JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

DESIGN AND ANALYSIS OF ALPHABET RECOGNITION USING YOLOV5

[1]Praveen Gupta, [2]Sevvana.K.K. Sai Nischal, [3]Midatana Anusha,

[4]Rayudu Lashya Sri Lakshmi and [5]AbhisekhMahapatra

[1]Assistant Professor,[2,3,4,5]Student

[1] Dept. of CSE, GITAM University, Visakhapatnam, India

***Abstract*:** Image processing can be used to solve some daily applications. Sign language recognition can assist people who are deaf or have problem in communicating with the world. Recognition of hand gestures is widely used in various industries such as entertainment, acting, gaming, automobiles etc. Results for real-time, image-based hand gesture identification are improved and more successful as a result of recent advances in machine learning and object detection techniques.[1] We worked on two different hand gestures for recognition using the object detection model YOLOv5. All Data were collected, labelled using Labelimg and used Roboflow to create a custom trained computer vision model and deployed in Yolov5 to get the desired output and are compared how accurate these models were.

Keywords—Image Processing, YOLO, Labelimg, Object detection

1.INTRODUCTION

Humankind places a great deal of importance on communication. Both sides need a means of communication for people with hearing or speech impairments. As a result, they cannot integrate into society. I try to show myself well in various situations. In some cases, you may need to interact using visual aids or translators. Sign language use is of utmost importance. As a result, any combination of primitive components can be used to describe complex values.

Unlike sounds transmitted through hearing, sign language includes hand gestures and gestures to convey messages. It is primarily used by people who are deaf or hard of hearing, translators, hearing friends, and hearing families. Most people still have trouble communicating in one of the widely used sign languages. This is because some sign languages require an interpreter or translator to communicate. Digital translation is essential for smooth communication. This is because it is very difficult and unlikely that everyone who is hearing impaired will always be accompanied by an interpreter.

Translating sign language using machine learning will greatly facilitate communication between deaf and hard of hearing people. It is often thought that written language is the most practical way for deaf people to communicate, but that is not the case. This assumption is not true for people who are born deaf, but may hold true for people who later become deaf. By associating the shapes of words and the alphabet with sounds, people learn to read and write. It is difficult for people who have never heard these sounds to learn to read and write. As a result, they find it difficult to read and write, two essential skills. This problem is solvable, and people with congenital hearing loss will find it much easier to read thanks to a way to read text from pictures and translate it into sign language. In many parts of the world, people with speech disabilities speak more than 120 different sign languages, including American Sign Language (ASL), Indian Sign Language, Arabic Sign Language, British Sign Language, and Chinese Sign Language. Sign language is spoken by over 70 million people. It is a predominantly spoken language worldwide, including about 10 million people in India alone. One of the most widely used sign languages around the world is American Sign Language or ASL. Approximately 2 million deaf community members live in the United States and Canada and use ASL as their primary form of communication.

The American Manual Alphabet, often known as American Sign Language symbols, consists of 26 letters that can form many traditional English words. Indian Sign Language has 1.5 million users mainly in India. ISL uses the same set of 26

alphabets as ASL, with static gestures for each alphabet. Like most other modern languages, sign language consists of a predetermined alphabet, each letter represented by a specific hand position or gesture. Sign language is usually spoken by single people, but there are many different languages. For this reason, this essay will use the name NAD (Association of the Deaf). The form of sign language that is the subject of our study is known as American Sign Language (ASL). Each of the 35 English-based American Sign Language (ASL) gestures is assigned a letter or number from the alphabet. The case of the letter doesn't matter.

This study determines the accuracy of the alphabet displayed by the fingers and talks about the accuracy of this alphabet. I am using Yolov5, Py-torch, Roboflow and labelimg for this research.

To train the data, I created my own dataset tagged with the Pascalvoc format required by Yolov5. After labeling, we used Roboflow to create a specially trained computer vision and training model in Yolov5 that trains the dataset using the Yolov5n model. As a result, we will be able to use sign language to recognize the alphabet and increase the accuracy of the alphabet.

2.RELATED WORK

Tejashri J. Joshi presented a feature Extraction method using Principal Component Analysis (PCA), which is a method for dimension Reduction [2]. This tells us about the importance of data preprocessing speed where it is necessary to keep the properties of high dimensional data and also reducing the noise and unwanted elements.

H N Saha [3] have worked on sign language by capturing the images using the webcam and preprocessing the images. Used a convex deconstruction and polygon to estimate the figure and further divide various hand's convex regions is noted. The singularities are used to create recovered feature vectors. This calls for practice using the learned traits, which are essentially exclusive to different hand movements. Hence, sign languages are acknowledged.

In 2014, Ching-Hua Chuan described about American Sign Language recognition system. This used a 3D motion sensor. It focuses on a Leap Motion sensor that is significantly more inexpensive and smaller. The 26 English alphabet letters in ASL were classified using ML techniques, k-nearest neighbour and support vector machine.[4]

In 2017, S. M. Lee introduced the smart sign Language interpretation system using a wearable hand device.[5] Three major modules make up the overall system: a wearable device which has a sensor, processing module, and mobile application module for the display unit. An embedded support vector machine classifier is used to gather and evaluate sensor data, identifies alphabets then Bluetooth is transmitted to the mobile device. A text-to-speech smartphone application with an Android platform was created.

