RESEARCH ARTICLE OPEN ACCESS

Sign Language and Emotion Recognition Using Convolution Neural Networks

Mohammed Imran SK[1], Thota Ashmitha Reddy[2], Mandala Koushitha reddy[3], Kondam Sanjana Reddy[4]

1(Mallareddy Engineering College For Women(Autonomous),Hyderabad,India
Email: smimran.it@gmail.com
1(Mallareddy Engineering College For Women(Autonomous),Hyderabad,India
Email: ashmithar2509@gmail.com)
2(Mallareddy Engineering College For Women(Autonomous),Hyderabad,India
Email: koushithareddy.m@gmail.com)
3(Mallareddy Engineering College For Women(Autonomous),Hyderabad,India
Email: ksanjanareddy1306@gmail.com)

Abstract:

In this study, a real-time sign and emotion identification system is developed utilising a deep learning toolkit for sign language and a traditional neural network for emotion recognition.We aim to combine a manual sign language recognition system with an emotion recognition system that can recognise both emotions and gestures in real time because sign language is not an international language and cannot be used as a medium of communication between an impaired person and a normal person.

KEYWORDS:CNN, facial expressions, emotions, gestures, sign language, and open cv.

I.INTRODUCTION

There are not many technologies available that can help deaf and communities to connect with the rest of the world. so, the only medium they can use is sign language.

The practice of sign language may be viewed as a technique for sensory people and deaf people to communicate freely with one another. However, sign language also encompasses facial emotions in conjunction to motions. Face expressions accomplish the same thing as prosody in sign language as intonation does in spoken languages.This is how signers place emphasis on words as well as convey semantics such as sarcasm. In our project, we combine a gesture recognition system with an emotion recognition system to create a prototype that can recognize both sign language and emotions in real time. Here we will demo a few scenarios of our system.

Our current setup is compatible with any computer equipped with a 2D camera. Sign language Recognition uses the Open CV Deep Learning Toolkit using a model trained on 10 different signs. Convolutional neural network trained on four emotions plus neutral serves as the basis for the emotional recognition modelachieving a high accuracy final result of 83%. The results from these two are then combined into afinal output. In future, we'd be looking to integrate this system into a robot in areas such as areceptionist or some other people facing roles, and the emotion output can serve as extra information to inform the robot's response. Even though the communication can be done through writing but when immediate response is needed in emergency conditions sign language is more efficient and quicker.

There are more than 200 sign languages spoken throughout the world, with Chinese, Irish, British, Spanish, and American being the most often used ones.

System for sign language recognition remains a challenging issue. Additionally, it calls for the ability to recognise signs in human body postures, facial expressions, hand gestures, and stances, as well as emotions. In addition, sign languages include tens of thousands of phrases, including hand gestures that are very similar.

Language recognition is divided into two categories: recognitions of relaxed hand motions and movements of the hand that are in motion. It is vital to note that the focus of this project paper is sign language restrained finger spelling. As we utilise it in various languages, it is a crucial component of language recognition. Disability-related issues are eliminated via sign language.

METHODOLOGY PHASE-1

1.1 Sign language recognition:

The sole means of communication for those who are deaf or dumb is sign language; however, this problem can be solved by building a computer that can quickly translate all sign languages into acceptable language, raising the status of deaf and dumb individuals in society. In many aspects of our daily lives, such as when referring to names, brands, locations, and traffic signals, we also utilise sign language.

One of the underutilised ideas that may assist a huge portion of the population is this one. Static and dynamic gestures are the two categories into which sign language detection may be divided.

The motions may be recognised using machine learning and image classification.Static pictures captured by a 2D camera are included in the image data sets utilised here. In terms of dynamic pictures, deep learning approaches may be used to do the following, which are shown to be effective. This will enable us to develop a Deep Belief Networks model that will be able to recognise movements in the continuous video stream (DBN). In addition, continuous sign language identification using HMM is advantageous and achieves an overall accuracy of 99%. Feed-forward networks were used in another study on deep learning approaches for sign categorization. All of the works mentioned above include instructions on how to extract hand signals before supplying them to the network. CNN uses deep learning to give the precise and effective information.

