SIGN LANGUAGE TO SPEECH CONVERTER FOR INDIAN LANGUAGES

S. Dhamodaran, M.E.,
Assistant Professor,
Department of Computer Science and Engineering, AVS Engineering College, Ammapet, Salem,
Tamilnadu, India.

V. M. C. Venkatraja,
Department of Computer Science and Engineering,
AVS Engineering College, Ammapet, Salem,
Tamilnadu, India.

S. Vignesh,
Department of Computer Science and Engineering,
AVS Engineering College, Ammapet, Salem,
Tamilnadu, India.

S. Naveen,
Department of Computer Science and Engineering,
AVS Engineering College, Ammapet, Salem,
Tamilnadu, India.

A. Vaseekaran,
Department of Computer Science and Engineering,
AVS Engineering College, Ammapet, Salem,
Tamilnadu, India.

Abstract : Sign language is a very essential communique tool for plenty deaf and mute human beings. So we proposed a version to understand sign gestures the usage of YOLOv5 (You only look as soon as version 5). This version can locate sign gestures in complicated surroundings also. For this model we were given the accuracy of 88.4% with precision of 86.6% and don't forget of 87.2%. The proposed model has evaluated on a labeled dataset Roboflow. Moreover we added some photographs for schooling and checking out to get higher accuracy. We as compared our model with CNN (convolutional neural network) in which we were given accuracy of 91.98%. We checked this model for actual time detection also and got the accurate consequences.

1. Introduction

In our surrounding we are able to see there are people having various disabilities and a few of them are located to be deaf and mute. To talk with others, those humans want to learn sign language and ordinary people are unable to apprehend signal language. This problem reasons miscommunication among humans. due to this miscommunication mute human beings can live remoted from society. they can't able to take part in social occasions or any discussion. This create big hole among ordinary human beings and people with disabilities. we can reduce this gap through using technologies like pc vision, deep gaining knowledge of and so on. So this is the primary motive to choose this undertaking. Our assignment built the model to understand the sign language from the person proposed CNN algorithm for sign language device but CNN is considerably slower due to an operation which include maxpool and we checked CNN for actual time detection but it gave misguided outcomes. Hand detection and(may be mute or normal) and translated it into the comprehensible text. There are many item detection algorithms in deep getting to know. This paper includes the evaluation between popular CNN and
YOLO and why YOLO is better. There are special versions of YOLO like YOLO v1,

v2, v3, v4 and v5. In our model we have used the brand new version of YOLO that is YOLOv5. YOLO v5 version runs approximately 2.5 instances faster than other versions at the same time as handling better performance in detecting smaller objects. Our model can hit upon the static pictures as well as gestures from on camera (video). Image processing needs to be carried out in CNN which will increase processing time. In assessment, YOLO is specially evolved for actual time machine. We used version of YOLO i.e. YOLOv5 in our task.

2. Problem Statement

Convolutional neural community has max pooling layers which cause gradual processing. CNN has many layers for training, so the computer takes a whole lot of time for training the version. CNN calls for loads of facts factors for training the model. In contrast to CNN, coordinate frames can't be used. those coordinate frames are the component of pc vision. these frames are used to keep music of the orientation and extraordinary capabilities of an item. In actual time detection we need to define the body for detection of objects. it will stumble on images most effective in confined region. So this is the primary disadvantage of CNN. YOLO can locate pics at any position with rapid processing. So that is the primary purpose why are we choosing YOLO.

3. PROPOSED SYSTEM

In our project we essentially consciousness on producing a model which can recognize in finger spelling based totally hand gestures in order to form a entire phrase with the aid of combining each gesture.

ADVANTAGES

- Excessive accuracy
- Automated
- No human power
- High efficient and consumer pleasant

4. Methodology

First, we select the dataset language and we get images from this data. We use Image Processing to convert these images to pixels. We made this picture for CNN. Then in a separate dataset, we split this data for training and testing purposes. Using this training and testing model, we trained and tested our models. Finally, we create a user interface for actual detection time . If the images/features in the training data are skewed or rotated, the CNN will have difficulty classifying the images.

