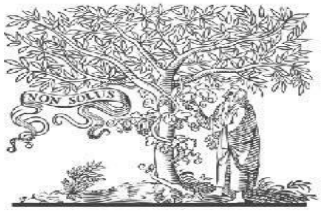
International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

COPY RIGHT



ELSEVIER SSRN

IJIEMR Transactions, online available on 10th Apr 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 04

10.48047/IJIEMR/V12/ISSUE 04/105

Title SIGN LANGUAGE RECOGNITION USING CONVOLUTIONAL NEURAL NETWORKS

Volume 12, ISSUE 04, Pages: 836-843

Paper Authors

N.Ashok, Pulivarthi Shashank , Shaik Muneer Ahamed , Tanneeru Srikanth

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

SIGN LANGUAGE RECOGNITION USING CONVOLUTIONAL NEURAL NETWORKS

N.Ashok[1], Assistant Professor, Department of IT,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
Pulivarthi Shashank [2], Shaik Muneer Ahamed[3] , Tanneeru Srikanth [4]
[2,3,4] UG Students, Department of IT,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
[1]nutalapati.ashok@gmail.com,[2]pulivarthishashank401@gmail.com
[3]muneer.toney@gmail.com ,[4]tannerusrikanth123@gmail.com

Abstract:

The ability to communicate will determine how successfully people navigate both their personal and professional lives. We may express ourselves thanks to it. In order to communicate, sign language users combine different hand gestures, positions, and movements with their arms, hands, and bodies. With the aid of sign languages, deaf-dumb persons can interact with hearing people. It comprises of word level signs, numbers, and human relations as well as fingerspelling, which spells out each letter in a word. Deaf and dumb persons, however, find it extremely challenging to communicate with regular people. So, it is difficult for them to interact with us until and unless others like us acquire the ability to communicate through sign language.The suggested system analyses and converts hand motions, which are sign language, into text using deep learning and the Python libraries OpenCV and Keras. We develop a sign detector that recognises some signs and can be readily expanded to recognise a huge variety of additional signs and hand gestures, such as the alphabets. This is broken down into three steps: building the dataset, using it to train a CNN, and predicting the records.

Key words: Sign language, ASL, CNN, Keras, Hunspell, Tkinter, and Matplotlib.

Introduction:

The most common sign language is American sign language. People who are deaf and unable to speak are limited in their ability to communicate using spoken language. As a result, they rely on sign language to exchange ideas and messages with others. Sign language involves using hand gestures, arm movements, facial expressions, and lip patterns to convey information visually. This nonverbal form of communication allows those who are deaf and unable to speak to express themselves and understand others without the use of sound. Contrary to what many people think, sign language is not universal. They differ from one place to the next.

Reducing the verbal exchange gap among D&M and non-D&M folks evolves into a goal to make certain productive

communication among everybody. For those with hearing loss, sign language translation offers the most natural form of communication and is one of the fields of study that is expanding the fastest.

Literature survey:

The literature review examined various initiatives on sign language recognition,

[1]The 1997 paper by Pavlovic et al. on hand gesture interpretation for human-computer interaction. The benefits and drawbacks of employing 3D models or image appearance models of the human hand to decipher motions were compared by the authors. Although 3D models provided more precise gesture modelling, they required a lot of computation and were unsuitable for real-time HCI.

[2]Hidden Markov Models (HMMs) for a real-time system designed to recognise continuous Mandarin Sign Language were provided with the data from Jiyong et al. as input (CSL). A 3-D tracker, two Cyber-Gloves, and raw data were collected. The training sentence was divided into basic units using the dynamic programming (DP) technique, and the Welch-Baum algorithm was utilised to estimate. Results of tests utilising 220 words and 80 sentences revealed 94.7% recognition rates for the system.

[3]In their study, Gunasekaran et al. suggested using a microcontroller system to recognise sign language and translate it into voice. The proposed model was made up of four modules: a wireless communication unit, a speech storage unit, a processing unit, and a sensing unit. The PIC16F877A and the flux sensor and APR9600 were integrated to obtain the desired result. Gloves with flux sensors that react to the gesture are used. APR9600, a microcontroller, and an appropriate circuit were used to transmit the sensor's response to the microcontroller, which then played the recorded speech. High responsiveness and reliability were features of this technology.

[4]An image processing, deep learning, and computer vision-based real-time two-way sign language communication infrastructure was proposed by Tanuj Bohra et al. For better outcomes, methods like hand detection, skin colour segmentation, median blur, and contour detection are applied to images in the collection. The CNN model, which was trained on a sizable dataset for 40 classes, predicted 17600 test images with a 99% accuracy rate in just 14 seconds.