1.2 PROPOSED MODEL:

The proposed model emphasizes on a deep network architecture which will detect the signs through hand gestures those which are numeric gestures using convolutional neural network from the deep learning and with the help of open CV.In order to train and forecast the data set, the CNN model is employed, and in order to record the hand motions, open CV is used. The hand motions for our system are captured using a camera. In order to normalise the size of the captured photos, the open CV library of Python is utilised. Using a variety of methods, including the colour extraction algorithm, the undesirable areas and backgrounds of the image are removed. Let's say we have a data set of 2000 photos, and we decide to utilise 1600 of them for training and the remaining 400 for testing. This will result in an 80:20 ratio.

Having followed the extraction of the digital pixel from each input frame, CNN approaches are used for further computation, conditioning, and categorization. The system that we advocate has the block diagram below:

1.3 **CNN AND ITS LAYERS**:

A thousand photographs may be tested and trained on by CNN, which can then anticipate and categorise them properly. Here, we have taken into consideration the 26-alphabet American sign language. Finally, we were able to convert the signs into their corresponding alphabets.

For humans it is quite easy to differentiate between various objects in real time, in contrast it is a lot harder for computer to do this. A CNN consists of several layers, including convolutional, pooling,

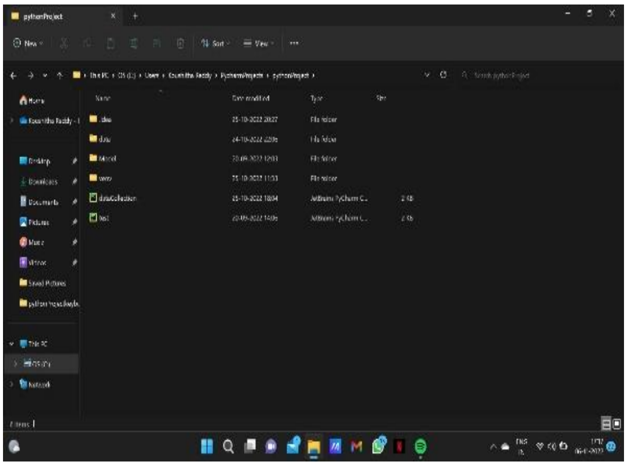
and fully linked layers. The core component of CNN is the convolutional layer, which produces a straightforward matrix by extracting features from the input pictures. The size of the resultant matrix will be further reduced by pooling.It will obtain a smaller matrix with less weights and less required training. And in fully connected layers,
the classification takes place. The CNN classifies the image using some activation functions
.

1.4 MODEL OF WORKING

(i) Developing actual data sources: Although it is possible to obtain datasets online, for this project we will generate the dataset ourselves.

Every frame that identifies a hand in the region of interest, also known as ROI, created will be placed in a directory that includes two folders, train and test, each of which has 10 folders holding pictures gathered using the generate _gesture data.py programme.

We obtain the live camera feed using OpenCV to generate the dataset, and we then establish a ROI, which is just the area of the frame where we want to recognise hands for motions.



Name	Date modified	Type	Size
.idea	25-10-2022 20:27	File folder	
data	24-10-2022 [illegible]	File folder	
Model	[illegible]	File folder	
venv	[illegible]	File folder	
dataCollection	25-10-2022 18:04	JetBrains PyCharm C...	2 KB
test	[illegible]	JetBrains PyCharm C...	2 KB

Fig-1:Directory

The ROI is highlighted by the pink box, and this window enables you to retrieve the webcam's video feed.

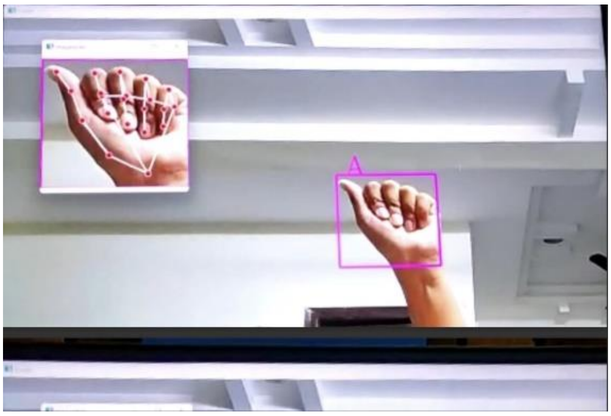


Fig-2: Recognition of an alphabet using sign language

To discriminate between the background and foreground, we compute the background's cumulative weighted average and then remove it from the frames that include a distinguishable foreground item in front of the background.

