Sign Language Detection using Deep Learning Architecture in Cloud computing Environment

MSc Research Project
Cloud Computing

Monika Malik
Student ID: x20149611

School of Computing
National College of Ireland

Supervisor: Aqeel Kazmi

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Monika Malik
Student ID:	x20149611
Programme:	Cloud Computing
Year:	2021
Module:	MSc Research Project
Supervisor:	Aqeel Kazmi
Submission Due Date:	16/12/2021
Project Title:	Sign Language Detection using Deep Learning Architecture in Cloud computing Environment
Word Count:	6560
Page Count:	20

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:	Monika
Date:	31st January 2022

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):	

Sign Language Detection using Deep Learning Architecture in Cloud computing Environment

Monika Malik
x20149611

Abstract

In order to facilitate the deaf community and general public, the usage of sign-language detector plays an important role. However, detecting the hand gesture movements in real-time is a challenging task and still an open area of research. Deep learning based architecture has provided significant outcomes in terms of pattern recognition, image processing and object detection. Therefore, in this work we have 4 different deep learning based neural network architectures. Among them VGG-19 and ResNet-50 are the transfer learning models and 3-Layer and 5-layers of convolutional architecture are the custom models. After evaluating the results over the test data based on accuracy, loss, precision, recall and F1-score, we have obtained the better outcomes using 5-layers of Convolutional architecture for Sign language detection.

1 Introduction

Individuals who have less to no capacity of hearing are called as hearing-impaired people. Most of the hearing-impaired people use the technique of lip reading to communicate, but it is a bit harder to learn. Since there is a difficulty in hearing, the usage of vocal communication has no effect. So, the language of actions came into picture which was later called as Sign Language. It consists of specific hand gestures and movements. Just like all other languages sign language is a mode of effective communication among the hearing-impaired people. The sign language also makes use of the facial expressions and bodily reactions also. Among the sign languages there are also different patterns and system. The most prevalent are the British and American as well as the Indian sign language and hand movements. Although the hand gesture is quite simple and effective most non-hearing-impaired people find it difficult to use when establishing communication with the hearing-impaired people. Both the static and dynamic traits influence the method of hand gestures. The shape of the gesture represents the static trait and the type of movements and actions are considered to be dynamic. It should be balanced correctly when communicating in the sign language with the hearing-impaired people. Most of the non-hearing-impaired people find it very difficult to correctly get the sign language as it reaches perfectly only through practise. So, there is a gap in communication. Because of ineffective communication the hearing-impaired may be isolated and left on their own.

Knowing and practising the sign language should become easier to bridge the communication gap between the non-hearing impaired and hearing-impaired people. In this research, we aim to develop a very efficient and smart model that can help in easing the communication between these two types of individuals. Sign language detector will be of great use for the hearing-impaired individual to communicate with the regular well hearing person without the aid of an actual human interpreter. The scope of this work will help interpret and understand the sign language using advanced technological approaches. To achieve this objective, we use the benefits of deep learning algorithms. By taking in the inputs of several hand gestures this project is designed. The dataset taken for analysis also contains several hundreds of sign languages gestures and images. Once the sign language symbols are identified in the perfect manner, then the classification process is carried out by analysing all the different positions of sign gestures from a video sample. For achieving this, multiple data processing and feature extraction methods are used by employing the deep learning algorithms. For real time identification of sign language, the multiple number of frames are extracted from the dataset. For the feature extraction process, pre trained deep neural network model such as RestNet-50, and VGG-19 are used. In this work, we utilize the Pre-trained models such as VGG-19 and ResNet-50. Other than that, we have created custom models which contains the 3 layers and 5 layers of convolutional neural network. To identify the best model, we have calculated evaluation metrics such as Accuracy, Precision, Recall, F1-Score and validation Loss.

The novelty of this research is to improve the classification accuracy for detection of Hand Gestures for Sign Language detection over American Sign Language Data (ASD) by training the custom model with different layers of CNN. The custom model proposed by our research are also compared with pre-trained deep neural network architectures in terms of accuracy, precision, recall and f1-score.

1.1 Research Question

In this work we will try to solve the following research question:

- How efficiently the Sign-language can be recognised using the deep learning algorithms?
- Does Pre-trained Model provides the better results for Sign-language detection?
- Which deep neural network architecture achieves the better outcomes for Sign language detection over test data ?

2 Literature Review

This section of the research paper analysis all the predominant existing research work in the area of sign language detecting. The outcomes of the research are also discussed in a clear manner.

2.1 Machine Learning Technique for Forecasting

In the research carried out by Dabre and Dholay (2014) the authors employ the machine learning languages to predict, interpret and understanding the sign learning language from the webcam images. The Indian sign language is taken for representation here and they are analysed using the techniques of image processing, the neural network algorithms and the ideologies of the computer-based vision. To effectively point out the hand gestures obtained from the webcam images these algorithms and techniques are employed. The usage of Haar cascade amplifier is also observed in this approach. At last, the speech synthesis is brought in to covert the detected and analysed gestures into a full-blown speech. In another similar research done by Saquib and Rahman (2020) the practice of machine learning algorithms are used in the real time detection of sign language and its application using the sensors approach. A framework has been created to distinguish and predict the American as well as the Bengali Sign Language utilizing the information containing some appropriately situated sensors. The systems utilized can be utilized even in asset compelled conditions. The framework is able to do precisely distinguishing both static and dynamic images in the letter set. The framework shows a promising exactness. Moreover, this work presents a clever way to deal with play out a ceaseless appraisal of images from a surge of run-time information.

The use of motion controller are seen in the research done by Chong and Lee (2018). The majority of the characters that are present in the sign language are static however few of the characters are dynamic too as they need specific developments. Consequently, the paper likewise intended to separate components from the gesture and hand movements to separate between the two types of signals. The exploratory outcomes uncovered that the communication via gestures acknowledgment rates of all the letters utilizing a SVM and a neural organization called as DNN are used individually. In the interim, the acknowledgment rates for a blend of the group of letters and 10 digits are somewhat lower, roughly around the half of the whole. Accordingly, the communication via gestures acknowledgment framework has incredible potential for decreasing the ratio among the networks and others. The proposed model could likewise fill in as a mediator in regular day to day existence in assistance areas.