5. Components Used

Hardware Requirements:

System : Intel 6.0

Hard Disk : 250 GB

Ram : 2 GB

Monitor : 14" Color Monitor

Mouse : Optical Mouse

Software Requirements:

Operating System : Windows 8(64 bits)

Front End : Anaconda, Python

Database : Dataset

Python: Python is a high-level programming language designed to be easy to read and easy to enforce. it's miles opensource, which means it is unfastened to use,even for commercial programs. Python canrun on Mac, windows, and Unix structuresand has also been ported to Java and .netdigital machines. Python is a reasonably oldlanguage created via Guido Van Rossum.The layout started within the late Eightiesand turned into first launched in February1991.

Deep Learning: Deep studying strategies intention at mastering characteristic hierarchies with functions from higher stages of the hierarchy formed by means of the composition of decrease stage features. Mechanically mastering functions at multiple tiers of abstraction permit a system to examine complicated features mapping the enter to the output directly from records, with out relying completely on human crafted functions. Deep learning algorithms are seeking to make the most the unknown structure within the input distribution so that you can discover exact representations, frequently at more than one levels, with better-level found out capabilities defined inphrases of lower-degree functions.

OpenCV: OpenCV (Open source computer vision Library) is open supply pc vision and gadget learning software program library. OpenCV was built on provide common infrastructure laptop imaginative and prescient programs and to accelerate using gadget belief commercial products. Being a BSD certified product, OpenCV makes it easy groups to apply and alter the code.

Tensorflow: TensorFlow is a free and open-source software library for dataflow and differentiable programming across a range of obligations. It's miles a symbolic math library, and is likewise used for device getting to know programs which include neural networks. It is used for both studies and manufacturing at Google, TensorFlow is Google brain's second-generation device.

ALGORITHM

RCNN: R-CNN is a two-stage detection set of rules. The first stage identifies a subset of regions in an image that could comprise an item. The second one stage classifies the item in each area. Programs for R-CNN item detectors.

Include: Independent driving.

Mask RCNN: Mask R-CNN is a popular deep learning instance segmentation technique that

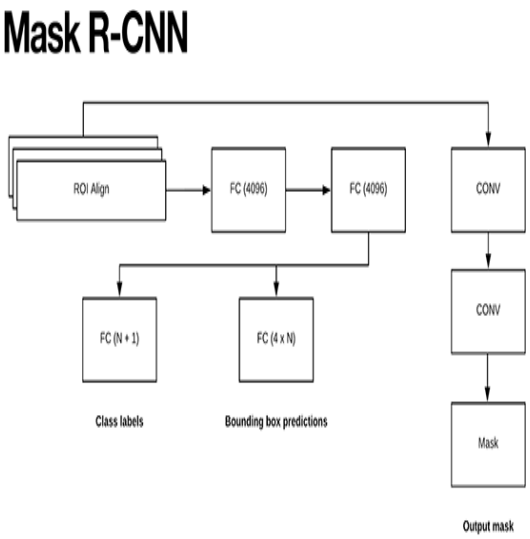
etected

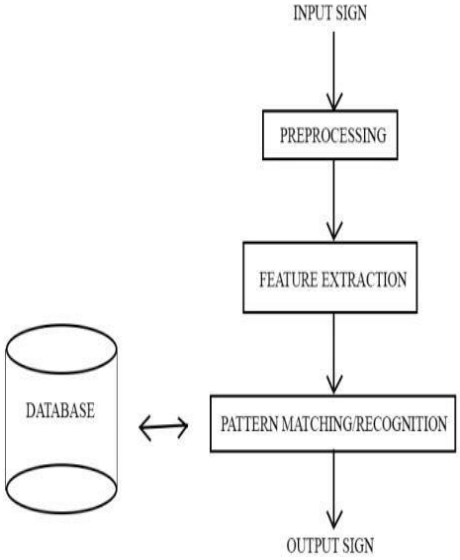
m can

apping

ask R-

CNN network using the Mask R-CNN object.





