Computer-Vision-Powered American Sign Language Alphabet Translator Using a Convolutional Neural Network

Brian G. Rodiles Delgado

The University of Texas at El Paso

AI4ALL Discover AI Fall 2022

Introduction and Motivation

Artificial Intelligence is a tool that has the potential to create a more inclusive world.

5%

Of the world’s population experience disabling hearing loss

700 mill

Number of people that will experience disabling hearing loss by 2050

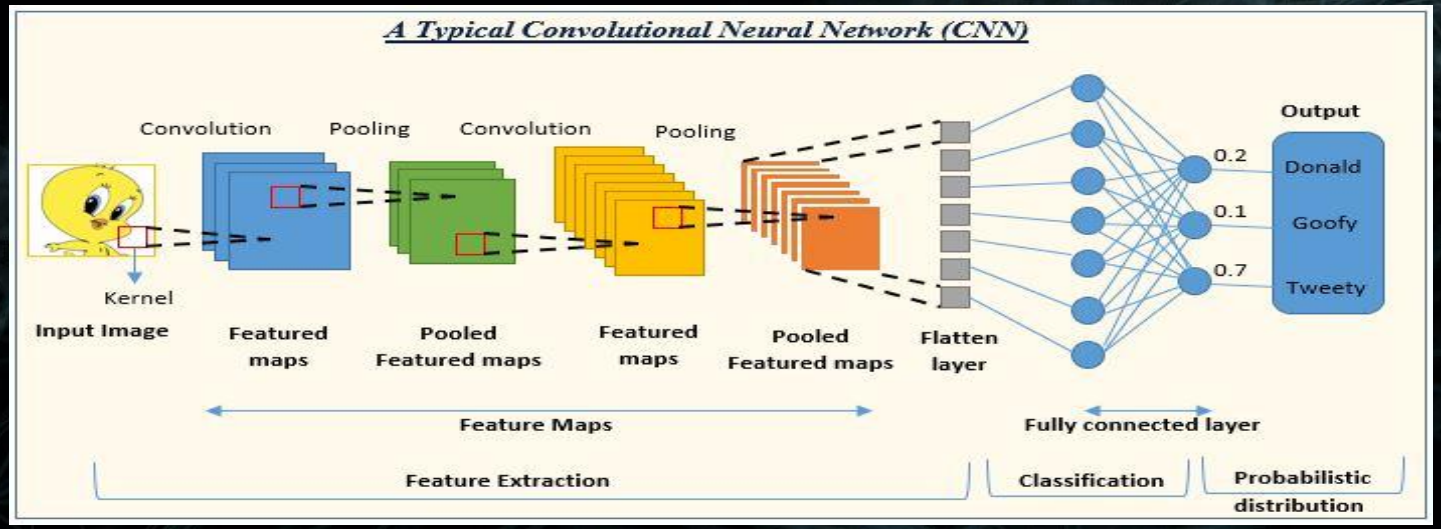
Introduction and Motivation

- Computer-vision-powered ASL alphabet translator that is able to classify hand gestures used for fingerspelling in sign language

- Potential to remove communication obstacles between those who are hard of hearing and those who wish to communicate with them

Background (Convolutional Neural Network)

- CNN is an image analysis neural network that uses filters to recognize patterns within the image
- Kernel Convolution
- Deep Learning

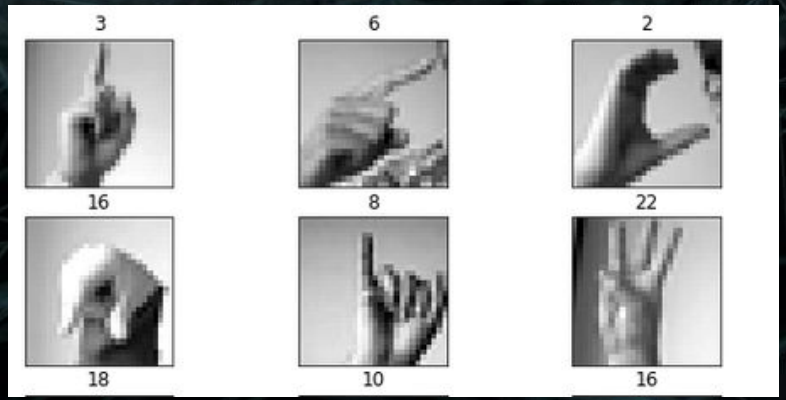


Background (CNN - Kernel Convolution)

