National College of Ireland

American Sign Language Recognition and Translation using Deep Learning and Computer Vision

MSc Research Project
Data Analytics

Sruthi Chandrasekaran
Student ID: x19233159

School of Computing
National College of Ireland

Supervisor: Christian Horn

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Sruthi Chandrasekaran
Student ID:	x19233159
Programme:	Data Analytics
Year:	2021
Module:	MSc Research Project
Supervisor:	Christian Horn
Submission Due Date:	16/08/2021
Project Title:	American Sign Language Recognition and Translation using Deep Learning and Computer Vision
Word Count:	5239
Page Count:	18

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	22nd September 2021

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

American Sign Language Recognition and Translation using Deep Learning and Computer Vision

Sruthi Chandrasekaran
x19233159

Abstract

Humans communicate using the language that they are familiar with. There are people who have hearing impairment and speaking issues. Its not necessary that everyone should know sign language and communicate in sign language. The Recognition and Translation of Sign Language in real-time would benefit people to have a effective communication and bridges the communication gap between people who do not know sign language. This work addresses the issue of " sign language recognition and sign language translation". The dataset used here is American Sign Language(ASL) Alphabets, which contain 26 letters and 3 special characters that are space, delete and nothing. These special characters help in real-time recognition. Convolutional Neural Network(CNN) was chosen for the ASL recognition and Translation. Since the dataset is images, Image augmentation, color conversion, size reduction are implemented. The model could predict the letter and display in text and convert to audio with the help of python library called 'gTTs'(Google Text to Speech). The CNN model with Image Augmentation have achieved an accuracy of 94% with 10 epochs. In terms of accuracy the model is compared with all other ASL translation techniques, where most of them are using glove and sensors. However the model is overfit, and future work should address this while improving.

1 Introduction

1.1 Background

To express what we feel, we need to communicate with other person. Communication plays a major role when it comes to human. Everyone communicate with the language that they know. So for a person who cannot speak and hear, the communication happens through Sign Language. Not everyone is aware of sign language. Its a non-verbal communication with hand gestures and facial expression. There is no common Sign Language Xue et al. (2017). There has been an increase in the number of hearing-impaired students who use sign language as their primary mode of communication attending higher education institutions around the world in recent years. To support everyone who is having speaking and hearing impairments, improvement in sign language translation would help Oliveira et al. (2019). In the beginning stage, skeleton data of a person's hand gesture is used for sign language translation Fang et al. (2017). Most of the research in sign language translation is using sensors and wires attached to the gloves, which makes it more complicated. The other type includes hardwares connected to phones, there are very few research which is purely software based.

1.2 Motivation

Any movement of body part such as face, hand is a form of gesture. There are lot of people who are having speaking and hearing issues. The only way of communication is through sign language. They face a lot struggles in communicating to people who don't know sign language in their day to day life. Seeing this many researchers have stepped to help them by Sign Language Recognition and Translation, still many researches are on going. The field American Sign Language Recognition and Translation was explored to help people who cannot speak and hear communicate to people who cannot understand sign language. This research builds a system for American Sign Language Recognition and Translation using deep learning method, which helps the people with speaking and hearing problem to have a easier communication to people who do not know Sign Language. The previous researchers were mostly into a hardware based solutions, that is it had sensors and wire attached to gloves making the design complex as well as expensive. The proposed model is cost effective and easy, it is able to recognize the letter in less time. The Fig. 1 shows the basic ASL recognition and translation.

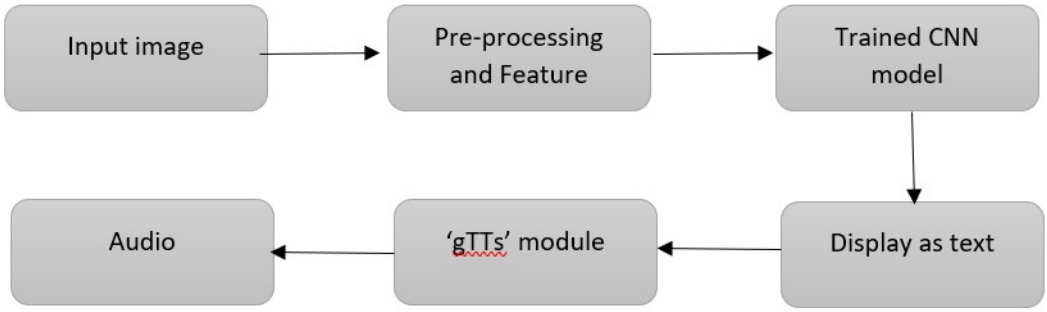

Figure 1: Basic ASL Recognition and Translation

1.3 Research Question

How Convolution Neural Network with Image Augmentation can be employed in American Sign Language Recognition and Translation ?

1.4 Research Objective

The main objective of this research is that, let's assume that a person with hearing impairment and speaking issues wanted to convey something to a person who doesn't know sign language. So it becomes very difficult for the person who is deaf and dump to communicate. The following are the objective

- Extract, Pre-process and analyze the datase
- To build a efficient model that is capable of recognizing and translating the American Sign Language with less expense.
- To provide a valid comparison to the study, implementing Convolution Neural Network

1.5 Outline of Research Paper

The content of the research paper is divided into the following sections. Section 2 is Related work, followed by 3 sections under that. Section 2.1 is Sensors based translation, Section 2.2 is Software based translation and Section 2.3 is Recognition. Section 3 is Methodology followed by Data Acquisition and Data Pre-processing. Section 4 is Design Specification, Section 5 is Implementation, Section 6 is Evaluation, Section 7 is discussion and last Section 8 is Conclusion and Future Work.

