IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138
INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG
An International Open Access, Peer-reviewed, Refereed Journal

INDIAN SIGN LANGUAGE TO SPEECH CONVERSION FOR HSI PEOPLE

[1]Dr. Swetha P, [2]P Mourya Chiraag Balaji, [3]Prajwal S, [4]Preetham K, [5]Sai Chethan P

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

[1]Department of Computer Science and Engineering
[1]Global Academy of Technology, Bengaluru, Karnataka-560098

Abstract : According to the All India Federation of the Deaf, approximately 4 million Indians are deaf and more than 10 million have difficulty hearing. According to studies, one in every five deaf people worldwide is Indian. In India, more than 2 million deaf people use Indian Sign Language (ISL) to communicate. Sign language is a language that conveys meaning through visually conveyed sign patterns. It consists of a variety of hand shapes, alignment and movement of the hands, arms, or body, as well as facial expressions. Our system can recognise sign language symbols, which can be used to help HSI people to communicate with people around them who are unaware of Indian Sign Language. Hand signs and body language are used rather than vocal or tonal patterns to convey meaning in sign language. It all comes down to the combination of hand shapes, alignment, and movement. Many techniques and algorithms in this field have been developed using image processing and artificial intelligence. Any hand gesture recognition system is trained to recognize characters and convert them into the appropriate patterns. The proposed system is designed to provide speech to people who are unable to speak by detecting objects, in this case hand signs.

IndexTerms - Indian Sign Language Recognition; Object Detection; Sign Language Recognition; Object Detection; Depth wise Separable Convolutional Neural Networks; Transfer Learning; SSD Mobilenet V2

I. INTRODUCTION

Indian Sign Language (ISL) is a sign language used in India by the hearing and speech impaired to communicate with others. ISL owns the research presented in this article. Gestures are used in ISL to represent complex words and phrases. There are 36 hand poses in total, with 10 numbers and 26 letters. The system is trained using ISL hand positions, as shown in Fig.1. Most people struggle to understand ISL gestures. This has helped to bridge the communication gap between those who understand ISL and those who do not. It is not always possible to locate an interpreter to translate these gestures. It consists of a webcam for capturing hand gestures and poses and a backend for processing the webcam images. The system's goal is to implement a fast and accurate detection technique. This article's system successfully classifies ISL alphabets and numbers. The following section of this document discusses related work related to sign language translation. Section III explains how to process each frame and translate a pose/hand gesture. Section IV is about engineering.

II. LITERATURE SURVEY

The hand position was recognised using Indian Sign Language identification with the help of Microsoft Kinect in 2020 [1] , the hand region was carefully segmented, and the hand features were retrieved using Accelerated Robust Functions, Gradient Histogram Oriented, and Local Binary Patterns. The system combines all the three Support Vector Machine-trained classifiers to improve average recognition accuracy. The benefits of this system include the ability to transform an order of recognised hand sign movements into the most approximate and detailed English sentences with a precision of 100 for the characters A, F, G, H, N, P and 9. The response time was not adjusted to meet the needs of real-time hand sign translation, therefore additional work was required in this direction.

"Hand Gesture Recognition Using a Support Vector Machine," published in 2015 [2], For feature extraction, Canny's edge detection and gradient histogram were used, as well as the support vector machine which is commonly used for classifying and for testing the regression. The SVM method creates a model that predicts whether a new sample falls into one of two groups. Furthermore, when data points from examples that are classified into their particular categories, the classification algorithm learns from them. This method has the advantage of translating hand movements into 20 equivalent numbers in real time. The system's drawbacks included a limit of twenty numbers and less work in model evaluation.

"Visual-Based Sign Language Translation Device," published in 2013 [3]. Using LABVIEW software, proposed a mobile sign language translation device system for automatic translation of Indian Sign Language into English. With just one hand, this system can distinguish the representation of letters and numbers. The outcomes are remarkably constant and precise. The disadvantage is that the results are light sensitive and depending on the background.

The paper by Dardas N H and Georganas N D [4] uses bagoff features and support vector machine approaches for detecting real time hand signs and recognising them develops a system that includes detection of the skin as well as hand pose contour comparison after face subtraction from the image for detecting as well as tracking the bare hand in a cluttered background using bagoff features and multiclass SVM for hand gesture recognition, and creating a grammar for hand sign which is generated to manage an application. It achieved real-time performance of any image resolution, and has a high classification accuracy which includes a variety of scaling, orientation, lighting, and cluttered backdrop situations. However, this technology was restricted to a set of video game gestures.

