SIGN LANGUAGE INTERPRETER

K. Keerthi, K. Gnapika, P. Mythili Reddy, B. Ishitha Reddy

K. Veena (GUIDE)

Hyderabad Institute of Technology and Management

---***---

Abstract - Every average person has the ability to see, hear, and respond to their surroundings. Some unfortunate people are without this significant blessing. These people primarily include the deaf and dumb, who interact with others by using sign language. Yet, since not all common people can understand their sign language, communicating with regular people is a significant limitation for them. Also, this will make it difficult for members of the deaf and dumb communities to engage with others, especially when they want to participate in educational, social, and professional settings. The goals of this research are to create a sign language translation system for hearing or speech challenged individuals utilising deep learning and neural networks to help those who are deaf or have trouble speaking or hearing communicate with others. The model utilised in this classification assignment is a pretty simple Convolutional Neural Network implementation for the methodology (CNN). This exploits the convolutional property, which was primarily developed for the analysis of visual imagery. Three layered CNNs were trained and tested in real time using segmented RGB hand motions. With the help of a personal device, such as a laptop webcam, simple photos of the hand were used to construct the image dataset for each gesture. I was able to get training accuracy of roughly 89% and testing accuracy of 98.5% with this CNN model.

Keywords : Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Gesture Recognition, Open CV, TensorFlow

1.INTRODUCTION

Every average person notices, pays attention to, and reacts to everything around them. Nonetheless, some people who are less fortunate are excluded from receiving this significant blessing. These people, who are primarily dumb and deaf, rely on sign language to communicate with others. Statistics indicate that approximately 9 billion individuals worldwide are stupid and deaf [1]. Dealing with interactions between deaf-dumb people and average people has never been easy. In general, not every average person can understand the sign language used by the weak to communicate. They find it quite challenging because communication is one of life's most essential needs. Also, this will make it difficult for the deaf and dumb groups to engage with others, especially when they want to organise into educational, social, and employment settings. A sign language recognition system must be created in order to eliminate the distinction between the average person and the disabled person in order to solve this problem. The primary objective of this project is to create a system for translating sign language into text. Since not every regular person is taught how to communicate by signing, this system will assist them in understanding the language of the deaf and dumb, providing them with useful information for carrying out their daily tasks..

2. SYSTEM ANALYSIS

2.1 EXISTING SYSTEM

Using various sensors to collect the hand gesture and motion, such as using a data glove, is an efficient technique to increase the robustness of hand gesture recognition [3]. Such sensors, as opposed to optical sensors, are typically more dependable and unaffected by background clutter or poor lighting.

Fig -1: Data glove

Disadvantages

- As the user has to wear a data glove which sometimes requires calibration, it not only is inconvenient for the user but also may hinder the naturalness of the hand gesture.
- Also, these data gloves are quite expensive.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | April - 2023 Impact Factor: 8.176 ISSN: 2582-3930

2.2 Proposed System

In a changing environment, neural networks are adaptable. Systems that use rules or are programmed are only effective in the circumstances under which they were intended. Conditions become invalid when they alter. Although it might take some time for neural networks to adjust to a rapid, abrupt shift, they are very good at doing so.

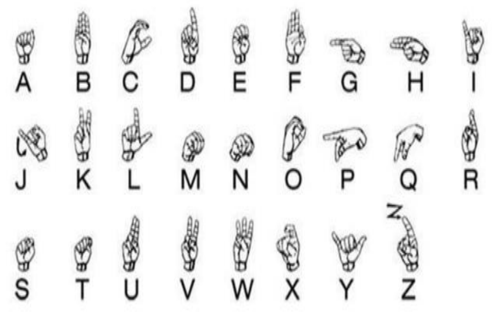


Fig -2 : Indian Sign Language Symbols

Advantages

- Can build informative models where more conventional approaches fail.
- Can handle very complex interactions. They can easily model data which is too difficult to model with traditional approaches, such as inferential statistics or programming logic.
- Build models that are more reflective of the structure of the data in significantly less time.

2.3 STRUCTURE

The project is structured into 3 distinct functional blocks :

- Data Processing
- Training
- Classify Gesture

2.3.1 Data Processing

The load data.py script contains functions to load the Raw Image Data and save the image data as numpy arrays into file storage. The process data.py script will load the image data from data.npy and preprocess the image by resizing/rescaling the image, and applying filters and ZCA whitening to enhance features. During training the processed image data was split into training, validation, and testing data and written to storage. Training also involves a load dataset.py script that loads the relevant data split into a Dataset class. For use of the trained model in classifying gestures, an individual image is loaded and processed from the file system.

2.3.2 Training

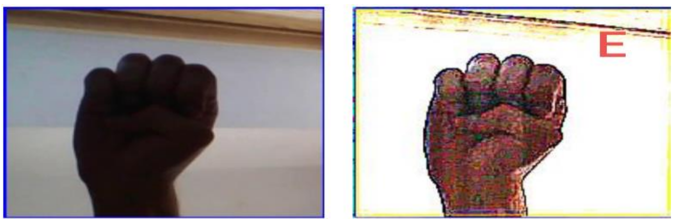
The train model.py file contains the model's training loop. The learning rate, batch size, image filtering, and number of epochs are listed in a configuration file that contains the hyperparameters used to train the model. The architecture of the model and the configuration used to train it are both stored, allowing for further evaluation and optimisation of the model's performance. The training and validation datasets are loaded as Data loaders during the training loop, and Adam Optimizer with Cross Entropy Loss is used to train the model.. The model is evaluated every epoch on the validation set and the model with best validation accuracy is saved to storage for further evaluation and use. Upon finishing training, the training and validation error and loss is saved to the disk, along with a plot of error and loss over training.