2 Related Work

Sign Language is a form of communication which is used by people who are having speaking issues and hearing. This type of communication helped the deaf and dump community to have a better communication. Many researchers have encouragement towards sign language recognition and translation and help this community to step forward. Firstly the deaf community didn't involve for the testing. This was because, the people with hearing and speaking issues see this as an cultural identity rather the point of view of others as an disease or deficiency. Later with informed consent from the students who are enrolled in the National Deaf Community, step forward to help the researchers. First of all there is no common sign language, but there are slight similarities between different sign languages. Such as India has Indian Sign Language, Germany has German Sign Language etc.. As the Sign Language is fully non-verbal communication that is either single or double hands are used for showing sign and along with facial expressions. It varies for different sign language.

The literature review for the Sign Language Recognition and Translation is divided to three sections. Firstly it discusses about the sign language translations based on sensors, next section is about vision, machine learning and deep learning based translation, and after that its about how recognition alone is executed. The is no unique sign language and sign language dataset varies from what sign language chosen and what needs to be recognized. Here the Sign Language used is American and the dataset type is images. Here, different sign languages such as Indian Sign Language, Sri Lankan Sign Language, Chinese Sign Language etc.. are discussed along with the various pre-processing techniques of dataset like size compression of images, convert to greyscale, skin detection etc... Many Sign Language Translation have carried out using sensors like Raspberry pi, Arduino with external sources like gloves for capturing sign, speakers for audio. Existing models of Sign Language based on recognition and translations using sensors, vision based, machine learning and deep learning algorithms are also reviewed here. Various type of translations that is Sign Language to text, audio, GIF's are explained. Overall, in this chapter focuses on the literature review conducted for American Sign Language using Open CV, sensors and their combination with preprocessing techniques in order to perform recognition and translation.

2.1 Sensors based Translation

A Convolution Neural Network based Bidirectional Sign Language Translation system is build that is it translates Sign Language to audio and text and converts the audio to sign language. This model is build using sensors and hardware glove. Since it is a bidirectional

translation system there are two approaches in it. The first approach is sensors based approach where flex sensors are attached to the microcontroller are connected to the glove. In the second approach the actual bidirectional translation is taking place, the pre-processed images are fed into CNN model and text is displayed. For translation OpenCV is used. For the conversion of audio to sign language, at first audio is converted to text and then it is taken as an input. According to the letters the hand gestures are displayed one after other. Overall the model, could predict but in the hardware appraoch due to combination of sensors few letters are predicted incorrectly. The limitation of the model is that its expensive and complex. Hence the software approach was efficient and better for translation. Fernandes et al. (2020)

A Sign Language interpreter is build using Microsoft's Kinect v2 that is for windows. This model has dual mode of communication that is it converts the hand gesture to audio as well as audio to sign language. This model acts like an translator and interpreter. In order to convert sign language to speech the person has to perform hand gesture in-front of the Kinect's field of view (FOV). The system receives the hand gesture performed through the sensors and compares it with the trained sign language which is stored in database. When a match is found, it maps with the keyword it is corresponding with and then send as a text before being converted to audio. Following that, the person must speak in native English into the Kinect FOV sensor in order for speech to sign conversion to occur.With the help of API it is converted from audio to sign language. The keywords from the audio or speech is taken and mapped with previously stored animated video. Adaboost and RFRProgess are used for recognition of hand gesture. The model have obtained about 84 % accuracy when tested with three person on a test data of 100. The accuracy also depends on the factor of number of sign language trained and the number of overlapping hand gestures. Abraham and Rohini (2018)

An android phone with GSM (Global System for Mobile) module is used for the real-time sign language conversion to speech. An android application is build for conversion and other purposes. Flex sensors are attached to the wearable gloves, each hand gesture has certain values. The corresponding messeges are sent to the android application using GSM module via Bluetooth. The Arduino board has connected with flex and GSM. Artificial Neural Network (ANN) algorithm is used for prediction when the values are detected. Forward Back-propogation algorithm is used for the layered feed. The weight of the neural network is adjusted accordingly to get prediction accurately. The android application makes it easier for the sign language users to communicate without any hassle. Ahmed et al. (2016)

A sensor and glove based Lebanese Sign Language translator were developed. It has 2 gloves with sensors on it and a phone with mobile application on it. Where this app can convert sign language to spoken words. The sensors attached to the glove is Flex. According to the gestures showing in hand, the flex sensors collect it and transmit to the application via bluetooth. The back-end of the application is Artificial Neural Network (ANN). It is build with 16 functions on a single layer and consist of 2 hidden layer and one output layer. The values received in App is forwarded to the back-end, predict the letter using Artificial Neural Network (ANN) and sends back to the application. The cost function for these hidden layers is cross entropy. ANN uses back propagation algorithm for the process and then gives the result. Abou Haidar et al. (2019)

Various sensors such as Contact, Flex, Simplified tilt were used to create a American Sing Language Translator. The entire system is developed using a single Arduino UNO board.For each letter, a binary gesture state table is created. The experiment was carried out in combination with Flex sensor, that is Flex alone is tested, then Flex and Tilt, Flex and Contact Sensor combinations. These sensors are connected to the gloves and from the glove to Arduino Board. The purpose of using different sensors and combination is to find a optimal solution. The flex sensor alone is used for the letters that is having bend like A, B, K, L etc.. The combination of sensors were unsuccessful when comparing to the results by Flex sensors alone. Flex and Contact sensor combination had higher frequency than Flex and Tilt combination. The translation was completed with the aid of processing software, and the sign was converted to text. Quiapo and Ramos (2016)