III. EXPERIMENTAL SETUP

A. Model Used

An object detection model which is MobileNet SSD that computes the bounding box and category of an object based on an input image. Using Mobilenet as a backbone, this Single Shot Detector (SSD) object detection model can perform quick object detection optimized for mobile devices. For the pre-training model, where Single Shot Detector MobileNet (SSD MobileNet v2 fpv lite 320x320 coco17 tpu 8) is downloaded and deployed, Tensorflow Object Detection API is deemed best practice. SSD Mobilenet is an object detection model that uses a captured image to generate the frame outline and classification of an object. As demonstrated in Figure 1.6, the Single Shot Detector (SSD) object detection model starts with MobileNet and then adds many layers of convolution. The functionalities of object location and categorization are accomplished in a single network evolution step. The RCNN series, on the other hand, uses the Regional Proposal Network (RPN) approach and requires two shots, one for producing region proposals and the other for determining the proposal's goal. In comparison to two-tier RPN-based approaches, SSD is believed to be substantially faster. This pre-trained model will enable us to complete the work of transfer learning considerably more quickly.

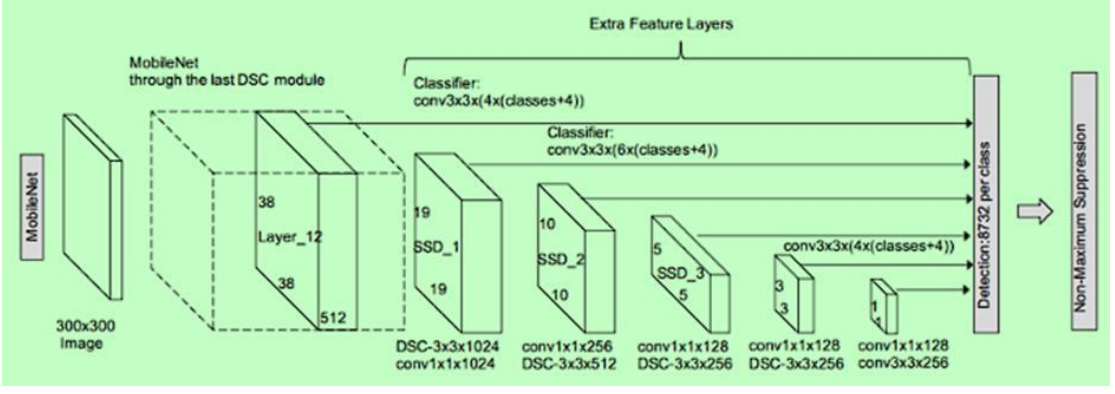

Fig 3.1: SSD MobileNet layered architecture

B. Design and Implementation

Speech recognition and speech-to-text recognition have benefited greatly from deep learning. Deep learning models require a large amount of data to be trained. Transfer learning is a strategy for solving a given job with insufficient data by employing a related task with a vast quantity of accessible data. Pre-training a network on the collected data set is used to apply transfer learning. For transfer learning, the SSD MobileNet pre-training model is faster and uses Depthwise Separable Convolution which is more efficient. The following are the four segments that make up the sign language recognition architecture:

1. Create our own collection of data.
2. Captured photos should be labeled.
3. Use Tensorflow object recognition to train the models.
4. Recognize the gestures with your hands.

For improved communication, the proposed system can recognise static word characters and display their labels

IV. SYSTEM DESIGN

System design is one of the most crucial stages of software development. The purpose of the design is to come up with a solution to the problem stated in the requirements paper. To put it another way, project design is the initial stage in solving an issue. System design has the greatest impact on software quality. The purpose of the design phase is to develop the overall look and feel of the software. The system focuses on Indian Sign Language (ISL), creating a dataset that includes all 26 alphabets and 10 digits (English). Processing and labeling images to highlight attributes Following that, CNN concepts on object detection are

presented. In addition, transfer learning will be studied by experimenting on models that have achieved state-of-the-art performance.

