Sign Language Recognition and Horizontal Voting Ensemble Implementation Using CNN Algorithm

Dr.Rohini Hanchate[1], Mr. Parth Nitin Jaiswal[2], Miss. Saniya Yogesh Gapchup[3], Mr. Rushikesh Rajendra Dhawale[4]

Computer Engineering Department[1,2,3,4]
Nutan Maharashtra Institute of Engineering and Technology, Pune, Maharashtra[1,2,3,4]

***Abstract*—**The research analysis on sign language crosses multiple fields and academic areas. These days, the two primary fields of research in gesture recognition are data glove use and visual sign language recognition. While the latter records the user's hand features with the camera for the purpose of identifying and translating sign language, the former uses the information collected by the sensor for these purposes. Deaf and hard to hearing individuals typically employ sign language as a form to interact both within and outside of their own community. In this language, communication is facilitated through hand gestures, which is particularly essential for individuals who are deaf and mute. The goal of SLR is to recognize these hand signals and translate them into spoken or written language.Within this domain, hand signs are classified into types: dynamic and static. While recognizing static hand gestures is generally easier, the recognition of both dynamic and static gestures is valued by the community. Hand gestures could be recognized using Deep Learning Computer Vision and Deep Neural Network concepts (Convolution Neural Network designs). The model will learn to recognize the hand gesture photos over the course of an epoch.

Keywords— ***Hand gestures, computer vision, text-to-speech, convolution neural networks, and recognition of sign language***

INTRODUCTION

The prime tool used by deaf and hard to hearing people to communicate with one another and within their own society through gestures of their hands and bodies is called sign language (SL).The grammar, vocabulary, and meaning are all different from those of written or spoken language. Sound waves are arranged into words and grammatical structures in spoken language to produce meaningful messages. Contrarily, sign language is a visual language that uses motions with the hands and the body. This is particularly vital for the approximately 7 million deaf individuals. Educating the deaf and mute in sign language poses challenges due to the current shortage of qualified sign language interpreters. Translating these hand gestures into the proper language of speech or writing is the aim of sign language recognition. Deep learning and computer vision are of great interest these days, and It is possible to create a variety of state-of-the-art (SOTA) models. Text matching can be produced via deep learning and image processing to classify motions. Examples of written and spoken English letters beginning with "A".

Convolution neural networks, or CNNs, are the most widely used neural network technique in deep learning and are frequently employed for image and video tasks. In order to get State of the Art (SOTA), we may use state-of-the-art neural network convolution (also known as CNN designs like LeNET-5 and MobileNetV2. These architectures are all applicable, so we can merge them with neural network ensemble methods. We are able to develop a model that can identify hand motions with almost 100% reliability by doing this. This method will be utilized for transforming hand gestures recorded by a live camera into text in standalone applications, embedded systems, and web frameworks such as Django. Dumb and deaf people will find it easier to converse because to this tech.

THEORY:-

To achieve top-tier results, we can utilize advanced neural network architectures such as LeNET-5 and MobileNetV2. By integrating these models through neural network ensemble techniques, we can create a system capable of identifying hand gestures with nearly perfect accuracy. This approach will be implemented to convert live camera-captured hand gestures into text for use in standalone applications, embedded systems, and web platforms such as Django. This technology will significantly enhance communication for individuals who are deaf or mute.

ALGORITHM AND MATHEMATICAL MODEL:-

In deep learning technique called Convolutional Neural Networking, or CNN, can take an input image, give various components or items in the image a variety of value (learnable weights and biases), and then discern between them. It is important to dissect the procedure into its constituent parts in order understand the mathematical concept underlying CNNs: convolutional layers, activation functions, pooling layers, and fully connected layers.[1]

1. **Convolutional Layers**

A convolutional layer's primary job is to execute convolutions. Sliding an image filter of size ×k×k over an input picture I, calculating the filter's dot product and the surrounding area it covers, and creating an output matrix (feature map) is the process of convolution. Mathematically, this can be written as (ref eq.01)

$$(I * F)_{(i,j)} = \sum_{k-1}^{m=0} \sum_{k-1}^{n=0} I(i+m, j+n).F(m,n) \quad \ldots\ldots(1)$$

Where (i,j) are the coordinates in the result feature map, and (m,n) repeat over the filter dimensions.

