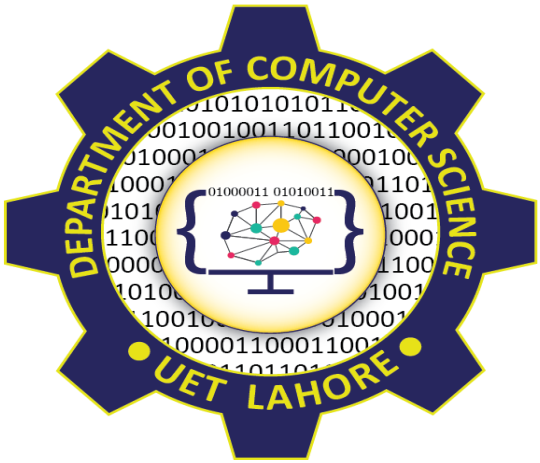
Sign Language Recognition

(Sign to Speech)



SUBMITTED TO

Dr. Talha Waheed

SUBMITTED BY

Umair Manzoor 2020-CS-24

Musawir Ahmed 2020-CS-46

Zain Hassan 2020-CS-134

Department of Computer Science

University of Engineering and Technology, Lahore

Pakistan

1. Introduction:

i. Overview:

Deaf people rely completely on sign language, a non-verbal language, to communicate with the people around them. The hands, eyes, face, lips, and body are used as visual clues. In sign language, the motions or symbols are arranged linguistically. It combines a variety of nonverbal communication techniques, such as finger spelling, hand gestures, body language, facial emotions, timing, touch, and anything else.

ii. Background:

The first academic to recognize that hearing is not necessary for learning was possibly the Italian mathematician and physician Geronimo Cardano. In the 1500s, he learned that written words might be used to educate the deaf. He educated his deaf son using his techniques. While instructing deaf youngsters in Spain, Spanish monk Pedro Ponce de Leon had great success with his techniques of instruction. During this time, Cardano was also teaching his deaf kid.

iii. Motivation:

The hope of lowering the communication gap between the deaf and hearing communities is what drives this initiative. It is commonly known that deaf persons have difficulty communicating with the general public.

iv. Objectives:

Understanding and sensitivity to the deaf and hard of hearing community are improved via study of signed language. If you know signed language well, you'll have a deeper understanding of deaf culture and be able to encourage others to understand and respect the sign language.

v. Scope:

The main reason sign languages are created is to help the deaf and dumb. To express certain information, they use a coordinated and precise combination of hand movements, hand forms, and hand orientation.

2. Problem Statement:

Normal people have trouble understanding the hand signs used by deaf individuals because they use them to communicate. Therefore, there is a need for systems that can identify various indications and communicate information to regular people.

3. Technology Stack:

- Python
- TensorFlow object detection API
- OpenCV
- MATLAB
- Text-to-speech API

4. Literature Reviews:

i. Previous Research:

Most of the research in this sector is conducted with a glove-based technique. Sensors like potentiometers, accelerometers, and other devices are mounted to each finger in the glove-based system. The corresponding alphabet is shown in accordance with what they read.

A glove-based gesture recognition system created by Christopher Lee and Yangsheng Xu was able to recognize 14 of the hand alphabetic letters, learn new gestures, and update the model of each gesture in the system in real-time. The primary issue with this glove-based system is that it needs to be calibrated each time a new user is added on the fingertip level in order for the Image Processing unit to recognize the fingertips.

ii. Innovation:

Using image processing, we are carrying out our project. Our project's key benefit is that it is not limited to use with black backgrounds. With any background, it can be used. Additionally, our method does not mandate the wearing of color bands.

5. Methodology:

A labeled map is created which is a representation of all the objects within the model, i.e., it contains the label of each sign (alphabet) along with their id. The label map contains 26 labels, each one representing an alphabet. Each label has been assigned a unique id ranging from 1 to 26. This will be used as a reference to look up the class name. TF records of the training data and the testing data are then created using generate_tfrecord which is used to train the TensorFlow object detection API. TF record is the binary storage format of TensorFlow. Binary files usage for storage of the data significantly impacts the performance of the import pipeline consequently, the training time of the model. It takes less space on a disk, copies fast, and can efficiently be read from the disk.

During the training, the model has some losses as classification loss, regularization loss, and localization loss. The localization loss is mismatched between the predicted bounding box correction and the true values. The formula of the localization loss. The real-time detection is done using OpenCV and webcam again. For, real-time detection, cv2, and NumPy dependencies are used. The system detects signs in real-time and translates what each gesture means into English. From here then we can use any publicly available Text-to-speech API to get this text converted into a speech for the understanding of ordinary people.

6. Discussion:

i. Limitations:

Recognition of sign language is complicated by issues such precise monitoring of hand gestures, hand occlusion, and high computing cost. There are no sizable, general data sets that could be used to suggest more accurate statistical or neural machine translation models for translating sign language to natural language. All translation systems' accuracy is extremely low and has to be raised.

ii. Challenges:

In order to accommodate persons of all skin tones and in various situations, the dataset must be sufficiently diversified. A bias in the data may hurt deaf people of a certain ethnicity. The largest obstacles to deaf youngsters integrating into society are a general lack of understanding of deaf concerns and, in particular, a lack of sign language users in regular public places. Moreover, there is no general datasets available for signed gestures.