Sign Language Hand Gesture Identification Using Yolov3

Dhruv Karotra[1], Harshal Tamboli[2], Shashwat Gaikwad[3] , Brijesh Mali[4], Prof. Suhas J. Lawand.[5]

[1-4]*Department of Information Technology & Pillai College of Engineering, Navi Mumbai, India*
[5]*Department of Information Technology & Pillai College of Engineering, Navi Mumbai, India*

---***---

Abstract - The community of hearing and speech-impaired people has been dependent on sign language as a medium to communicate with people. It has always been difficult to communicate between a verbally impaired person and a normal person. The Sign Language Identifier will be very essential in communicating with deaf people. It will help people to communicate with verbally impaired people by detecting the signs and providing the output in real time. The model is based on You Only Look Once(YOLO). We would be building the model with Yolov3 version. The existing recognition systems are less reliable as they detect letters and take time to make a single word or they detect very few words. Our model will solve this problem as it will recognize words directly. The dataset will be built using LabelImg software and detect sign language through a live feed. The system is a Graphical User Interface (GUI) that is built by drag and drop functionality through PAGE tool. The identification of signs will take place by two methods, one by uploading an image and the other through a live feed. These images and live inputs will be taken by using OpenCV.

***Key Words*:** Sign Language Identifier, YOLO, LabelImg, PAGE, Live Feed, OpenCV.

1. INTRODUCTION

With the advent of innovation and technology, people's lives have become much easier. The drastic growth in technology has always given good fruit to society. It has always solved the problem of disabled people but people with speech or hearing impairment have always found it difficult to communicate with the general audience.

Communication will be always an important factor to share knowledge and socially interact with a person. The main problem that hearing or speaking-impaired people faced was that the public could not understand sign language as they were not aware of this language. So, this was the major barrier that has been broken with the help of technology.

For normal people, translators were the must to communicate with hearing or speaking-impaired persons. The translator is not always available and even it was a costlier option. Texting was fine up to some extent, but it could not solve the bigger problem. So, to overcome the above-mentioned problems, a very efficient system was necessary; hence, we have developed a Sign Language Identifier using Deep Learning techniques. This system enables the impaired person to communicate easily with sign language, and the application will convert this sign into text which will be displayed to the user who does not know sign language.

For the validation of the hand signs which are recognized by this Deep Learning method, they are referred from the official website of Indian Sign Language called indiansignlanguage.org. This website contains many visual hand signs for various words and phrases.

Welcome to Indian Sign Language Portal
ONLINE PORTAL
INDIAN TECHNICAL SIGN LANGUAGE
DICTIONARY
॥ आत्मनो मोक्षार्थं जगद्धिताय च ॥
by RKMVERI, Faculty of Disability Management and Special Education, Coimbatore
ISL Intro
0:00 / 2:44
YouTube
search the visual indian sign language dictionary here......
SEARCH

Indiansignlanguage.org offers a huge collection of Indian Sign Language (ISL) signs. Each sign has an image, running video and threaded discussions. It is an ideal resource to use while you learn/teach Indian Sign Language. Each sign has an image, running video and threaded discussions. We are continually adding more signs and designing new services to empower the Deaf.
Please share your ideas and comments and help us make this service better.

Fig 1 : Indian Sign Language Website [7]

2. LITERATURE SURVEY

Author Aman Pathak [1] proposes a method for real-time sign language detection using machine learning techniques. The system is designed to identify signs from the American Sign Language (ASL) and display the corresponding text in real-time. The proposed system uses a combination of hand tracking and convolutional neural networks (CNNs) for sign language detection. The hand tracking algorithm is used to detect the hand in the video stream, while the CNNs are used to classify the hand gestures into corresponding ASL signs. The authors trained the CNNs using a publicly available dataset of hand gestures, and evaluated the system using a separate dataset of ASL signs. Custom gestures can be added easily. Increase in images at different angle and frames will improve the accuracy of the model. This system shows recognition rate of 70-80%. Overall, the proposed real-time sign language detection system has the potential to be a useful tool for the deaf and hard-of-hearing community, enabling them to communicate more effectively and efficiently.