A Arabic Sign Language Translator is designed with the help of glove. The translator converts Arabic sign language to both text and speech. Arabic Language is widely used in Gulf Countries. Five Flex sensors are connected to the gloves, the hand gestures are detected. Arduino software is used to program Arduino, where it has two parts: one for the receiver side and one for transmitter side. When the signals from the glove receives at the receiver side that is RF and then passes to Arduino board. A LCD display and speaker is attached to the Arduino. The output of text is displayed in screen and the audio in speaker. No algorithms are used in this design, which is fully implemented using Arduino and a glove. Since there are 5 flex sensors the results are satisfactory. Abdulla et al. (2016)

A Convolutional Neural Network(CNN) and Customized Region of Interest (ROI) Segmentation based Bangala Sign Language Translator device were developed. The dataset with 5 hand gestures were created, labelled and further resized to 96 96 pixels. Convolutional Neural Network(CNN) is used to train the model. Batch normalization is applied to normalize the layers. The activation function Rectified Linear Unit (ReLu) is used for all layers except last layer. Distinct features from each image of Sign Language is extracted from the layers of CNN and are used for predicting the test set. In Raspberry Pi, the model is deployed and a small screen is attached to it to display the text. The time to detect the sign in real-time is reduced because of ROI segmentation method. When the sign is detected, for prediction it is sent to CNN model. Raspberry pi take less time to predict. Khan et al. (2019)

2.2 Software and Vision Based Translation

American Sign Language translator to detect hand gestures on hand sign using combination of AdaBoost and Haar-like classifiers.28000 samples of hand gesture dataset were used for training, from that 1000 positive samples of hand gestures images were captured in a variety of lighting, scale, and background settings. To find feature vector an algorithm by image processing technique is used. To classify features that are extracted Haar-cascade algorithm is applied. The conversion from text to audio is executed with speech SAPI 5.3 windows SDK powered by Microsoft. Adaboost were used for training, which aided in the collection of critical features at each stage. Thus the combination of both algorithm have increased speed of detection in less time. Truong et al. (2016)

Mobile phones are very common in these days and everyone has mobile with them. So an android application is developed to help dump and deaf people. All smart phones have front camera in it and speakers. The android application is divided into 2 sections: one for recognizing hand gesture and for adding new hand gesture to database. Three skin detection methods are used, they are YCbCr, RGB and HIS. Image resizing is executed in order to reduce processing time. The match in sign recognition is by histogram matching.The accuracy is 70 %. The limitation is that few letters and digits are predicted incorrectly. Mahesh et al. (2017)

A Support Vector Machine (SVM) algorithm along with an android application were proposed for the American Sign Language real-time translation to text and speech. A very small dataset is used that is 26 alaphabets of ASL and a customised symbol for 'space'. The small data-set is handled by Histogram of Gradient(HOG) by Support Vector Machine, which are useful in edge direction , detecting shape of an object. In pre-processing canny edge detection and contour masking is also done along with HOG descriptors.The limitation is that model is trained with less data and an accuracy of 91% is achieved. The app's main features were to add alphabet, backspace, clear, and audio. Tiku et al. (2020)

Sign Language are primary language for people who have hearing impairment and speaking issues. A German Sign Language Translation from videos of sign language is taken and converts to spoken language. The dataset used in this case is a German sign language video of a weather broadcast, as well as German sign vocabulary. The sign language is trained using Convolution neural network and attention-based encoder-decoders. Word embedding is required for better translation. Comparing to other translation models, this model requires plenty of training data. The output are average and the accuracy is 44%, some errors can be seen in ground truth table. Camgoz et al. (2018)

A support Vector Machine along with 3D object detection technique is used for German Sign Language translation. The 3D object detection has three network architectures that is localisation, elevation and segmentation. Thirty hand gestures are trained in this model. The dataset of hand gesture is labelled with 3D points, these are helpful in learning the pose . A three-network pipeline was used to recognize the hand gesture. Three network pipelines are PoseNet, HandSegNetand and PosePrior.For the 3D estimation, no depth cameras were used. After obtaining a hand pose representation,sign language is recognized. The classification model used here is the Support Vector Machine in conjunction with the Radial Basis Function. The model of Support Vector Machine and 3D object detection were tested and successful. The rate of error for the Support Vector Machine classifier alone was 0.58. The rate was reduced to 0.39 by combining features. Mohanty et al. (2020)

With the help of a webcam, real-time American sign language recognition and translation to English text is possible. A GUI (Graphical User Interface) is used to capture the sign in order to collect the dataset. The pre-processing of image includes edge detection, image size is reduced by morphological filtering algorithm. Morphed image does not contain any characterize information about gesture. Finally morphed image and the pre-processed image whose edges are extracted are combined together. When the person

performs sign language, the features are detected and check for matching by image cross-correlation. This model recognized and translated alphabets as well as a few English words. Joshi et al. (2017)

Minimum Eigen value algorithm is used for Double- Handed Indian Sign Language Translation to text and audio. Three types of features of minimum eigen value algorithm are selected, from that one feature is called corners are used. Each gesture's feature point is detected by this algorithm. The value of two eigen vectors are calculated and that score is evaluates each pixel of image. Pixel marks as a corner when the score exceeds certain value. Once the text output is obtained, it is analyzed and searched for a match in the letter to sound dictionary. It is translated to speech when match is found. Dutta et al. (2015)

Region-Based Convolution Neural Networks (RCNN) and Convolution Neural Networks (CNN) were proposed for Sri Lankan sign language translator. Three images for each letter is collected. In tensor flow, faster RCNN is build. Labels are created for the classes that are found. The image size is reduced to 800600 pixel, this enables efficient storage and training in a reduced time. Pascal VOC tool are used for label and get in XML format. It is trined in Faster RCNN. The accuracy obtained is 95%. To convert into speech, the word is analyzed using NLTK (Natural Language Toolkit) in Natural Language Processing by splitting the sentence into multiple word, when a match if word is found in database the appropriate signs have been ordered properly and will generate GIF images. Kumar et al. (2020)

