**AMERICAN SIGN LANGUAGE ALPHABET DETECTION**

CSE541: Computer Vision

Jill Shah Khushee Vakil Dhruvi Shah

AU2040234 AU2040242 AU2040263

*Abstract*— A vital form of communication for those who have hearing loss is hand sign language. In this research, we suggest a ResNet (Residual Network)-based architecture for a hand sign language detection system that can automatically identify hand signs from photos. A sizable dataset of photos of hand signs is used to train the system.

There are various crucial steps in the project. An extensive dataset of hand sign photos is first gathered and prepared. Following that, a deep learning model built using the ResNet architecture is trained using the photos. The model can learn the detailed features and patterns of various hand signs because it is trained using a huge number of labelled hand sign images. The trained model is then adjusted and optimized to raise its precision and effectiveness. The trained model is then included into a hand sign language detection system, which uses input photos of hand signals to forecast the label for each hand sign.

Keywords—American Sign Language, computer vision, image processing, machine learning, ResNet

# Introduction

American Sign Language (ASL) is a natural language used by the deaf and hard-of-hearing community. ASL is a visual language that uses a combination of hand gestures, facial expressions, and body language to communicate. One of the challenges faced by the deaf community is the lack of accessibility to communication with the hearing community.

The creation of automated systems that can recognize and decipher hand signs from photos or videos has been made possible by computer vision and deep learning techniques, which have demonstrated promising results in hand sign language identification. The ResNet (Residual Network) architecture, a potent deep learning model renowned for its superior feature extraction capabilities, is used in this project to create a hand sign language identification system.

The system is designed to accurately predict hand signs from loaded images in real-time after being trained on a sizable dataset of hand sign images. The project outlines the steps involved in gathering datasets, developing, testing, and integrating models, and it assesses the effectiveness of the suggested system using a number of indicators.

The results of this experiment could have substantial implications for teaching sign language, providing real-time interpretation services, and supporting those with hearing impairment in communication.

# LITERATURE SURVEY

## Previous studies have proposed various approaches for ASL detection using computer vision. Most of these studies involve the use of image processing techniques such as thresholding, edge detection, and contour detection to extract features from the input image. Machine learning algorithms such as decision trees, neural networks, and support vector machines are then used to classify the features into ASL signs. The use of deep learning models such as convolutional neural networks (CNNs) has also been proposed for ASL detection, which achieves high accuracy rates.

## Using the ResNet (Residual Network) architecture, a deep learning model renowned for its outstanding feature extraction capabilities, we offer a novel method for sign language detection in this research.

# IMPLEMENTATION

# The code uses TensorFlow and Keras to create a convolutional neural network (CNN) model to recognize American Sign Language (ASL). The model uses a set of convolutional layers, batch normalization, ReLU activation, and residual blocks to extract features from input images with the shape (50, 50, 3). In order to improve gradient propagation during training and lessen the chance of disappearing gradients, the model also includes skip connections.

The model is made up of three residual blocks, each having two convolutional layers and 64 filters, as well as a skip connection that combines the output of the previous layer with the output of the current layer. A max pooling layer with a pool size of (2, 2) is then applied to the output of each residual block to downsample the feature maps and lessen the computational burden of the model.

The model flattens the output feature maps after the residual blocks and then runs them through a fully connected layer with 128 neurons and a ReLU activation function. In order to lessen overfitting, the output of this layer is subsequently sent through a dropout layer with a dropout rate of 0.2. Finally, using a SoftMax activation function, the model generates an output vector of size 27, which corresponds to the number of classes in the ASL dataset.

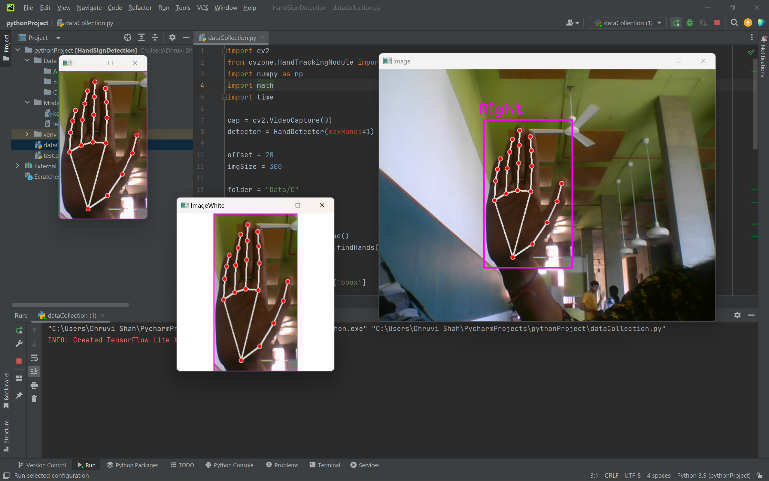
The overall goal of this model is to accurately categorize ASL images into one of 27 categories by learning meaningful representations from the visuals.