2.3.3 Classify Gesture

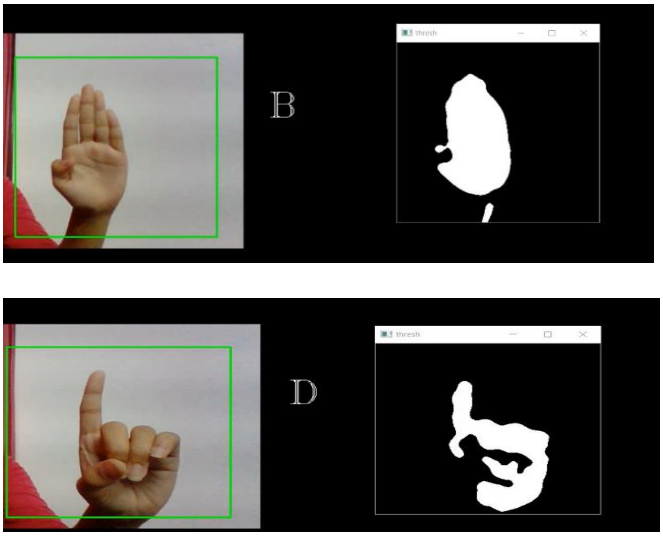
After a model has been trained, it can be used to classify a new sign language gesture that is available as a file on the file system. The user inputs the file path of the gesture image and the test data.py script will pass the file path to process data.py to load and pre-process the file the same way as the model has been trained.

2.4 .Implementation

• Importing modules and libraries required for the project.
• Setting up of hand histogram for creating gestures.
• Once a good histogram is achieved, saving it in the code folder.
•Adding gestures and labelling them, using OpenCV which uses webcam feed.
• Splitting all the captured gestures into training, validation and test set.
• View all the gestures.
• Train the CNN model using Keras.
• Opening up the gesture recognition window which will use the webcam to interpret the trained Sign Language gesture.



Figs. An instance of real time implementation of the Sign Language Recognition Model



Figs. An instance of real time implementation of the Sign Language Recognition Model

2.4.1 Sign Language Recognition Model

The model used in this classification task is a fairly basic implementation of a Convolutional Neural Network (CNN). As the project requires classification of images, CNN is the go-to architecture.

CNN

Convolutional Neural Networks (CNN), are deep neural networks used to process data that have a grid-like topology, e.g images that can be represented as a 2-D array of pixels. A CNN model consists of four main operations: Convolution, Pooling, Flattening and Classification (Fully-connected layer).

i. Convolution: Convolution is used to take features out of the input image. With the use of small squares of input data to learn image attributes, the comes after it.

ii. Pooling: Pooling (also called down-sampling) reduces the dimensionality of each feature map but retains important data.

iii. Flattening: Our pooled feature map will be reduced to the size of a single column. You are left with a lengthy vector of input data after the flattening stage, which you then run through the artificial neural network to continue processing.

iv. Fully-connected layer: It is a multi layer perceptron that uses SoftMax function in the output layer. Its purpose is to use features from previous layers for classifying the input image into various classes based on training data.

3. CONCLUSION

The objective of this project is to anticipate in real time the alphabetic hand movements used in Indian Sign Language. The work mentioned above demonstrates that when we take into account the segmented hand-gestures, it may be solved more accurately. We eliminate the overheads of dynamic background by using depth-based segmentation. It facilitates focusing on the primary element of the image, which is the hand gesture, and eliminates the extraneous background from the picture. Three layered CNN received the segmented RGB hand-gestures and trained and tested it in real-time. I was successful in achieving 89.30% training accuracy and 98.5% testing accuracy. My model was very accurate at predicting the outcomes.

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to our mentor Mrs. K Veena for providing us with the fantastic chance to work on this amazing project on this subject. This project also enabled us to do extensive research, which allowed us to learn a lot of brand-new information. They have our sincere appreciation.
As a follow-up, we also want to express our gratitude to my friends, who greatly aided us in completing this job in the allotted time.

REFERENCES

[1] Mukesh Kumar Makwana, " Sign Language Recognition", M.Tech thesis, Indian Institute of Science,Bangalore.

[2] Pigou, Lionel, et al. ”Sign language recognition using convolutional neural networks.” Workshop at the European Conference on Computer Vision. Springer International Publishing, 2014.

[3] Escalera, Sergio, et al. ”Chalearn looking at people challenge 2014: Dataset and results.” Workshop at the European Conference on Computer Vision. Springer International Publishing, 2014.

[4] KuznetsovaAlina, Laura Leal-Taix, and BodoRosenhahn. ”Real-time sign language recognition using a consumer depth camera.” Proceedings of the IEEE International Conference on Computer Vision Workshops.2013.