- Convolutional layers of the neural network
- Mostly used in computer vision
- Pattern detector “filter” functionality
- Dot product of a small kernel matrix (filter) by a pixel matrix from the input (image)
- Gaussian blurs, mean blurs (Photoshop)

Background (CNN) Recognizing Shapes

- Convert pixel data of input images from 0-255 range to 0-1 (color to a grey scale).
- Reshaping input images to 28 x 28 pixels.
- Performing Kernel Convolution operations to filter input images and create convolution layers (BatchNormalization, Dropout).
- Deep learning shape label assignment (0 - 24 letters of the alphabet, omitting j and z as they require motion).



Background (TensorFlow)

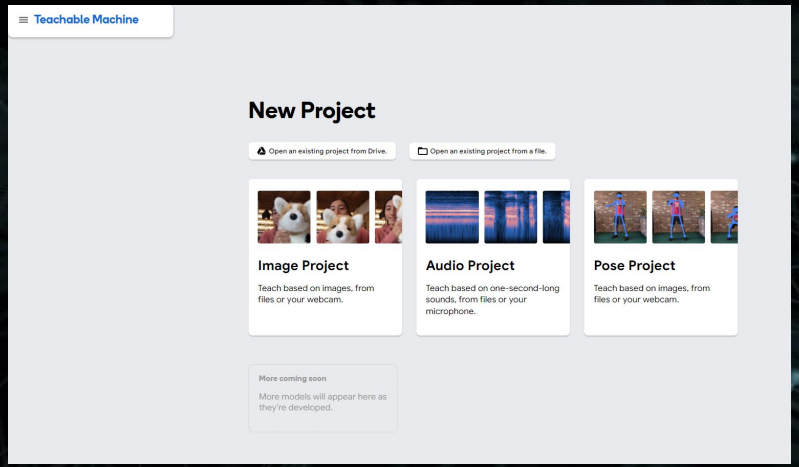
- Tensorflow (TF) was developed by Google Brain for Deep Learning applications in research.
- It has largely expanded since its inception and now includes optimizers such as Adagrad and SGD. Along with a plethora of Neural Network operations like Softmax, Relu and Max-Pooling.

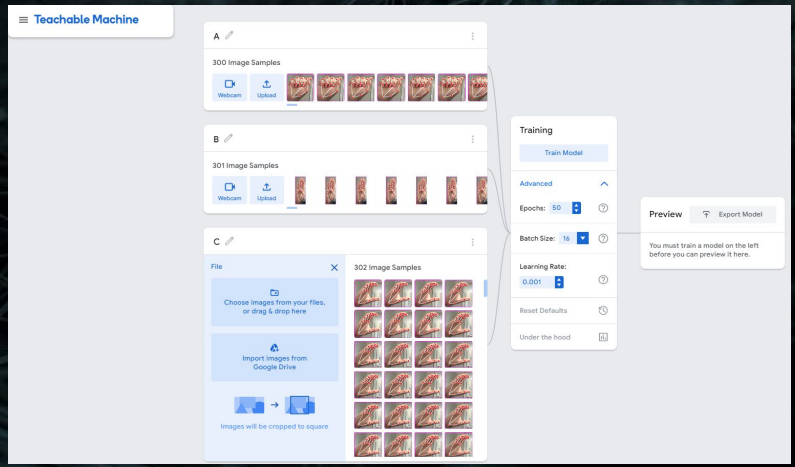
- Additionally, numpy is commonly utilized with TF in order to compute the matrix/mathematical operations needed for models.

Background (Teachable Machine)

- The project needs a classification model to differentiate every letter in the ASL alphabet when doing the live showcase.

