AutoGstr: Relatively Accurate Sign Language Interpreter

Devashish Gupta, Jaganath Prasad Mohanty, Ayas Kant Swain, Kamalakanta Mahapatra
Department of Electronics and Communication Engineering
National Institute of Technology, Rourkela, India

***Abstract*—Hand gesture recognition is the one of the method to identify the hand position, its pattern and then translate it to the corresponding meaning or purpose. This paper contributes a real time sign language interpretation of hand gestures based on deep convolutional neural networks with focus on development of a cost-effective and efficient hardware prototype for communication ease with deaf and dumb people.**

***Index Terms*— Deep learning, hand gesture recognition, sign language interpreter, ASL, human-machine interface, convolutional neural network, raspberry pi.**

I. INTRODUCTION

HAND gesture recognition is an important aspect of sign language interpreter which plays an important role in Human Computer Interaction (HCI). In recent years, various ASL interpreters have developed and it has been witnessed that research in the areas of HCI have increased and various algorithms have been proposed [9]. Researchers focus on computer vision identification for gesture recognition category [10]. In this paper our motive is to build a cost-effective and efficient hardware prototype AutoGstr for communication at ease with deaf and dumb people with American Sign Language (ASL) and simulated in Figure 1.

II. RELATED WORK

In order to recognize hand gesture various methods and techniques are already available. The principle component involved for hand gesture recognition system is data acquisition, hand localization, hand feature extraction and identification based on the features extracted.

A. Glove and Vision based methods

Gestures can be recognized using embedded sensors like electronics gloves using flex sensors which detects gestures based on resistance change, copper-plate based glove, optical marker which detects reflected IR rays. These devices are expensive and wired which gives way to adopt new techniques.

B. Image Processing and Machine Learning

Various techniques have been proposed for hand gesture recognition based on features such as motion, skin color [1][2], trajectory-based information [3][4].

III. CNN MODEL DEVELOPMENT

We present a system based on deep convolutional neural networks, capable of identify and classify all the 26 ASL hand gestures accurately.

A. Image Dataset

In order to recognize the input image accurately neural network needs a huge amount of data. As it is already know the more the data, better the performance of the CNN model. For our case we are classifying the input image into 29 classes (26 ASL alphabets + 2 other gesture like space character and OK signal + background).

B. Identification of pattern's in image

A CNN consist of numerous convolutional layers that restrict the connection between layers (acting as local identifiers) and shares the same weight inside a layer. Let Conv(X+X0, W) = Conv(X,W) + X0, where X is the image and W is the filter, X0 is the translation factor and Conv is the convolutional operation [5]. Hence the convolutional layer (filters + ReLU layer) together works as feature identifier to identify the patterns in an image.

C. Hand region detection

In general the hand region may or may not occupy the whole input image. To keep high classification performance, a reasonable ratio between feature map area and hand gesture area should be kept [5] which are met by adopting the proposed CNN architecture

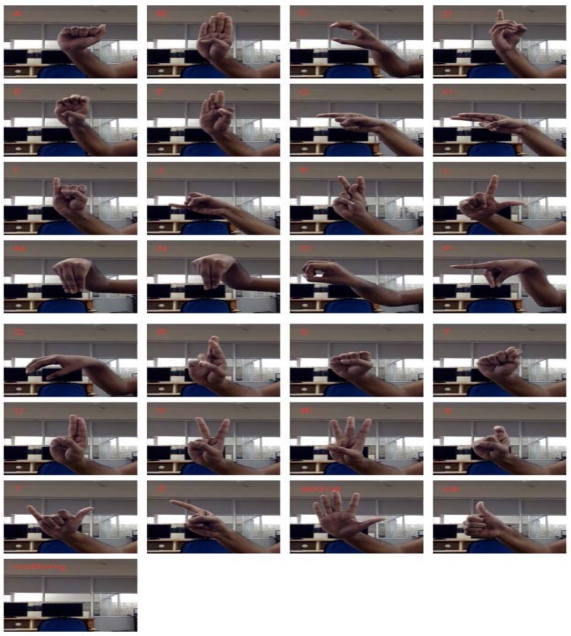

Figure 1: Simulation results for ASL alphabets and other gesture

D. ANN Classification

CNN is the extended version of Artificial neural network (ANN) where convolutional layers and pooling layers are present before ANN architecture.