Francois published a paper on Human Posture Recognition in a Video Sequence. Methods used are 2D and 3D appearance. It uses 3D to build posture for recognition after applying PCA. Then identifies silhouettes from camera. This method uses intermediary gestures which is a disadvantage. This results in uncertainty in training and reduces prediction, accuracy.[6]

Wu and Zhang talk about Faster R-CNN-based gesture recognition. This uses the DisturbIoU method to reduce network overflow and boost gesture detection accuracy.[7]

3.METHODOLOGY

I. MODEL

In order to find the accuracy of the object in this case we are using the hand gesture to find the accuracy of the alphabets. In order to find the accuracy, we need to train the data so that it gives the best possible outcome. The data needs to highly accurate so that it gives valid outcomes. Here we are using the YOLO algorithm which basically uses the CNN (Convolutional Neural Network) algorithm.

CNN is an algorithm that take in an input image, assign importance to various aspects/objects in the image, and distinguish between them. Before going deep into CNN, Neural Network have three types of layers they are:

1. Input layer – Input layer basically gives input to the model.

2.Hidden layer - a layer that lies between the input and output layers, where artificial neurons receive a set of weighted inputs and then use an activation function to generate an output.

3. Output layer – The final layer where the prediction is obtained.

II. ROBOFLOW

In this study we are using Roboflow for managing, preprocessing, and augmenting large datasets. Roboflow is a cloud-based platform, The platform provides a suite of tools for training machine learning models, including object detection, segmentation, and classification, with minimal code.

III. YOLOv5

"You Only Look Once version 5". It is an improvement on the previous versions of the YOLO algorithm and has achieved significant improvements in accuracy and speed. The YOLOv5 algorithm has a small memory footprint and can run efficiently on both CPU and GPU. It is available in several pre-trained models of different sizes, allowing users to choose the model that best suits their needs

4.DATASET

We are creating our own custom dataset to train and evaluate the model. The dataset contains 358 colour images. These images are taken by 3 individuals with various lighting conditions and different hand positions and a camera is used to take images. The images contain 2 alphabets C and O.

Fig.4.1 C and O

5.RESULT AND ANALYSIS

The results displays that the maximum mAP_0.5 obtained is 0.95 and minimum obtained is 0.18. The mAP_0.5:0.95 displays that maximum is 0.66 and minimum is 0.05.

The highest precision is 0.93 and highest recall is 0.95. These were all taken for 120 iterations. The graph is shown below in Fig.5.1.

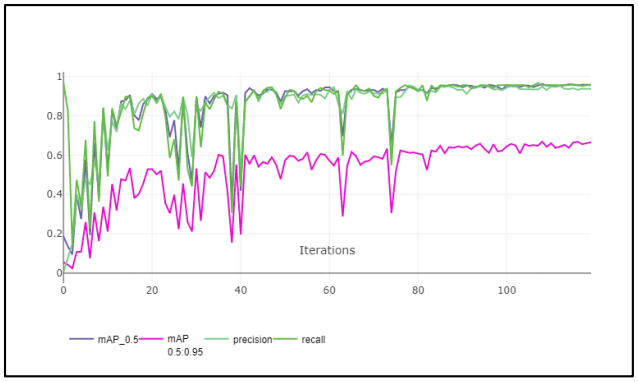

Fig 5.1 Precision and Recall Graph

6.CONCLUSION AND FUTURE WORK

We can conclude that with very few data we can at least detect 2 objects at a time and there are many factors such light, distance and different hand posture etc. factors are coming into play during the detection. On an average we have got an accuracy of 0.83 during identification which is pretty much decent keeping in mind that we have used very less amount of data. Both the accuracy and the frame rate must be optimal for real-time gesture detection. Therefore, this can be used for the Real Time detection upon adding more gestures and more datasets

REFERENCES

1. https://www.techtarget.com/searchenterpriseai/definition/convolutional-neural-network

2. P. S. Zaki, M. M. William, B. K. Soliman, K. G. Alexsan, K. Khalil, and M. El-Moursy, “Traffic signs detection and recognition system using deep learning,” 2020.

3. Arshad, Habiba, et al. "A multilevel paradigm for deep convolutional neural network features selection with an application to human gait recognition." Expert Systems (2020): e12541.

4. “Deep Learning for American Sign Language Fingerspelling Recognition System” Huy B.D Nguyen and Hung Ngoc Do, 2019 26th International Conference on Telecommunications.

5. Amrutha, K., and P. Prabu. "ML Based Sign Language Recognition System." In 2021 International Conference on Innovative Trends in Information Technology (ICITIIT), pp. 1-6. IEEE, 2021.

6. Q. Areeb, Maryam, M. Nadeem, R. Alroobaea, and F. Anwer, “Helping Hearing-Impaired in Emergency Situations: A Deep Learning-Based Approach,” IEEE Access, vol. 10, pp. 8502–8517, 2022.

7. Shweta. K. Yewale and Pankaj. K. bharne, “Hand gesture recognition system based on artificial neural network”, Emerging Trends in Networks and Computer Communications (ETNCC), IEEE, 22-24 April, 2011.