The Author R Rumana [2] provides an overview of the current state of research on sign language recognition systems. It then reviews the existing techniques for sign language recognition, including vision-based approaches, data glove-based methods, and machine learning-based techniques. The system has two modules, namely: data acquisition, pre-processing and feature extraction and sign language gesture classification. The paper also highlights the challenges in sign language recognition such as variation in sign language, background noise, and lighting conditions. Two layers of algorithms are used to verify and predict symbols which are similar to each other. Improved input image size and dataset provide more accuracy. The final accuracy of 92% is achieved on the dataset. Overall, the paper provides a comprehensive review of the current state of research on sign language recognition and identifies areas for future research.

The author Amrita Thakur [3] proposes a system that can recognize and translate sign language gestures into spoken language in real-time. The video of hand signs was taken and was split into frames to improve the classification. The system uses a wearable device equipped with sensors to detect hand movements and a deep learning algorithm to classify the gestures. The proposed system was evaluated on a dataset of American Sign Language (ASL) gestures, and the results showed an accuracy of 99.62% in recognizing 26 ASL hand signs. The system also includes a speech generation module that converts the recognized signs into spoken language. System also supports text to sign conversion this helping in two way communication. The research paper concludes that the proposed system could have important practical applications, such as helping individuals with hearing impairments communicate with people who do not know sign language. This system could be improved by expanding the dataset to include more sign language gestures.

3. SYSTEM ARCHITECTURE

A. Proposed System Architecture

The proposed system uses a Deep Learning technique to predict the hand signs. It uses the YOLO (You Look Only Once) algorithm and the version used is Yolov3.
The system includes the following steps :
a. Image Capture and Labelling : Images are labelled using the LabelImg application. The annotations and Numeric ID's are saved in a .txt file.
b. Darknet Yolo : The official darknet repository is loaded from AlexeyAB's github profile and changes are made to the configuration file of Yolov3 according to our requirements.
c. Training the model : The initial yolo weights used is darknet53.conv.74 which is necessary to train the darknet yolov3 model.
d. Save and Import weights : The weights file generated after training the model is saved and imported to backend for prediction.
e. GUI : PAGE tool is used for building the GUI of our system. It uses drag and drop mechanism to build the GUI.
f. Input : After logging in in the GUI, the user is asked to two inputs - one by image upload and other by live feed through the webcam.
g. Output : The predicted result is displayed to the user along with the accuracy of the predicted hand sign.

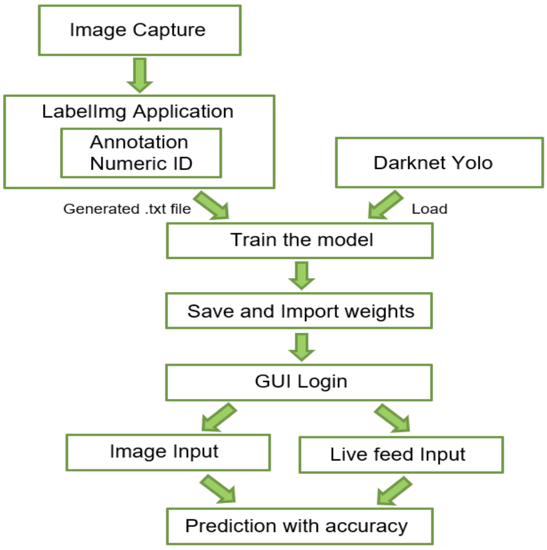


Fig 3.1 : Proposed System Architecture

B. Dataset

The Yolo model was created using our own dataset. The dataset consists of 10 hand signs, each of which has an average of 85-90 images. There are a total of 870 images in our dataset. Some of the images were taken from the laptop's webcam and some from the mobile's camera. The application used for labelling the images is LabelImg. All of the images were labelled and transformed into a Yolo text file. This text file contains two important things - one is the class name and the other is the dimensions of the labelled region. Once all the labelling is done, all the images are saved in a zip file and uploaded to Google Drive. The training is done on Google Colaboratory where the zip file is exported and trained using Yolov3. For training purpose, the free version of Google Colaboratory is used as it provides sufficient memory and space for the training of our dataset.

