International Journal of Research in Engineering, Science and Management
Volume 4, Issue 6, June 2021
https://www.ijresm.com | ISSN (Online): 2581-5792

Finger Spelled Signs in Sign Language Recognition Using Deep Convolutional Neural Network

Reshmi Rajendran[1*], Sangeeta Tulasi Ramachandran[2]
[1]*Student, Department of Electronics and Communication Engineering, Sree Buddha College of Engineering, Pathanamthitta, India*
[2]*Professor, Department of Electronics and Communication Engineering, Sree Buddha College of Engineering, Pathanamthitta, India*

***Abstract*: Sign languages are a form of communication by deaf people. A person who knows sign language can easily communicate with them. It is not possible to learn sign language without the help of an expert and through continuous practice. Many versions of sign languages exist in our world. For a normal person to communicate with the deaf, they need to learn sign language that requires interest and guidance. Without continuous practice, learning signs seems difficult. This need triggers many inventions as well as innovations in the sign language learning area. Technology-based tools exist in different forms which depend on external sensors. But most of them are costlier and unaffordable. This paper discusses a GUI-based system that facilitates self-learning of Static American Sign Language (ASL). Web-based Graphical User Interface (GUI) allows operations easier and intuitive. Deep Convolutional Neural Network is used to classify finger-spelled signs that are captured via webcam. The accuracy of the system is also much better than many existing systems.**

***Keywords*: accuracy, costlier, external sensors, self-learning, sign language.**

1. Introduction

Hearing-impaired people use a natural language for interaction, known as sign language. 'Deaf' people are those who have less ability to hear. The principal means of communication of deaf to others is sign language. The communication process between them becomes less complex. At school and home, the instructions in spoken language are very difficult for the deaf to follow, which hinders the learning of deaf children. Fingerspelling is an important component in sign language learning. Fingerspelling is the form of representation of the letters and numeral systems. Sign language has numerous regional variations in the world. In the United States, the principal language used by deaf people is American Sign Language (ASL). American Sign Language (ASL) can show English alphabets A-Z using fingerspelling, which can be one-handed or sometimes two-handed style. ASL is an extremely effortful language with all the distinctions of a spoken language. Beyond a basic level, it is not able to understand easily. Memorizing requires extensive not able to understand easily. Memorizing requires extensive exposure and practice. Training a sign language would be an arduous task for the teachers. The need to observe the gestures of the teacher carefully and repeating them accurately will be a difficult task for students. Teaching signs requires keen observation of the gestures of the students for accuracy, which may lead to avoidance of moderate learners. In this case, a tutor needs to provide assistance, observation, and encouragement to students.

There are various technology-based tools for learning sign language that was developed to curb this situation. But most of them are costlier and not affordable. So, the sign language learning tools are not enough to provide a better learning process. Therefore, the problem in learning signs still exists in our society. Previous researches have been focused on glove-based solutions which are irrelevant and exorbitantly priced. These innovations also improved the problems faced by the deaf to some extent. But the innovations and inventions must be a focus on developing a system that makes the learning activity of deaf children interesting and fast-paced. Here a system is required that will work on a basic computer configuration such as a laptop or desktop with a webcam. This is achieved in our thesis; we are proposing a GUI-based approach for training and detecting the American Sign Language. The system is developed using an algorithm called Deep Convolutional Neural Network (DCNN). Deep Convolutional Neural Networks (DCNN) is said to be a Deep Learning (DL) method that shows differences in normal Convolutional Neural Network (CNN) in terms of the number of hidden layers or the number of nodes in the hidden layer. Hand gestures are available in two forms: static or dynamic. Our system is used to detect and recognize static signs. The proposed system detects input images, check the correctness of sign and assign score. The image is captured via webcam which is a cheap and economical method since it requires no external sensors. The system appears as a Graphical User Interface

*Corresponding author: reshmirajendran31197@gmail.com

based system which adds visually appealing sensation to the system.

2. Literature Survey

Our proposed system is based on image recognition using the benefits of a neural network. In order to gain information regarding the needs for the implementation of the system, we had made thorough research on the sign language recognition field of technology.

