Proceedings of the 4th International Conference on Computing and Data Science (CONF-CDS 2022)
DOI: 10.54254/2755-2721/2/20220585

Sign Language Translation using Tensor Flow Model Zoo

Chunduru Anilkumar[1], Sathishkumar V E[2], and Pallavi Pareek[1]

[1] Dept of Information Technology, GMR Institute of Technology, Rajam, Andhra Pradesh-532127
[2] Department of Industrial Engineering, Hanyang University, 222 Wangsimini-ro, Seongdong-gu, 04763, Seoul, Republic of Korea

*srisathishkumarve@gmail.com

Abstract. Communication is very essential for all humans because it allows us to share or express our sentiments, emotions, and other thoughts. Humans communicate with one another via natural language (for example words), body language (hand gestures, facial motions, and so on) or writing etc. People who do not have any impairment can easily converse with each other in natural language. However people with impair-ments, such as deafness or blindness, suffer a communication hurdle. They converse with common peo-ple by adopting sign languages, which are difficult for normal people to grasp or understand. People with hearing and speech impairments often have a very difficult time conversing with other people without any translator or interpreter. As a matter of fact, this research is being undertaken that transforms sign language into text that is easily understood by the ordinary individual. This system will identify numbers, alphabets, and hand gestures as well. Our primary purpose is to remove the obstacles that exist between the deaf, dumb, and the rest of people.

Keywords: Sign Language, Impaired Persons, Hand Gestures, Text etc.

1. Introduction

The WHO (World Health Organization) conducted a survey/research in the year 2021 and discovered that approximately 5-10% of the world's population suffers from hearing and speaking impairments. India is the world's second-largest country, with around 63 million people suffering from deafness or dumbness. These people interact with normal people in a variety of methods, including writing, text messaging, sign language, and so on. However, Sign language is widely considered as one of the most prevalent ways for People who have difficulty hearing or are deaf to communicate with others.

Sign/Gesture language is a nonverbal communication method that replaces spoken words with hand gestures and movements, body language, and facial expressions (oral communication) [1]. Most of the people communicate using both words and signs, but Deaf and Dumb persons only use signs to convey their emotions and sentiments. In today's modern world, there are numerous types of sign languages available such as ASL (American Sign Language), BSL (British Sign Language), ISL (Indian Sign Language), and so on. People in today's modern world utilize/use their regional sign language to communicate and convey feelings so that others can readily comprehend. A sign language is made up of three key components [2].

DOI: 10.54254/2755-2721/2/20220585

Finger-Spelling is one of the most important or significant aspects, which implies that there is a symbol for each and every letter of the alphabet. This style of communication is primarily used to spell names and, on occasion, geographical/location names. This is sometimes used to convey words/terms that do not have any particular signs, or to emphasize or explain a certain term/word.

The other/second element is word based sign vocabulary, which is one of the most often utilized forms of communication among individuals with disabilities. It indicates that in sign language, there is a matching related sign for each word in the vocabulary.

Non-manual elements are the next and the last important aspect of any sign communication. This communication style makes use of facial expressions, lips, body posture, as well as the tongue, among other things.

Sign Language recognition can be possible by two approaches: vision-based and glove-based. In glove based technique gloves are worn which consist of sensors. The major disadvantages of these glove based techniques are, data gloves are relatively costly and these gloves need to be worn continuously which is very difficult hence the modern systems use vision based approach. The vision based is again classified into two various types such as static approach and dynamic approach. Below figure shows the two types of techniques to language recognition.

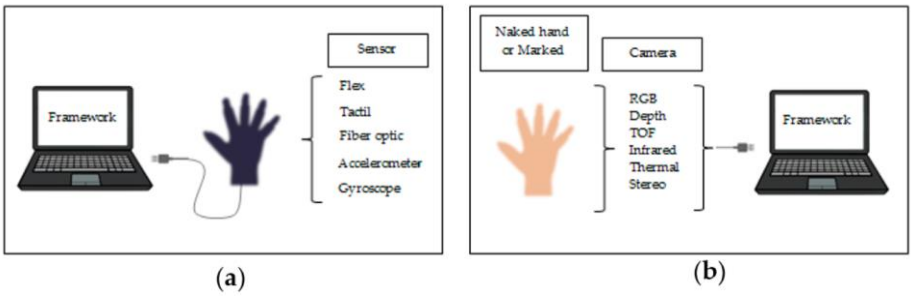

Figure 1. Sign Language Recognition Techniques, [a] Glove Based Approach [b] Sensor Based Approach.