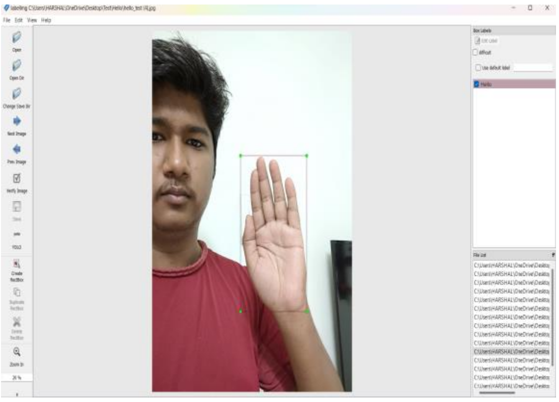
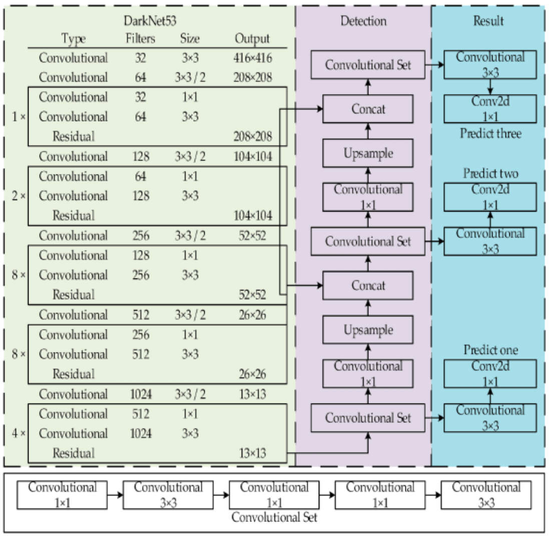


Fig 3.2 : Labelling images using LabelImg

C. Yolov3

The YOLOv3 model is based on the following concepts :

1. Darknet-53 : Yolov3 is a Deep Learning model which is based on Darknet-53. The Darknet-53 consists of a 53 layer network which is trained on ImageNet dataset. For the prediction, it adds more 53 layers resulting in a 106 layer of fully convoluted network.



DarkNet53

	Type	Filters	Size	Output
	Convolutional	32	3×3	416×416
	Convolutional	64	3×3 / 2	208×208
1 ×	Convolutional	32	1×1	
	Convolutional	64	3×3	
	Residual			208×208
	Convolutional	128	3×3 / 2	104×104
2 ×	Convolutional	64	1×1	
	Convolutional	128	3×3	
	Residual			104×104
	Convolutional	256	3×3 / 2	52×52
8 ×	Convolutional	128	1×1	
	Convolutional	256	3×3	
	Residual			52×52
	Convolutional	512	3×3 / 2	26×26
8 ×	Convolutional	256	1×1	
	Convolutional	512	3×3	
	Residual			26×26
	Convolutional	1024	3×3 / 2	13×13
4 ×	Convolutional	512	1×1	
	Convolutional	1024	3×3	
	Residual			13×13

Fig 3.3 : Darknet-53 Layers [8]

The detection of Yolov3 happens at three layers - the 82nd layer, the 94th layer, and the 106th layer. In our case, the size of image is 416 x 416 which is downsampled for the first 81 layers. Then at the 82nd layer, the first prediction is made resulting in a feature map of size 13 x 13 due to downsampling. Then few layers are upsampled by 2x to get a resultant feature map of the size of 26 x 26 in the 94th layer. The same procedure is followed again, by upsampling which gives a feature map of size 52 x 52 in the 106th layer.

2. Intersection over Union (IOU) : Intersection over Union (IOU) is the ratio between the area of overlap and the area of union. The purpose of IOU is to benchmark the predicted accuracy of the trained model. The higher the overlap of the predicted bounding box with the ground truth bounding box, the better is the performance of the trained model. There can be more than one predicted bounding box for an image. So, according to the IOU score, the best bounding box is displayed during the NMS procedure.

