# 

# **Final Report**

# 

# **Team Members: Shlok Jethwa, Purvi Kakkasagar, Ankita Mittal, Harsh Sharma**

# 

# **ASL Gesture Recognition**

# 

# **BUDT758B(Big Data & AI)**

# 

# **December 18th, 2020**

# 

# **Fall 2020**

**Introduction**

ASL(American Sign Language) has been around for quite some time now and started with the most obvious use - to help the voiceless and those who have impaired hearing to help communicate with other people. It can be called a standardized method of sign language across the USA and Canada.

Today’s world is driven by technology; there is so much more we can do with ASL if we combine it with the advanced science and technology that we currently have. This project aims to create a working model of American Sign Language/Gesture Detection. We also want to highlight some practical applications where this model can improve the quality of life of the deaf and hard-of-hearing and help them communicate better with community members.

There is a lot of research going on to develop an effective gesture recognition system. In the gesture recognition system, hand gesture recognition is the most researched area. In this paper, we will be narrowing our view to American sign language (ASL) recognition. Sign Language (SL) is the only communication method between the hearing-impaired community and the shared community. Sign Language Recognition (SLR) system, which is necessary to recognize SL, has been widely studied for years.

The goal of recognising the ASL is to interpret human gestures via mathematical algorithms. ASL gesture originates from the finger position of the hand. Here we focus on Convolution Neural Network (CNN) based ASL recognition. Human-computer Interaction(HCI) is a multidisciplinary field. It focuses on designing and developing computer technology, which is used for interaction between computers and humans. Humans talk to each other with their mouths or through sign language.

The HCI-dependent hand movement is the most intuitive way to connect humans and computers. Identifying hand gestures and what they mean is an acceptable form of HCI and is made using different methods. HCI based gesture recognition has tremendous scope for growth. Here we are proposing gesture recognition using Convolutional Neural Networks(CNN) for American Sign Language(ASL).

CNN is the most efficient technique that we can use, involving feature extraction and classification. CNN provides image recognition with more accuracy in comparison with other methods. Our approach involves implementing concepts like back projection, contour generation, binary threshold, and segmentation during preprocessing, which will aid the feature extraction process. The user’s actions will be collected using a live web camera, and the corresponding sentence will be displayed on the screen.

**Background**

There are around 500 million people in the world who are suffering from hearing and speaking disabilities. These people communicate using lip-reading, gestures, or facial expressions. However, COVID-19 pandemic has proved to be a significant problem with everyone wearing masks or turning to video communication. Besides, sign language is not common in our society. We are hoping to use this project to help them communicate better through the help of human-computer interaction.

There are some obvious challenges that would come our way:

* Sign language is not a standard language; many versions can cause communication difficulties between 2 globally separated people.
* Accurately separating one-handed apart from two-handed signs.
* We might not always get high-resolution images, which can impact accuracy.
* CNN model on images would require a lot of memory, and we would train them as much as possible on our laptops

**Methodology**

The most effective deep learning model is Convolution Neural Network(CNN) to detect the gestures from the images. CNN can do much better than the classic neural networks while dealing with images because it takes advantage of images' inherent properties. A simple feedforward neural network doesn't see any order in inputs, but CNN takes advantage of the images' local spatial coherence.

CNN is fast and decreases the number of operations needed to process the image using convolution patches of adjacent pixels. We propose a gesture recognition method using convolution neural networks (CNN) for American Sign Language (ASL), the most efficient feature extraction and classification technique.

Image-based gesture recognition uses CNN because it provides more accuracy than others, and CNN's are very good feature extractors. The procedure involves the application of Back Projection, contour generation, Binary Threshold, and segmentation during preprocessing, in which they contribute to better feature extraction. Our model recognises the class of our gesture (from 0-43 class folders). After recognizing the class, our model will look for the gesture corresponding to the class id in the gesture db. The results are then printed on the screen.

**Data Description**

The Data we gathered is labeled data. We had 50\*50 pixels grayscale images. We stored the images of 44 gestures (in 0-43 folder names). Each folder has 2400 images. The images are 26 letters (A-Z) and ten numbers (0-9), and seven other gestures, which we kept empty for future use. We trained the model on these images 80% (approx.), validated on 10% (approx.), and tested on the remaining 10% (approx.) images.

**Approach**

The images we have are 50\*50 grayscale images. First, we load the images using pickle. Pickle is used to convert Python objects to byte streams. After loading, we shuffled the data, and we divided the data into training data, validating data and testing data.

Then we created our CNN network. Our CNN network has three convolution layers and three max-pooling layers. After this, we shall flatten the image into a column vector. The flattened output is fed to a feed-forward neural network, and backpropagation is applied to each coaching iteration. After probing the above process, we've successfully enabled the model to grasp the features. Our model has a batch size of 500 and 10 epochs. We trained 80% of the data on this model. After that, we ran our model on a test dataset. We got an accuracy of 99%.

After this, our model is ready for capturing live images frame by frame and predicting the gesture. We used the OpenCV library for computer vision. We have to pre-process the image for each frame to extract only the image’s essential features and remove noise. The basic idea behind all this method is to concentrate on hand features, regardless of background and intensity. Later these images are fed to CNN to get the result/ labels of the hand gestures.

**Results**

Our image processing methodology converts the images into a greyscale. A database consists of 43 gestures, and for each gesture, we have 2400 images. Our solution to convert ASL gestures to text involves a real-time vision-based hand gesture recognition system. The Convolution Neural Network (CNN) architecture achieved high success rates at a relatively low computational cost, and we achieved an accuracy of 99%. This shows that our model is well trained and robust. The architecture we proposed is comparable to the current industry standards. Even though this is relatively simpler and has lower computational costs, it can be scaled to a bigger scale as per need.

**Practical Applications**

Taking the present situation into perspective, this pandemic has forced everyone to approach and tackle different light problems. Everything has moved to a virtual setup - education, management, daily jobs and even healthcare. It brought perspective on how even areas that seemed impossible/impractical to shift online can survive, shifting to the virtual world.

There have been instances where those who have impaired hearing have had trouble communicating with first responders when in need. A system like this can help people who are not familiar with ASL communicate and get them the help they need. This can maybe be incorporated in a mobile application since everyone always has their phones on them.

It can help the impaired communicate effectively with healthcare providers/educators in the virtual setup where the system can be a plug-in for video conferencing applications.

If we reach accuracy and speed where gestures can be recognized in real time without lag, this can be super beneficial and will have a much higher user acceptance rate as people will find more use to this technology.

**Appendix**

The code and database file can be found at the below URL.

<https://github.com/shlokjethwa/ASL-Recognition>