Problem identification:

Those who are mentally impaired frequently lack access to regular social interaction. Because so few of their gestures are recognised by most people, it has been noted that they can have a tremendously

hard time interacting with regular people.People who are deaf or have hearing loss often depend on visual communication because they cannot communicate verbally. They may experience isolation, which can lead to feelings of loneliness and sadness. The deaf community may have a harder time integrating into the hearing culture, which can lower their quality of life due to the communication gap. This can lead to limited access to information, difficulty forming social connections, and challenges in adapting to society. The increased level of support from the public and financial support for international projects highlights the significance of sign language. In the modern era of technology, people with speech disabilities require a computer-based system. Researchers have been focusing on this issue for some time, and the results are encouraging. While there are many exciting advancements being made in speech recognition technology, there are currently no commercial products available for sign language recognition.The goal is to create user-friendly human computer interfaces and enable computers to comprehend language (HCI). Having a machine comprehend speech, human gestures, and facial expressions are some steps in that direction. The information that is exchanged nonverbally is gestured. A human is capable of making countless gestures at once. As human motions are seen visually, computer vision experts are particularly interested in this topic.

Methodology:

The system uses a vision-based approach, which eliminates the need for any artificial devices for communication as all signs are represented using bare hands.

Data set generation:

However, the team was unable to locate any pre-existing datasets in the form of raw photos that met their specifications, with only RGB value datasets available. Consequently, they decided to create their own dataset using the Open Computer Vision (OpenCV) library. The process they used to generate the dataset is outlined below.

For testing, around 200 photographs of each American Sign Language (ASL) symbol were taken, while approximately 800 images of each symbol were captured for training. The team started by capturing an image of each frame generated by their computer's webcam, and each frame had a Region Of Interest (ROI) marked by a blue square boundary, as shown in the illustration below.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Fig 1 defining the ROI region

Then, we use the Gaussian Blur Filter to extract different information from our image. After using Gaussian Blur, the image appears as follows:

Fig 2 Image after applying Gaussian blur filter

Gesture Classification:

To forecast the user's final sign, our method uses two levels of algorithm.

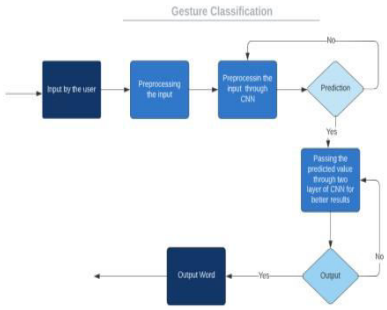


Fig 3 Gesture classification

The following is Algorithm 1:

1.The image captured by openCV is processed using the Gaussian Blur filter and threshold to extract features, resulting in input images.

2.The processed image is fed into a CNN model for prediction. If a letter is detected in over 50 frames, it is displayed and taken into consideration while constructing a word.

3.The space between two words is taken into account using the blank symbol.

The following is Algorithm 2:

1. Different sets of symbols are identified that yield comparable results when recognised.

2. Classifiers that are customised for each set are utilised to differentiate between them.

Layer 1:

Layer 1 of the model involves a CNN architecture:

1.The input image, with a resolution of 128x128 pixels, is processed by applying 32 filter weights in the first convolution layer, with each filter being 3x3 pixels in size. This produces a 126x126 pixel image with one output per filter weight.

2.The output of the first convolution layer is downsampled to 63x63 pixels using max

pooling of 2x2, where the maximum value in each 2x2 square is preserved.

3.The 63x63 pixel output from the first pooling layer is fed into the second convolution layer, which applies 32 filter weights, each being 3x3 pixels in size.

4.The output of the second convolution layer is then downsampled to 30x30 pixels using max pooling of 2x2.

5.The output of the second pooling layer is flattened into an array of 30x30x32 = 28800 values, which is then fed into a fully connected layer with 128 neurons. The output of this layer is then sent to another fully connected layer with 96 neurons, and a dropout layer with a value of 0.5 is used to prevent overfitting.

6.Finally, the output of the second fully connected layer is fed into the final layer, which has as many neurons as there are classes being classified, including the blank sign.

• Activation Function:

The activation function used in all layers is ReLU, which calculates the maximum value between zero and the input pixel value. This non-linear function helps to learn complex features and reduces computing time, thus avoiding the vanishing gradient problem and speeding up training.

• Pooling Layer:

Additionally, a pooling layer with a pool size of (2, 2) is applied after the ReLU activation function. This helps to reduce the number of parameters, computing cost, and overfitting.

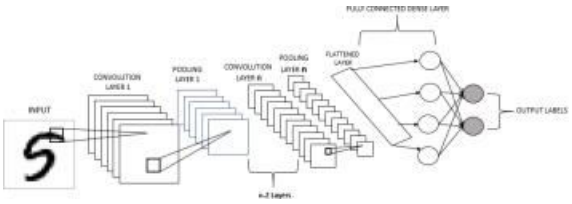


Fig 4 basic architecture of CNN

Implementation:

A. Data collection:

Data is collected manually which consists of 13,000 pictures . Dataset consists of 26 different classes for 26 alphabets. Each alphabet includes 500 individual pictures representing its sign. Because the data is raw and is unable to be utilized in the model directly, it needs to be preprocessed.

