International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 01 | Jan 2020 www.irjet.net p-ISSN: 2395-0072

Hand Sign Recognition using Convolutional Neural Network

Kumud Alok[1], Anurag Mehra[2], Ankita Kaushik[3], Aman Verma[4]

[1]Assistant Professor, Dept. of Computer Science & Engineering, Inderprastha Engineering College, Uttar Pradesh, India
[2,3,4]Dept. of Computer Science & Engineering, Inderprastha Engineering College, Uttar Pradesh, India

---

Abstract - *Sign Language is a type of communication that relies on hand signs, gestures and expressions, commonly used in deaf community. But people outside from deaf community find it hard or almost impossible to communicate with deaf people. They have to rely on an Interpreter which can both be expensive and hurt privacy of the people trying to communicate. In this paper we are trying to propose a method which uses the power of Convolutional Neural Net to identify and recognize hand signs which is captured in real time through a laptop's webcam. Since there is no universal Sign Language, we chose our model to specifically be trained on ASL (American Sign Language) alphabets. But this method could also work for other sign languages.*

Key Words: Sign Language Recognition, ASL recognition, American Sign Language, CNN, Convolutional Neural Network, TensorFlow

1. INTRODUCTION

Sign language is a highly visual-spatial, linguistically complete language. It is typically the first language and the main means of communication for deaf individuals [1]. Very few people understand sign language. Moreover, contrary to popular belief, it is not a universal language. Thus, further increasing communication gap between the Deaf community and the hearing majority. Written communication is time consuming and slow and only possible when people are sitting or stationary. Written communication can be awkward while walking or moving. Also, the Deaf community in general is less skilled in writing a spoken language [2]. The main aim of hand sign recognition system is to make an interaction between human and CNN classifier where the recognized signs can be used for conveying meaningful information or can be used for giving inputs to a machine without touching physical knobs and dials on that machine [3]. In our paper we proposed a model made using Convolutional Neural Net. We have taken 5 alphabets from American Sign Language due to limited computation power. The alphabet we chose were: E, F, I, L, V. This model is then connected to a webcam which captures hand sign in real-time and give the output based on the current hand sign.

2. TECHNOLOGY USED

2.1 Convolutional Neural Networks (CNN)

Convolutional Neural Network is a class of Artificial Neural Networks, commonly used in object recognition tasks. They are made up of neurons that have learnable weights and biases [4]. A CNN consists of various layers like pooling, convolutional, dropout, fully connected, The weights and biases are adjusted using Gradient Descent and Backpropagation method.

2.2 TensorFlow

TensorFlow is an open source software library developed by Google Brain Team. Mainly used in machine learning tasks. TensorFlow is written in C++, Python, CUDA. We can use both CPU and GPU to execute code in TensorFlow. Although executing the code with a GPU is much faster. A word of caution TensorFlow only work in CUDA enabled GPUs. Which only consists of Nvidia GPUs. Since CUDA is a proprietary application programming interface made by Nvidia.

2.3 Keras

Keras is a free and open source neural net library written in Python. It is a high-level API which could use TensorFlow, Theano, CNTK, MXNet or PlaidML library in backend. The primary reason why Keras appeals to the masses is its user friendliness and ease of use. It provides a high-level abstraction to backend libraries so we won't have to define each and every function. Thus, lowering the bar for people who are not very familiar with coding.

3. LITERATURE REVIEW

Convolutional Neural Networks based Sign Language Recognition [4]: This paper aims to recognize the hand sign through camera and predict the alphabet corresponding to a particular hand sign with the help of Convolutional Neural Network (CNN) in real time. The system which implements predictive model that gives result the overall 91% accuracy in real-time recognition.

Using Deep Convolutional Networks for Gesture Recognition in American Sign Language [5]: The use of Deep Convolutional Networks to classify images of

American Sign Language. This paper has also mentioned Background Subtraction Technique for removing the background from each of the images. They achieved an accuracy of 82.5% on the alphabet gestures and 97% validation set accuracy on digit gestures.