Jestin Joy, Kannan Balakrishnan, and Sreeraj M [1] proposed a system, Sign-Quiz, which was an advanced innovation in the field of sign language recognition. They used pre-trained models like Inception V3 and Nasnet; and the Indian Sign language dataset for making their system.

Helene Brashear, [2] proposed Copy-Cat which is an American Sign Language (ASL) learning tool that helps deaf children to practice ASL sign language within the style of a game. Copy-Cat is a prototype that was an after-effect of the research that combines interactive game technology with the process of sign language recognition.

E. Carpenter, N. Adamo-Villani, and L. Arns, [3] have proposed a system that provides a virtual environment for learning mathematics and its related words in the form of a sign language. The system uses Maya 6.5 and motion capture technology for modeling and animating the characters.

E. Efthimiou et al., [4] describes Dicta-Sign, a project that geared towards developing the technologies required for transforming available sign language tools into web-based. Four different European sign languages such as Greek, British, German, and French are involved in this three-year consortium research project called Dicta-Sign.

Ching-Hua Chuan, Caroline Guardino, and Eric Regina, [5] have introduced an affordable 3D palm-sized leap motion sensor-based ASL sign recognition system. They employed a support vector machine and k-nearest neighbor classifiers to classify the English alphabets in American Sign Language.

P. Kumar, P. P. Roy, and D. P. Dogra, [6] had proposed a brand new intermodal structure for sign language recognition system by assimilating facial expression with sign gesture using Kinect and leap motion sensors. The detection operation is accomplished by employing Hidden Markov Model (HMM) and then they have involved the Independent Bayesian Classification Combination (IBCC) approach for combining the decision of various paradigms for enhancing the recognition process.

Ajay Kumar Sharma, A Pardasani, Vaibhav Garg, et al., [7] designed a system which allows mute people to use their sign language in order to interact with machines. The system was implemented using CNN. The image of the hand gesture will be captured and then this data will go towards the Chabot.

Geethu G Nath, Arun C S, [8] had designed an ARM CORTEX A8processor board-based sign language recognition system with two algorithms such as convex hull algorithm and template matching algorithm for sign language recognition.

3. System Model

The proposed system shown in fig. 1 is a web-based application that has features such as user account, sign display and quiz option. On the webpage, a new user can create an account to store his details and score obtained by playing a quiz. There is an option to show the listed sign language alphabets (A to Z); clicking on each alphabet will display the corresponding sign. Also, there is an option for playing quiz; the quiz appears with questions asking signs of random alphabets.

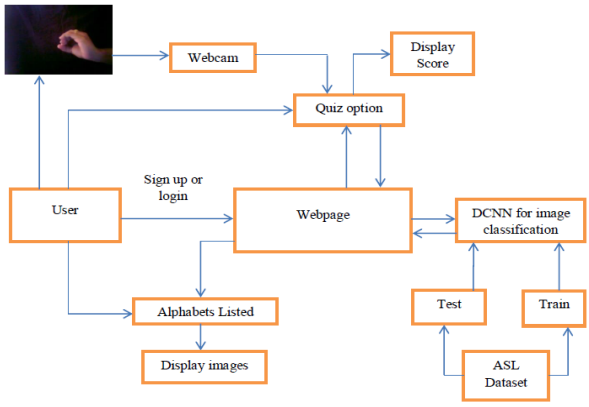

Fig. 1. Block diagram of the proposed system

On the quiz page, the user needs to click on the capture button to activate the camera and after capturing they must click on submit button. The image classification takes place with help of the Deep Convolutional Neural Network (DCNN) Algorithm and based on the accuracy of the image the score is displayed to the user.

A. Dataset

The American Sign Language in fig. 2 is a sign language that is expressed through hand gestures. The brain analyses the linguistic information through the eyes while signing. The shape of the hand, location, and movement, as well as facial emotions and body motions, all play a role in communicating information.

