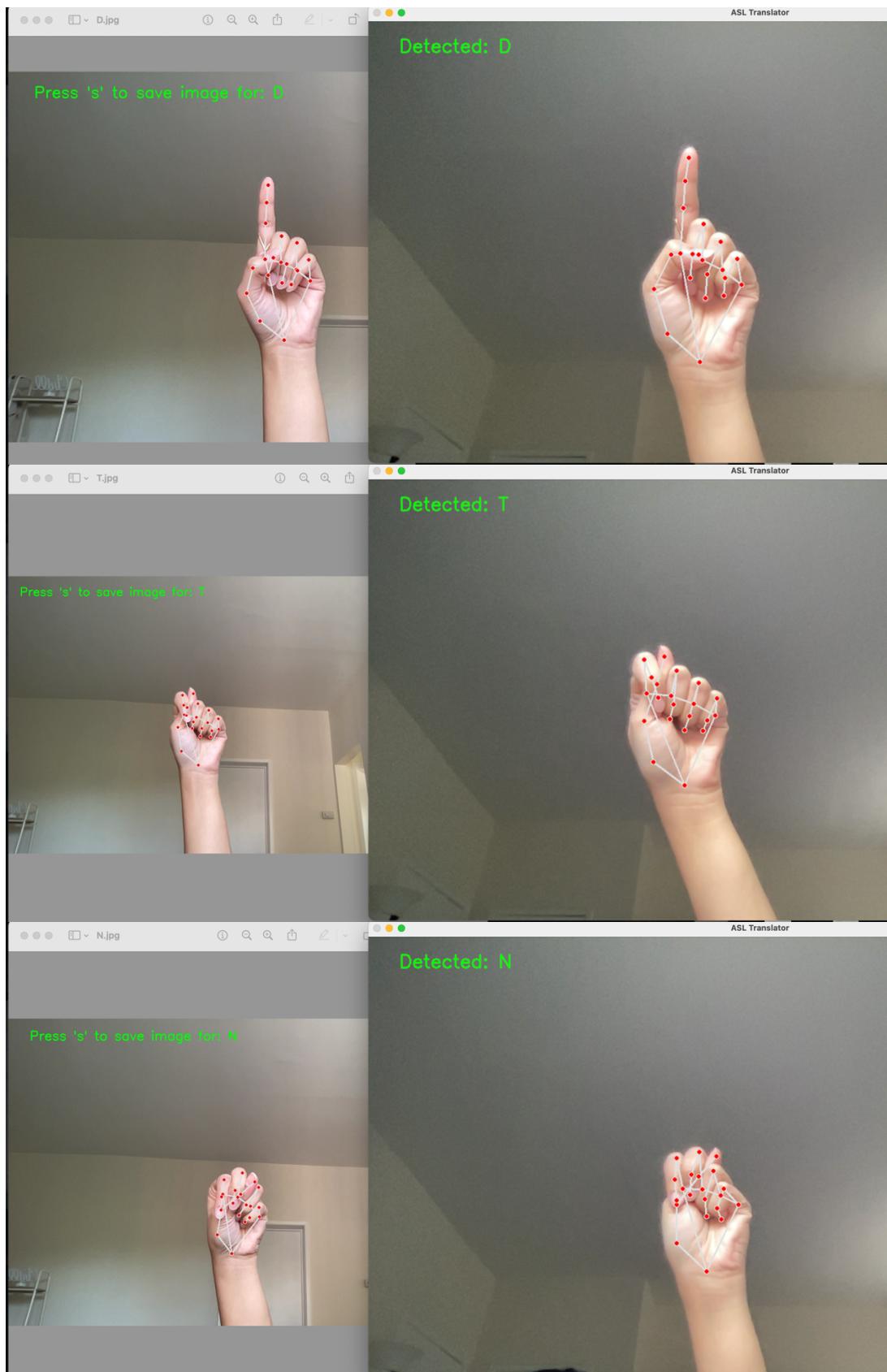
I started by capturing images for each ASL letter and naming them according to their respective letters. It took a considerable amount of time to figure out how to create a file and develop a script that would prompt me to do the gesture and automatically save images for each letter. Fortunately, it worked, and I successfully created a small dataset of ASL images, which I used to train the model.



Finally, I can now successfully use MediaPipe for live-hand gesture recognition based on the trained dataset. (Detected Letter in green on top left corner)



In the following screenshots, the left side shows a part of the dataset for the respective letter, while the right side displays the live camera feed with real-time letter detection.



Outcomes and Aspirations:

I successfully created a two-way communication model with a Text-to-ASL translator and a live letter detection model. Initially, I only planned to create a Text-to-ASL translator, but during my research, I discovered MediaPipe and realised it was a great way to add live hand tracking and letter recognition, transforming it into a two-way communication model.

I had initially planned to include gestures using pre-trained models and will continue working on this until it is complete. A possible spin-off of this project is creating gesture-based AI control systems for AR/VR experiences, which would enhance overall user interaction in media consumption and make it more inclusive.

I also aim to create a website to host this model. I successfully created a virtual environment and wrote the HTML but faced persistent errors related to version incompatibility. I plan to continue working on this over the break.

Learning:

I have gained a comprehensive understanding of how to start and manage a coding project from research to implementation, despite having zero background in coding. I also learned how to integrate AI into media to enhance user experience and recognised the significance of accessibility features for specially-abled communities.

Ethical Considerations:

This project's goal is to make media consumption and user experience more accessible for the deaf and Hard of Hearing communities. However, I must acknowledge the limitations of this model in its current state, as it has only been trained on alphabets rather than the full scope of the language. Considering this is a work in progress and that inclusivity is at the core of this project, I recognize the importance of ensuring that the model does not oversimplify or misrepresent ASL. Therefore, I will be incorporating the gestures, expressions and cultural nuances that the ASL language encompasses.

Another major ethical consideration is data privacy, as this model uses real-time video input using an AI tool and this becomes an importance factor if the model were to go live. Such tools could be misused on a large scale, making it the developer's responsibility to consider implementing safeguards to protect user's data.

I also understand the importance of collaboration with ASL experts, and members of the community to ensure the project's usability and alignment with their needs.