Another very different approach is carried out by Hasan et al. (2016). The author in his research tries to solve and analyse the sign language of Bangladesh. To work with the discussion, change of communication via gestures to sound is exceptionally important. This paper focuses on change of communication via gestures to discourse so that handicapped individuals have their own voice to speak with the overall individuals. In the research, the sign acknowledgment is performed utilizing HOG gradients for extraction of elements from the signal picture as a classifier from the support vector machine. At last, foresee the signal picture with yield text. This yield text is changed over into perceptible sound utilizing the text converter. Another unique research done by Elakkiya (2021) presents the advanced concept of recognition using the AI techniques and machine learning algorithms. It researches the effect of cutting-edge writing on communication via gestures acknowledgment and grouping. It features the issues looked by the current acknowledgment framework for which the examination on communication through signing acknowledgment expects the arrangements. Here around 200 distinct methodologies have been looked at that investigate communication through signing acknowledgment for perceiving multi-design signs. The examination done by different writers is likewise considered, and a portion of the significant exploration articles are additionally talked about in this work. In view of the discussions from these methodologies, this work examined how AI techniques could help the field of programmed communication through signing acknowledgment and the potential holes that AI approaches need to address for the ongoing gesture-based communication acknowledgment.

In the study carried out by Mustafa (2021), the authors presents the detection of sign languages using machine learning algorithms. It is taken into count to audit the gesture-based communication acknowledgment framework dependent on various classifier procedures. For the most part the Neural Network and certain other classifiers were used to perceive the diverse communications via gestures and this study is proposed to survey the best classifier model to address communication via gestures acknowledgment. The authors zeroed in fundamentally on the detection strategies and furthermore on Arabic gesture-based communication acknowledgment frameworks. Various classifiers and more were carried out to detect the sign framework. Every classifier is explored with the acknowledgment precision, wherein the learning-based classifiers executed the ideal acknowledgment result as differentiated to different kinds of other techniques. A very smart and intelligent sign language detection system is presented by Rosero-Montalvo et al. (2018). Here the authors carry out various techniques and methodologies of machine learning algorithms to detect and interpret the sign language gestures. A very self-automotive electronic gloves is used in this process. The aim of the research presented is to text and make into speech the signs interpreted to make the communication a lot easier. The steps of data balancing and comparison of the prototype is also done here. The use of KNN is also employed in this research. The performance of this model is also found to be comparatively better than the others.

2.2 Deep Learning technique for forecasting

One such research done by Gattupalli et al. (2016) gives a brief idea of the sign and gesture detection process when it comes to the identification and interpretation of such languages. This task is a bit challenging when it comes to taking into account all the essential information of the vision-based identification process. There is a combination of algorithms used to perform this. Among all those the predominantly used are the deep learning approaches. The dataset taken for analysis is thoroughly processes and observed as the main process of the approach. The use of deep learning algorithms is found to be comparatively better when considering the normal traditional machine learning algorithms. Another research is done by Taskiran et al. (2018) in the area of analysing and predicting the sign languages using the deep learning algorithms. It is nothing but an ongoing communication through signing acknowledgment framework for individuals who don't realize communication via gestures to discuss effectively with hearing-disabled individuals. The gesture-based communication utilized in this paper is American gesture-based communication. Here the neural organization was prepared by utilizing a dataset gathered from the University of Mathematical Sciences, and the test precision was also gotten. After network preparation is finished, the organization model and organization loads are recorded for the ongoing framework. In the continuous framework, the skin not set in stone for a specific casing for hand use, and the hand still up in the air utilizing the arched body calculation, and the hand signal is characterized progressively utilizing the enlisted neural organization model and organization loads. The exactness of the continuous framework is found to be excellent.

Another similar research on the same principles of sign language detection and interpretation using the deep learning algorithms are clearly explained in the research work performed by Konstantinidis et al. (2018). This given sign analysis and learning-based system, in which we inspect and examine the commitment of video which are picture and optical stream and skeletal highlights in the difficult undertaking of segregated features, in which each marked video relates to a solitary word. Additionally, they also utilize different combination plans to distinguish the ideal way of joining the data acquired from the different component portrayals and propose a strong approach. The research on two gesture-based communication datasets and the examination with cutting edge strategies uncovers the prevalence of ideally consolidating highlights for the sign and position errands. A very extensive research is performed by Abiyev et al. (2020) in knowing and analysing the hand gesture signals. It is fundamentally utilized for correspondence with individuals who are hearing impaired or almost deaf. To see such correspondence rapidly and precisely, the plan of an effective gesture-based communication interpretation framework is considered in the given research. The proposed framework incorporates object location and characterization levels. Right off the bat, the sign Detection strategy and design is used for hand discovery, then, at that point, a deep learning structure dependent on the Inception in addition to having a SVM that joins highlight extraction and characterization stages is proposed to valuably interpret the recognized hand motions. A gesture-based communication hand sign dataset is utilized for the plan of the devised plan. The acquired outcomes and relative examination exhibit the productivity of utilizing the proposed mixture structure in gesture-based communication interpretation.

Some of the high research on the deep learning algorithms done to detect and interpret the sign language is identified and presented in the research done by Wadhawan and Kumar (2020). To identify the Indian sign language, one has to have the option of differentiating it from the American sign languages. The dataset of comparison is used in this process. The feature selection and separation is employed and used in this research. This research work handles the strong demonstrating of static signs with regards to communication through signing acknowledgment utilizing these models of deep learning-based convolutional neural organizations. Many of the dynamic, as well as the static signs, are gathered from various clients. The effectiveness of the proposed framework is assessed roughly through all the models. The outcomes are additionally assessed based on various streamlining agents, and it has been seen that the proposed approach has accomplished the most noteworthy preparing high precision on hued and grayscale pictures, separately. The presentation of the proposed framework has likewise been assessed based on accuracy, and review. The framework additionally exhibits its viability over the previous works wherein a couple of hand signs are considered for acknowledgment.

Analysing the American sign language using the extensive deep learning algorithm is presented in the paper done by Kulhandjian et al. (2019). The sets of signs and the motions gestures are captured to perform the research with the help of an assured dataset. A special technique called as Doppler radar is also employed in this research work. For this process of analysis, the MATLAB program is also used. The usage of deep neural networks is also carried out in this research. From the outcomes it is evident that the working and results of deep learning algorithms are much better and effective than the machine learning algorithms.

2.3 Other techniques for forecasting

The research that establish the identification and detection of sign language is carried out by Konwar et al. (2014). The vision-based approach is performed in this method. The HSV model is employed here. From the image and the input dataset a complete analysis is done in term of the position, sign as well as colour tones and hues present in the image pattern. The detection of the sharp edges is also observed by checking the patterns and colour deepening of the image set. Many steps are carried out in terms of analysing the refined output that are given out by this research.