2. Activation Functions

An activation function is used following the convolution operation to provide the model non-linearity. According to equation 2, the Rectified Linear Unit (ReLU) is defined as is a popular option.The convolution's output is subjected to this procedure element by element.[2]

$$F(x) = \max(0, x) \ldots\ldots..(2)$$

3. Pooling Layers

Pooling, often max pooling, is employed to decrease the spatial dimensions of the convolutional layer's output. The max pooling operation is set (eq. 3) for a pooling window of size p×p that is,

$$P(i,j) = \max 0 \leq m, n < p\ I(i.p + m, j.p + n) \ldots.(3)$$

Where (i,j) are the coordinates in the pooled result P, and (m,n) iterates on the pooling window dimensions.

4. **Fully Connected Layers**

Every neuron in an entire linked layer is attached to all the others in the layer above it. The output y of a fully connected layer can be written (cf eq. 4) below if the weights and biases of the current layer are denoted by W and b, respectively, and the output from the one before it (or flattened feature maps) is denoted by x.

$$y = Wx + b \ldots.(4)$$

The output y is then generally put through another activation operation, such as softmax for tasks such as classification, which is defined for the first element (i) of a vector z of length K as:

Softmax: (ref eq. 5)

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad \ldots.(5)$$

PROPOSED WORK ARCHITECTURE:-

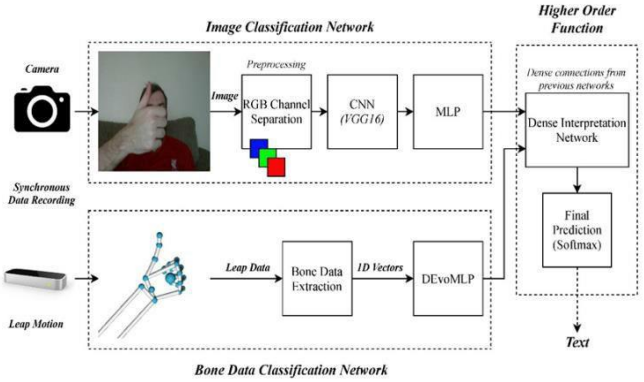

Fig. 1 System Architecture

Detailed Interpretation:

The following primary components and processes appear in the architectural diagram for sign language recognition using the (CNN) Convolutional Neural Network algorithm:

1. **Image Input**: The first step of the procedure involves feeding the system with images in sign language. These photos act as the raw material that CNN will use to train its recognition system to identify particular indicators.[4]

2. **Convolutional Layers**: The CNN's central components are these layers. They use a variety of filters on the input images to identify subtle characteristics like colors, textures, and edges. Usually, each convolutional layer builds on the outputs of the one before it to identify ever more complex characteristics.

3.Pooling Layers: During convolutional layers, the spatial dimensions (height and breadth) of the input volume decrease for the subsequent convolutional layer by using pooling (often max pooling) layers. By abstracting the features retrieved by each of the convolutional layers, this down sampling approach helps to minimize computation and regulate over-fitting by making feature detection equally scale- and orientation-invariant.

4. Fully Connected Layers: The fully-connected neural network(NN) layer are where in high-level reasoning happens followed a number of convolutional and pooling layers. As in ordinary neural networks, neurons in these layers are fully connected to every activation in the layer before it. They are responsible for combining the high-level characteristics that the convolutional and pooling layers extracted to create the final output.

5.**Classification Output**: The likelihoods of each sign language letter or action are frequently given by the softmax layer, typically the last layer in a CNN architecture. The recognized sign is the action or letter that has the highest probability.

TensorFlow:

An open-source software library called TensorFlow is intended for numerical calculation. The actual computations are carried out inside a session after the nodes in the

computation graph have been first set up. In the field of machine learning, TensorFlow is widely used.[5]

Keras:

A high-level Python wrapper for TensorFlow for neural networks is called Keras. It consumes less lines of code and is especially helpful for quickly building and experimenting with neural networks. Neural network components that are mostly utilized, including as layers, objectives, activation functions, and optimizers, are implemented in Keras. It also provides features that make managing text and visual data easier.[6,20]

OpenCV:

An open-source library devoted to real-time computer vision applications is called OpenCV (Open Source Computer Vision Library). It mainly creates image processing, video recording, and analysis easier. This includes features like object and face identification/recognition. The primary interface was initially built in C++, however for Python, Java, and MATLAB/OCTAVE compatibles are also available.[14]