- Google’s Teachable Machine allows for training a model that uses a Convolutional Neural Network with a TensorFlow.js library.





Methodology

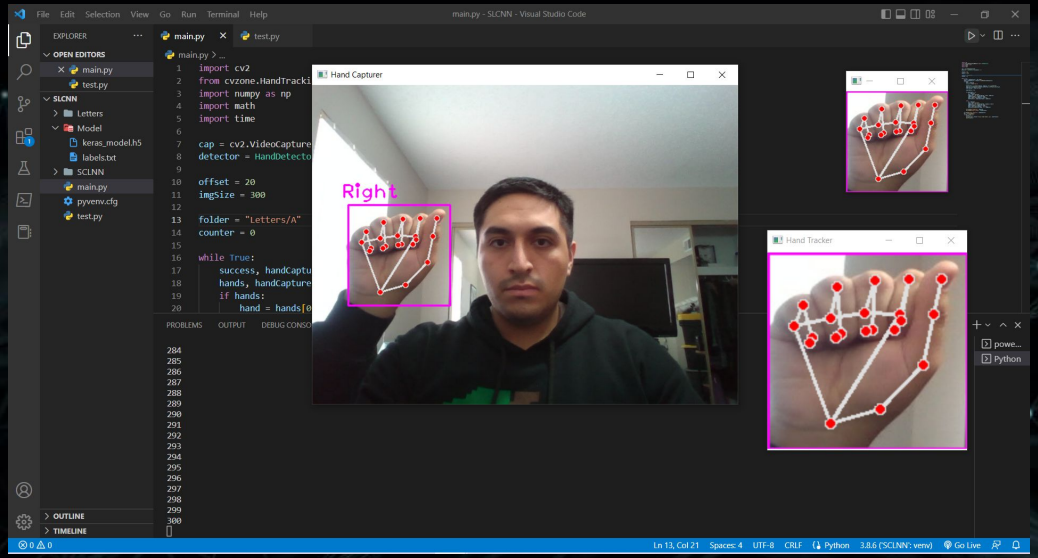
- Develop a program to capture hundreds of images using hand recognition for the representation of ASL letters.

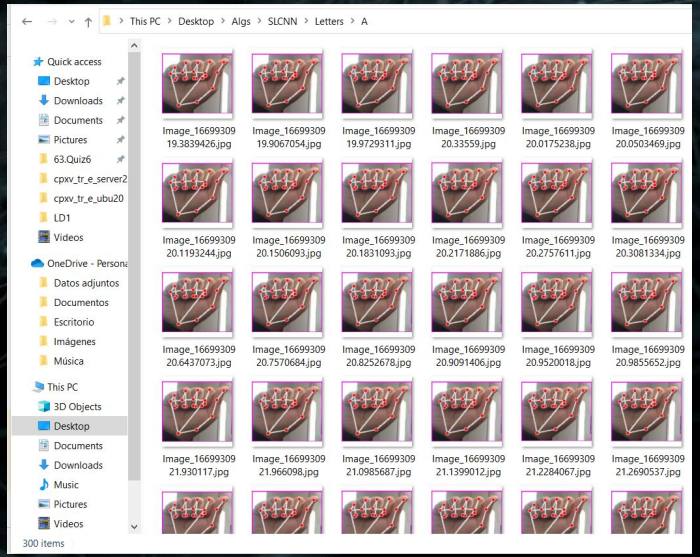
- Train a Tensorflow model that uses a Convolutional Neural Network (CNN) to classify the characters present in the alphabet.

- Develop a second program to test the obtained model in a live environment to showcase the correct recognition of letters and the prediction percentages in the terminal to know the model’s efficiency.