4.1 System Architecture

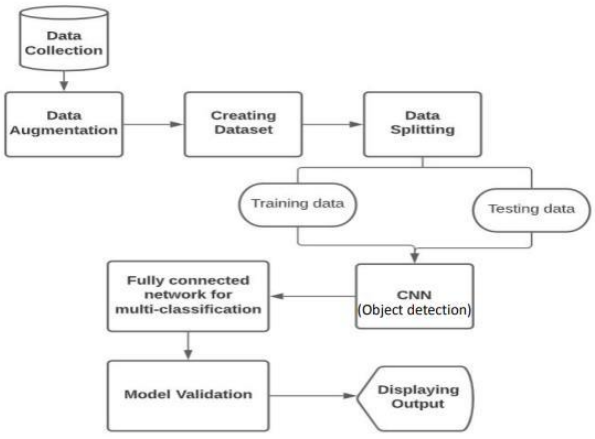

Fig 4.1.1 System Architecture

This illustration depicts a broad diagram that depicts the project's actions. Every type of architecture was represented. Data Architecture is a framework built to transfer data from one location to another, efficiently. It is full of models and rules that govern what data is to be collected. It also controls how the collected data should be stored, arranged, integrated and put to use in data systems of an organisation.We will now be discussing the system architecture of our data model. The first step is the data collection, where we gather or collect the data to provide an input to the model. Then we increase the amount of data by adding more similar copies of hand signs. Using all these data, we create a dataset used as an input to the model. Then we split the data accordingly into training data i.e.80%, and testing data i.e.,20%. Then we use the CNN model for object detection purposes. The functionality of CNN is to get an image designated by some weightage based on the different objects of the image, and then distinguishing them from each other. Then the data is sent to a multi-classification layer where every image or neuron in the previous layer connects to every image in the next layer. The next step is model validation where we confirm that the outputs of a statistical model are acceptable with respect to the real data-generating process. Then we display the output accordingly.

4.2 Data Flow Diagrams

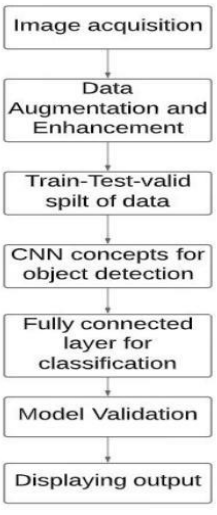

Fig 4.2.1 Data Flow Diagram

The above figure is the representation of the data flow diagram. A data-flow diagram is a way of representing a flow of data through a process or system. The DFD also provides information about the outputs and inputs of each entity and the process itself. Here, in DFD, the first step is image acquisition, which aims to transform any real-world data to an array of numerical data which could be later manipulated on a computer. Then we improve the amount of data of similar hand sign symbols and enhance them accordingly, known as data augmentation and enhancement. Then we split the dataset into training and testing data i.e.,80% and 20%. Then we detect the objects to distinguish from other data with the help of CNN model. Then we send the data to the classification layer to establish a connection between the previous and present layer. Then we check whether the model is valid for processing or functioning of real-data generated. Then we display the output images.

4.3 Use Case Diagram

The interactions between the various components are modeled using use case diagrams, which are dynamic in nature. There are two types of agents which are internal and external and both are called as actors and their relationships are depicted in use case diagrams. This diagram is used to represent an application's system and subsystems. A single use case diagram encapsulates much of the system's functionality.

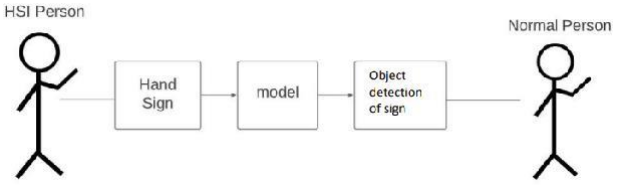

Fig 4.3.1 Use Case Diagram

In UML, use-case diagrams model the behavior of a system and help to capture the requirements of the system.Use-case diagrams describe the high-level functions and scope of a system. These diagrams also identify the interactions between the system and its actors. The use cases and actors in use-case diagrams describe what the system does and how the actors use it, but not how the system operates internally. Our model portrays the interaction between an HSI person and a Normal person. An HSI person communicates to a normal person with the help of hand sign symbols (in our case they use Indian symbols for communication). At first, an HSI person shows a hand sign symbol to communicate with a normal person. That symbol is shown to the model which consists of trained and tested hand sign symbols, and it then recognizes the value of that symbol. Then those symbols are sent for the next process i.e., object detection, where it detects a particular object if it is in a dataset and displays the correct meaning or exact symbol of that hand sign symbol. This object detection is done irrespective of the background. Then the recognized symbol is sent to the normal person and the communication is established between them.