The proposed system will surely help society by supporting physically impaired people because it integrates multiple sign languages, words, characters and numbers etc. This system facilitates daily communication between impaired people (deaf,dumb) and the entire world [3].

2. Literature Review

The author of this research paper suggested/proposed an automated approach for identifying sign language that uses PCA (Principal Component Analysis) and one-vs.-all SVM (Support Vector Machines) classification. For detecting skin sections color information is used, hand segmentation is accomplished using morphological operations and filters, and feature extraction in hand regions is achieved using PCA, and classification is done with SVM. The system was tuned to function with the five vowels, resulting in a testing accuracy of more than 80% and a frame execution time of 59 milliseconds. [4].

Sign languages are made up of hand signals and face emotions. To recognize those signs, OpenCV's skin segmentation technique is utilized to recognize and track the ROI (Regions of Interest). To train and predict hand gestures, the FCM machine learning technique (Fuzzy c-means clustering) is used. The FCM beats traditional clustering algorithms in terms of efficiency and dependability in many circumstances [5].

This research describes two approaches to the recognition/identification of hand gesture/sign: glove/sensor-based and vision-based. A glove-based approach cannot be extensively adopted since data gloves are somewhat expensive. The suggested system employs a vision-based non-invasive recognition approach. Vision-based recognition can be accomplished using either static or dynamic recognition [6].

DOI: 10.54254/2755-2721/2/20220585

Sign gestures are captured using the Python OpenCV package through the computer's camera. The dataset for different signs has been gathered. For each sign, approximately 3000 photos are captured. To analyze and classify visual images, a convolutional neural network is used. CNN's are multilayer perceptron variants that have been regularized. [7].

This system is developed and implemented in such a way that hand movements are identified by recognizing and tracking hand gestures with OpenCV's skin segmentation feature. The image frames are changed to maintain uniformity across all films, and OpenCV is used to extract features. Dual communication has been accomplished in this suggested system, and sensors are not necessary [8].

The proposed system is based on a kind of supervised machine learning known as Support Vector Machines. K-mean clustering on acquired pictures, segmenting hand region, feature extraction, and image classification are among the methods used in this approach to get the text for related signs. The term object detection refers to a set of computer vision algorithms that locate and categorize items. Object detection techniques can be used on both static and dynamic pictures. In sports, computer vision techniques are already widely used. The pertained model was trained and evaluated using our own data, which consisted of images extracted from surveillance video.

In contrast to R-CNN-based techniques, the Single Shot Multibox Detector (SSD) does not require an operation of per-proposal categorization in the second stage. As a result, it is fast enough for real-time detection applications. It's worth noting that "SSD with Mobile Net" refers to a model with SSD as the model meta architecture and Mobile Net as the feature extractor type.

TFOD (Tensor Flow Object Detection) API is an open-source framework built on top of Tensor Flow that is designed to make it simple to build, train, and deploy object detection models. The Tensor Flow Object Detection API accomplishes this by providing the user with numerous pre-trained object detection models as well as instructions.

3. System Architecture

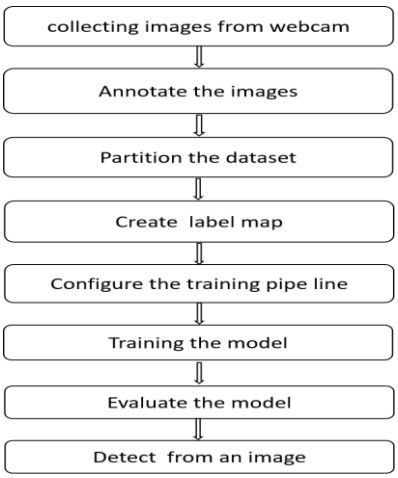

Figure 2. System Architecture.

This paper is broken into seven sections listed below. Section I provides the introduction to this paper, Section II comprises a Literature Review (study) of previous works, and Section III represents a system design. In Section IV the methodology is described. The output screens are included in Section V. Section VI contains the conclusion and future scope.

4. Methodology

A. Collection of images: Image collection is one of the very crucial or critical stages in any ML (Machine Learning) research. The images in this research were taken with the assistance of OpenCV.