Results (Hand Capturer)

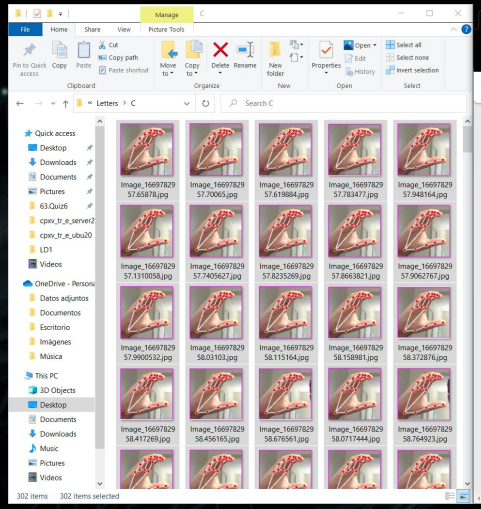
The first program captures hundreds of hand screenshots while using a hand tracking system. The precedent goal was accomplished by using Python and the OpenCV, NumPy, and MediaPipe libraries.

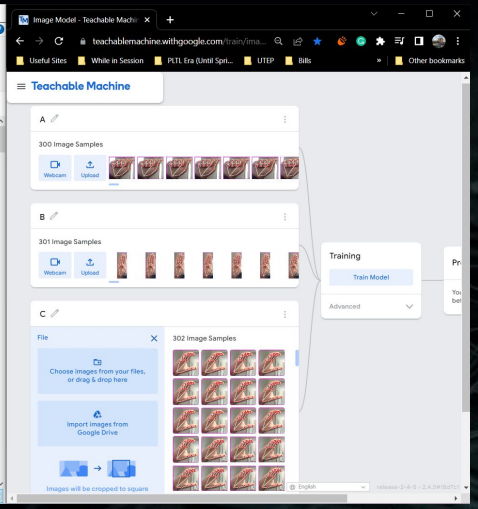


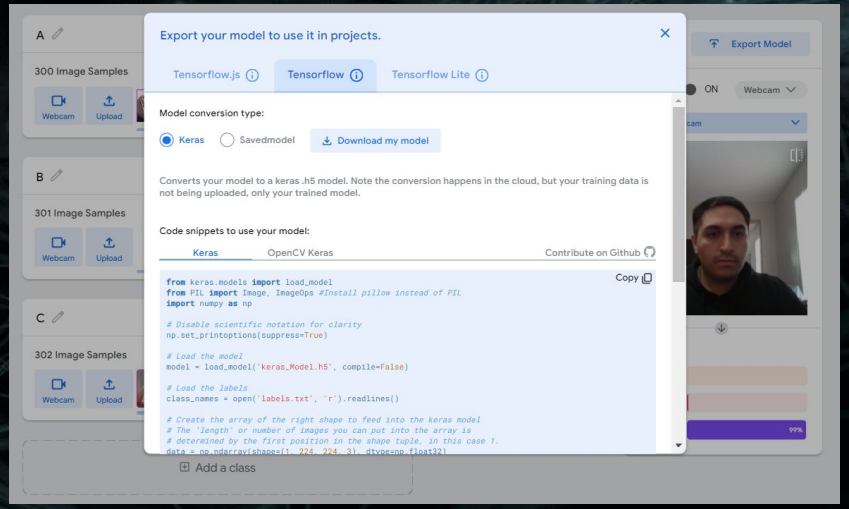


Results (Model with CNN)

The next step is to train a Machine Learning model that uses a Convolutional Neural Network as the tool that will allow us to classify every hand position with an ASL letter in the alphabet. Then, the model will be exported as a