A real-time sign language translation using Times Series Neural Network were proposed. This model is full and fully based on OpenCV and Neural Network. The goal is to convert sign language video into English sentences. ASL gloss are recognized from continuous sign language video. The person needs to perform Sign Language infront of webcam wearing dark cloth and also dark background. In further stage hand and face is extracted from frame, hand provides lots of information. A Recurrent Neural Network (RNN) with time series is used. It only works on one frame at a time. Encoders and decoders are used for gloss-to-speech translation. Overall, this approach was successful in providing a new approach to translation. Kumar et al. (2018)

2.3 Recognition Based

An Indian Sign Language Recognition were proposed using Deep Convolution Neural Networks. A standard dataset is not used in this system. The dataset was made up of 200 signs in five different backgrounds. Each sign took up 60 frames in the video. Video of sign language were continuously recorded using A selfie mode and can be viewed on a mobile device. Three different sample sizes were used to train a Deep CNN. The input used here is selfie sign images. To improve computational speed, each image size was reduced to 128 128 pixels. The CNN was trained with various sample sizes, and one batch with three sets of data and 180000 video frames achieved the highest recognition rate. Finally the accuracy obtained is 92 %. This model is successful, and more easier comparing to other models. Rao et al. (2018)

Character Recognition of American Sign Language were implemented by using Capsule Network. The images of hand gesture are trained. The amount of training data varies for each letter that is letter 'a' has 1126 of training images where as letter 'b' has 1010. There is a imbalance in data. Image size is reduced to 32 x 32 pixels. The features that are learned from convolution layer is combined by capsule network. Compute loss function for each letter and predict letter based on this loss. The capsule network achieved a 95%. accuracy. Bilgin and Mutludoğan (2019)

A real-time recognition for American Sign Language using Convolution Neural Network (CNN) and various pre-processing techniques such as image category, skin detection were implemented. The color detection of skin is based on the YCbCr and RGB. The background is removed in order to get better accuracy. It is followed by bounding box and mask filtering. A multi-layer perceptron is used by CNN. The model accuracy is 94.7 %. The limitation of the model is that, when there is something in background it couldn't recognize the sign language. Shahriar et al. (2018)

A real-time recognition for American Sign Language using Convolution Neural Network (CNN) and various pre-processing techniques such as image category, skin detection were implemented. The color detection of skin is based on the YCbCr and RGB. The background is removed in order to get better accuracy. It is followed by bounding box and mask filtering. A multi-layer perceptron is used by CNN. The model accuracy is 94.7 %. The limitation of the model is that, when there is something in background it couldn't recognize the sign language. Shahriar et al. (2018)

Dataset	No.of test images	Classifier used	Feature extraction	Accuracy
American Sign Language Mahesh et al. (2017)	30000	Histogram Matching	YCbCr, RGB and HIS	70%
American Sign Language Tiku et al. (2020)	3000	Support Vector Machine	contour masking,canny edge detection, HOG	91%
German Sign Language Kumar et al. (2020)	90,000	Convolution Neural Network and Attention-based encoder and decoder	Hand gesture detector	44%
American Sign Language Joshi et al. (2017)	260	Graphical User Interface	Edge Detection,cross co-relation	94%

Table 1: Summary of the related work on Software and Vision Based

3 Methodology

Sign Language is one of the common language used by the people who have hearing impairments and speaking issues. The most commonly used sign language is American Sign Language(ASL), but there are other sign languages such as Indian Sign Language(ISL), SriLankan Sign Language(SSL). Everyone does not understand sign language, so it makes it difficult for them to communicate. The researchers around the globe wanted to help them. The aim of this research is to help people who are having hearing and speaking problem to communicate in a easier manner. Here Convolutional Neural Network with lots of pre-processing techniques, a model is developed to Interpret American Sign Language to both text and audio. For the conversion of audio no external devices are connected to it, a python library called 'gTTs' (Google Text to Speech) is used for the conversion from text to audio. The model follows KDD(Knowledge discovery in Database) methodology.

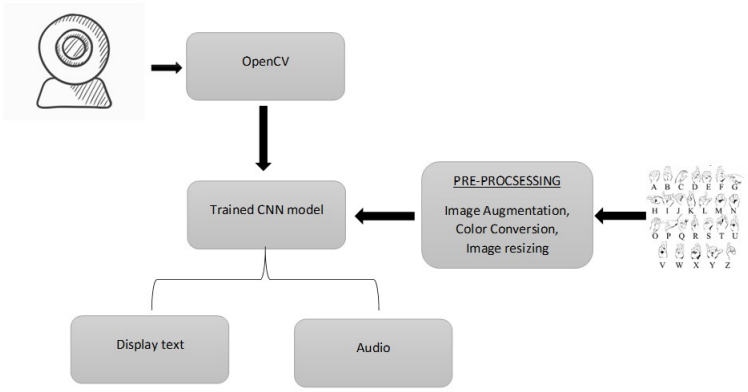

Figure 2: ASL Recognition and Translation Process

3.1 Data Acquisition

The dataset for American Sign Language(ASL) Translation and Recognition is taken from Kaggle which is from National Institute on Deafness and Other Communication Disorders(NIDCD). This dataset is a collection of alphabets images from American Sign Language, organized into 29 folders, which represent each classes. The dataset has separate folder for test and train data. The training image data contains 87,000 images of 200 x 200 pixel. There are extra 3 more characters other than 26 Alphabets, they are Delete ,Space and Nothing. So in total 29 classes. These three classes are extremely useful for application and classification in real-time. The training data is splitted into 90% of the total image for training and 10% of it for testing. Instead testing the model in real-time, it is tested with the images.