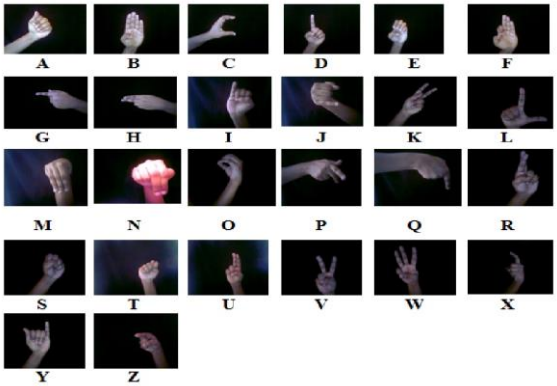

Fig. 2. American Sign Language

In our system, we have a collection of self-generated American Sign Language alphabet images that is about 150 images of each alphabet. A total of 3900 were stored in our

folder for testing and training purposes. All images are captured in a black background with the help of a webcam and are about 640x480 pixels. About 90% of our dataset is used for training the DCNN model and 10% for testing. Though the actual sign of J and Z are dynamic, we use a static image of these alphabets for our system.

B. Data preprocessing

The data gathered for training may not be formatted efficaciously or contain missing or null values. Such issues can be solved by executing the data preprocessing step. Using a well-processed dataset to train the intended model will improve its efficiency and accuracy. Two libraries used for preprocessing are numPy and matplotlib which allows us to work with arrays and helps in plotting graphs and charts respectively. OpenCV module is used for loading images as well as for performing preprocessing of images. Each image must be labeled by integer encoding to make the objects recognizable to a system that can detect similar objects and accurately predict the results.

The color images are converted into grey-scale images and resize into 227x227 pixels. Then Adaptive Gaussian Thresholding is applied for segmentation to create a binary image from a gray-scale image. If the object in the image is brighter than the background then standard thresholding is done. If it is darker then inverse thresholding is done. Binary images may have a variety of flaws due to the presence of noises and textures in them. Morphological Operations in Image Processing aim to eliminate these flaws by taking into consideration the image's form and structure. In our case, we use a morphological opening operation which is achieved by eroding and then dilating an image using the same structuring element for both operations. Opening removes any narrow connections and lines between two regions. The morphological opening operator is effective for deleting small things from a picture while maintaining the shape and size of the image's larger components. Finally, apply image normalization that converts an input image into a range of pixel values from the range of 0-255 to the range 0-1 preferred for neural network models.

C. System Architecture

A Deep Convolutional Neural Network (DCNN) is one of the machine learning algorithms typically with plenty of layers which can be a plain multilayer perceptron or CNN, used for image classification with large image datasets. Our system model is formulated with an input layer, three convolutional layers, three rectified linear units (ReLu), three Max pooling layers, two dense layers, and one SoftMax output layer.

Tensorflow is an open-source library, which uses data flow graphs to create neural network models and Keras acts as a Python interface for the TensorFlow library to perform image classification. We will split the dataset preprocessed into test data and training data. We will divide the data like 90 percent of it for training and 10 percent for testing. The preprocessed images will be fed into the Keras Deep CNN model.

We had mentioned that the Deep Convolutional Neural Network has many layers, next let's have a deep look into it. From fig. 3, we can understand clearly what is a Deep Convolutional Neural Network? It is just repeating a certain set of layers. i.e., here convolution layer, ReLu layer, and Maxpooling layer are repeating thrice to perform a particular purpose. These entire layers' ultimate goal is feature learning. Let's start with the convolution layer; this layer is used for deriving attributes or features from input data. The convolutional layers are designed to form 64, 96, and 256 feature maps respectively after convolution operation between the image matrix and filter matrix of 11x11, 5x5, and 4x4 respectively.

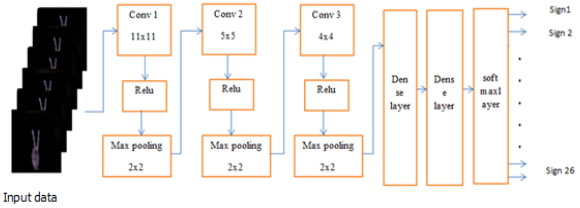

Fig. 3. The Architecture of proposed Deep CNN

The operations such as edge detection, blur, and sharpen can be enacted by applying convolution operation with the filters. The number of pixel shifts over each image matrix (is known as strides) provided is 2x2. ReLU stands for Rectified Linear Unit which is intended for executing a non-linear operation to eliminate the linearity in images that leads to gradient vanishing. Max pooling of 2x2 matrixes is added in every three layers to minimize the number of parameters by taking the largest element from the rectified feature map since the images are too large.

