IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 4, April 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Reseh in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004123|

American Sign Language Recognition Using Convolutional Neural Network

Omkar Bhandari, Ashraf Shaikh, Zishan Shaikh, Siddharth Dhumal, Prof. Prasanna Kandekar

UG Students, Dept. of C.S., Keystone School of Engineering, Pune, Maharashtra, India

Assistant Professor, Dept. of C.S., Keystone School of Engineering, Pune, Maharashtra, India

ABSTRACT: Sign language is the only communication tool for a person who cannot speak and hear anything. Sign language is a boon for physically challenged people to express their thoughts and emotion. Sign language Interpreters are challenging to come by. We have developed a real-time method using neural networks for fingerspelling-based American Sign Language. And the vital purpose of our suggested system is to progress toward an intelligent system that can turn into a translator between ordinary and deaf or dumb people and can be the communication path between people with speaking deficiency and everyday people in both practical and efficient forms. The presented system is created with the camera, and the result is presented in the text's formation. This report proposes an image processing technique to determine the dormant alphabets of sign language prevalent and widely used sign languages, American Sign Language. To detect gestures, we are using a Convolutional Neural Network (CNN) based on the deep learning algorithm for adequate extraction of convenient properties to identify the American Sign Language (ASL) for classifying the hand sign. A CNN is highly efficient in tackling computer vision problems and can detect the desired features with high accuracy upon sufficient training.

KEYWORDS: CNN, ASL, Opencv, training

I. INTRODUCTION

American Sign Language (ASL) is natural syntax that has the same etymological homes as being speaking languages, having completely different grammar, ASL can be express with destiny of actions of the body. People who are deaf or can't see, it's a reliable source of absurdity. There is not any formal or familiar form of sign language. Different signal languages are speculating in particular areas. For a case, British Sign Language (BSL). Some nations adopt capabilities of ASL of their sign languages. Sign language is a way of verbal exchange via human beings diminished by speech and listening to loss. Over 5% of the world's population – or 430 million people – require rehabilitation to address their 'disabling' hearing loss (432 million adults and 34 million children). It is estimated that by 2050 over 700 million people – or one in every ten people – will have disabling hearing loss. Hearing impairment extra than 40 decibels in the better listening to ear is referred as disabling listening to loss. Thus, with growing range of people with deafness, there is moreover a rise in demand for translators. Minimizing the verbal exchange gap among listening to impaired and regular humans turns into a want to make certain effective conversation among all. Sign language translation is one of the amongst most growing line of research nowadays and its miles the maximum natural manner of communication for the humans with hearing impairments. A hand gesture recognition gadget can offer an opportunity for deaf people to talk with vocal humans without the need of an interpreter. The system is built for the automated conversion of ASL into textual content and speech. A massive set of samples has been utilized in proposed device to understand isolated phrases from the same old American Sign Language which may be concerned about the use of virtual camera. Considering all the sign language alphabets and terms, the database includes one thousand special gesture images. The proposed system intends to understand some very fundamental elements of signal language and to translate them to text and audio. American Sign Language is a visible language. Along with the signing, the thoughts techniques linguistic data through the vision. The form, placement, motion of hands, in addition to facial expressions, frame movements, every play essential factor in convey facts. It Has its very own signal 6 language, and areas have vernaculars, like the numerous

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004123|

languages are spoken anywhere inside the globally speaking language, the detection rate by the ASL language as in compare to the grammatical accuracy is of 90 % percentage of institutions commonly use Indian sign language. The amazing elements of India it [ISL] has a bit difference in signing however the grammar is identical at a few stages in the U.S.A. The Deaf humans in India remember the fact that it's plenty better than one-of-a-kind sign languages on the grounds that it's far a natural method for them, they observe via the herbal interaction with the human beings around them. The stages of sign language acquisition are equal as spoken languages, the toddlers begin with the aid of rambling with their hands. Since India doesn't have many Institutions for growing Indian sign language [other than ISLRTC which is established in year 2015: would be future of ISL] there is lack of understanding a number of the human beings and some Institution indicates to select ASL over ISL without right knowledge.

II.LITERATURESURVEY

The review of literature focuses on techniques used for gesture recognition and the color spaces used while detecting different colors in the surrounding environment.