(ii) CNN Testing:

We now build a data set and train a CNN.

We obtain the information using the Image Data Generator of Keras, which allows us to flow from directory function to gather, train, and test set data. Each of the folder names' titles corresponds to a class name for the imported photos.

The validation dataset and formula are used to calculate the accuracy and loss for each input.

When the acceptance loss is undiminishing, the model's LR is determined using Reduce LR to prevent the model from overshooting the loss minima. Additionally, we are using the early stopping algorithm so that the training is stopped if the acceptance accuracy continues to decline for a number of epochs.

(iii) Forecasting the set of data:

In this step, we create a limiting box to collect data from the ROI (region of interest) and assess the accumulated average in the same way that we did when creating the dataset. In order to specify any desired item, this is done.

Given that we now know the maximum contour and since the presence of a contour indicates the presence of a hand, the ROI threshold may be thought of as a test set.

Using Keras, we load the previously stored model, get it, and save the threshold picture of the ROI containing the hand as an input to the model for prediction. gathering the necessary data for item for gesture.py

1.4 Findings and discussion

With the use of photos from a device's camera, a CNN algorithm was found to recognise novel sign language hand gestures. In order to create a written system that can be used as an input system for computers with the use of any computer camera, deaf hand motions are translated into text outputs in this project. This system demonstrated improved results by dominating a deep learning method. All of the test results that were obtained are discussed in this section.VGG Net was employed to create a multi-class recognition system. Every ASL sign was maintained as a separate category in the recognition system, which was acted upon by CNNs. A grade between 0 and 27 would be the item's output, which would be one of 28 grades. The system successfully identified 10 ASL letters as a first step (A, B, C, D, E, F, G, H, I, J). In order to achieve a threshold of fewer than 2000 images, the system was trained using just 10 packets of training data, as shown in table 1:

alphanumeric	classification	Quantity of samples
A	0	1984
B	1	2098
C	2	1931
D	3	2022
E	4	2186
F	5	2046
G	6	2135

Table:1-training data for each class

The letter G earned the highest result, 99.9567%. With a 99.9533% accuracy rating, Delete came in second, and the letter F had a 99.9069% accuracy rating. The letter A, however, had the lowest accuracy (97.3182%), while the letter N had the second-lowest accuracy (97.3793%), as seen in Table 4. The reason is because the letters M and N seem alike and can only be distinguished by how the thumb is positioned. The system was also able to recognise the letters J and Z despite the fact that they include movement by analysing each movement from three separate angles.

PHASE 2

RECOGNITION OF EMOTION:

Effective human-computer interaction depends on research in the area of emotion recognition. Currently, computers are becoming more adept at reading emotions. We could need to put up a second layer of protection where, in addition to the face, the emotion is also sensed. This can be considered as the second stage after face detection. To identify the individual in front of the camera, this might be beneficial. Body language, spoken signals, nonverbal cues, and electroencephalography may all be used to identify human moods (EEG). Seven categories of emotions may be characterised: joyful, sad, afraid, disgusted, furious, neutral, and astonished.

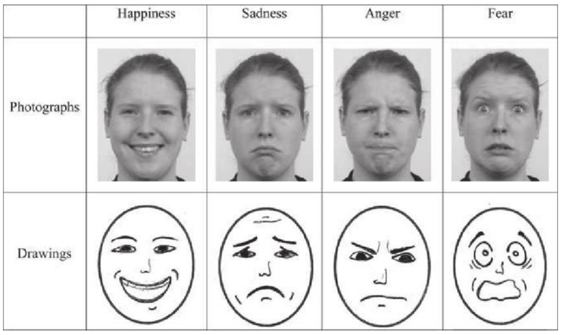


fig- 3:types of emotions

Due to the limited amount of facial muscle contortions, it can be difficult to distinguish between them because even a slight variation can cause a change in expression. Emotions are very context dependant, therefore even the same feeling may be expressed differently by various persons or even the same person again. But neural networks and machine learning have been used to these tasks with successful outcomes. Pattern recognition and classification have been shown to benefit greatly from machine learning techniques.