3. Objectness and Co-ordinate Loss : With each bounding box prediction, there is associated a predication called Objectness. The Objectness loss is used to predict the correct class and the Co-ordinate loss is used to predict the better bounding box. By default, the predicted bounding boxes with very low objectness score (<0.005) are removed and do not make it to the NMS. So most of the boxes are eliminated and very few are passed to the NMS.

4. Non-Maxima Suppression (NMS) : Yolov3 also uses a technique called Non-Maxima Suppression (NMS). This technique selects the best bounding box according to the IOU and threshold limit and suppresses the other bounding boxes predicted. The remaining bounding boxes which are passed in the Co-ordinate loss are then passed to the NMS. In case, if there are two bounding box predicted for the same image then the one with the best IOU will be considered and the weaker bounding box will be suppressed. In our case, the threshold value is set to 0.3.

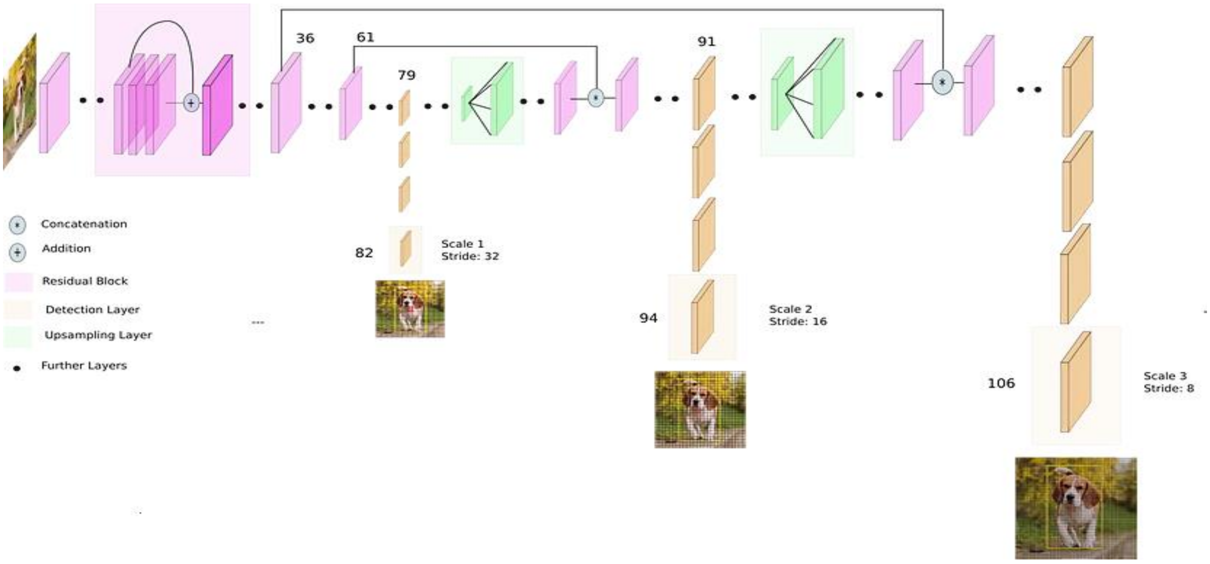


Fig 3.4 : Yolov3 Network Architecture [9]

4. RESULT AND DISCUSSION

The proposed Yolo model predicts 10 different words which are labelled by using the LabelImg application. These words are Hello, Yes, No, I Love You, Good, Ok, Pray, Call, Thank You and House. The user can login to the system and then can use the prediction modules of the Sign Language Identifier through the system's dashboard.
The results and the sample screenshots of the proposed system are provided in the given section.