Title*: Sign Language Alphabet Recognition Using Convolution Neural Network.*

Author: Kumar, M., Gupta, P., Jha, R. K., Bhatia, A., Jha, K., & Shah, B. K

This algorithm is useful to recognize it as a deep network, which is expected for the ASL alphabet classification task. Pre-Processing steps of the MNIST dataset are done in the first phase. After the first phase, different important features of pre-processed hand gesture image are computed. In the final phase, depending on the properties computed or calculated in the initial phases, the accuracy and AUC score of the network model with which it can recognize the sign language Alphabets were detected. The proposed CNN network has achieved an AUC score of 0.9981 and an accuracy of 0.9963**.[1]**

Title*: Sign Language to Text and Speech Translation in Real Time Using Convolutional Neural Network.*

Author: AnkitOjha, Ayush Pandey, ShubhamMaurya Abhishek Thakur, Dr. Dayananda P

We are implementing a finger spelling sign language translator. To enable the detection of gestures, we are making use of a Convolutional Neural Network (CNN). A CNN is highly efficient in tackling computer vision problems and is capable of detecting the desired features with a high degree of accuracy upon sufficient training**.[2]**

Title*: Deep learning-based sign language recognition system for static signs.*

Author: AnkitaWadhawan ,Prateek Kumar

In this research, total 35,000 sign images of 100 static signs are collected from different users. The efficiency of the proposed system is evaluated on approximately 50 CNN models. The performance of the proposed system has also been evaluated on the basis of precision, recall and F-score. The system also demonstrates its effectiveness over the earlier works in which only a few hand signs are considered for recognition.**[3]**

Title*: Development of hand gesture recognition system using machine learning.*

Author: PriyankaParvathy ,KamalrajSubramaniam ,G. K. D. Prasanna Venkatesan P. Karthikaikumar Justin Varghese ,T. Jayasankar

The developed system is to be trained and tested using Sebastian Marcel static hand posture database which is available online. Discrete wavelet transform (DWT) along with modified Speed Up Robust Feature extraction technique has been used to extract rotation and scale invariant key descriptors. Then Bag of Word technique is used to develop the fixed dimension input vector that is required for the support vector machine. The classification accuracy of class 2 and class 4 which corresponds to the 'No' and 'grasp' gesture has reached 98%. The overall classification accuracy of the HGR system using SVM classifier is 96.5% with a recognition time of 0.024 s. Due to fast recognition time, this system can be employed in real time gesture image recognition system. Our HGR system addresses the

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004123|

Complex background problem and also improves the robustness of hand gesture recognition.**[4]**

Title*: Recognition of static hand gesture with using ANN and SVM.*

*Author: Julius BAMWEND ,MehmetSiraç ÖZERDEM*I

In this paper, we propose a static hand gesture recognition system in real time using two machine learning methods namely Support Vector

Machine and Artificial Neural Networks. We use of the newly launched

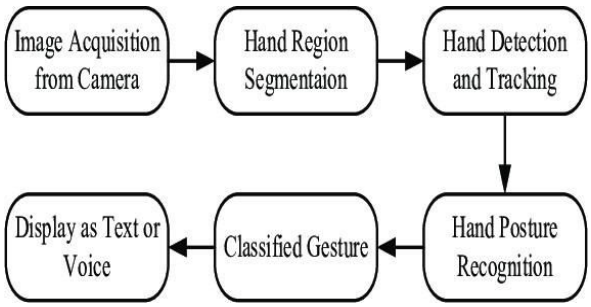
Microsoft Kinect sensor for image extraction. The sensor helps us to extract the hand images. We implement the system on aMtlab platform for reasons that Matlab is widely used by researchers in different fields and that can handle complex computations. In the training of the model, we collect a hundred depth-based Histogram of Oriented Gradient features per alphabet from the hand gesture images which we trained, tested and validated using Artificial Neural Networks (ANN) and Support Vector Machine (SVM). From this dataset, we can generate the generalized gesture model for each alphabet image. For the proposed system, the classification with ANN proves a higher performance then SVM.**[5]**

III.PROPOSED METHODOLOGY

Real time signal language to textual content and speech translation, specifically: 1. Reading man or woman signal gestures 2. Training the system learning model for image to textual content translation three. Forming words 4. Forming sentences 5. Forming the entire content 6. Obtaining audio output.

A.Flow Diagram

The flow chart explains the steps occurring to accomplish the objectives of the project. These steps have been explained in a greater detail below:



1.Image Acquisition

The gestures are captured through the web camera. This OpenCV video stream is used to capture the entire signing duration. The frames are extracted from the stream and are processed as grayscale images with the dimension of 50*50. This dimension is consistent throughout the project as the entire dataset is sized exactly the same.