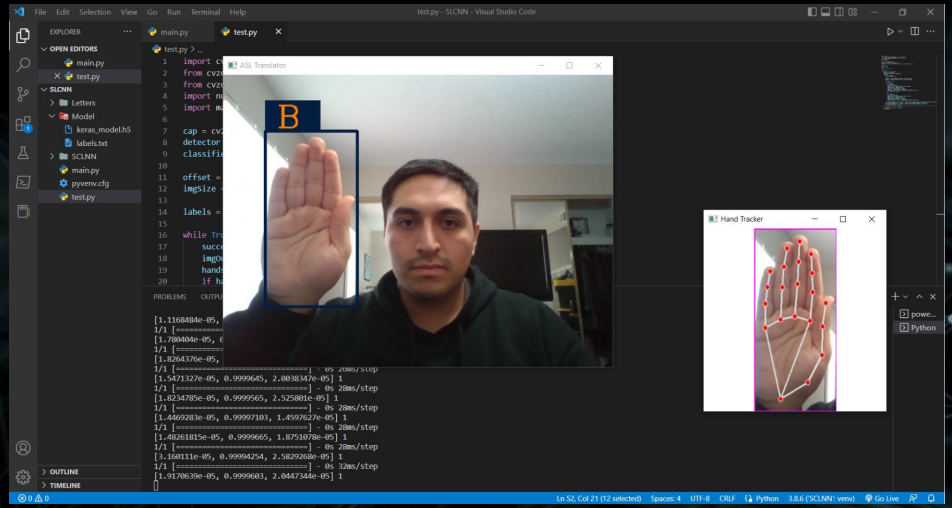


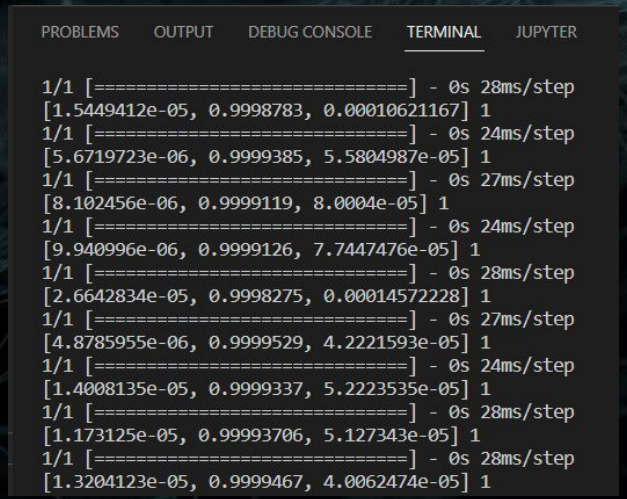




Results (ASL Translator)

Finally, the second Python program uses the trained model to track a hand and identify the ASL letter represented by it. The prediction value is displayed in the terminal, which helps in the assessment of the model's efficiency.





Conclusions and Future Work

- Using Python and its Machine-Learning-focused libraries allows for the rapid creation of Artificial Intelligence implementations.

- Applying a model that is pre-built with algorithmic and training capabilities is fast, easy, and accessible with web-based tools like Google’s Teachable Machine.

- In the future, the model can be improved for it to have a better accuracy when predicting values. However, the use of a CNN of our own would be of need to have more freedom when manipulating the training algorithm’s parameters.

References

"Deafness and Hearing Loss." World Health Organization, World Health Organization, https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss.

Hensley, Priscilla, and Michael Hansen. "Students with Hearing Loss Need Individualized Support, Access to Deaf Community." Brookings, Brookings, 18 Oct. 2022, https://www.brookings.edu/blog/brown-center-chalkboard/2022/10/06/students-with-hearing-loss-need-individualized-support-access-to-deaf-community/.

Priester, Jarred. "Sign Language MNIST CNN, Accuracy 100%." Kaggle, Kaggle, 11 Sept. 2022, https://www.kaggle.com/code/jarredpriester/sign-language-mnist-cnn-accuracy-100/notebook.

Taylor, Brandon, et. al., "Real-Time Depth-Camera Based Hand Tracking for ASL Recognition." Association for Computing Machinery, Association of Computing Machinery, 19 Oct. 2017, https://doi.org/10.1145/3132525.3134777.

Jirathampradub, Hassaluk, et al., "A 3D-CNN Siamese Network for Motion Gesture Sign Language Alphabets Recognition." Association for Computing Machinery, Association for Computing Machinery. https://doi.org/10.1145/3406601.3406634.

Acknowledgements

This project was possible thanks to the mentorship and guidance of Marilu Duque and Ricardo Godoy. We also thank the AI4All program and The University of Texas at El Paso for the opportunity to discover AI, ML, and be part of this project.