3.2 Data Pre-processing

The dataset 'ASL Alphabet' were downloaded. In Jupyter lab import cv2, keras, tensorflow, numpy, matplotlib and necessary functions. From module called

'tensorflow.keras.preprocessing.image' ImageDataGenarator is imported. Image Augmentation is a technique where various transformations of the original image is applied

to the original image. This results in multiple transformed copies of the original image. However each copy differ from each other in certain ways depending on the augmentation technique that is applied such as shifting, flipping, rotating etc.. Here augmentation technique applied is shifting, zooming, flipping and rotation. Keras ImageDataGenerator class provides a simple and quick way to enhance your images. Also, it ensures that the model receives new image variation at each epoch.Then image whose colour that are converted from 'BGR2RGB' are loaded to the Jupyter notebook.

4 Design Specification

4.1 Deep Learning Model

Deep Learning is a branch or a subset of machine learning, that includes algorithm that function closely to human brain. These structures are called as neural networks. Convolutional Neural Network is one of the Deep Learning Algorithm, it is widely used for image classification. In this research, the input is image.

4.1.1 Convolutional Neural Network

As the technology is improving day by day, the improvement and advancement in the field of Computer Vision and Deep Learning enabled machines to see this world as humans do.That's how Convolutional Neural Network(CNN) has been constructed. CNN has been widely applied on various fields such as medical image classification, natural language processing, recommendation system, video recognition etc.. The whole idea of image classification is like this, taking an image as input and gives the output as a class. For example if the input images is of rat and cat then the result is the classes of rat and cat.

A convolution tool separates and recognize image's several features, this is called as feature extraction. A fully connected layer that uses the convolution process output to predict the image's class based on the features that were separated from in prior stages. Convolutional neural network is made up of layers and they are categorized into three layers that are Convolution, Pooling and Fully Connected. While training a image it has to pass through these sequence of layers. We finally get output after passing through these layers. Other than these two layer there are two more important layer, one is activation and other is dropout layer.

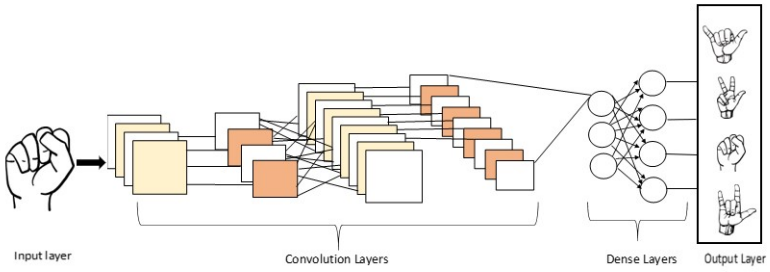

Figure 3: Basic CNN Architecture

The Fig. 4 shows the architecture of Convolutional Neural Network, as you can see from the image there are three layers. The Alphabet of American Sign Language is the input image, which is of 200 * 200 pixel. It passes through Convolution, Pooling and Fully connected layers and finally get the classes of American Sign Language. Initially for training the model, the images are pre-processed by Image Augmentation. The class-mode is set to categorical and the image size is targeted to 64 x 64 pixel. The proposed Convolutional Neural Network Model consist of 3 convolution layers and two fully connected layer. The image after pre-processing first passes through the convolution layer. This layer extracts the features from image. The first Conv layer of the architecture has 16 filters each with a kernel size of 3 x 3. After the first layer the output shape is 21, 21, 32. The second Conv layer has 32 filters with each of 3 x 3 kernal size and the output shape is 7, 7, 64. The thrid Conv layer has 128 filters with 3 x 3 kernel size of each and the output layer is 2,2,256. All the three Convolution layer has padding value as 'same' and activation function as 'relu'. Each layer has max-pooling function. After that batch normalization is applied. It normalises the output from the prior layer, also it helps to control the overfitting. Dropout of 0.5 % nodes are dropped of the network. The layers are flatten. The last layer is fully connected layers which contains the 2 dense layers of 512 and 29 units. Finally the output shape of 29 classes are shown.

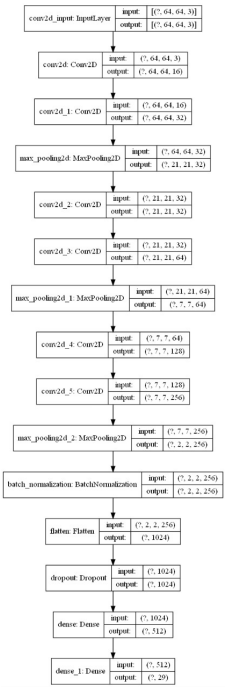

Figure 4: CNN Model Plot

5 Implementation

This section provides the overall implementation of the model with detailed process flow, experiment setup, objective of the research, final output. In section 3 Data Acquisition, Data Pre-processing is discussed and in section 4, the proposed CNN Architecture is discussed. In this research Convolution Neural Network with extra layers and various pre-processing techniques like Image Augmentation, image size reduction, colour conversion are preformed.