Another similar research based on the models of the same colour comparison and the checking of hues and sharp edges of the signs are done by PURNAMASARI and Erwin (2019). The authors taken in account of all the signs as a dataset before the analysis and then make a comparative learning model. The signs are read and analysed using the video sample that is taken for the research. The difference in the colour of the background, tips and flat surfaces including many other focal points are analysed and then incorporated in analysing and predicting the pattern of the sign into the text. The K-means clustering algorithms is employed for the working of this process. From the outcomes of this research work, it is established that by employing this method, one can effectively detect the sign language in an accurate manner. A very peculiar research work based on the sign language identification is done by Wazalwar and Shrawankar (2017) using the NLP technique. It focuses on translation of gesture-based communication in an appropriate English. Very Distinctive NLP methods are utilized notwithstanding sign acknowledgment. Information is offered as a video of a hint and gestures followed by outlining and division on record. CamShift calculation is utilized for following here. Haar Cascade classifier is utilized for the sign of the design. Once the sign acknowledgment is done, the consistent words for separate sign are given as contribution to POS labelling module. Finally, an LALR is utilized to outline the sentence. In this manner proposed gesture-based communication translator gives the yield in a significant English. The outcomes of the research are found to be very beneficial.

A research work that is found to very useful in the detection of sign language is done using the temporal analysis presented in the analysis done by Kindiroglu et al. (2019). All the research is based on the HSV analysis and all the deep patterns are carried out to find the predictive outcomes. The given strategy is a HSV based collective video portrayal where the complete analysis is dependent on the etymological development models are addressed by various shadings. The authors likewise use and shape data and utilizing a limited scale neural organization, show that consecutive displaying of collective components for phonetics enhances benchmark arrangements. The sign language detection process is done using analysing the pattern and model of the hand and gesturing which is presented in the paper done by Jalilian and Chalechale (2013). The strategy for face and hand division that assists with building a superior sign of the approach is done using a language acknowledgment framework. The strategy proposed depends on getting the shading space, using the applications of the Bayes rule and also the physical and sign activities. It distinguishes areas of face and turns in the complex foundation of the research work. This strategy tried on various levels of pictures of the signs that are done with the position of the hands. Exploratory outcomes show that the strategy has accomplished a great execution for pictures with a solid foundation.

3 Methodology

Hand gestures is one the efficient way for non-verbal communication used in sign language. Sign language is the way through which the person having the hearing complication can communicate with the normal people. Auto translation of sign language will facilitate the communication between the normal people and specially-abled people. Machine learning and deep learning are such emerging technologies which can perform object detection, image recognition and pattern recognition. Therefore, in order to solve such issues, we have proposed a deep learning architecture based framework for sign language detection. Our proposed work can identify the character from hand gestures in real-time as well as from the images. In order to achieve the optimal outcomes, there are certain set of steps has been followed in this research. The steps includes data collection, data pre-processing, model training, evaluation etc. We will discuss about the each of the step-in detail. The developed framework has been executed on the cloud platform, where the video frames are sent from the client end in order to predict the hand gestures. The proposed framework of methodology is represented with the help of flow diagram in Figure 1 .

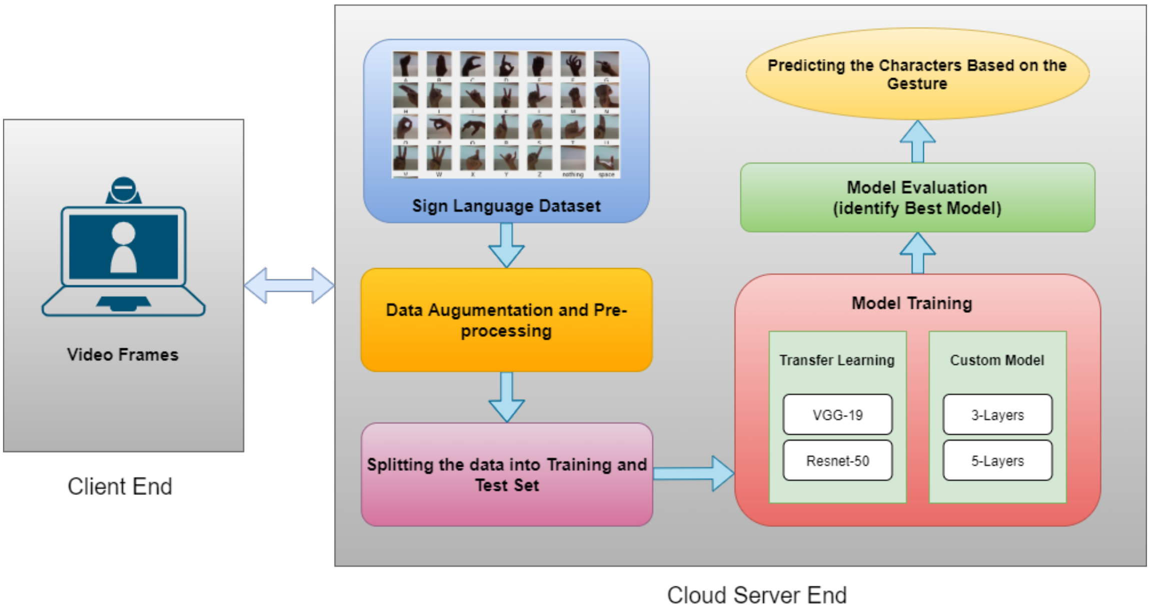
Video Frames
Client End
Sign Language Dataset
Data Augumentation and Pre-processing
Splitting the data into Training and Test Set
Predicting the Characters Based on the Gesture
Model Evaluation (identify Best Model)
Model Training
Transfer Learning
VGG-19
Resnet-50
Custom Model
3-Layers
5-Layers
Cloud Server End

Figure 1: Proposed Framework for Sign Language Detection

3.1 Data Collection

In order to predict the hand gestures for the sign language, the dataset collected from American sign language data which is available on Kaggle *ASL Alphabet* (n.d.). The dataset contains the 87,000 images, which can be categorized into the 29 classes. Where the 26 classes represent the letters from A-Z along with other classes such as the Space, delete and nothing. The size of each image is 200X200 pixels. The test dataset contains the images of 29 characters.

3.2 Data Pre-processing

The obtained images need to be pre-processed where we are applying multiple data augmentation techniques to process the image data. The data augmentation techniques include the operations such as image rotation, different zoom ratios, re scaling, horizontal flip etc. The data pre-processing task helps the model to learn efficiently in order to generate the better results. In order to extract the features from the images the pre-trained weights of ImageNet, has been used for extracting the features from the images in an efficient way. After Pre-Processing the obtained dataset is shown in figure 2.