6. System Architecture

Working Modules:

- Data Set Generation
- Acquire the gestures
- Gesture Classification
- Compared with stored gestures
- Generated corresponding text

Modules Description

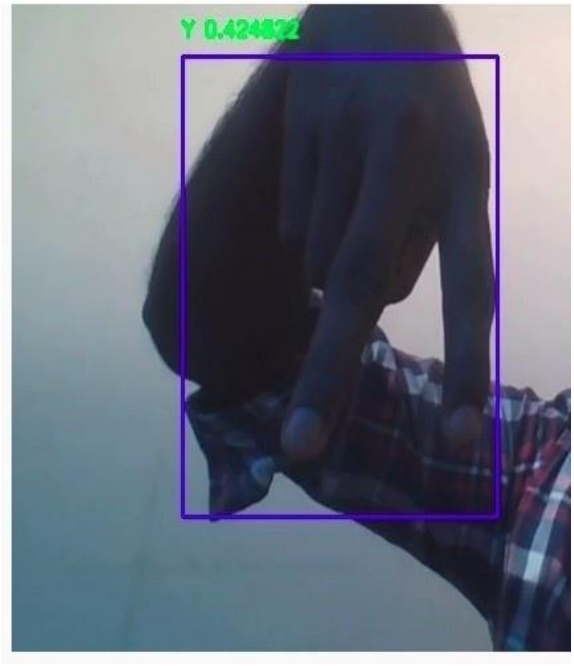
Data Set Generation:

All we should realise had been the datasets within the sort of RGB values. Steps we tend to followed to make our information/data set are as follows. We tend library as a supply to offer our dataset. Firstly, we have a tendency to seize around 1500 pictures of every of the image in ISL to use Open computer vision (OpenCV) for coaching functions and around 200 pictures per image for testing purpose. 1st we generally tend to seize each body proven by the digital camera of our machine. In every body we have a tendency to define a segment of interest (ROI) that is denoted bya blue delimited square.

Gesture Classification

In [2] Hidden Markov Models (HMM) is hired for the class of the gestures. This model deals with dynamic aspects of gestures. Gestures are extracted from a sequence of video pictures by pursuing the skin-color spots comparable to the hand into a body-face house targeted on the face of the user. The aim is to well known 2 classes of gestures: deictic and symbolic.

The image is filtered using a quick look-up categorization table. when filtering, coloring pixels are gathered into spots. Spots are applied mathematics objects supported the situation (x, y) and also the colourimetry (Y, U, V) of the coloring pixels so as to see undiversified areas.



Compared with stored gestures

Gather Gesture, this node acquires/detects/recognizes any hand gestures which are being made inside the vicinity of the body of the digital cameraand captures any hand gestures that are seenby way of the digital camera and feeds it tothe ML version. Evaluate, this node takesthe received gestures captured in the previous node and the ML version compares it with the dataset that it has trained on.

Generated corresponding text

Generating text, if the precedingnode returns that made gesture has matched and access inside the dataset, the version returns the corresponding cost associatedwith the made gesture and presentations it to the user. Forestall, applications stops its function.

7. Literature Survey

1. Anup Kumar; Karun Thankachan; Mevin M. Dominic, “Sign Language Recognition”

This paper provides a singular machine to aid in speaking with those having vocal and hearing disabilities. It discusses an stepped forward approach for sign language reputation and conversion of speech to signs. The set of rules devised is capable of Extracting symptoms and symptoms from video sequences below minimally cluttered and dynamic historical past the use of skin colour segmentation. It distinguishes among static and dynamic gestures and extracts the best characteristic vector. Those are categorized using guide Vector Machines. Speech recognition is built upon widespread module - Sphinx. Experimental consequences display nice segmentation of symptoms below diverse backgrounds and comparatively excessive accuracy in Gesture and speech recognition.