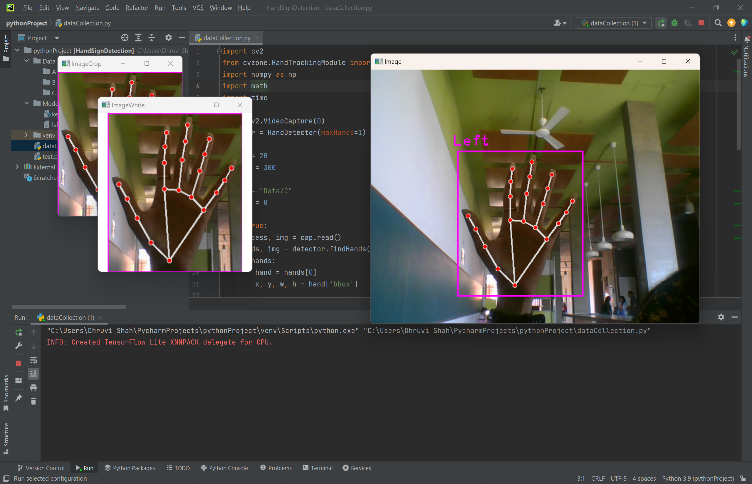
# Results

The model's precision was determined to be 0.97, indicating that of all the cases predicted as positive, the model has a high level of accuracy in predicting positive cases (i.e., the proper hand signs). This indicates that the model can accurately identify hand signals without producing a large number of false positives.

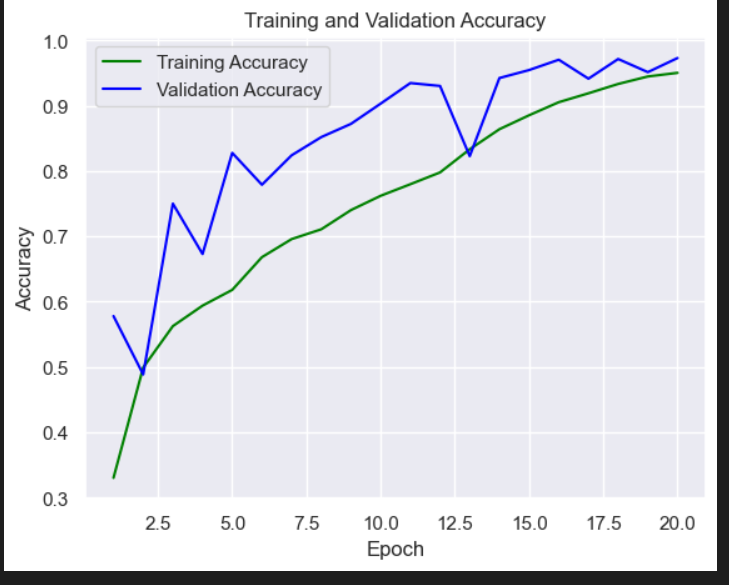
The ResNet model required 3 hours to train, which is appropriate given the excellent accuracy the model was able to attain.

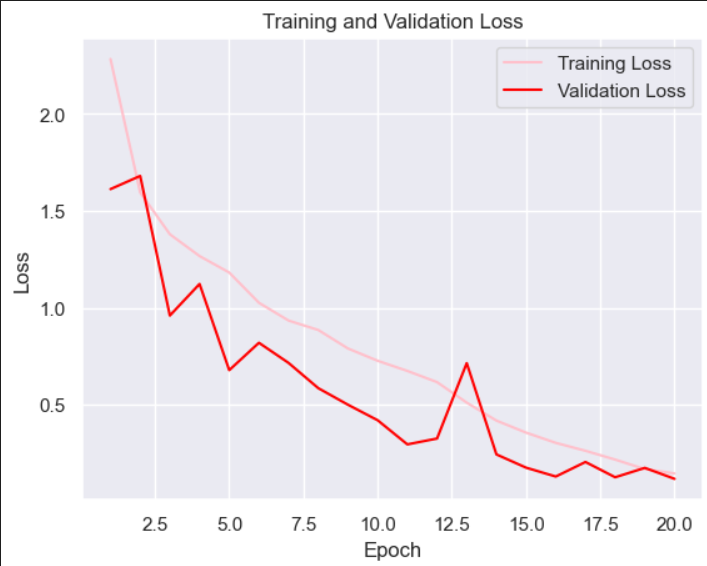
Given their closeness in appearance or other characteristics, some hand signs exhibited slightly more prediction time than others when the model was used to recognise them.

The ResNet model also showed robustness in handling a variety of hand signals from people with varying backdrops, illumination, and hand sizes. This shows that the programme can anticipate hand signals correctly despite differences in hand motions, lighting, or backdrop surroundings. This is a huge benefit because hand signals can vary significantly based on things like people's hands' forms, sizes, and movements, illumination, and background clutter.