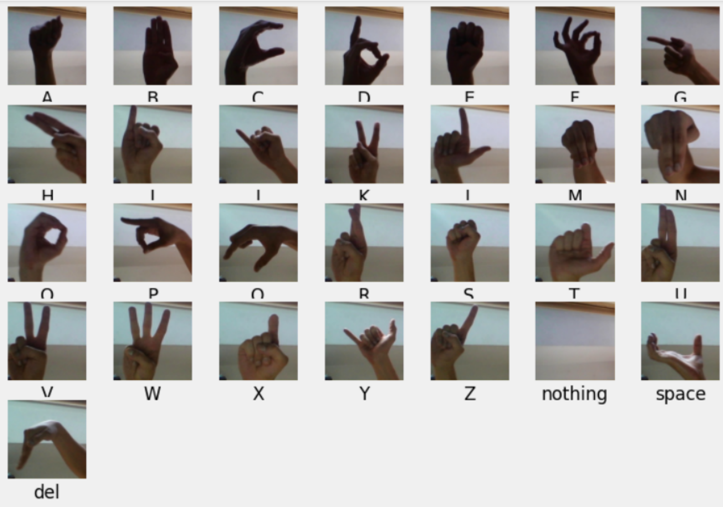

Figure 2: Pre-Processed image dataset

3.3 Model training

After successful splitting the data into train and test set, the model with different deep learning architectures has been trained over the American sign language dataset. Total we have utilized the 4 different deep learning architectures for training the model, which can be categorized into the two types Transfer Learning model and custom models. In the transfer learning model, we are using the two different algorithms that are VGG-19 and ResNet-50. On the other hand, for custom model, we have developed the model from scratch with different set of convolutional layers.

3.4 Model Evaluation

As the current dataset contains the 29 different set of classes, the sign language detection from hand gestures becomes a classification problem. For classification, the models can be evaluated based on the certain metrics such as Accuracy, Validation Loss, Precision, Recall and F1-Score. The model with minimum loss and maximum accuracy and PRF score will be considered as the most efficient model. For each model we will calculate these scores in order to perform the comparative analysis.

4 Design Specification

Deep learning algorithms plays an important role in making the prediction and analysing the pattern behaviour. Therefore, for sign language detection, four deep learning architectures are utilized. Among them, two transfer learning model for which we are utilizing the pre-trained weight for accurate results. The remaining two architecture are the CNN based Custom architectures where 3 and 5-layers of convolutional neural network is utilized.

CNN is a deep learning algorithm that works similar to the human brain, it is designed to mimic the working of the visual cortex. What it does is image recognition, it considers the image as a matrix having multiple rows and columns of 0s and 1s. It arranges this matrix in a linear manner and then starts processing each part of the image by assigning value to that part, it basically is adding importance to each section of the image which will help in recognition later. This helps the CNN model in differentiating one image from another, and the best part is it uses minimal computational power in doing that as compared to other classification models.

4.1 VGG-19 Model

It is a convolutional neural network that consists of 19 layers out of which 16 layers are convolutional layers and three are fully connected ones. As its database it uses Image Net database, it is a trained model of about a million images that has the capability of classifying an image on 1000 image categories like a car, bike, vehicle, utensil, all types of animals, numbers etc. It takes a 224 X 224 size of the image for processing and has a good categorization of images. This model is used with its pre trained weights which evolves the concepts of Transfer Learning. As the name suggests transfer learning transfers the knowledge of one model to the other model. What it does is use one model as the starting point of another model, this saves a lot of time as we don't have to start from scratch every time we are training a model and also a lot of computational power. Many a time in a lot of the models there exists a pre-trained one that can be used as the starting point or its outcome can be fed to the new one for better accuracy and less processing. For example, if you have a trained model that recognises animals through eyes you can use the same model for training a new one that uses ears or say nose for the recognition of animals. The architecture of VGG-19 is shown in Figure 3. We used the Transfer learning technique during this project and as pre-trained models, we used two CNN's namely, VGG19 and ResNet15.

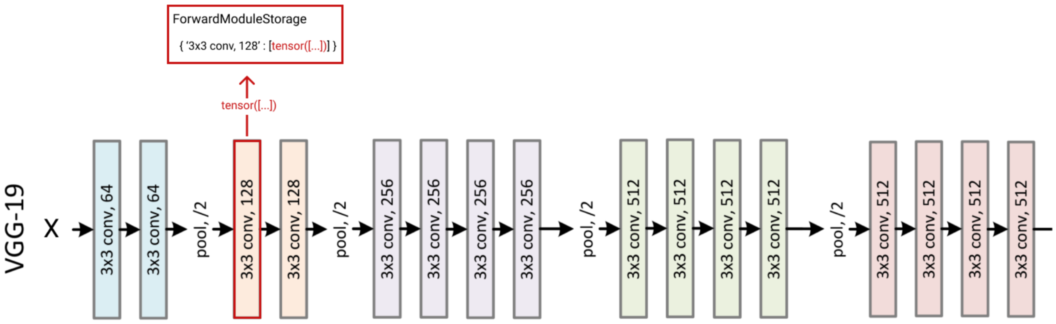

Figure 3: VGG-19 Architecture *A easy to use API to store outputs from forward/backward hooks in Pytorch* (2019)

4.2 ResNet-50 Model

ResNet is also known as residual network is a very popular neural network that solves the problem of training deep neural networks, with the introduction of ResNet it became quite feasible to train neural networks of more than 140 layers very easily. Before ResNet, there was a problem of vanishing gradients after each additional layer as gradients are back linked, every time a layer is introduced its value decreases and performance saturates after a certain number of layers. ResNet does this by using something called identity connections that solves the problem of vanishing gradients. The architecture of ResNet-50 is shown in Figure 4.

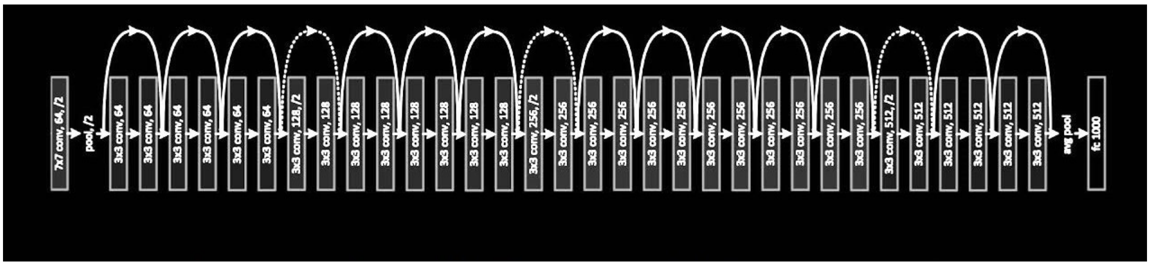

Figure 4: ResNet-50 Architecture *English: ResNet50* (2021)

4.3 Custom Model-1 (3 Layer CNN)

We are using 2 custom models here one has 3 layers and the second has 5 layers. It is a sequential model with 3 layers the first layer is a convolutional two-dimensional layer having 64 neurons with a (3, 3) filter keeping the padding set to same. It has the activation keyword set to “relu”. This layer is followed by a max pool layer with the pool size set to (2, 2). The 2nd layer is similar to the first one only it has 128 neurons with a (3, 3) filter keeping the padding set to the same. It has the activation keyword set to “relu”. This layer is followed by a max pool layer with the pool size set to (2, 2). The 3rd layer is again similar to the other two layers only here there are 256 neurons with a (3, 3) filter keeping the padding set to the same. It has the activation keyword set to “relu”. This layer is followed by a max pool layer with the pool size set to (2, 2). This model is compiled with Adam Optimizer and categorical Cross entropy is used as loss function. This model is trained on 10 epochs.