In order to perform the classification operation of signs, we choose two dense layers and one (activation function) softmax layer. The output from the final Pooling Layer must be flattened i.e. output matrix is converted into vector form and then fed into the fully connected (FC) layer. At the fully connected layers, we will combine the features extracted together to create the desired model. At last, the activation function named softmax classifies the outputs.

The optimization technique to change the attributes of our neural network for reducing loss is the Stochastic Gradient Descent (SGD) algorithm. This algorithm is a part added to enlarge or prolong the Gradient Descent algorithm to overcome its disadvantages. To avoid the drawbacks of requiring a large amount of memory to load the dataset of n-points at a time in the Gradient Descent algorithm, the SGD algorithm was developed. In the SGD algorithm, the derivative of the loss function is calculated by taking one point at a time. But in SGD, the updates take more iterations than gradient descent to reach minima. The algorithm has an iterative nature because the search process occurs over multiple discrete steps and each step slightly enhances the model parameters. The update procedure is based on backpropagation.

During Backpropagation, the gradient will begin to backpropagate through the derivative of the loss function with

respect to the output of the Softmax layer, and then it flows back to the whole network to calculate the gradients with respect to weights and bias. Initial weights are set by default as a value greater than zero and not than one. Because of the large amount of data, datasets are usually grouped into batches and the batch size chosen is 24 and the number of times that the learning algorithm will work through the complete training dataset is 30 epochs.

The Softmax layer in DCNN is a layer where the classification of images takes place based on computed probabilities. They make use of the cross-entropy loss function which is also called Softmax Loss. Keras provides a lot of loss functions, among them; we choose sparse categorical cross-entropy loss function to compare the predicted label and original label for calculating the loss.

D. Web based application

The entire system is implemented as a web-based application so that deaf, as well as normal users, can access it anywhere through the internet. The webpage is created with a template downloaded from colorlib, a popular WordPress theme online provider for free and premium. Their template named "ETrain" is used for the Web app, there is an option called Register for creating a user account as shown in Fig. 4 below. Then they can gain access to a web app by validating and verifying themselves by logging in.

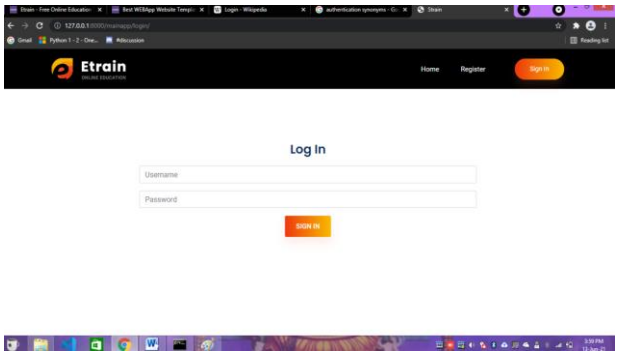

Fig. 4. The login page of sign learning app

When we enter into the account, we can see the profile details, icons like alphabets, train, and sign out. The option 'Alphabet' is designed to display icons corresponding to English alphabets as shown in fig. 5; clicking it shows the image of sign language to the user.

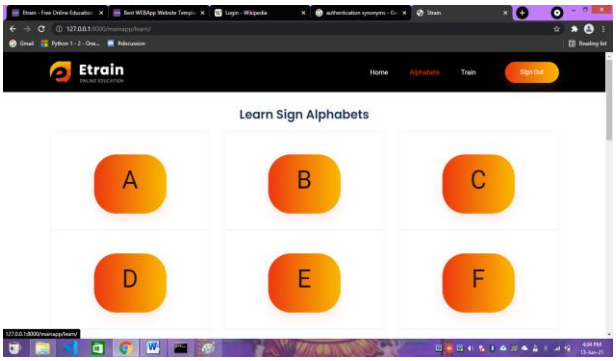

Fig. 5. Icon for displaying sign

After learning it, they can practice it by attending the quiz built-in option named 'Train'. The test page in fig. 6 comes with questions and two buttons: one button for activating the camera and the other is for submitting the answer.

