International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 05 | May 2022 www.irjet.net p-ISSN: 2395-0072

SILINGO – SIGN LANGUAGE DETECTION/ RECOGNITION USING CONVOLUTIONAL NEURAL NETWORKS

Rudransh Kush[1], Tanisha Chaudhary[2], Shubham Gautam[3], Sai Suvam Patnaik[4]

[1,2,3,4]*Student, Department of Computer Science and Engineering, Bennett University, Greater Noida, Uttar Pradesh, India*

Abstract – *Sign Language is a medium for conversation used by the deaf and mute people that focuses on hand gestures, movements, and expressions. The hearing and speech impaired individuals have difficulty in conveying their thoughts and messages to the people. Recognizing a Sign Language is a topic of deep research and will help people who can't understand sign language and break down this communication barrier between deaf/dumb/mute people with other people. Sign Language Recognition using Hand Gesture is a System which presents a novel, organic, interactive, and easy to use method of engaging with computers that is more recognizable to homo sapiens. Human-machine interface, language of sign, and immersive game technology are all examples of applications for gesture recognition. People who are not deaf, on the other hand, find it difficult or impossible to converse with deaf people. They must depend on an interpreter, which is both costly and inconvenient for the persons trying to converse with deaf/mute/dumb people. This project aims to provide a method that can employ the abilities of layers of Convolutional Neural Network (CNN) to detect and identify hand signs taken in real time using a device with camera.*

Key Words: Sign Language, Convolutional Neural Network (CNN), sign, gestures, deaf and mute people, TensorFlow, smoothing, normalization, feature extraction, ASL, classification

Impact Statement - Sign Language Recognition and Detection using AI is one of the most demanding tools which can help in creating a smooth communication using AI and ML between two or more human beings using CNN (Convolutional Neural Network) algorithm. It is very helpful tool, and it will help in boosting the number of mute and deaf people to be actively part of workforce. It's often come to known that a deaf/mute person was rejected from a job because of difficulty in communication. This tool will break all such barriers for deaf and mute community and will let them put forward their thoughts without any hesitation. Nothing can limit talent and ideas from such creative minds to be discarded and thus, this tool gives them a voice which they were wanting since birth.

1. INTRODUCTION

Sign language is a lingually complete and highly visual-spatial language. For deaf and mute people, it is usually their first language and primary mode of communication. Sign Language is still not used in India at huge scale and is very limited to communities and NGOs working for empowerment of blind and deaf people. It is also not a universal language as many countries have different sign language interpretations contrary to common perception.

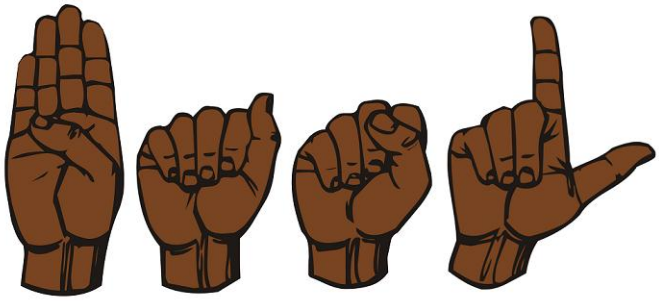

Figure 1: Sign Language

As a result, the ineffective communication between the hearing majority and the Deaf and silent people is expanding decade after decade. Written communication takes time and is only possible when people are seated or standing still or both sides have knowledge of a common language. Like, if a person is Bengali and the other person tries to communicate in Hindi then it would result in a complete failure. So, considering English as not a regional language and is an official language of India and is taught in every school it, it would be good to make a sign language recognition project which translates sign language to English words. While walking or moving, written communication might be awkward. In addition, the Deaf and mute community is less adept at writing a spoken language. The fundamental goal of a hand sign recognition system is to create a human-CNN classifier interface in which the recognized signs can be utilized to convey meaningful information or to give inputs to a machine without having to touch physical knobs and dials on that machine. Our project's main goal is to raise the bar and make progress in the field of sign language recognition. We're concentrating on detecting signals and motions as quickly as possible. There are two essential steps that must be followed in order to build an automated recognition system for human behaviours in spatial-temporal data. The first and most important step is to take the frame sequences and accurately extract characteristics from them. As a consequence, we'll obtain a representation