4.4 Custom Model-2 (5 Layer CNN)

This model is also a sequential model having 5 layers, the first layer is a 2D convolutional layer having 64 neurons with a (3, 3) filter, padding set to same. It has the activation keyword set to “relu”. This layer is followed by a max pool layer with the pool size set to (2, 2). The second layer is again a 2D convolutional layer having 128 neurons with a (3, 3) filter, padding set to same. It has the activation keyword set to “relu”. This layer is followed by a max pool layer with the pool size set to (2, 2). Here, we have also added a batch normalization function and a dropout function with a parameter set to the value of 0.5. The 3rd layer is same as the 2nd one with the only difference being that it has 256 neurons instead of 128. The fourth and fifth layers are the same as the first layer only with a different number of neurons having 256 and 512 neurons respectively. After these 5 layers, we have used a flatten() function followed by a dropout function with a parameter passed with the value of 0.5. It also has a dense function with 2 parameters which are value=1024 and activation=sigmoid and last layer model is activated by soft max function to predict 29 classes. This model is trained for 10 epochs.

5 Implementation

The deep learning model plays an important role in predicting the hand gestures of sign language. To implement the deep learning models the steps involved are data preprocessing and model training. There are 4 different deep learning architectures have been utilized which includes the VGG-19, ResNet50, 3-layer convolutional model and 5-layers convolutional model. All the implemented Models utilized the ADAM as an optimization function and categorical cross entropy as loss function. There are numerous number of libraries has been utilized in order to implement the proposed work which includes pandas, numpy, matplotlib, seaborn, tensorflow, keras, opencv, mediapip etc. The mediapip and opencv has been used to draw the bound box around the hand gestures and print the predicted label besides the hand gesture. Whereas, the matplotlib, pandas and seaborn library has been utilized for visualizing the results after training the models. This whole experiment is performed on google colab platform for training with python as programming language, After evaluating all model we have found Custom model (5 layer of CNN) as best optimal model to detect the hand gesture movement for sign language detection. This model has dumped into .h5 file for further use. The model has been utilized for for real time prediction system Our proposed system can identify the different hand gestures of sign-language in real time. Some of the gestured predicted by custom model test images is shown in Figure 5.

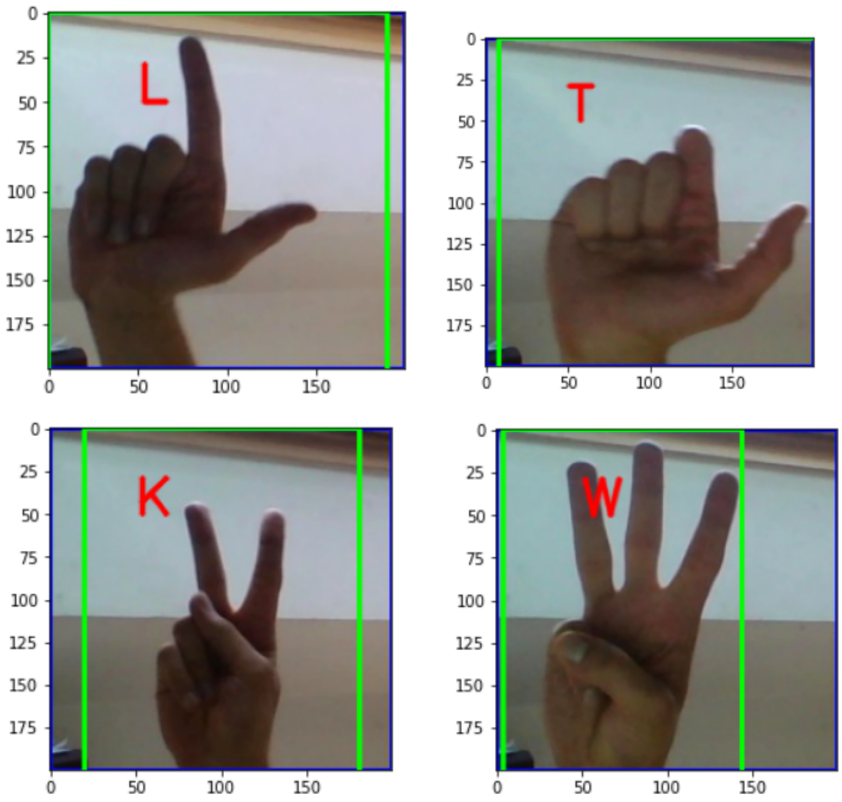

Figure 5: Detection of Hand Gestures on Test Image

The developed framework can also be deployed on cloud and also can be used in the local machine. As our task involves the image processing, GPU (Graphical Processing Unit) is the essential need to train the models. Therefore, in this work we have used tol Google colab platform for training the model. After training and evaluating the performance of each model, the best model can be saved for prediction, which can be either stored in cloud system or in a local machine.In order to implement the sign-language detection system, system with following specification is required as shown in Table 2.

Platform	Google Colab (for training)
GPU	Nvidia Tesla K80
RAM	12 GB
Hard Disk	65GB
Programming Language	Python
Python Libraries	OpenCV, Tensorflow, Keras, Pandas, Numpy, matpltolib, mediapipe

Table 2 : System Specification

6 Results and Evaluation

There are total 29 classes of different hand gestures in our dataset, which needs to be classified accurately. Since, it is a classification task we have utilized the classification models for predicting the the results. The different deep learning nerual network architectures are trained over the American sign language data. The architecture that will be used for sign language prediction are VGG-19, ResNet-50, Custom model (with 3 layers and 5 layers of CNN). The architecture of each deep neural network is different from one another. Therefore, in order to identify the most optimal model for sign language detection the each model needs to be evaluated. Before the evaluation the dataset has been splitted into training and test set. Each model has been trained over 10 number of epochs. All the models will be trained over the training data and their performance will be evaluated on test data. Over Test data, we will calculate the Accuracy, Precision, Recall and F1-Score for each model. The model having highest Accuracy and PRF score along with minimum validation loss will be considered as the most optimal model for sign language detection. In the upcoming subsections we will discuss about evaluation methods of each model and compared the results with line graphs of model with different metrics.