The following are generalisations of the emotion detection procedures:
1. Gathering data (cropped face and correct label)
2. Extraction of features and face detection
Educating the data set 3.
4. Assess the training's outcomes

2. Detecting faces and extracting features

The picture may be normalised in vector form and many sorts of characteristics can be extracted from it. Some of the most popular characteristics that may be utilised to train machine learning algorithms are as follows:

2.1.1 Landmarks

The Dlib package includes a 68 facial feature detector that recognizes the placements of 68 points on the face. Facial landmarks are very fundamental and could be employed in face detection and recognition. This accomplishment can be used with expression.

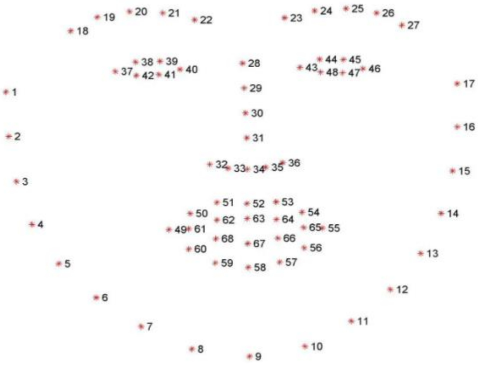
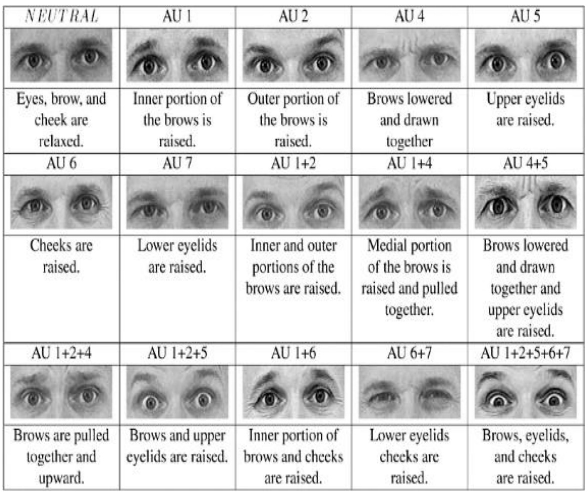


fig-4:landmarks to identify expressions

These are the 68 points, which are further separated into the left eye, right eye, left brow, right brow, mouth, nose, and jaw. Each face point's coordinates (x, y) may be extracted using the dlib library.

2.2 FACS

Each and every visual information is assigned a unique number by using **Facial Action Coding System**. The action unit is the number that has already been stated. An action unit can be employed to represent the subtle alterations in the muscles.



NEUTRAL	AU 1	AU 2	AU 4	AU 5
Eyes, brow, and cheek are relaxed.	Inner portion of the brows is raised.	Outer portion of the brows is raised.	Brows lowered and drawn together	Upper eyelids are raised.
AU 6	AU 7	AU 1+2	AU 1+4	AU 4+5
Cheeks are raised.	Lower eyelids are raised.	Inner and outer portions of the brows are raised.	Medial portion of the brows is raised and pulled together.	Brows lowered and drawn together and upper eyelids are raised.
AU 1+2+4	AU 1+2+5	AU 1+6	AU 6+7	AU 1+2+5+6+7
Brows are pulled together and upward.	Brows and upper eyelids are raised.	Inner portion of brows and cheeks are raised.	Lower eyelids cheeks are raised.	Brows, eyelids, and cheeks are raised.

fig-5:FACS

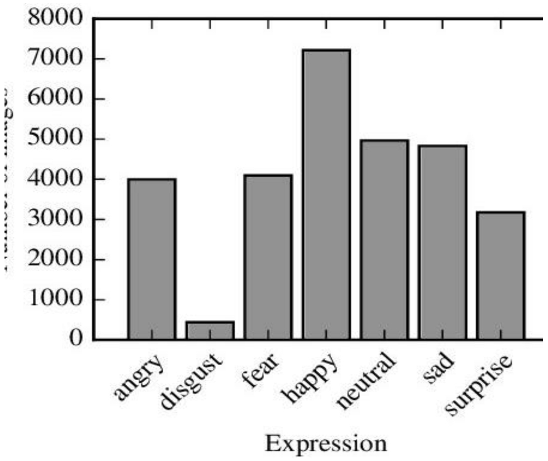
As an illustration, the definition of a stunning face in terms of action units is 1 + 2 + 5, which just indicates that the movement of the AU 1+2+5 and AU 6 muscles causes a joyful face.