E. Generalization of CNN Model

Overfitting is a common problem in deep learning [6][7]. Hence, we need to generalize our model. To reduce the effect of overfitting we have used 2 dropout layers each with a dropout ratio of 20% (which randomly disables 20% of neurons in the hidden layers).

F. Training of the proposed model

The proposed sign language interpreter system based on a deep CNN uses open source framework like keras and tensorflow. Training of the CNN architecture was performed on Google cloud NVIDIA Tesla T4 with 2560 cores and a memory of 16GB. The dataset was prepared by collecting

4300 images for each of the 29 classes. 8 different backgrounds were incorporated in each of those images. For training 77% of the dataset was used i.e. 3300 images out of 4300 were used for training and remaining 1000 images were used for validation purpose.

IV. WORKING OF THE SYSTEM

The overall working of the system is explained below as shown in Figure 2 to establish the communication between normal person and the specially abled individual.

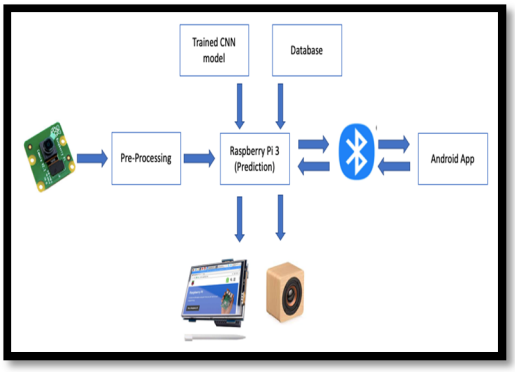

Figure 2: Block diagram for proposed system

A. Prediction of ASL alphabets: The hand gesture is first captured using the camera. The input image is pre-processed using filters to remove noise and is further fed to the trained model for classifying it into the one of the 29 categories (26 ASL alphabets & 2 more hand gestures).

*B. Framing words from hand gestur:*The proposed deep CNN model is capable of predicting the ASL alphabets in real time.

C. Text to voice translation: On completion of framing the words, the user needs to show 'OK' gesture to the camera.

D. Two-way communication using Bluetooth: In order to communicate with normal person the deaf and dumb person can uses the voice output. But for the other case our system uses inbuilt Bluetooth of the raspberry pi.

E. Time efficient communication using database: In our day to day life a person uses few sentences more frequently; hence forming them each time will be inefficient.

F. Graphical user interface: For user friendly interaction with the system a GUI is developed which feeds input from the camera, the words formed after the feed is predicted using the trained model, received messages and the phrases stored in the database.

G. Audio output: Once the user indicates the system to translate the message into voice output, the espeak library converts the text message into voice output through the raspberry pi aux jack.

V. RESULTS

The training results are shown in Figure 3. As shown in the figure, the accuracy of the proposed model gradually increases upto 93% within 40 epochs. Also it can be infer that the variance between the validation accuracy and training accuracy also reduces as the number of epochs increases.

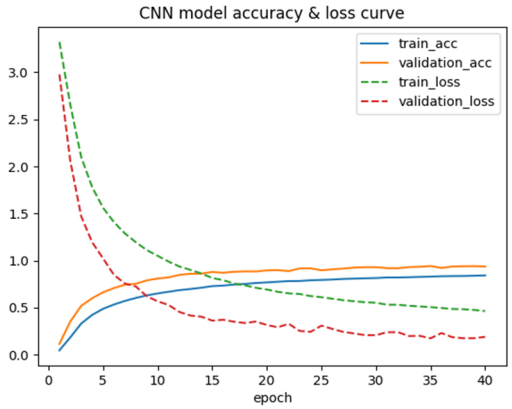

Figure 3: Loss and accuracy curve

Figure 4 shows the confusion matrix obtained for the validation dataset. From the matrix it can be inferred that each letter is predicted accurately.