6.1 Evaluation

In order to identify the most optimal algorithm for sign language detection, each algorithm needs to be evaluated based on the certain different metrics. After reviewing the prior work, we have identified that pre-trained models such as VGG-19 and ResNet-50 has outperformed than traditional approaches. Therefore, in this work we have considered these models for evaluation. Other than that, in this research we have introduced the custom deep learning architectures based on convolutional neural network for improving the classification accuracy, where we have proposed 3-layer and 5-layer architectures. All of these models will be compared and evaluated based on the certain metric such as Accuracy, Precision, recall, F1-Score and validation loss.

6.1.1 Experiment 1 / Evaluation of VGG-19 Model

In this experiment the performed of VGG-19 architecture will be evaluated over American sign language data. This model is trained over 10 epochs on training data. Accuracy, Recall, precision, f1 score is calculated on test data which is used to compare various implemented models. over every epoch the accuracy and PRF score of model has been analyzed. For VGG-19 the weights of pre-trained weights of imagenet are used where the activation function is used as ReLU and on dense layer the softmax activation function is used. The results obtained after calculating the metrics for VGG-19 will be discussed in Result section.

6.1.2 Experiment 2 / Evaluation of ResNet-50 Model

The Resnet-50 Model also has been trained over the American sign language data, which contains the 87,000 images training data. The performance of model will be evaluated over the test image data. Like VGG-19, the resnet-50 architecture also will be trained over 10 number of epochs. For each epoch we will evaluate the model performance in terms of accuracy, precision recall and F1-Score. Pre-trained weights of imagenet has been utilized for training the ResNet-50 architecture. The results obtained after training the architecture will be discussed in the result section.

6.1.3 Experiment 3 / Evaluation of Custom Model (3 layers)

For improve the performance measure, we have developed the custom model from scratch, which utilizes the 3-layers of Convolution neural network, where the activation function is used as ReLu for each convolutional layer. Whereas, the two layers densenet is utilized with signmoid and softmax function. Same as other models, this model also has been trained over the 10 number of epochs and metrics such as Accuracy, PRF score are evaluated for every epoch. The results obtained using Custom Model (with 3-layers) will be discussed in the results section.

6.1.4 Experiment 4 / Evaluation of Custom Model (5 layers)

Another custom model, we have developed from scratch in order to improve the detection rate. In this model, we have used the 5-layers of convolutional neural network. The activation function used for convolutional layers is ReLu and for dense layer sigmoid and softmax function is used. The model will be trained over the 10 epochs and over every epoch the Accuracy and PRF score of model will be calculated. The results obtained by 5-layer CNN model will be discussed in next subsection.

6.2 Results

The results of each deep neural network architecture is calculated using the following metrics: Accuracy, Precision, Recall and F1-Score. As we already described that each model will be trained over 10 epochs, the results of every model will be visualized in the form of line graph.

6.2.1 Accuracy Score

Accuracy of any model can be calculated by the ratio of correctly predicted observation with total observations. Accuracy informs about the overall performance of the model. On analysing the graph as shown in Figure 6, it has been observed that over the 10th epoch, the accuracy score provided by VGG-19 architecture is 41%. Where the other transfer learning model, ResNet-50 achieved the accuracy score of 23.85%. Discussing about the custom models for 3-layer architecture we have obtained the accuracy score of 54.48%. Whereas, for 5-layer architecture the accuracy score is 58.33%. After successfully calculating the results we can say that our custom model with 5-layers has outperformed as compared to the other models. Where the very poor results has been obtained using ResNet-50 architecture followed by VGG-19.

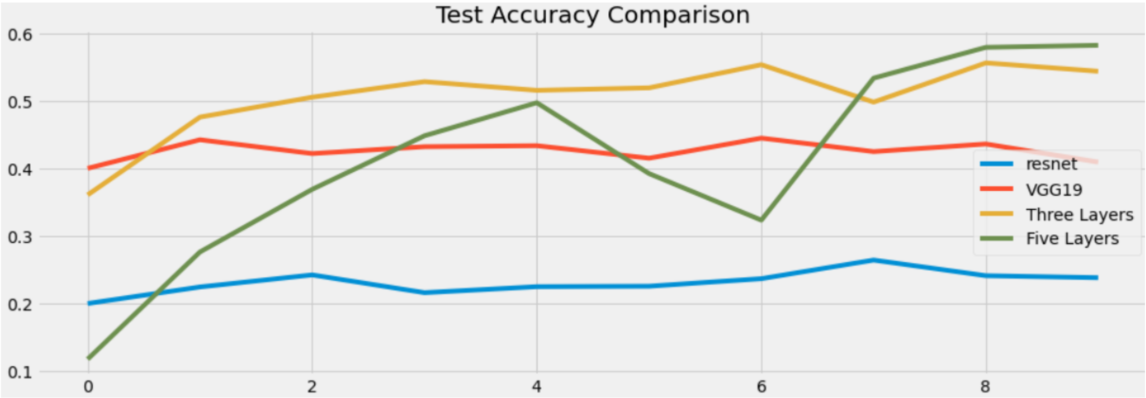

Figure 6: Comparison of models based on Testing Accuracy

6.2.2 Precision Score

Precision score is used in the sign language detection for evaluating the model performance, this score mainly informs about false positive values available in dataset. Higher precision score, less number of false positive values. The precision score obtained over test data by VGG-19 model is 0.425 and ResNet-50 is 0.5708. For custom mode the precision score obtained over test data is 0.6162 and 0.6479 for 3 layers and 5-layer architecture. The high precision score also has been obtained using 5-layers of convolutional architecture.

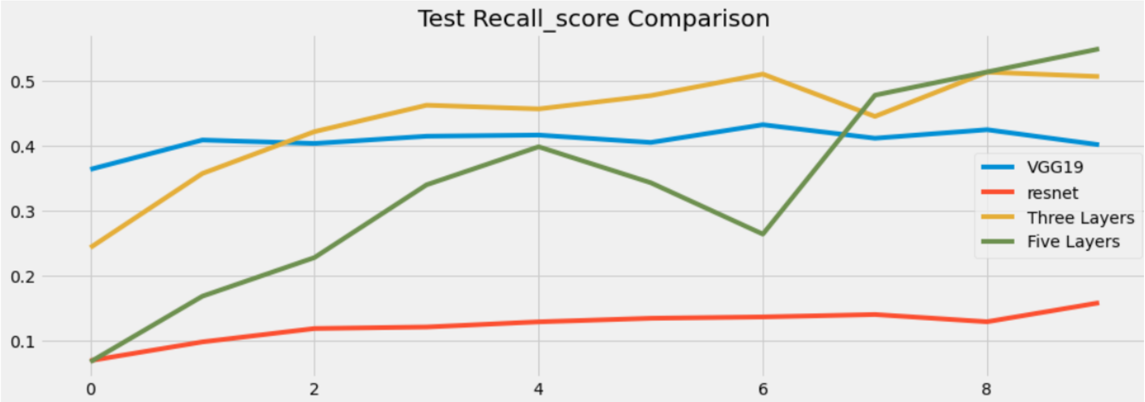

Figure 7: Comparison Of models based on Precision

6.2.3 Recall Score

Recall score is used in the sign language detection for evaluating the model performance, this score mainly informs about false negative values available in dataset. Higher Recall score, less number of false negative values. The recall score obtained over test data by VGG-19 model is 0.4007 and ResNet-50 is 0.157 For custom mode the recall score obtained over test data is 0.5057 and 0.5482 for 3 layers and 5-layer architecture. The high recall score also has been obtained using 5-layers of convolutional architecture.