2.Hand Region Segmentation & Hand Detection and Tracking

The captured images are scanned for hand gestures. This is a part of preprocessing before the image is fed to the model to obtain the prediction. The segments containing gestures are made more pronounced. This increases the chances of prediction by many folds.

3.Hand Posture Recognition

The preprocessed images are fed to the keras CNN model. The model that has already been trained generates the predicted label. All the gesture labels are assigned with a probability. The label with the highest probability is treated to be the predicted label.

4.Display as Text & Speech

The model accumulates the recognized gesture to words.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004123|

The recognized words are converted into the correspondingspeech using the pyttsx3library. The text to speech result is a simple work around but is an invaluable feature as it gives a feel of an actual verbal conversation.

B.Convolutional Neural Network for Detection

CNN are a class of neural network that are highly useful in solving computer vision problems. They found inspiration from the actual perception of vision that takes place in the visual cortex of our brain. They make use of a filter/kernel to scan through the entire pixel values of the image and make computations by setting appropriate weights to enable detection of a specific feature.

The CNN is equipped with layers like convolution layer, max pooling layer, flatten layer, dense layer, dropout layer and a fully connected neural network layer. These layers together make a very powerful tool that can identify features in an image. The starting layers detect low level features that gradually begin to detect more complex higher-level features.

C.The CNN Architecture functioning

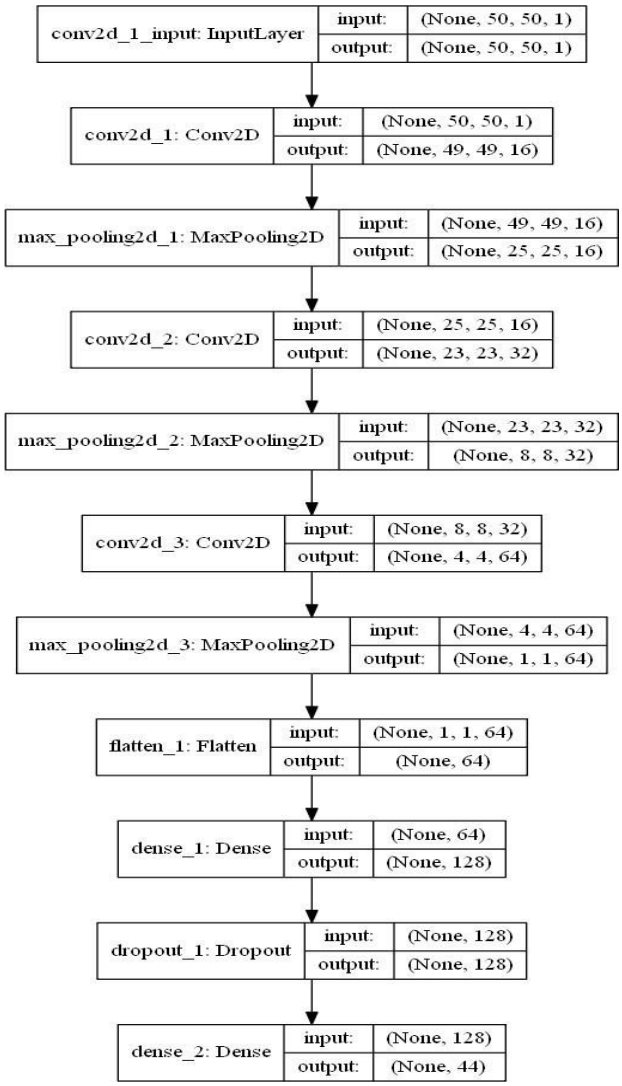
The CNN model for this project consists of 11 layers. There are 3 convolutional layers. The first convolutional layer, which is responsible for identifying low level features like lines, accepts an image with 50*50 size in the grayscale image. 16 filters of size 2*2 are used in this layer which results in the generation of an activation map of 49*49 for each filter which means the output is equivalent to 49*49*16. A rectifier linear unit (relu) layer is also added to eliminate any negative values on the map and replace it with 0. A maxpooling layer is applied which reduces the activation to 25*25 by only considering maximum values in 2*2 regions of the map. This step increases the probability of detecting the desired feature. This is followed by a second convolutional layer. It is responsible for identifying features like angles and curves. This layer has 32 filters of size 3*3 which results in the generation of an activation map of 23*23 which means the output is equivalent to 23*23*32. A maxpooling layer further reduces the activation map to 8*8*32 by finding the maximum values in 3*3 regions of the map. A third convolutional layer is used to identify high level features like gestures and shapes. 64 filters of size 5*5 reduce the input to an output of 4*4*64. A maxpooling layer reduces the map to 1*1*64. The map is flattened to a 1d array of length 64. A dense layer expands the map to an array of 128 elements. A dropout layer drops out random map elements to reduce overfitting. In the end, a dense layer reduces the map to an array of 44 elements which represent the number of classes.