made up of one or more feature vectors called descriptors, which may be used to store signs. This illustration will help the computer identify between the many action classes of the sign in the library. The categorization of the action is the next stage. These representations will be used by a classifier to distinguish between the various activities and signs. Convolutional neural networks are basically used to automate feature extraction and we are using this for same in our research (CNNs).

2. TECHNOLOGY USED

A. CNN-Convolutional Neural Networks (ConvNet)

Convolutional Neural Networks are a type of Neural Network and a type of Artificial Neural Network that is commonly used in object recognition. They are made up of artificial neurons with trainable weights and biases. One or many layers such as: pooling, convolutional, and dropout, followed by one or more convolution layers, make up a ConvNet. Their purpose is to find patterns locally i.e., from within images, which is then used to detect local patterns utilizing a part of the input data to the connected layer and classify the image or sections of it. Gradient Descent and Backpropagation method are used to adjust weights and biases. CNNs are feature extraction models utilized by various industry leaders such as Google, Facebook, Instagram, Oracle, and Amazon.

B. TensorFlow

Developed by the Google Brain team, TensorFlow is an open-source software package that allows you to train and construct machine learning models from start to finish . TensorFlow is coded in C++, Python and CUDA. Both CPU and GPU can be used to execute code on TensorFlow. It has a comprehensive ecosystem of libraries, tools, and community resources that allow academics to push the frontiers of machine learning and developers to rapidly construct and deploy ML-powered apps utilizing high-level APIs like Keras. [7].

C. Keras

Keras is a free, open-source Python-based neural network library. It is a deep learning API that can make use of TensorFlow, Theano, CNTK, MX Net or PlaidML libraries in the backend [7]. Keras's popularity stems primarily from its user friendliness and ease of usage. It gives a high-level interface to backend libraries. It minimizes developer's cognitive load and provides industry strength performance and scalability. It is used by organizations and corporations including NASA, YouTube, or Waymo.

3. PROPOSED METHODOLOGY

A. METHODS

1. COLLECTION

The data would be made manually by using the camera to capture images and then using the image labelling from the TensorFlow library. We can also download images and then label it using the same Image Labelling process.

2. PRE-PROCESSING

The image would be processed and labelled manually, the first step would be, to crop the hand from the body, if both hands are used, they are usually mirroring each other's actions and furthermore the hand which is higher is the one whose gestures are being used for communication, this helps in reducing the noise for the training. First starting with the alphabet and then training for small phrases, TensorFlow and Keras library would be used to train the network. The command prompt would be used after the images have been labelled and there is a set of images which are used for trial testing, on which the dataset would be giving results. About 20000 iterations would be used for a set of 4 signs, the greater the number of iterations the better it would be able to recognize.

3. CONVOLUTIONAL NEURAL NETWORK

CNN is an algorithmic approach which majorly involves of two steps which include Feature Extraction and Classification. CNN has an advantage as it has an architecture which enables 2D structure reading. CNN is also easier to train as there are not many parameters as compared to a fully connected network.

4. MODELLING AND TRAINING

While using the CNN architecture, we would get best results as the input is a 2D image. The training would be done using the manually created dataset, which was created using the TensorFlow image labelling. A file along with each photo would be created which would include all the details, about what the image is and what it represents. TensorFlow would then use this data and test the data using some test images. Data augmentation is important to produce good quality results. While training, the more is the number of iterations the better is the accuracy of the output, loss of 3 anything less than 0.1 is good enough, if you train more the loss score can get lower, the lower the better.