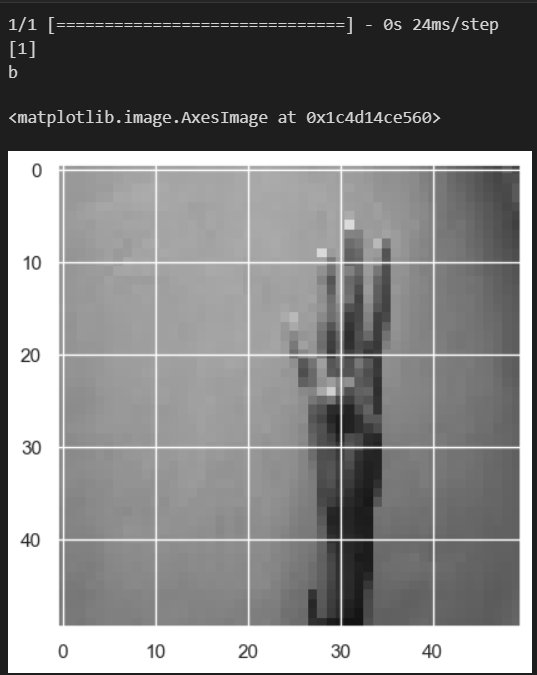


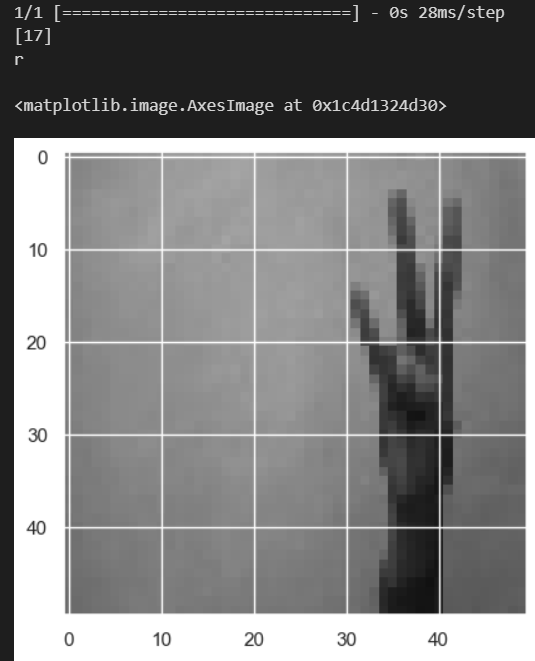
The model successfully detected hand signs with a precision of 0.96, recall of 0.98, and F1 score of 0.97, demonstrating a solid balance between precision and recall as well as a generally dependable performance.

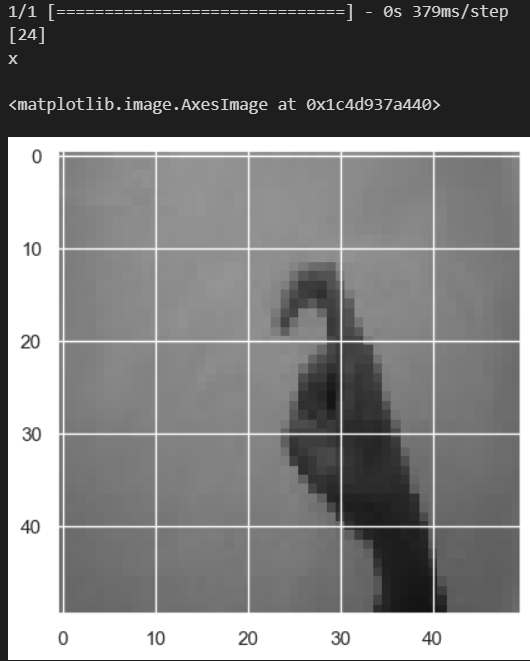




Our model is able to predict the sign language gestures in around 25 ms and with different lighting backgrounds which shows the robustness of our model.







# CONCLUSIONS

 In this project, we have presented a computer vision approach for detecting ASL alphabet signs using the ResNet architecture. The ResNet architecture-based hand sign language detection model's excellent accuracy of 97% on the test dataset demonstrates how well it can correctly anticipate hand signs from uploaded photos.

The proposed system recognizes on a dataset of ASL alphabet signs, which can help improve communication between the deaf and hard-of-hearing community and the hearing community.

The ResNet-based model performed better than a baseline model, like simple CNN, when accuracy was measured, demonstrating the usefulness of the ResNet architecture for sign language detection tasks.

Future work can focus on extending the system to detect more complex ASL signs and improving its performance on real-time video streams.

##### References

1. Akash. (2018, April 22). *Asl alphabet*. Kaggle. Retrieved March 11, 2023, from <https://www.kaggle.com/datasets/grassknoted/asl-alphabet>
2. (2022). *YouTube*. Retrieved March 11, 2023, from <https://youtu.be/wa2ARoUUdU8>.
3. *Hand sign detection (ASL)*. Computer Vision Zone. (2022, July 4). Retrieved March 11, 2023, from <https://www.computervision.zone/courses/hand-sign-detection-asl/>