PREVIOUS WORK ARCHITECTURE:-

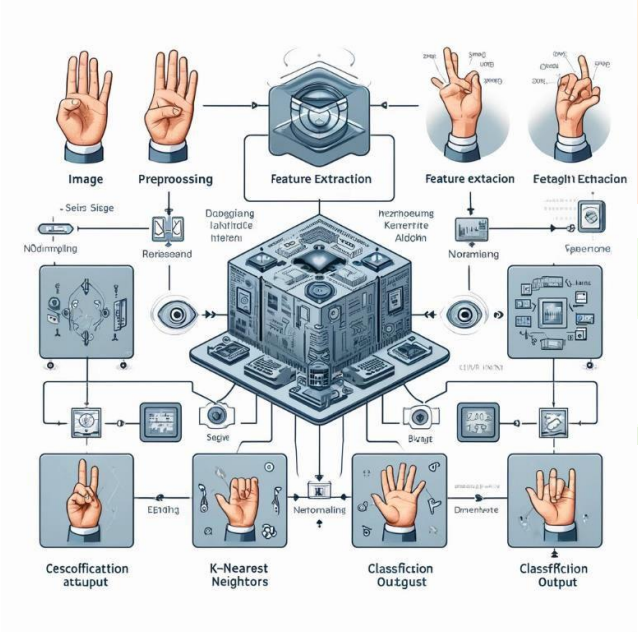

Fig. 2 Previous Architecture

The following primary parts and procedures are shown in the schematic diagram for sign language recognition using the k-nearest-neighbors (KNN) algorithm:

1. Image preprocessing: In order for roughly lighting and color conditions across all of the input pictures, this step entails downsizing the images to a typical size and normalizing them. Using this standardization is vital to extracting features correctly.[6][18]

2.Feature Extraction: On this stage, relevant characteristics from the previously processed pictures were identified and extracted. They could be edges, forms, or particular patterns that identify a motion in sign language. In order to focus on the most important portions of the photos and reduce the dimension of the data, the procedure of extraction is essential.[15]

3. KNN Algorithm: The next section shows the basic concepts of the verification procedure. Using a selected distance metric, the extracted characteristics from an input image are compared with those in a dataset to determine the 'k' nearest neighbors, or the most comparable samples. Next, the input is classified by the algorithm using the prevailing grouping among its closest neighbors.[17]

4.Classification Output: Ultimately, an established sign language signs or letter produce a result of the analysis and classification procedure carried out by the method known as KNN. The algorithm's prediction of the signaling gesture shown in the input image is represented by this result.