3. APPLICATION (training)

The usage of convolution neural networks for emotion recognition is employed. employing machine learning methods to anticipate and extract features using Python and image processing packages. There are seven classes in this database to be classified.

The following key emotions are contained within this dataset: (0=Angry, 1=Contempt, 2=Afraid, 3=Delighted, 4=Sad, 5=Unexpectedness, and 6=Neutral). Events such as:

- Configuring the database
- Pipeline enabling image processing
- Extraction of visual appearance



FUTURE WORK:

We will work to expand and improve this project so that it is capable of employing machine learning to create an accurate and effective model and learning new symbols and letters on its own.

CONCLUSION:

We believe that sign language alone cannot produce the greatest results to comprehend the handicapped folks properly since the way their express their feelings differs from others. However, sign language will become one of the effective instruments that will aid people with hearing and speech disorders. In order to achieve the greatest results, this model should be used with both sign and emotion recognition.

REFERENCES:

[1]. *Mehreen Hurroo , Mohammad Elham, 2020, Sign Language Recognition System using Convolutional Neural Network and Computer Vision, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 12 (December 2020),*

[2]. *Ahmed KASAPBAŞI, Ahmed Eltayeb AHMED ELBUSHRA, Omar AL-HARDANEE, Arif YILMAZ, DeepASLR: A CNN based human computer interface for American Sign Language recognition for hearing-impaired individuals,Computer Methods and Programs in Biomedicine Update,Volume 2,2022,*

[3].*International Conference on Trendz in Information Sciences and Computing (TISC). : 30-35, 2012.*

[4]. *Herath, H.C.M. & W.A.L.V.Kumari, & Senevirathne, W.A.P.B & Dissanayake, Maheshi. (2013). IMAGE BASED SIGN LANGUAGE RECOGNITION SYSTEM FOR SINHALA SIGN LANGUAGE*

[5]. *Huang, J., Zhou, W., & Li, H. (2015). Sign Language Recognition using 3D convolutional neural networks. IEEE International Conference on Multimedia and Expo (ICME) (pp. 1-6). Turin: IEEE.*

[6]. *Kuehne, H., Jhuang, H., Garrote, E., Poggio, T., & Serre, T. (2011). HMDB: a large video database for human motion recognition. Computer Vision (ICCV), 2011 IEEE International Conference on (pp. 2556-2563). IEEE*

[7]. *M. Geetha and U. C. Manjusha, , A Vision Based Recognition of Indian Sign Language Alphabets and Numerals Using B-Spline Approximation, Inter- national Journal on Computer Science and Engineering (IJCSE), vol. 4, no. 3, pp. 406-415. 2012.*

[8]. *Raut, Nitisha, "Facial Emotion Recognition Using Machine Learning" (2018). Master's Projects.632.DOI:https://doi.org/10.31979/etd.w5fs-s8wdhttps://scholarworks.sjsu.edu/etd_projects/632*

[9]. *H. Ebine, Y. Shiga, M. Ikeda and O. Nakamura, "The recognition of facial expressions with automatic detection of the reference face," 2000 Canadian Conference on Electrical and Computer Engineering. Conference Proceedings. Navigating to a New Era (Cat. No.00TH8492), Halifax, NS, 2000, pp. 1091-1099 vol.2.*

[10].*K. M. Rajesh and M. Naveenkumar, "A robust method for face recognition and face emotion detection system using support vector machines," 2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT), Mysuru, 2016*

[11]. *X. Jiang, "A facial expression recognition model based on HMM," Proceedings of 2011 International Conference on Electronic & Mechanical Engineering and Information Technology, Harbin, Heilongjiang, China, 2011*