B. PLANS AND DIAGRAMS

1. FLOW DIAGRAM

Recognizing sign language using hand gesture movement has stages through which it must go through. Stage one comprises of Image taken by the device with camera within conditions such as scaling, translation, and rotation. The second stage comprises of processing where smoothing; edge detection and filtering occurs. Feature Extraction happens next where hand contour and Complex Movement are used to examine image. And in last we do classification using CNN and the result of evaluation & Comparison hand contour-based CNN and Complexed moments-based CNN are carried out and resulted. Below is the flowchart which comprises of all the above stages in a precise manner.

2. ARCHITECTURAL DIAGRAM

A Convolutional Neural Network (CNN) compromises of certain layers of convolution which are in a standard multilayer neural network. It detects some local connections and the tied weights in the captured image which are followed by pooling and results in translation invariant features. CNN helps in training data efficiently and have many fewer parameters. A series of Convolutions and Pooling operation is done to extract features. Pooling will be done to maximize the value for a local neighborhood and pool them by mapping different features. 2-D convolution will provide a good result in this case and the results are matched with the output with highest probability. Hence, Image of hand sign gestures is taken as input using a camera and the output of recognized gesture using classification is produced as word or sentence

3. USE CASE DIAGRAM

Use case diagrams represents a better picture how the user is going to interact with the environment. The user starts the interaction by giving input as image using a device with camera and the system further responds using the methods provided to it to recognize/ interpret the sign to the nearest possible word/alphabet in English. The CNN is used for extracting features of the image provided by the user. It further has certain level of convolution layers and after pooling and convolution repetitively we reach the closest probable word which user was trying to commute.

4. RELATED WORK

The discoveries of Roel Verschaeren constitute the cornerstone of our work. Using Microsoft Kinect, offers a CNN model with 2.5% inaccuracy that recognizes a collection of 50 different Flemish Sign language signs. Unfortunately, because it only considers one individual in a certain context, this technique is constrained. A vocabulary of 30 words is used in an American language recognition system. They created appearance-based representations and a manual tracking system that uses a Hidden Markov Model to classify objects. In the RWTHBOSTON50 database, an error rate of 10.91 percent has been reported. Mekala et al. used the advanced extraction function and a 3-layer neural network to convert the ASL letter video to text. Features of two types were retrieved: movement and hand position

Figure 2: ASL Recognition System

The placement of six "points of interest" in the hand are reviewed prior to ASL classification: each of the fingertips and the palm's center . The photos' Fourier transforms are obtained, as well as the region of the frame where the hand is detected. Although using the framework, it claims to be able to accurately classify100% of the photos, there is no mention of achieving this in training, testing, or validation sets [8] [11]. Using a feedforward Neural Network, Admasu and Raimond correctly classified Ethiopian Sign Language 98.5% of the time. Extensive use of image pre-processing techniques was made such as image background subtraction, image size normalization, contrast correction, and image segmentation. Gabor filter and Principal Component Analysis were used by Admasu and Raimond to extract features [9]. L. Pigou et al's CNN's applications to classify 20 Italian gestures from the 2014 ChaLearn Looking at People competition for gesture identification is the most relevant study to date. Microsoft Kinect is used on full-body images of people performing movements and achieves 91.7% cross validation accuracy. The Kinect, like the 3D glove, described previously, can acquire depth data, which greatly assists in the classification of American Sign Language signs [8][10].

5. RESULT AND ANALYSIS

Our project runs on Convolutional Neural Network, with 2 convolutional layers whose pooling size was 2x2 where in the first layer and second layer 32 and 64 kernels were present which helped us in making the code highly capable and making it somewhat like 3x3 CNN system. Not much of optimization was required for getting the optimized and desired result from running the network but somewhat of regularization is the basic need. For every pair of convolutional layer and pooling layer the regularization of dropout is to 25% and 50% and which simply eliminates 1 input for every 4 and 2 for every 2 in different layers. And these layers then use ReLu and SoftMax Activation functions for complex patterns to be learnt by the network. The system was trained multiple times to lower down the error and repetitive epoch lead us to have a loss rate of less than 0.1,

and we strongly believe that continuous epoch can further work in to making the project more efficient. The best results have been observed when the loss becomes less than 0.2. Using different hyperparameters we come to a conclusion. Further the program was also able to run even when there were multiple actions under a given time period. The Loss went down from 0.752 to 0.099.