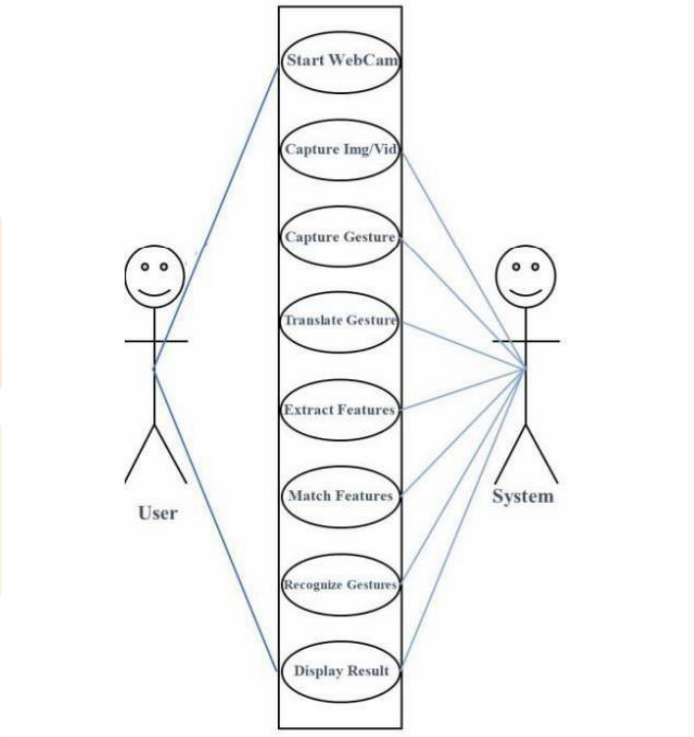

Fig. 3 User Case Diagram

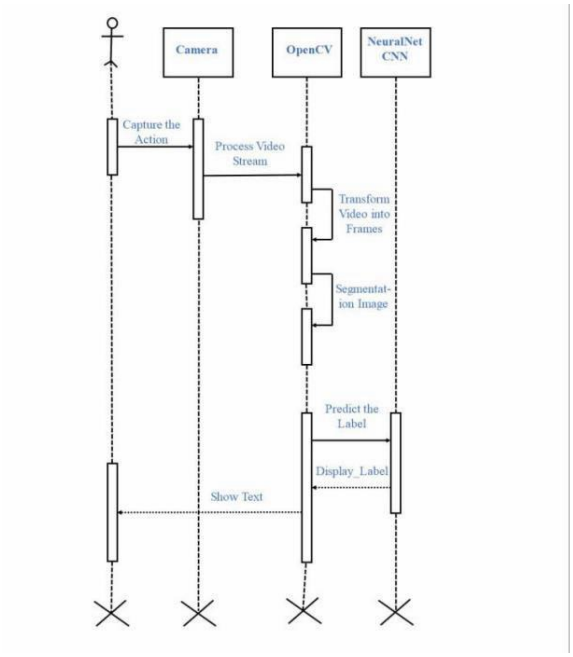

Fig. 4 System Diagram

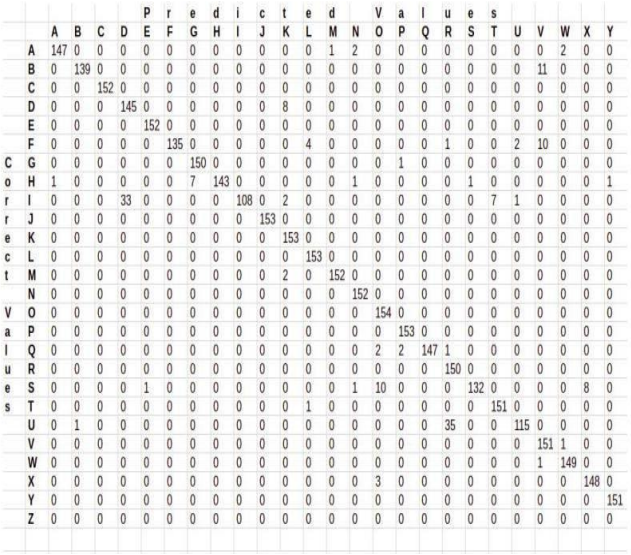

Fig. 5 Confusion Matrix

BELOW IS THE OUTPUT FOR THE WORKING GUI –

Fig.6 Working GUI

Fig.7 Working GUI

RESULTS

Our model's accuracy is 95.8% when layer 1 of our method is used separately; when layer 1 and layer 2 combine, the accuracy is 98.0%, topping the most of currently accessible study on sign language in America. A large number of research articles address hand detection with Kinect-like sensors. In [7,21] they create a Flemish sign language recognition system with a 2.5% error rate utilizing convolutional neural networks(CNN) and Iris.

In [8], a thirty-word vocabulary and a hidden markov chain model classifier are used to create a recognition model with a 10.90% error rate. For 41 Japanese static sign motions, they gives a mean accuracy of 86% in [9].

An accuracy of 83.58% and 85.49% was obtained for new system users, and 99.99% for seen new users using a depth sensors map [10]. CNN was also utilized by them for their recognition technologies.

Note that unlike several of the models shown above, our model does not relay any background subtraction techniques. As a result, the accuracy may vary when we try to use background removal in our project.

However because most of the above stated apps utilize Kinect devices, our primary goal is; our main goal was to develop a project that could be executed with efficiently accessible resources. [19,22]

FUTURE SCOPE

•**Better Reliability and Real-time Processing:** As CNN designs and instruction methodologies advance, we ought to anticipate more advances in the accuracy of sign language recognition systems. Rapid processing capabilities will continue to improve, bringing to more responsive and useful platforms for practical real-world applications.

•**Integration of 3D and Depth Data:** This boosts the system's ability to comprehend and comprehend complex signals, such ones that move a lot or are performed in three dimensions. This may indicate using 3D CNNs, or hybrid models, that are capable of successfully evaluating depth information in addition to 2D images.

•**Uses of VR (Virtual Reality) and Augmented Reality (AR):** These technologies offer intriguing possibilities for the use of sign language will be providing a platform for dumb and deaf people to communicate with the outside world.

CONCLUSION

The quest for a more accessible and inclusive society has grown considerably with the invention and use of sign language recognition technologies. the concepts, the difficulties, and uses of machine learning towards sign language recognition are all addressed in this study. Deaf and hard of hearing users may be able to connect with one other and the larger community more easily if sign language recognition is put in place. The uses of the technology are numerous and extensive, ranging from accessibility features to instantaneous translation to support for learning.

REFERNCES

[1] Prof. Radha S. Shirbhate, Mr. Vedant D. Shinde, Ms. Sanam A. Metkari, Ms. Pooja U. Borkar, Ms. Mayuri A. Khandge" Sign language Recognition Using Machine Learning Algorithm," International Research Journal of Engineering and Technology.