4.4 Sequence Diagram

Sequence diagrams are used to indicate how the objects in the application interact with each other to finish a task. This is the most common type of interaction diagram where it is used to see a system’s interaction behavior.

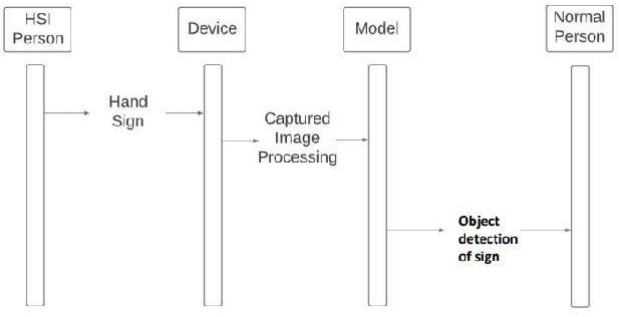

Fig 4.4.1 Use Case Diagram

Sequence diagrams are used to indicate how the objects in the application interact with each other to finish a task. This is the most common type of interaction diagram where it is used to see a system’s interaction behavior. It consists of a group of objects that are represented by lifelines, and the messages that they exchange over time during the interaction. A sequence diagram shows the sequence of messages passed between objects. Sequence diagrams can also show the control structures between objects. Firstly, an HSI person shows a hand sign to the device, where the device captures the hand sign with the help of a webcam. Then the device sends the captured image to the data model. The captured image is processed before sending it to a model. At last, the model detects a particular hand sign irrespective of background, and then sends it to a normal person by identifying the proper or valid hand sign if it consists in the respective dataset.

V. METHODOLOGY

A. Image Acquisition

Image Acquisition is a task phase where we capture the images to make a dataset which comprises the various hand signs in Indian Sign Language (ISL). The number of character images taken for this system is 36 which are as follows: A-Z , 0-9.

B. Data augmentation and Enhancement.

Data augmentation and Enhancement means improving the data set by collecting more data. This is the most important phase as this is the place we can take control over our dataset. As our approach is object detection we need to label the image but it takes a

lot of time to label a huge image which will be inefficient but do require a good amount of image for each hand sign. One of the ways to improve the diversity of the dataset is by changing some aspects of image like brightness,opacity, adding noise etc for the same label image and automate the whole system for each image.

C. Data Splitting

In data partitioning, the data is divided into two or more subsets. In our case it is a two-part split, one part is used to train the model and the other is used for testing. We can also make another split for validation. This phase required us to program a system where we assign the same unique name for the image and its respective label file with extensions like jpg and xml accordingly.

D.Object Detection Model training(Using SSD MOBILENET)

The approach of our system is object detection and we are using SSD mobilnet V2 which is a lightweight state of the art object detection model. Firstly, we configured our model to our custom labels and created pipelines accordingly. Then we train the model with the dataset we created using the above three methods. We did need to loop around on parameter aspects like epochs,early stop etc based on loss,learning rate etc.

E. Model Validation

After model training, the trained model is evaluated against a test data set in a process known as model validation. The test data may or may not be derived from the same data set as the training data and validating the model's results to ensure their accuracy. A large amount of training data is used when training the object detection model, and the primary goal of verifying model validation gives us insights to improve the model.

VI. RESULTS AND DISCUSSION

This is the phase where we check the model's actual output i.e to manually make hand signs in front of the camera where the frames are captured and the object is predicated by our trained object detection mode since we also require the detected sign in speech we return the character with is read by the python speech output module.

Fig 6.1: Result of accuracy of alphabet

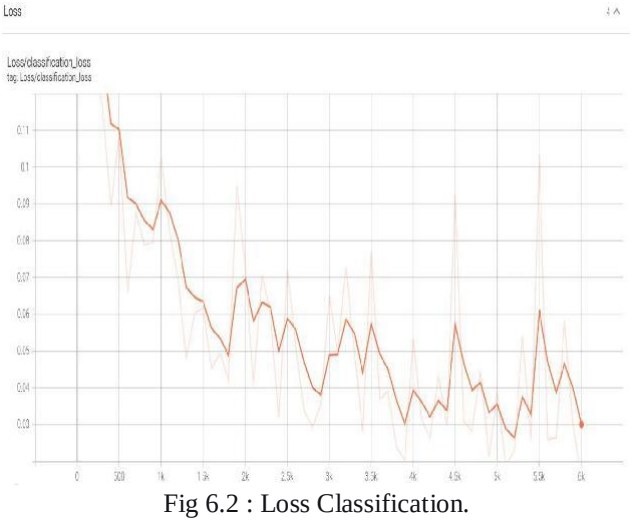

Fig 6.2 : Loss Classification.