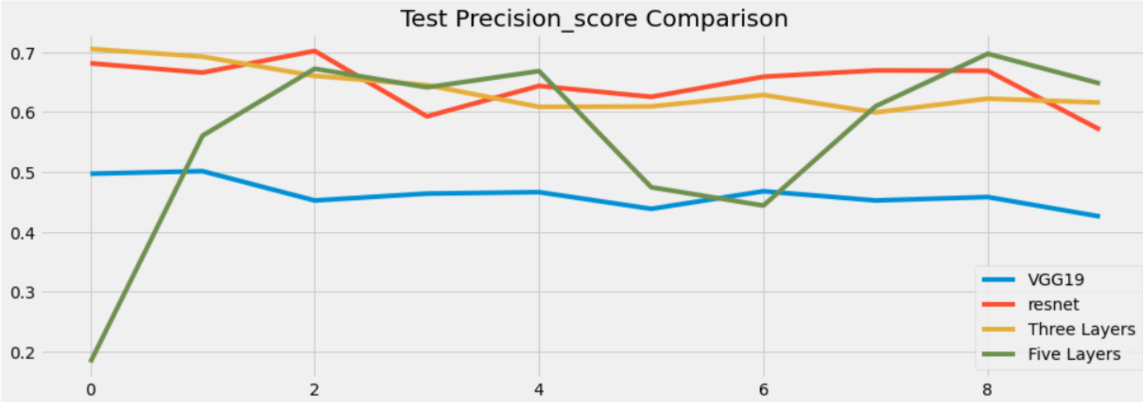

Figure 8: Comparison Of models based on Recall

6.2.4 F1-Score

F1-Score is the combined score obtained from Precision and Recall it can be derived by the following formula.

$$F1 = 2 \text{ x } \frac{Precision*Recall}{Precision+Recall}$$

The F1-Score obtained by VGG-19, ResNet-50, 3-layer architecture and 5-layer architecture are 0.4009, 0.2165, 0.5410 and 0.5910. The highest F1-Score has been obtained using 5-layer architecture, which is custom developed model.

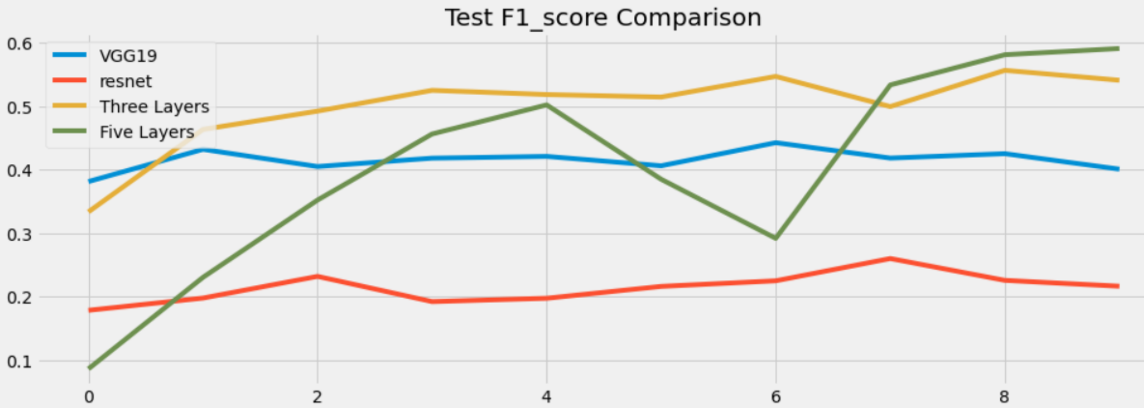

Figure 9: Comparison Of models based on F1-score

6.2.5 Validation Loss

The validation loss represents the behaviour of the model over validation set. Model with minimum loss will be considered as the best model. As per our analysis, we have obtained the higher validation loss using VGG-19. On observing the graph it has been found that after certain number of epoch the loss of VGG-19 architecture was increasing which represents the overfitting nature of model. The loss comparison of all the models is shown in Figure 10

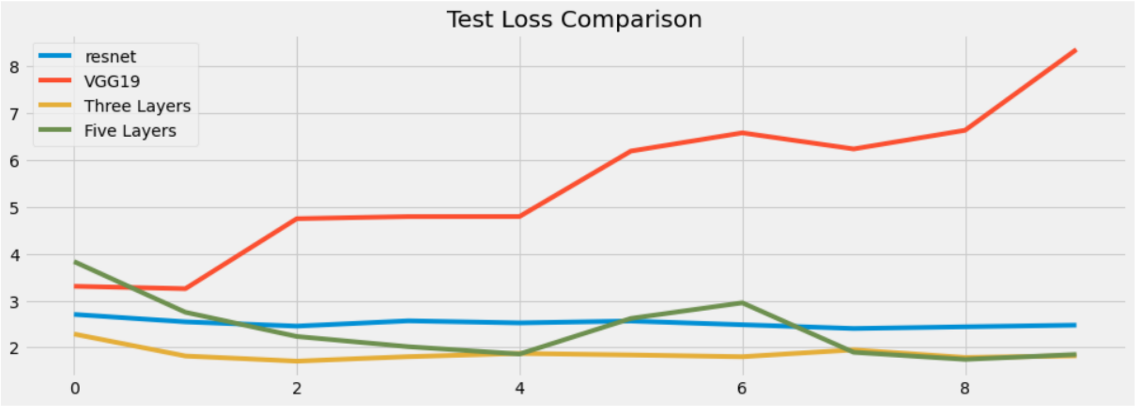

Figure 10: Comparison Of models based on Validation Loss

However, the minimum loss has been obtained using 3-layer and 5-layer architecture, which are the proposed custom models. ResNet-50 in terms of loss performs better than VGG-19, however due to overfitting VGG-19 provides the better accuracy and PRF Score.

7 Discussion

Thereafter performing the 4 experiments on deep learning models it is observed that the our custom model with 5 layers has outperformed than all other models. We have also used pre-trained models such as VGG-19 and ResNet 50 with image net weights but still our custom has performed well in comparison of those. These applied models are analysed based on validation accuracy, Precision, recall and f1 score, so that we can make a comparison between the model. Using custom model the obtained validation accuracy is 58.33%, Precision is 0.6479, Recall is 0.5482, F1-score is 0.5910. Since this model is selected as best optimal model hence this model has been used further for prediction. We have utilized the best model for real-time prediction of hand-gestures for sign language detection. We also identified poor performance of ResNet-50 model and the overfitting in VGG-19 model after certain number of epochs. Therefore, these models can not be considered as an efficient model for our prediction.