[2] Madhuri Sharma, Ranjna Pal and Ashok Kumar Sahoo(2014)" Indian Sign Language Recognition Using Neural Networks And Knn Classifiers", ARPN Journal of Engineering and Applied Sciences.

[3] Le, Trong T. et al. (2023) "Deep Learning for Hand Sign Language Recognition," IEEE Transactions on Image Processing.

[4] Starner, Thad et al. (2022) "Real-time American Sign Language recognition from video using hidden Markov models," ACM.

[5] Futane, P. R. and R. V. Dharaskar. 2011. HastaMudra: An Interpretation of Indian sign hand Gestures.International Conference on Electronics Computer Technology (ICECT). : 377-380.

[6] Athitsos, Vassilis et al. (2020) "The challenging CLEVR-KiDS database: Tools and methodology for hand gesture recognition," Proceedings of the 10th ACM international conference on Multimodal interfaces.

[7] Li, Ying et al. (2020) "Sign Language Recognition Using LSTM and CNN," IEEE Access.

[8] Pigou L., Dieleman S., Kindermans PJ., Schrauwen B. (2015) Sign Language Recognition Using Convolutional Neural Networks. In: Agapito L., Bronstein M., Rother C. (eds) Computer Vision - ECCV 2014 Workshops. ECCV 2014. Lecture Notes in Computer Science, vol 8925. Springer, Cham .

[9] Zaki, M.M., Shaheen, S.I.: Sign language recognition using a combination of new vision based features. Pattern Recognition Letters 32(4), 572–577.

[10] N. Mukai, N. Harada and Y. Chang, "Japanese Fingerspelling Recognition Based on Classification Tree and Machine Learning," 2017 Nicograph International (NicoInt), Kyoto, Japan, 2017, pp. 19-24. doi:10.1109/NICOInt.2017.9 .

[11] Byeongkeun Kang , Subarna Tripathi , Truong Q. Nguyen "Real-time sign language fingerspelling recognition using convolutional neural networks from depth map" 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR).

[12] Dr. Rashmi D.Kayatanvar, Prof P.R. Futane "Comparative study of sign language recognition systems", June 2020.

[13] Rohini Hanchate, Rameshwari Khamkar, Kshitij Thakre, Aditya Kotkar, Priti Jadhav, Low rate DDoS Attack Identification and Defense using SDN based on Machine Learning Method International Research Journal of Engineering and Technology (IRJET) 8 (3), 6 2021.

[14] Anuradha D Thakare, Rohini S Hanchate Introducing hybrid model for data clustering using K-harmonic means and Gravitational search algorithms , Journal International Journal of Computer Applications, Volume 88 Issue 17, 2014/1/1.

[15] Pujan Ziaie, Thomas M¨uller , Mary Ellen Foster , and Alois Knoll"A Na ¨ı ve Bayes Munich,Dept. of Informatics VI, Robotics and Embedded Systems,Boltzmannstr. 3, DE-85748 Garching, Germany.

[16] Mohammed Waleed Kalous, Machine recognition of Auslan signs using PowerGloves: Towards large-lexicon recognition of sign language.

[17] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," (2020) IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 4510-4520, doi: 10.1109/CVPR.2018.00474.

[18] Pfister, Thomas et al. (2011) "Recognizing American Sign Language Using Kinect," International Conference on Computer Vision.

[19] S.Shirbhate1, Mr. Vedant D. Shinde2, Ms. Sanam A. Metkari3, Ms. Pooja U. Borkar4, Ms. Mayuri A. Khandge/Sign-Language Recognition-System.2020 IRJET Vol3 March,2020.

[20] Vamvakas, G., B. Gatos and J. Stavros. Perantonis. 2010. Handwritten character recognition through twostage foreground sub- sampling. Pattern Recognition. 43(8): 2807-2816.

[20] S. V. Joshi and R. D. Kanphade, "Deep Learning Based Person Authentication, Using Hand Radiographs: A Forensic Approach," in IEEE Access, vol. 8, pp. 95424-95434, 2020, doi: 10.1109/ACCESS.2020.2995788.

[21] Joshi, S.V., Kanphade, R.D. (2020). Forensic Approach of Human Identification, Using Dual Cross Pattern of Hand Radiographs. In: Abraham, A., Cherukuri, A., Melin, P., Gandhi, N. (eds) Intelligent Systems Design and Applications. ISDA 2018, 2018. Advances in Intelligent Systems and Computing, vol 941. Springer, Cham. https://doi.org/10.1007/978-3-030-16660-1_105.

[22] Discover compatibility: Machine learning way., PR Ahire, P Mulay -Journal of Theoretical and Applied Information Technology, . Vol.86. No.3, 2016