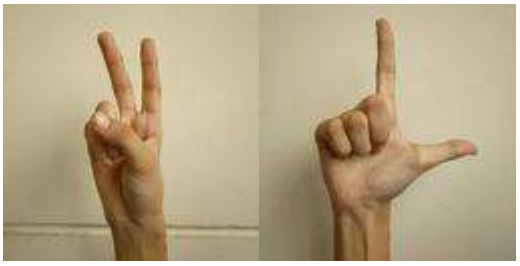
Hand Gesture Recognition for Sign Language Recognition: A Review [6]: Authors presented various approaches of sign language and hand gesture recognition. That were proposed in the past by various researchers for expressing emotions and thoughts to another person by sign language. There are broadly three approaches mentioned in this paper which are Approach based on Vision, Approach based on Instrumental Glove and Approach based on Colored Marker.

Hand Gesture Recognition Using a Convolutional Neural Network [7]: In this paper the approach towards hand sign recognition is based on Convolutional Neural Network (CNN) and Microsoft Kinect SDK. Microsoft Kinect Sensor is used for capturing the raw skeletal joint motion data. Authors presented the comparison of the proposed approach to previous work on hand-crafted features.

4. METHODOLOGY

4.1 Data Collection

Data or photos of hand signs were collected using a smartphone's camera. We took a total of 1600 photos, 320 photos for each hand sign. Photos of hand signs were taken in various environments with different lighting conditions having slightly different orientation in each photo. So that we could simulate all types of real-world conditions. Given below are the hand signs on which our model will be trained on. These letters were also used in training dataset.



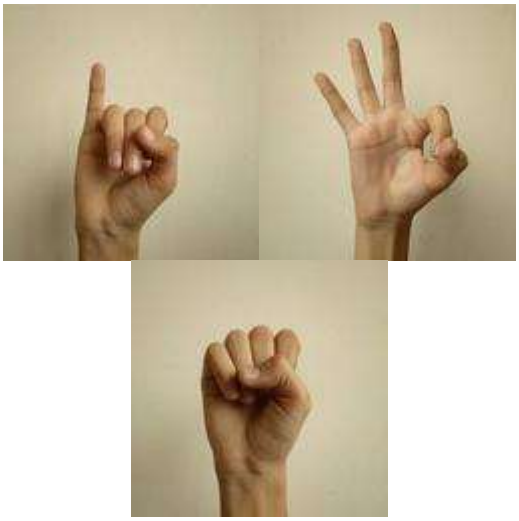


Fig - 1: ASL hand signs for Letters V, L, I, F, E respectively

4.2 Preprocessing

Every image in the dataset is converted to an image of 128x128 pixel. After that we have applied random transformation like shear, rotation, zoom, width shift and height shift using ImageDataGenerator class from Keras library. Due to small size of dataset we made a validation split of 80/20. Meaning 80% of images (which is equal to 1280 images) will be used for training the model. And the remaining 20% (320 images) will be used for model validation.

4.3 Model

Although there can be many different configurations of a Convolutional Neural Net model. After some trial and error, we settled for a model which consists of a total of 9 layers of which 4 layers are of convolutional layer, 2 maxpool layers, 2 fully connected layers and 1 flatten layer as shown in Fig-1. We have also added some dropout layers which randomly drop some firing neurons so that we could avoid overfitting of the model.