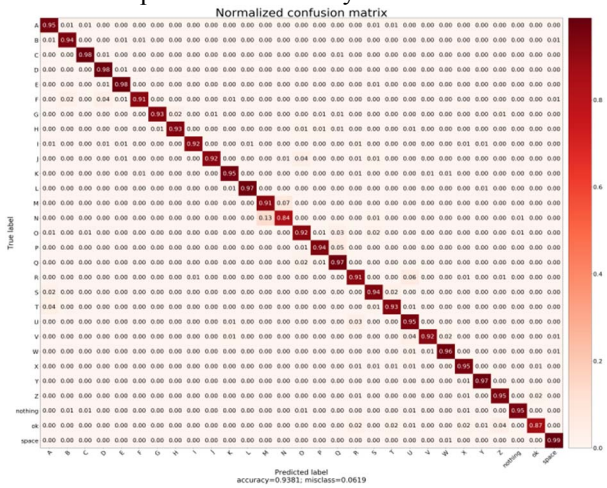

Figure 4: Confusion Matrix

VI. CONCLUSION

AutoGstr model results a relatively automatic gesture recognition using CNN which gives improved results than the traditional image processing approach using deep CNN based approach. The system performance of the hardware is decent and can be improved using dedicated processor and sufficient amount of RAM. The CNN Model accuracy, fine tuning of parameters like learning rate, batch size, and normalization needs to be taken care of in future activities.

VII. REFERENCES

[1] D. Lee and Y. Park, "Vision-based remote control system by motion detection and open finger counting," IEEE Trans. Consum. Electron., vol. 55, no. 4, pp. 2308-2313, 2009.

[2] F. Erden and A. E. Çetin, "Hand gesture based remote control system using infrared sensors and a camera," IEEE Trans. Consum. Electron., vol. 60, no. 4, pp. 675-680, 2014.

[3] S. Jeong, J. Jin, T. Song, K. Kwon and J. W. Jeon, "Single-camera dedicated television control system using gesture drawing," IEEE Trans. Consum. Electron., vol. 58, no. 4, pp. 1129-1137, 2012

[4] Y. Wang, and Y. Ruoyu, "Real-time hand posture recognition based on hand dominant line using Kinect," in Proc. ICME, San Jose, CA, 2013, pp. 1-4.

[5] P. Bao, A. I. Maqueda, C. R. del-Blanco and N. García, "Tiny hand gesture recognition without localization via a deep convolutional network," 2017 *IEEE Transactions on Consumer Electronics*, vol. 63, no. 3, pp. 251-257

[6]] S. J. Nowlan, and G. E. Hinton. "Simplifying neural networks by soft weight-sharing," Neural computation, vol. 4, no. 4, pp. 473-493, 1992.

[7] G. E. Hinton, Geoffrey, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov. (2012, July) Improving neural networks by preventing co-adaptation of feature detectors. Cornell University Library, NY.

[8] S. Hussain, R. Saxena, X. Han, J. A. Khan and H. Shin, "Hand gesture recognition using deep learning," *2017 International SoC Design Conference (ISOCC)*, Seoul, 2017, pp. 48-49.

[9] S.Mitra and T.Acharya,"Gesturerecognition:Asurvey," 2007 IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 37, no. 3, pp. 311–324.

[10] Y . Chuang, L. Chen, and G. Chen, "Saliency-guided improvement for hand posture detection and recognition," *Neurocomputing,* vol. 133, pp. 404-415, 6/10/ 2014.