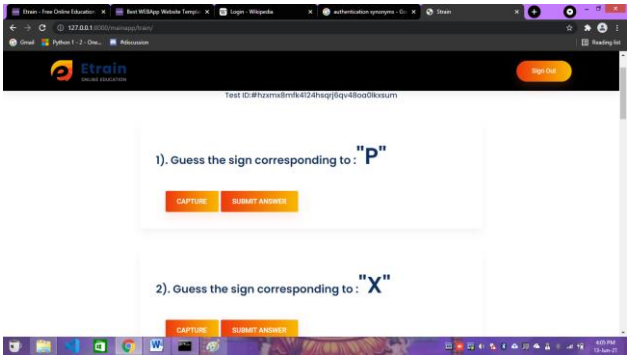

Fig. 6. Test or quiz page

On clicking 'Train', it displays Test Id and 10 questions. After attending 10 questions, the user must click on the 'end quiz' option to display the score. The scoreboard will show the Id and score of the present and previously attended quiz. For each correct answer, the system adds a score of '5'. So for 10 questions, the correctly signed user can obtain a total of 50 scores. 'Sign out' option is provided to get out of the account. The data update on a web page is communicated with the server using a method called Asynchronous JavaScript and XML (AJAX). Ajax is a client-side script that acts as a link to assist communication between browser and server without refreshing or reloading the entire web page. Django framework provides a web-based administrative interface that allows site administrators to create, manage site users, edit and publish content, etc. Django is the most popular, fully-featured server-side web framework, implemented in Python.

4. Result

The dataset of American Sign Language dataset trained with Stochastic Gradient Decent (SGD) optimizer at 30 epoch which results in accuracy, test accuracy, loss, and test loss at each epoch.

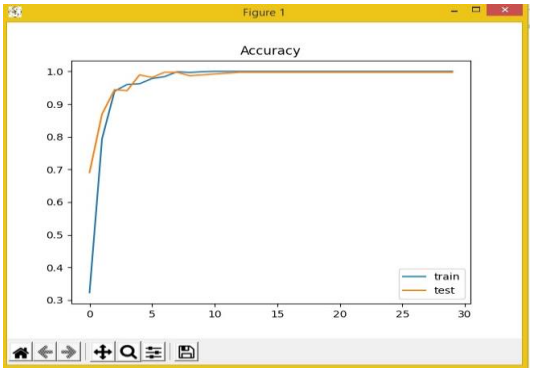

Fig. 7. Accuracy versus epoch

Accuracy obtained at the final epoch gives the entire accuracy of the trained dataset. The accuracy of DCNN versus epoch in Fig.7 and loss versus epoch is shown in Fig. 8.

Complete system performance is evaluated using the Sparse categorical cross-entropy loss function.

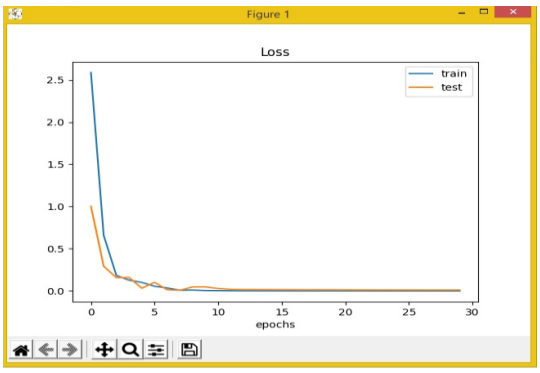

Fig. 8. Loss versus epoch

The accuracy of the system is nearly 98% which is better than many systems. Fig. 9 shows that the trained DCNN gives the correct value of label corresponding to the image of alphabet 'B'.