The proposed system is for Sign Language Recognition and Translation. That is it will recognize the hand gesture or the ASL Alphabet that is been shown and display the letter in text and also converts into audio. The input image is ASL Alphabet, the CNN model is trained with ASL Alphabets with three special characters that are space, delete and nothing. These three characters help in the real-time recognition and translation

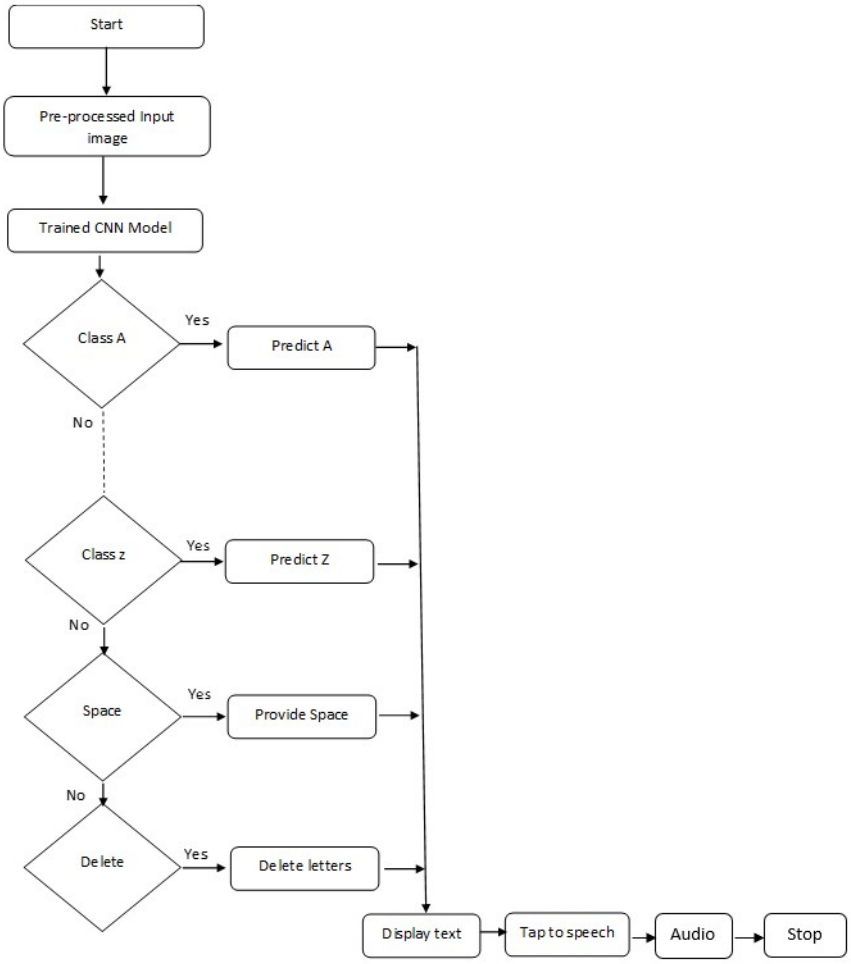

Figure 5: Flow Chart

5.1 Process Flow

The following are the modules for ASL Recognition and Translation.

1. Data Acquisition and Pre-Processing

   The dataset is taken from National Institute on Deafness and Other Communication Disorders(NIDCD). It has 29 classes, 26 letters and three characters. They are space, delete and nothing. The images are pre-processed by Image Augmentation that has, shifting, rotating, flipping. Then image size reduction to 64 x 64 pixel. The color conversion of image from BGR(Blue, Green, Red) to RGB(Red, Green, Blue)

2. Convolution Neural Network model

   The pre-processed images are fed into the CNN model. Pre-processed images are divided into train and test. The model has 2 Conv layers with max pooling on both of the layer, batch normalisation and dropout layer to control overfitting, then 2 fully connected layer which contain dense().

3. Predicting and categorizing the frame

   After training the model, it is saved. In OpenCV, the trained model is called and performs the prediction. A full frame is designed to show the hand gesture and the translated letter comes below and in order to convert the text into audio 'gTTs'(Google Text to Speech) is used and a small rectangle on top right corner is present to convert into audio.

5.2 Objective of the project

Assume a person who have speaking and hearing issues wanted to communicate. They use Sign Language for communication, and not everyone knows the alphabets in sign language. So this model helps the community of people who are having hearing and speaking problem communicate easily.

5.3 Experiment Output

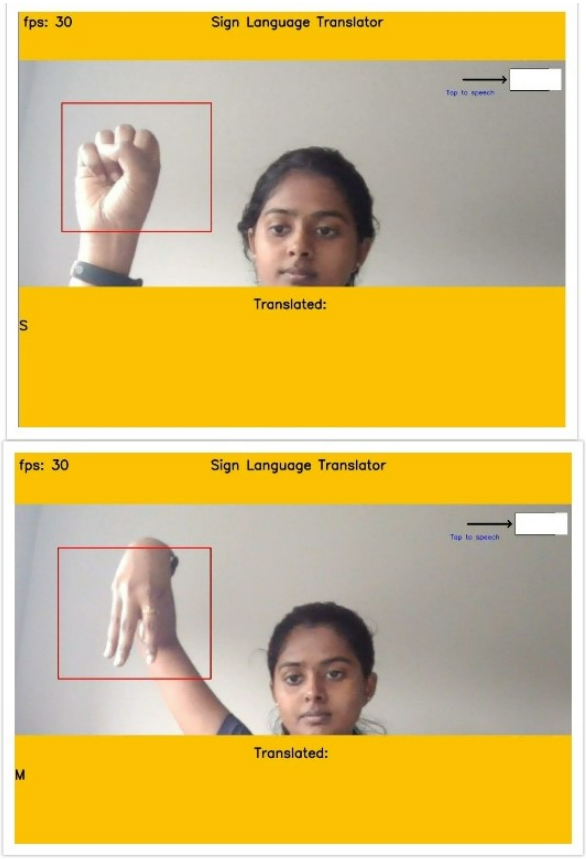

Figure 6: Real-Time implementation of American Sign Language

From above fig. 6, we can see that the model is able to predict the American Sign Language alphabet and display the text. For audio conversion, tap on the top right corner.