B. Data preprocessing:

To remove any noise, artefacts, or inappropriate movements that could impair the accuracy of recognition, the acquired data is preprocessed. The information is adjusted to take into account the differences in lighting, camera angles, and signers' hand sizes. At this stage, the data is scaled to a standard reference frame, such as the image's size or the hand's location in the frame. For feature extraction, the preprocessed data is passed

into a CNN. This entails identifying crucial visual cues in the captured photos or video frames, like the placement, shape, and motion of the hands and fingers. The continuous flow of sign language motions is divided into discrete components that represent various signs. This stage entails locating each sign's beginning and ending positions and segregating it from its surroundings. Every sign is identified with the class or category it belongs to, such as the phrase or word it stands for.

C. Model training :

The ability of CNNs to recognise sign language motions has been demonstrated, particularly when trained on extensive and varied datasets.The next stage is to specify the CNN's architecture, including the number of layers, their kind (such as convolutional, pooling, or fully connected), and their parameters (e.g., filter size, number of filters, activation functions).Using the training set(80% of dataset), the CNN is tuned by modifying its parameters to reduce the discrepancy between the predicted and real labels.The trained model is then put to the test on the test set(20% of dataset) to see how well it generalises. In this stage, the test data is fed into CNN, and the predicted labels are compared to the actual labels.

D. Classification:

The classification process involves a series of steps. Firstly, the preprocessed data is passed through convolutional layers that detect important visual features in the images or video frames. These filters are optimized during the training process to improve the accuracy of classification. Secondly, the output from the convolutional layers is downsampled using pooling layers to reduce dimensionality and improve efficiency. The output is then flattened and passed through fully connected layers with activation functions that introduce nonlinearity into the model. During the training process, a loss function is used to measure the difference between predicted and true labels, and the CNN is optimized using an algorithm such as stochastic gradient descent. Finally, the CNN can be used to predict the class of new data by passing it through the model and obtaining the output from the final layer.

Results:

1.Sign to text

Blank detection: When no sign is detected in the region then it specifies as blank.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Fig 5 blank detection

Letter detection: The appropriate letter is been predicted

Fig 6 letter detection

Prediction for word: It even predicts the word with the help of previous letters

Fig 7 word prediction

Word formation:

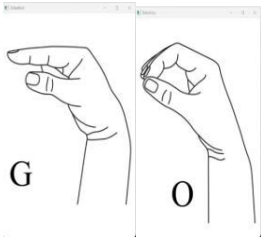
Fig 8 word formation

2. Text to sign

The system generates a slideshow format output that represents the signs of the individual letters in the word when a word is inputted.

Input : Go

Output:



Conclusion:

The study created a system in real-time for recognizing American Sign Language that could benefit individuals with speech and hearing impairments. By using a specific dataset, the system achieved an accuracy rate of 95.7%. Two layers of algorithms were used to enhance the system's accuracy by detecting and predicting symbols that were increasingly similar to

each other. The system can recognize the symbols almost all the time, provided that the symbols are visible without background noise and with sufficient lighting. The system displays the gestures as a slideshow when the user provides verbal input, enabling users to learn the fundamentals of sign language and better understand the hand movements of the deaf and mute.

Future scope:

The main goal of the project is to increase the accuracy of recognizing sign language by exploring various background subtraction techniques, especially in complex backgrounds, and by improving pre-processing methods to identify gestures accurately in poorly lit environments. The project can be made more user-friendly by creating a web or mobile application. While the current project only supports American Sign Language, it can be expanded to include other native sign languages with the appropriate data sets and training. The current project focuses on finger spelling translation, but to recognize contextual signing that involves gestures representing objects or actions, additional processing and natural language processing (NLP) would be required.

References:

[1] Gunasekaran, K., &Manikandan, R. (2013). Sign Language to Speech Translation System Using PIC Microcontroller. International Journal of Engineering and Technology (IJET), Vol 5 No 2.

[2] Kalidolda, N., &Sandygulova, A. (2018). Towards Interpreting Robotic System for Fingerspelling Recognition in Real-Time. HRI Companion: 2018 ACM/IEEE International Conference on Human-Robot Interaction Companion.

[3] Liang, R., & Ouhyoung, M. (1998). Real-time continuous gesture recognition system for sign language. Proc Third. IEEE International Conf: on Automatic Face and Gesture Recognition, pp. 558-567.

[4] “American Sign Language,” National Institute of Deafness and Other Communication Disorders, 14-Dec-2020. [Online].Available:https://www.nidcd.nih.gov/health/american-signlanguage.

[5]https://docs.opencv.org/2.4/doc/tutorials/imgproc/gausian_median_blur_bilateral_filter/gausian_median_blur_bilateral_filter.html