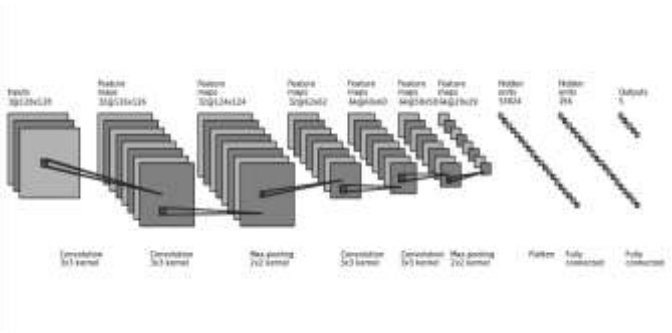


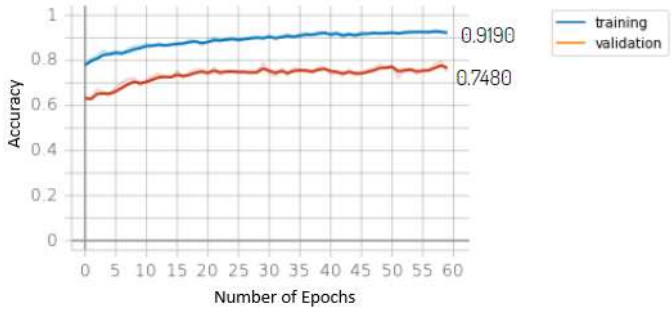
Fig - 2: Model Architecture

4.4 Training

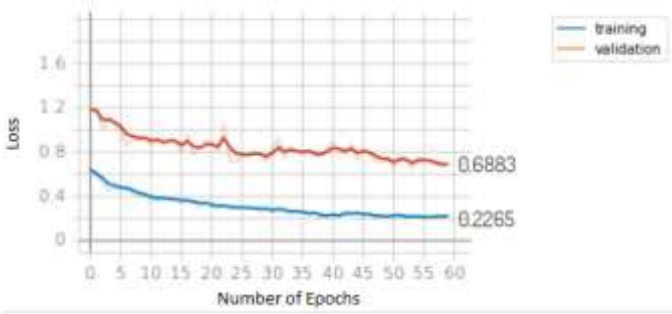
Adam Optimizer was used with a learning rate of 0.0001. Model was trained to 60 epochs. Post training Validation Accuracy was 0.7480 and Validation Loss was 0.6883.

Model Performance		
	Training	**Validation**
Model Accuracy	0.9190	0.7480
Model Loss	0.2265	0.6883

Table -1: Model Performance



Graph-1: Epoch Accuracy



Graph-2: Epoch Loss

5. RESULTS AND DISCUSSIONS

The model proposed in this paper is trained for recognition of only 5 characters. The model can be refined in the future and can be trained to recognize all the 26 alphabets from the sign language. The input sample for building the model was also sparse. An input with a lot of variations and large number of sample images can be effectively used to train the proposed model in order to increase its efficiency and accuracy as well.

6. CONCLUSION

The trained CNN model can be further developed to process the continuous input stream of sign language and convert them into their corresponding sentences. Our model's validation accuracy is 74.80% whereas training accuracy is 91.90%. High training accuracy and low validation accuracy means our model was not able to generalize well on the given data and there might be some degree of overfitting going on in our model.

7. FUTURE WORK

A. Develop the model to process the continuous input stream of sign language and convert them into their corresponding sentences.
B. Develop a mobile app based on this model.

REFERENCES

[1] Honggang Wang, Ming C. Leu and Cemil Oz, "American Sign Language Recognition Using Multidimensional Hidden Markov Models", Journal of Information Science and Engineering, January 2006.

[2] Lionel Pignou. Sander Dieleman, Pieter-Jan Kindermans, Benjamin Schrauwen, "Sign Language Recognition Using Convolutional Neural Networks", March 2015.

[3] Rafiqul Zaman Khan and Noor Adnan Ibraheem, "Hand Gesture Recognition: A Literature Review", International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3 , No.4 , July 2012.

[4] Karthick Arya, Jayesh Kudase, "Convolutional Neural Networks based Sign Language Recognition", International Journal of Innovative Research in Computer and Communication Engineering, Volume 5, Issue 10, October 2017.

[5] Vivek and N. Dianna Radpour, "Using Deep Convolutional Networks for Gesture Recognition in American Sign Language", Department of Computer Science, Department of Linguistics, State University of New York at Buffalo, 2017.

[6] Pratibha Pandey and Vinay Jain, "International Journal of Science, Engineering and Technology Research (IJSETR), Volume 4, Issue 3, March 2015.

[7] Eirini Mathe, Alexandros Mitsou, Evaggeles Spyrou and Phives Mylonas, "Hand Gesture Recognition Using a Convolutional Neural Network", 13th International workshop on Semantic and Social Media Adaptation and Personalization (SMAP), 2018.