6 Evaluation

This experiment was conducted on American Sign Language alphabet images. The data was split to train and test data. There are 29 folders which contain letters and three characters that are space, delete and nothing. The data is splited in 9:1, where 90% of the data is taken for training and 10% for testing. On the ASL alphabet dataset the model is trained for 10 epochs. The Fig. 7 represents the accuracy of the model. The model have achieved 94% of accuracy in 10 epochs. The model is not tested in real-time, rather it is tested with image data.

```
Epoch 10/10
1223/1223 [==============================] - 265s 216ms/step - loss: 0.1617 - accuracy: 0.9882 - val_loss: 0.3103 - val_accu
racy: 0.9407
```

Figure 7: Model Accuracy

A train learning curve is the curve that is derived from the training dataset, indicates how well the model is learning. From the loss plot, after analyzing the curve its been found that there is slight overfitting in data. One of the reason might be the model has additional capacity than what is needed which also results in the flexibility.

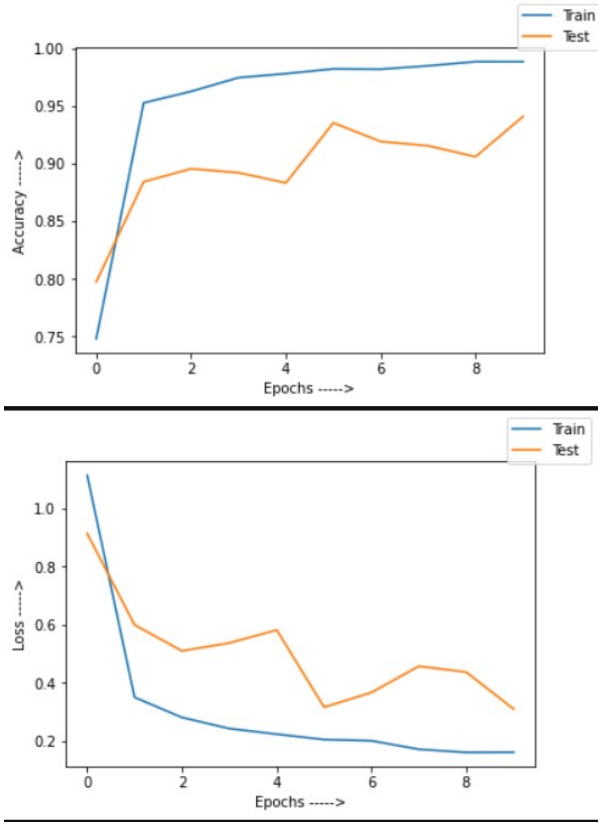

Figure 8: Accuracy and Loss plot using CNN

7 Discussion

In this research paper, one of the deep learning technique called Convolution Neural Network is proposed for the Real-time Recognition and Translation of American Sign Language(ASL). The main objective of this research is that, let's assume that a person with hearing impairment and speaking issues wanted to convey something to a person who doesn't know sign language. So it becomes very difficult for the person who is deaf and dump to communicate. The proposed system would make it easier, the person just has to show the ASL alphabet on the webcam and it displays the text associated to it below and for the conversion to audio a move hand towards a small rectangle where it shows 'Tap to Speech'.

The proposed system has acquired an accuracy of 94% for the American Sign Language(ASL) Recognition and Translation using Convolution Neural Network, and was evaluated by the Accuracy and loss plot curve. The result was achieved in 10 epochs. It was noticed in loss plot curve that there is a slight overfitting and it is may have caused because the model has sufficient capacity than what is actually needed which results in flexibility. The strength of this model is that, it is able to recognize the ASL alphabet correctly in very less time and predict it correctly and display as text and also converts into audio with the help of 'gTTs'. The background should be white. The limitation of this model is that, while showing the sign infront of camera, there should not be any over lighting which results in predicting the letter incorrectly.

For Bangala Sign Language Conversion to text, Khan et al. (2019) CNN and Region of Interset(ROI) Segmentation were used. This model were developed with Raspberry pi.To display text a screen is attached to it. The dataset contains only images for 5 sign. This model have acquired an accuracy of 93%. The limitation of this model is that hardware is required, the whole setup is quite large and few signs are predicting incorrectly. American Sign Langugage Translation were proposed using sensors like flex, contact and gloves Quiapo and Ramos (2016) , this model couldn't predict few letters, and the model is fully hardware based. The person has to wear glove which has sensors and wires connected to it. This model is not user friendly. ASL translation to text using Graphical User Interface with edge detection were proposed Joshi et al. (2017). Though the images are pre-processed and edge detection and cross-correlation were applied, the model could reach a accuracy of 80% only. The model which is proposed in this research have an accuracy of 94% comparing to other models using CNN, image augmentation and color conversion. Also most of the previous work are developed using sensors attached to glove which has wire connection to Arduino board or Raspberry Pi and a display screen and speaker attached to it Fernandes et al. (2020), Quiapo and Ramos (2016) , Abou Haidar et al. (2019). The advantage of our model is that its handy and predict the alphabet in very less time. Due to bad lighting, sometimes the letters are predicted incorrectly. When the sign is not properly shown, the letters predicted are incorrect.