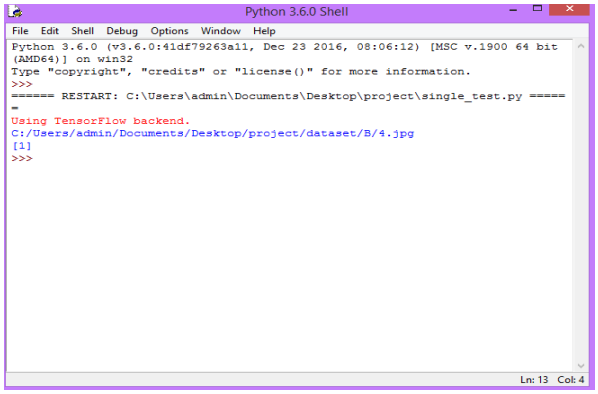

Fig. 9. Image recognition

The final output of the implemented system in fig. 10 is the score generation.

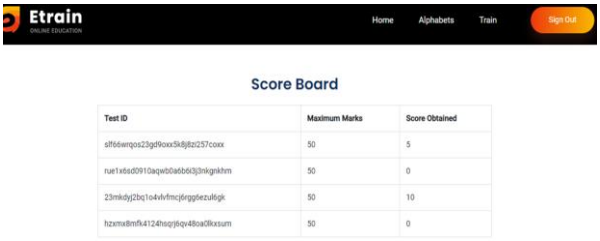

Fig. 10. Scoreboard

Our system is designed to assign a score of '5' to each correctly signed image and '0' to the incorrect signs.

5. Conclusion and Future Work

Sign language recognition by making use of Deep Convolutional Neural Networks (DCNN) is the main objective of our project. DCNN is also a machine learning algorithm that is simply a refined model of Convolutional Neural Networks. Our project uses a three-layer Deep CNN which provides 99.96% accuracy during training and 98% accuracy on testing. The loss during training and testing also decreases with accuracy in a great amount. A system is said to be efficient if the loss value is low with increasing accuracy. Our system also exhibits similar behavior. Compared to existing systems, our system comes with new features such as user account creation, logging in, logging out, alphabet display, sign test; scoreboard with test Id. Future work can include creating a mobile-based application, adding more sign languages, choosing sign language other than American Sign Language, changing the model used by adding more layers, or using RNN, ANN.

References

[1] Jestin Joy, Kannan Balakrishnan, And Sreeraj M, "SignQuiz: A Quiz Based Tool for Learning Finger spelled Signs in Indian Sign Language Using ASLR," *IEEE Access*, vol. 7, pp. 28363-28371, March, 2019.

[2] H. Brashear, "Improving the efficacy of automated sign language practice tools," in *ACM SIGACCESS Accessibility Comput.*, vol. 89, no. 1, pp. 11–17, Sep. 2007.

[3] U. N. Adamo-Villani, E. Carpenter, and L. Arns, ''An immersive virtual environment for learning sign language mathematics," in *Proc. ACM SIGGRAPH Educators Program*, Jul. 2006, p. 20.

[4] E. Efthimiou et al., "Sign language recognition, generation, and modelling: A research effort with applications in deaf communication," in *Proc. 4th Workshop Represent. Process. Sign Lang. Corpora Sign Lang. Technol.*, 2010, pp. 80–83.

[5] C. -H. Chuan, E. Regina and C. Guardino, "American Sign Language Recognition Using Leap Motion Sensor," *2014 13th International Conference on Machine Learning and Applications*, 2014, pp. 541-544.

[6] P. Kumar, P. P. Roy, and D. P. Dogra, "Independent Bayesian classifier combination based sign language recognition using facial expression," *Inf. Sci.*, vol. 428, pp. 30–48, Feb. 2018.

[7] Arjun Pardasani, Ajay Kumar Sharma, Sashwata Banerjee et al., "Enhancing the ability to communicate by synthesizing American Sign Language using image recognition in a chatbot for differently abled", *2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRIT)*, 2018.

[8] G. G Nath and C. S. Arun, "Real Time Sign Language Interpreter," *2017 International Conference on Electrical, Instrumentation and Communication Engineering (ICEICE2017)*, 2017.