A. Proposed System GUI Screenshots

1. Login Page

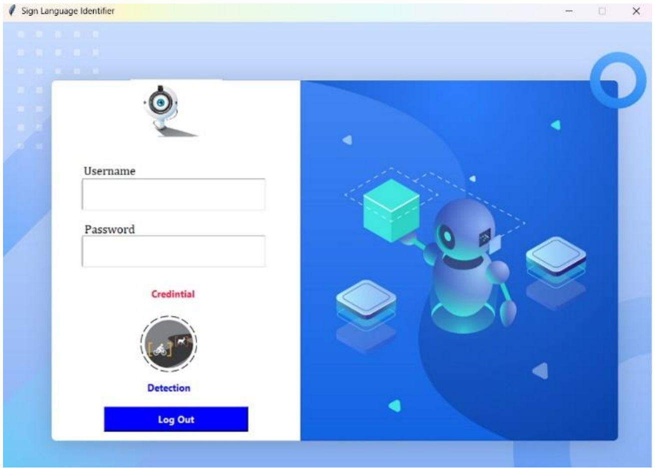


Fig 4.1 : Login Page

2. Dashboard

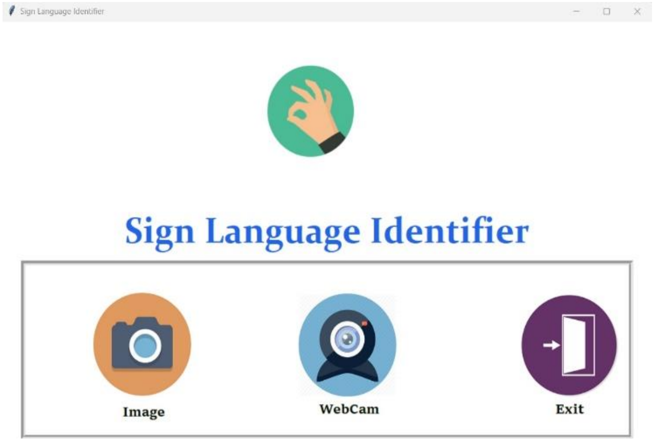


Fig 4.2 : Dashboard

B. Prediction Results

There are two prediction modules of the system- one is the image module and the other is the live feed module. These modules are available on the system GUI's dashboard.

1. Image Module Prediction

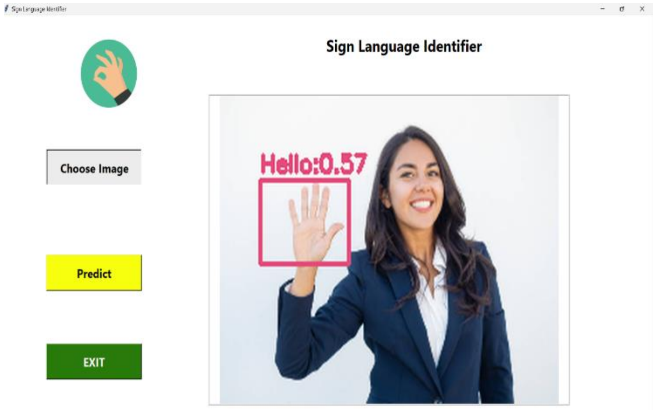


Fig 4.3 : Image Module Prediction

2. Live Feed Module Prediction

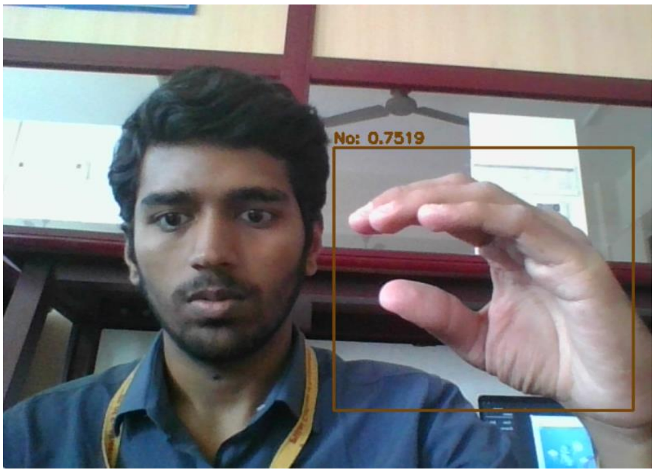


Fig 4.4 :Live Feed Module Prediction

C. Performance Evaluation

Here are some parameters which were initialized to train the model :

batches=64, subdivisions=16, width=416, height=416. To train Yolov3 the model, the formula for the number of maximum batches is given by :

max_batches=(No.of classes)*2000. In our case, the number of classes is 10. So the max_batches is set to 20000.