Figure 3: Roc Curve

6. COMPARATIVE ANALYSIS

On performing the first few iterations the results were not as impressive but when the iterations ran increasingly the results became better and the loss decreased. The accuracy of the applications has marginally overcome the already existing projects in this category. We have achieved a loss of 0.09 which can be a benchmark for other similar systems. Another system developed by Kshitij Bantupalli of the Kennesaw State University, achieved an accuracy of 83% with 300 samples of 150 signs, even though our data is lot less but observing the marginal difference between our system's and their accuracy, I would confidently say that our system could and possibly will perform better than that of Kshitij.M. M. Kamruzzaman developed a Sign Language Recognition system which reads and gives output regarding the Arabic Sign Language System. With Image Augmentation his loss is greater than 0.5 and without the image augmentation somewhere south of the 0.5 mark.

7. CONCLUSION

In this paper, it is shown how convolutional neural networks can be used to design an effective solution for efficiently recognizing distinct signs of a sign language without the need to include users or their environment in the training set.

REFERENCES

[1] Following are set of references used while writing this document. The sources are open and can be accessed for fair-use

[2] Cire¸san, D., Meier, U., Schmidhuber, J.: Multi-column deep neural networks for image classification. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. pp. 3642–3649. IEEE (2012)

[3] arrett, K., Kavukcuoglu, K.: What is the best multi□stage architecture for object recognition? Computer Vision, 2009 IEEE 12th International Conference on pp. 2146–2153 (2009)

[4] Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning (ICML-10). pp. 807–814 (2010)

[5] Goodfellow, I.J., Bulatov, Y., Ibarz, J., Arnoud, S., Shet, V.: Multi-digit number recognition from street view imagery using deep convolutional neural networks. arXiv preprint arXiv:1312.6082 (2013)

[6] Rafiqul Zaman Khan and Noor Adnan Ibraheem, "Hand Gesture Recognition: A Literature Review", International Journal of Artificial Intelligence & Applications (IJAIA), Vo

[7] Bachani, S., Dixit, S., Chadha, R. and Bagul, P., 2020. "SIGN LANGUAGE RECOGNITION USING NEURAL NETWORK". International Research Journal of Engineering and Technology (IRJET), Volume 07, Issue 04. pp.583-586.

[8] Alok, K., Mehra, A., Kaushik, A. and Verma, A., 2020. "Hand Sign Recognition using Convolutional Neural Network". International Research Journal of Engineering and Technology (IRJET), Volume 07, Issue 01. pp.1680-1682.

[9] Garcia, B., & Viesca, S. A. (2016). Real-time American sign language recognition with convolutional neural networks. Convolutional Neural Networks for Visual Recognition, 2, 225-232.

[10] Admasu, Y. F., & Raimond, K. (2010, November). Ethiopian sign language recognition using Artificial Neural Network. In 2010 10th International Conference on Intelligent Systems Design and Applications (pp. 995 - 1000). IEEE

[11] Pigou, L., Dieleman, S., Kindermans, P. J., & Schrauwen, B. (2014, September). Sign language recognition using convolutional neural networks. In European Conference on Computer Vision (pp. 572 - 578). Springer, Cham.

[12] P. Mekala et al. Real -time Sign Language Recognition based on Neural Network Architecture. System Theory (SSST), 2011 IEEE 43rd Southeastern Symposium 14 -16 March 2011.l.3 , No.4 , July 2012.