DOI: 10.54254/2755-2721/2/20220585

OpenCV is a comprehensive open-source library/toolkit for CV(computer vision), image processing etc. which is widely used in real-world scenarios/ applications. When paired with additional modules such as Numpy, Python can analyze the OpenCV array structure.

B. Annotate the Images Using Labeling tool: Image annotating is described as the task of labeling an image. The **process** of recognizing and categorizing distinct elements in a picture is known as image labeling. Labeling is a python module that is open source and used to label pictures. This is built with the help of QT. It is a simple and cost-free approach to recognize the huge number of photos. It's a tool for annotating graphical images which is developed in python. After annotation automatically xml files have been created.

C. Partitioning the Dataset: When separating a data set into a training set and a test set, it has been critical to successfully determine which piece of the data set is chosen as the training set and which percentage is chosen as the test set. However, there are two major downsides to this partitioning method: (a) class imbalance and (b) sample representativeness problems. Manually we can separate the collected images by placing 70% of the collected images in the training folder. Train images are stored in Tensorflow/workspace/collected images/train And 30% of collected images in the testing folder. Test images are stored in Tensorflow/workspace/collected images/test.

D. Creation of Label Map: Annotations referring to classes are common in computer vision datasets. The many signs for various hand motions are included in class labels. Annotation is the process of capturing the object of interest in the image to make it recognizable and understandable to machines through computer vision. For class annotations, the label map is an unique source of record. In SSD Mobilenet V2, the label map translates integers to a class list defined in the label map.

E. Creation of TF Records: Tensor flow's own binary storage format, the TFRecord file format, is a lesser-known component of the framework. Binary data takes up less disc space, takes less time to transfer, and can be read from disc considerably faster. Creation of Train.record and Test.record in annotations folder is done.

F. Download pre-trained model from Tensor flow Model Zoo: In this we have taken an already trained model from the Tensor Flow Model Zoo. The Tensor Flow Model Zoo is a collection of object identification architectures that have done exceptionally well on the COCO dataset. From this tensor flow model zoo we have taken ssd-mobilenet, this model is faster as compared to R-CCN. We chose SSD-mobilenet from the tensor flow model zoo since it is quicker than R-CCN. To detect many objects in a single picture, the ssd mobile net is used.

G. Update pipeline config file in models folder: Configuration files are just Protocol Buffers objects described in the research/object detection/protos.proto files. The top-level item in the pipeline is a Train Eval Pipeline Config.). Each object and field's significance may or may not be evident or well-documented, but you can always look at the source code to discover how each value is used (for example, check preprocessor.py to understand how data augmentation is done).

H. Training the Model: In the process of training the model it automatically creates the checkpoint files in the models folder. These files consist of the snap-shots of the model at each step of the model. If the loss is zero then the model perfectly fits the application. if loss is more then it doesn't fit the application. To verify the model we have taken the one of the test data and found whether the model detects the object correctly or not.

I. Evaluation of model: By default, the training process keeps track of several fundamental training metrics. While the training process is running, checkpoint files will be created in the

DOI: 10.54254/2755-2721/2/20220585

"Tensorflow/workspace/models" folder, which correspond to snapshots of the model at specific points. The steps for conducting the evaluation are as follows: First, we must download and install the metrics that we intend to use. Second, we must make changes to the configuration pipeline configuration file.

J. Detection from web-camera: When we run the following code, our built-in camera will open and take image frames as input tensors, which will be passed to the detect fn() function, which will preprocess the image, forecast the image's accuracies, and postprocess the image, returning a high-accuracy result. By drawing a bounding box around the motions, we will provide an image path as input and the movements' matching text as output.

5. Results

Figure 3. Labelling of Images.

Figure 4. Labelling of Images.

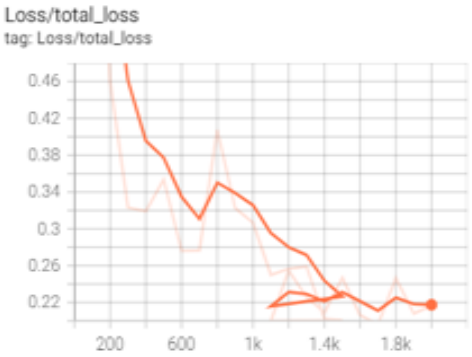

Figure 5. Tensor board Dashboard.