8 Conclusion and Future Work

The main objective of this research is to implement American Sign Language Recognition and Translation, for people who are having difficulties in speaking and hearing to have a easier communication. The idea is to build a model that can recognize and able to translate into text as well as audio. By developing this model people can communicate to others who do not know sign language. The model is build for ASL Alphabet dataset using Convolutional Neural Network with extra layers, different pre-processing techniques like Image Augmentation, color conversion, image resizing. The model have achieved an accuracy of 94%. The advantage of this model is that, it takes less time to recognize each letter, also this model is full and fully software and vision based hence it is not expensive. The limitation of this model is that the background while performing the sign language should be white background with proper lighting, and its not necessary that everyone will have a white background always. This model does not have any hand classifier to classify its hand, while performing we have to move hand to the red column and show the sign. While evaluating the loss curve plot, it is found that there is over-fitting of dataset.

As part of future work, skin detection technique can be involved to get more accuracy and off-course recognize the letter correctly. A hand classifier can be included, so that while performing the sign we don't need to move hand towards the rectangle and perform. The FPS in current model is very high due to the high resolution, try to reduce FPS for better recognition.

References

Abdulla, D., Abdulla, S., Manaf, R. and Jarndal, A. H. (2016). Design and implementation of a sign-to-speech/text system for deaf and dumb people, *2016 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA)*, IEEE, pp. 1–4.

Abou Haidar, G., Achkar, R., Salhab, D., Sayah, A. and Jobran, F. (2019). Sign language translator using the back propagation algorithm of an mlp, *2019 7th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*, IEEE, pp. 31–35.

Abraham, A. and Rohini, V. (2018). Real time conversion of sign language to speech and prediction of gestures using artificial neural network, *Procedia computer science* **143**: 587–594.

Ahmed, M., Idrees, M., ul Abideen, Z., Mumtaz, R. and Khalique, S. (2016). Deaf talk using 3d animated sign language: A sign language interpreter using microsoft's kinect v2, *2016 SAI Computing Conference (SAI)*, IEEE, pp. 330–335.

Bilgin, M. and Mutludoğan, K. (2019). American sign language character recognition with capsule networks, *2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, IEEE, pp. 1–6.

Camgoz, N. C., Hadfield, S., Koller, O., Ney, H. and Bowden, R. (2018). Neural sign language translation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7784–7793.

Dutta, K. K., GS, A. K. et al. (2015). Double handed indian sign language to speech and text, *2015 Third International Conference on Image Information Processing (ICIIP)*, IEEE, pp. 374–377.

Fang, B., Co, J. and Zhang, M. (2017). Deepasl: Enabling ubiquitous and non-intrusive word and sentence-level sign language translation, *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*, pp. 1–13.

Fernandes, L., Dalvi, P., Junnarkar, A. and Bansode, M. (2020). Convolutional neural network based bidirectional sign language translation system, *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, IEEE, pp. 769–775.

Joshi, A., Sierra, H. and Arzuaga, E. (2017). American sign language translation using edge detection and cross correlation, *2017 IEEE Colombian Conference on Communications and Computing (COLCOM)*, IEEE, pp. 1–6.

Khan, S. A., Joy, A. D., Asaduzzaman, S. and Hossain, M. (2019). An efficient sign language translator device using convolutional neural network and customized roi segmentation, *2019 2nd International Conference on Communication Engineering and Technology (ICCET)*, IEEE, pp. 152–156.

Kumar, D. M., Bavanraj, K., Thavananthan, S., Bastiansz, G., Harshanath, S. and Alosious, J. (2020). Easytalk: A translator for sri lankan sign language using machine learning and artificial intelligence, *2020 2nd International Conference on Advancements in Computing (ICAC)*, Vol. 1, IEEE, pp. 506–511.

Kumar, S. S., Wangyal, T., Saboo, V. and Srinath, R. (2018). Time series neural networks for real time sign language translation, *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, IEEE, pp. 243–248.

Mahesh, M., Jayaprakash, A. and Geetha, M. (2017). Sign language translator for mobile platforms, *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, IEEE, pp. 1176–1181.

Mohanty, S., Prasad, S., Sinha, T. and Krupa, B. N. (2020). German sign language translation using 3d hand pose estimation and deep learning, *2020 IEEE REGION 10 CONFERENCE (TENCON)*, IEEE, pp. 773–778.

Oliveira, T., Escudeiro, P., Escudeiro, N., Rocha, E. and Barbosa, F. M. (2019). Automatic sign language translation to improve communication, *2019 IEEE Global Engineering Education Conference (EDUCON)*, IEEE, pp. 937–942.

Quiapo, C. E. A. and Ramos, K. N. M. (2016). Development of a sign language translator using simplified tilt, flex and contact sensor modules, *2016 IEEE Region 10 Conference (TENCON)*, IEEE, pp. 1759–1763.

Rao, G. A., Syamala, K., Kishore, P. and Sastry, A. (2018). Deep convolutional neural networks for sign language recognition, *2018 Conference on Signal Processing And Communication Engineering Systems (SPACES)*, IEEE, pp. 194–197.

Shahriar, S., Siddiquee, A., Islam, T., Ghosh, A., Chakraborty, R., Khan, A. I., Shahnaz, C. and Fattah, S. A. (2018). Real-time american sign language recognition using skin segmentation and image category classification with convolutional neural network and deep learning, *TENCON 2018-2018 IEEE Region 10 Conference*, IEEE, pp. 1168–1171.

Tiku, K., Maloo, J., Ramesh, A. and Indra, R. (2020). Real-time conversion of sign language to text and speech, *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, IEEE, pp. 346–351.

Truong, V. N., Yang, C.-K. and Tran, Q.-V. (2016). A translator for american sign language to text and speech, *2016 IEEE 5th Global Conference on Consumer Electronics*, IEEE, pp. 1–2.

Xue, Y., Gao, S., Sun, H. and Qin, W. (2017). A chinese sign language recognition system using leap motion, *2017 International Conference on Virtual Reality and Visualization (ICVRV)*, IEEE, pp. 180–185.