Each class has a corresponding probability of prediction allocated to it. The class with the maximum probability is displayed as the predicted gesture.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004123|



D.Recognition of Alphabets

To discover bounding packing containers of various objects, as we used the Gaussian historical past subtraction which used a technique to version each history pixel with the resource of a mixture of K Gaussian set distributions (k varies from 3 to 5). The possibly historical past colorations are those that stays longer are greater the static. On those fluctuating pixels, we design a square bounding field. After Obtaining all the gesture and heritage, a Convolutional NN model has designed using those photos to apart the gesture symptoms and signs from the historical beyond. These function maps explain that the CNN can understand the common unexposed structures some of the gesture indicators within training set, and then therefore able to distinguish amongst a gesture and the past

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004123|

Fig 3.3 The Gesture Symbols for ASL Alphabets that will be in the training data

E.Algorithm

Algorithm Real time sign language conversion to text and Start

S1: Set the hand histogram to adjust with the skin complexion and the lighting conditions.
S2: Apply data augmentation to the dataset to expand it and therefore reduce the overfitting.
S3: Split the dataset into train, test and validation data sets.
S4: Train the CNN model to fit the dataset.
S5: Generate the model report which includes the accuracy, error and the confusion matrix.
S6: Execute the prediction file – this file predicts individual gestures, cumulates them into words, displays the words as text, relays the output. Stop

IV.CONCLUSION

The project is a simple demonstration of how CNN can be used to solve computer vision problems with an extremely high degree of accuracy. A finger spelling sign language translator is obtained which has an accuracy of 98%. The project can be extended to other sign languages by building the corresponding dataset and training the CNN. Sign languages are spoken more in context rather than as finger spelling languages, thus, the project is able to solve a subset of the Sign Language translation problem. The main objective has been achieved, that is, the need for an interpreter has been eliminated. There are a few finer points that need to be considered when we are running the project. The thresh needs to be monitored so that we don't get distorted grayscales in the frames. If this issue is encountered, we need to either reset the histogram or look for places with suitable lighting conditions. We could also use gloves to eliminate the problem of varying skin complexion of the signee. In this project, we could achieve

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 4, April 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1004123|

accurate prediction once we started testing using a glove. The other issue that people might face is regarding their proficiency in knowing the ASL gestures. Bad gesture postures will not yield correct prediction. This project can be enhanced in a few ways in the future, it could be built as a web or a mobile application for the users to conveniently access the project, also, the existing project only works for ASL, It can be extended to work for other native sign languages with enough dataset and training. This project implements a finger spelling translator, however, sign languages are also spoken in a contextual basis where each gesture could represent an object, verb, so, identifying this kind of a contextual signing would require a higher degree of processing and natural language processing (NLP). This is beyond the scope of this project.

REFERENCES

[1]Kumar, M., Gupta, P., Jha, R. K., Bhatia, A., Jha, K., & Shah, B. K. (2021). Sign Language Alphabet Recognition Using Convolution Neural Network. 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, 2021.

[2]AnkitOjha, Ayush Pandey, ShubhamMaurya, Abhishek Thakur, Dr. DayanandaP.Sign Language to Text and Speech Translation in Real Time Using Convolutional Neural Network, IJERT,21-09- 2020.

[3] AnkitaWadhawan, Parteek Kumar. Deep learning-based sign language recognition system for static signs. SPRINGER, 2021.

[4] PriyankaParvathy, KamalrajSubramaniam, G. K. D. PrasannaVenkatesan, P. Karthikaikumar,Justin Varghese & T. Jayasankar . Development of hand gesture recognition system using machine learning. IEEE. 16 July 2020.

[5] Julius BAMWEND1, Mehmet Siraç ÖZERDEM Recognition of static hand gesture with using ANN and SVM. IEEE. 2019.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details