DOI: 10.54254/2755-2721/2/20220585

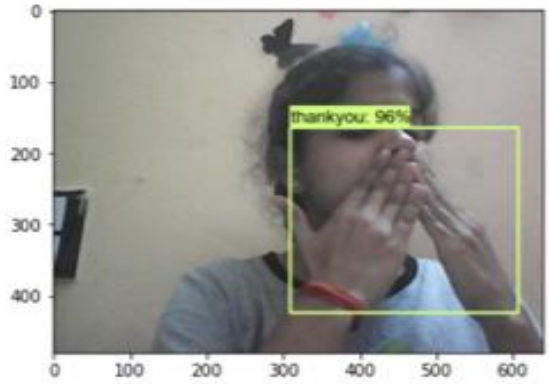

Figure 6. Detection of sign from image and webcam.

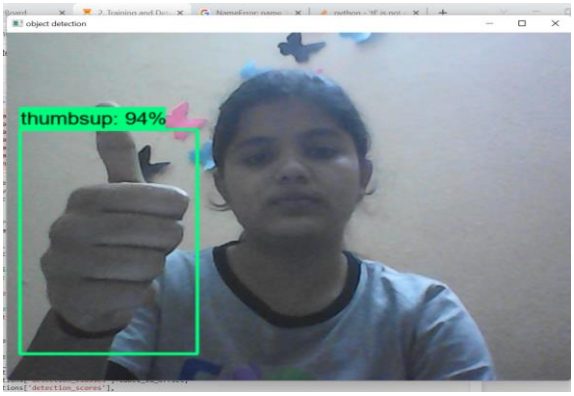

Figure 7. Detection of sign from image and webcam.

6. Conclusion and Future Scope

This research aimed to conceal some of the difficulties that disabled individuals face when communicating and listening. The proposed system can recognize the signs (alphabets, words, and numerals) gestured by the signer in real-time. The web camera is used to record/take input for this proposed system. Frames are retrieved/extracted from the supplied input. The system will employ image processing algorithms to process the extracted input images and then convert them into text using various models. This system has a significant impact on increasing casual conversation between people with hearing impairments and the general public. The proposed method's accuracy for different sign languages was tested and found to be greater than 90% for the majority of the sign words. In the future, we are planning to use CNN (Convolutional Neural Networks) to ameliorate the accuracy of the existing system. The system can be improved further by incorporating whole-body images for scanning hand gestures from photos. Furthermore, we can also add more hand motions, words etc and save the data from these gestures in the cloud. In the future, we hope to integrate facial expressions in addition to hand signals.

References

[1] Neyra-Gutiérrez, A., & Shiguihara-Juárez, P. (2020, September). Feature extraction with video summarization of dynamic gestures for peruvian sign language recognition. In 2020 IEEE XXVII International Conference on Electronics, Electrical Engineering and Computing (INTERCON) (pp. 1-4). IEEE.

[2] Mariappan, H. M., & Gomathi, V. (2019, February). Real-time recognition of Indian sign

DOI: 10.54254/2755-2721/2/20220585

language. In 2019 International Conference on Computational Intelligence in Data Science (ICCIDS) (pp. 1-6). IEEE.

[3] NB, M. K. (2018). Conversion of sign language into text. International Journal of Applied Engineering Research, 13(9), 7154-7161.

[4] Harini, R., Janani, R., Keerthana, S., Madhubala, S., & Venkatasubramanian, S. (2020, March). Sign Language Translation. In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS) (pp. 883-886). IEEE.

[5] Srivastava, S., Gangwar, A., Mishra, R., & Singh, S. (2021, December). Sign Language Recognition System using TensorFlow Object Detection API. In International Conference on Advanced Network Technologies and Intelligent Computing (pp. 634-646). Springer, Cham.

[6] Suthagar, S., Tamilselvan, K. S., Balakumar, P., Rajalakshmi, B., & Roshini, C. (2020). Translation of sign language for Deaf and Dumb people. Int J Recent Technol Eng, 8(5), 4369-4372.

[7] Mustamo, P. (2018). Object detection in sports: TensorFlow Object Detection API case study. University of Oulu.

[8] Divkar, A., Bailkar, R., & Pawar, C. S. (2021). Gesture Based Real-time Indian Sign Language Interpreter [J]. International Journal of Scientific Research in Computer Science Engineering and Information Technology, 2021, 387-394.