We observed that after 1000 epochs the graph for the avg loss was almost constant and was not further decreasing by a sufficient amount. So, the weights file generated at 1000 epochs was considered as after 1000 epochs the model was over-trained since there was no reduction in object loss.

The mAP (mean average precision) was calculated on the training data and mAP was observed to be 99.98%. While testing the model the mAP was observed to be 92.71%.

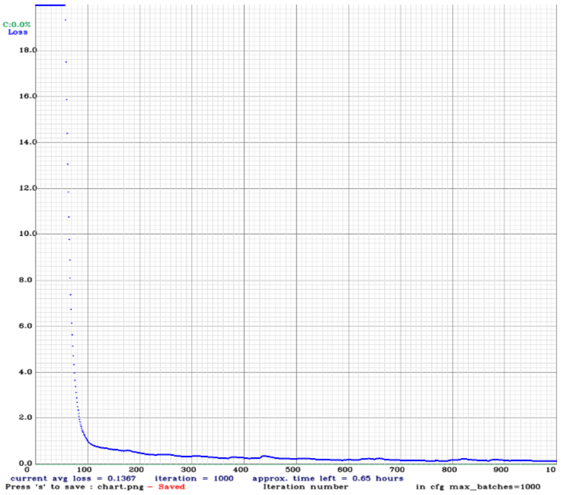


Fig 4.5 : Model Average Loss

Table 4.1 : Parameters and results

Parameter	Results
Technique	Yolov3
Number of Epochs	1000
Training Model mAP	99.98 %
Testing Model mAP	92.71 %
Model Average Loss	0.1367

5.CONCLUSION

Communication with the help of sign language is still a difficult task as there are too many phrases or words which need to be recognized for the one who is communicating with the deaf people. This system came with the solution of recognizing 10 words which are used frequently on a daily basis. For recognizing the signs, we created a YOLO (You Look Only Once) model which is trained on hand images and used GPU in order to increase of performance.

As the motive of this system is to help people understand the sign language in real time, the model was created to predict the signs in a real time environment so that the hand signs can be understood quickly during communication.

REFERENCES

[1] Aman Pathak, Avinash Kumar, Priyam, Priyanshu Gupta and Gunjan Chugh, "Real Time Sign Language Detection" 2022 International Journal for Modern Trends in Science and Technology, ISSN: 2455-3778, doi:10.46501/IJMTST0801006

[2] R Rumana, Reddygari Sandhya Rani, Mrs. R. Prema, "A Review Paper on Sign Language Recognition for The Deaf and Dumb" 2021 International Journal of Engineering Research and Technology , ISSN: 2278-0181, doi: 10.17577/IJERTV10IS100129

[3] Amrita Thakur, Pujan Budhathoki, Sarmila Upreti, Shirish Shrestha, Subarna Shakya, "Real Time Sign Language Recognition and Speech Generation" 2020 Journal of Innovative Image Processing, ISSN: 2582-4252 doi: 10.36548/jiip.2020.2.001

[4] Steve Daniels, Nanik Suciati, and Chastine Fathichah. "Indonesian Sign Language Recognition using YOLO Method" 2021 IOP Conference Series: Materials Science and Engineering, doi: 10.1088/1757-899X/1077/1/012029

[5] Ali Mahmood AL-Shaheen1, Mesut Çevik (Adviser) and Alzubair Alqaraghuli, "American Sign Language Recognition using YOLOv4 Method" 2022 International Journal of Multidisciplinary Studies and Innovative Technology., doi: 10.36287/ijmsit.6.1.61

[6] N. Mallikarjuna Swamy, H.S. Sumanth, Keerthi, C. Manjunatha, "Indian Sign Language detection using Yolov3", pp.157-168, doi: 10.1007/978-981-16-9885-9_13

[7] https://indiansignlanguage.org/

[8] saiuachandler.blogspot.com – Yolov3 architecture calculating loss.

[9] towardsdatascience.com - Digging deep into YOLO V3 - A hands-on guide Part 1