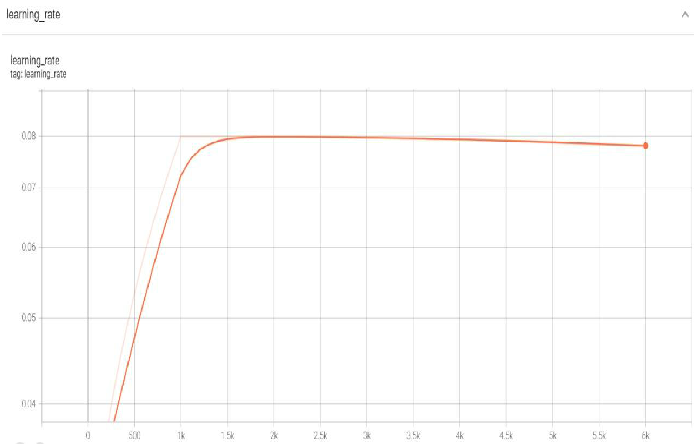

Fig 6.3 : Learning rate

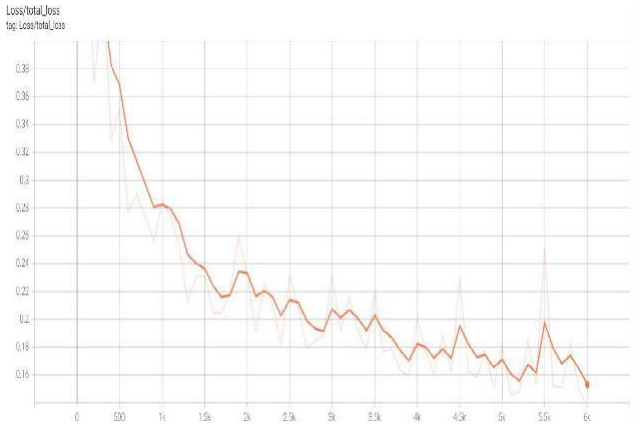

Fig 6.4: Total Loss

VII. CONCLUSION

This system helps and aids the hearing impaired and mute people to live independently. It develops confidence and will power to share their emotions, thoughts, ideas and difficulties with the normal people in the society. This eliminates the gaps among the people and achieves a better society. We have worked on this project based on the survey we made and to find a better way of communication approach for HSI people by eliminating the previous work limitation.

REFERENCES

[1] Raghuveera, T, R Deepthi, R Mangalashri, and R Akshaya. "A Depth-Based Indian Sign Language Recognition Using Microsoft Kinect." Sādhanā 45, no. 1 (2020). https://doi.org/10.1007/s12046-019-1250-6.

[2] Nagashree R N, Stafford Michahial, Aishwarya G N, Beebi Hajira Azeez, Jayalakshmi M R, R Krupa Rani, "Hand gesture recognition using support vector machine ", 2015, The International Journal Of Engineering And Science (IJES), ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805.

[3] Y. Madhuri, G. Anitha. and M. Anburajan., "Vision-based sign language translation device," 2013 International Conference on Information Communication and Embedded Systems (ICICES), 2013, pp. 565-568, doi: 10.1109/ICICES.2013.6508395.

[4] Dardas, Nasser H., and Nicolas D. Georganas. "Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques." IEEE Transactions on Instrumentation and Measurement 60, no. 11 (2011): 3592–3607. https://doi.org/10.1109/tim.2011.2161140.

[5] https://vidishmehta204.medium.com/object-detection-using-ssd-mobilenet-v2-7ff3543d738d

[6] https://jonathan-hui.medium.com/ssd-object-detection-single-shot-multibox-detector-for-real-time-processing-9bd8deac0e06

[7] Indian Sign Language Research and Training Center (ISLRTC) http://www.islrtc.nic.in/

[8] https://indiansignlanguage.org/