2. Singha; Joyeeta; Karen Das,“Automation Indian Sign Language Recognition for Continuous Video Sequences”.

Data Acquisition, Pre-processing, Characteristic Extraction, and Classification are the 4 primary modules in the proposed system. Pores and Skin filtering and histogram matching are carried out in the pre-processing step, observed by means of Eigen vector-primarily based characteristic Extraction and Eigen value weighted Euclidean distance-based type technique. In this paintings, 26 special alphabets have been considered. Obstacles of this paper were use of records gloves and popularity of unmarried hand gestures. Both static and dynamic hand gestures were being recognized in [10] with the use of assist Vector Machines and Hidden Markov model but this device has to be operated theusage of coloration gloves.

3. P.K. Athira, C.J. Sruthi, A. Lijiya, "A Signer Independent Sign Language with Coarticulation Elimination From Live Videos: an Indian Scenario"

The system produced through the authors acknowledges ISL gestures from cell camera videos without any additional sensors to hit upon hand regions. The benefit of this gadget is that it's far low- priced and can be carried out with a mobile digital camera, making it very consumer- pleasant. However the drawback is that it isn't efficient below cluttered backgrounds and exceptional illumination conditions.

8. Conclusion

In this record, a purposeful actual time vision based Indian signal language popularity for D&M human beings were advanced for ISL alphabets. We executed an accuracy of 95.7% on our dataset. Prediction has been progressed after enforcing two layers of algorithms in which we verify and are expecting symbols which are greater much like each different.

9. Future Enhancement

We are planning to acquire higher accuracy even in case of complex backgrounds withthe aid of attempting out various heritage subtraction algorithms. We are also takinginto account improving the preprocessing toexpect gestures in low light situations witha higher accuracy.

10. References

[1] Richmond Alake, "Understanding and Implementing LeNet-5 CNN Architecture", June 2020.

[2] Madeline Schiappa, "Understanding the Backbone of Video Classification: The I3D Architecture", June 2020, https://towardsdatascience.com/understanding-the backbone-of-video-classification- the-i3d-architectured4011391692

[3] Muttu Mariappan H , Dr. Gomathi V . "Real Time Recognition of Indian Sign Language" Conference: 2019 International Conference on Computational Intelligence in Data Science (ICCIDS).

[4] Neel Kamal Bhagat, Vishnusai Y, Rathna G N ," Indian Sign Language Gesture Recognition using Image Processing and Deep Learning,2019".

[5] T. Bohra, S. Sompura, K. Parekh and P. Raut, "RealTime Two Way Communication System for Speech and Hearing Impaired Using Computer Vision and Deep Learning," 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2019, pp. 734-739.

[6] Priyankur Sarkar, "What is LDA: Linear Discriminant Analysis for Machine Learning", September 2019.

[7] Kumar Mahesh, "Conversion of Sign Language into Text," International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 9 (2018)pp. 7154-7161.

[8] Suharjito, Suharjito& Gunawan, Herman

&Thiracitta, Narada & Nugroho, Ariadi. (2018). Sign Language Recognition Using Modified Convolutional NeuralNetwork Model. 1-5.

[9] K. Bantupalli and Y. Xie, "American Sign Language Recognition using Deep Learning and Computer Vision," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 4896-4899, doi: 10.1109 /BigData. 2018. 8622141.

[10] Sik-Ho Tsang, “Review: Inception-v3- 1 st Runner Up (Image Classification) in ILSVRC 2015”, September 2018, https://shtsang.medium.com/review- inception- v3- 1strunner-up-image-classification-in-ilsvrc-2015- 17915421f77c

[11] Sign Language Recognition Using Convolutional Neural Networks, International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume:5 Issue: 6,2017.