8 Conclusion

Detection of sign language from hand gestures is a challenging task and still a open area of research.In this research, we have implemented different deep learning models to identify and correctly predict the different hand gestures for sign language detection. The field of this research that is sign language identification is a vast and ever evolving field of work, where different deep learning models and algorithms are being used to accomplish this task with utmost accuracy. In this research, we have analysed that pre-trained model such as VGG-19 and ResNet-50 are not doable for predicting the hand gestures correctly. However, the custom model with 5-layers of convolutional has achieved the better outcomes as compared to the transfer learning models. The highest accuracy obtained after this research is 58.33%, which is not very high and can be improved in the future work. Our current proposed work is able to identify the characters from hand gestures. However, in the future work the complete sentence or words should be predicted based on the video frames data. Along with the model, dataset also plays an important role for optimizing the performance. A better quality of input data can make the predictions more better, this can be incorporated in the future work for obtaining the better results.

References

Abiyev, R. H., Arslan, M. and Idoko, J. B. (2020). Sign Language Translation Using Deep Convolutional Neural Networks, *KSII Transactions on Internet and Information Systems (TIIS)* **14**(2): 631–653. Publisher: Korean Society for Internet Information.
URL: *https://www.koreascience.or.kr/article/JAKO202011161036129.page*

A easy to use API to store outputs from forward/backward hooks in Pytorch (2019).
URL: *https://pythonawesome.com/a-easy-to-use-api-to-store-outputs-from-forward-backward-hooks-in-pytorch/*

ASL Alphabet (n.d.).
URL: *https://kaggle.com/grassknoted/asl-alphabet*

Chong, T.-W. and Lee, B.-G. (2018). American Sign Language Recognition Using Leap Motion Controller with Machine Learning Approach, *Sensors* **18**(10): 3554. Number: 10 Publisher: Multidisciplinary Digital Publishing Institute.
URL: *https://www.mdpi.com/1424-8220/18/10/3554*

Dabre, K. and Dholay, S. (2014). Machine learning model for sign language interpretation using webcam images, *2014 International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA)*, pp. 317–321.

Elakkiya, R. (2021). Machine learning based sign language recognition: a review and its research frontier, *Journal of Ambient Intelligence and Humanized Computing* **12**(7): 7205–7224.
URL: *https://doi.org/10.1007/s12652-020-02396-y*

English: ResNet50 (2021).
URL: *https://commons.wikimedia.org/wiki/File:ResNet50.png*

Gattupalli, S., Ghaderi, A. and Athitsos, V. (2016). Evaluation of Deep Learning based Pose Estimation for Sign Language Recognition, *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '16, Association for Computing Machinery, New York, NY, USA, pp. 1–7.
URL: *https://doi.org/10.1145/2910674.2910716*

Hasan, M., Sajib, T. H. and Dey, M. (2016). A machine learning based approach for the detection and recognition of Bangla sign language, *2016 International Conference on Medical Engineering, Health Informatics and Technology (MediTec)*, pp. 1–5.

Jalilian, B. and Chalechale, A. (2013). Face and Hand Shape Segmentation Using Statistical Skin Detection for Sign Language Recognition.

Kindiroglu, A. A., Ozdemir, O. and Akarun, L. (2019). Temporal Accumulative Features for Sign Language Recognition, *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, IEEE, Seoul, Korea (South), pp. 1288–1297.
URL: *https://ieeexplore.ieee.org/document/9022275/*

Konstantinidis, D., Dimitropoulos, K. and Daras, P. (2018). A Deep Learning Approach for Analyzing Video and Skeletal Features in Sign Language Recognition, *2018 IEEE International Conference on Imaging Systems and Techniques (IST)*, pp. 1–6. ISSN: 1558-2809.

Konwar, A. S., Borah, B. S. and Tuithung, C. T. (2014). An American Sign Language detection system using HSV color model and edge detection, *2014 International Conference on Communication and Signal Processing*, pp. 743–747.

Kulhandjian, H., Sharma, P., Kulhandjian, M. and D'Amours, C. (2019). Sign Language Gesture Recognition Using Doppler Radar and Deep Learning, *2019 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6.

Mustafa, M. (2021). A study on Arabic sign language recognition for differently abled using advanced machine learning classifiers, *Journal of Ambient Intelligence and Humanized Computing* **12**(3): 4101–4115.
URL: *https://doi.org/10.1007/s12652-020-01790-w*

PURNAMASARI, D. and Erwin, E. (2019). *ANALISIS STUDI KASUS MULTIVARIAN INTENSITAS DENGAN PERBANDINGAN METODE SEGMENTASI COLOR HISTOGRAM HSV, YCBCR, L*A*B (CIELAB) DAN K-MEANS CLUSTERING WARNA PADA FINGERSPELLING AMERICAN SIGN LANGUAGE (ASL)*, undergraduate, Sriwijaya University.
URL: *https://repository.unsri.ac.id/17127/*

Rosero-Montalvo, P. D., Godoy-Trujillo, P., Flores-Bosmediano, E., Carrascal-García, J., Otero-Potosi, S., Benitez-Pereira, H. and Peluffo-Ordóñez, D. H. (2018). Sign Language Recognition Based on Intelligent Glove Using Machine Learning Techniques, *2018 IEEE Third Ecuador Technical Chapters Meeting (ETCM)*, pp. 1–5.

Saquib, N. and Rahman, A. (2020). Application of machine learning techniques for real-time sign language detection using wearable sensors, *Proceedings of the 11th ACM Multimedia Systems Conference*, MMSys '20, Association for Computing Machinery, New York, NY, USA, pp. 178–189.
URL: *https://doi.org/10.1145/3339825.3391869*

Taskiran, M., Killioglu, M. and Kahraman, N. (2018). A Real-Time System for Recognition of American Sign Language by using Deep Learning, *2018 41st International Conference on Telecommunications and Signal Processing (TSP)*, pp. 1–5.

Wadhawan, A. and Kumar, P. (2020). Deep learning-based sign language recognition system for static signs, *Neural Computing and Applications* **32**(12): 7957–7968.
URL: *https://doi.org/10.1007/s00521-019-04691-y*

Wazalwar, S. S. and Shrawankar, U. (2017). Interpretation of sign language into English using NLP techniques, *Journal of Information and Optimization Sciences* **38**(6): 895–910. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/02522667.2017.1372136.
URL: *https://doi.org/10.1080/02522667.2017.1372136*