Hand Gesture Recognition System Using Deep Learning

Prof. Nilam Honmane[1], Ashitosh Nevase[2], Sourabh Pawar[3], Vaibhav Purane[4], Saurabh Sawant[5]

[1]Guide, [2345]Students

Department of Information Technology,
Zeal College of Engineering and Research,
Savitribai Phule Pune University, Pune,
Maharashtra, India

E-mail: zcoer.it.seb134@gmail.com

Abstract – Hand gesture recognition is an attractive research field with a wide range of applications. Another important application of hand gesture recognition is the translation of sign language. The importance of hand gesture recognition has increased due to the prevalence of touchless applications The software proposed in this project is created using Python, NumPy, Open-CV, Tk-inter, labelling, and TensorFlow. Recent research has proved the supremacy of Convolutional Neural Networks (CNN) for image representation and classification. the CNN model, the provided image or video will be classified as the respective Alphabet or number from the American Sign Language Set. The proposed system is evaluated on a very challenging dataset, which consists of 11 dynamic hand gestures. The results show that the proposed system outperforms state-of-the-art approaches, demonstrating its effectiveness. The model was trained on 800 images and tested on 400 images. The model with augmented data achieved an accuracy of 86.75%.

Keywords - Convolutional Neural Networks, OpenCV, Computer vision, Deep Learning, TensorFlow, Gestures.

I. INTRODUCTION

Gestures are an aspect of body language that can be conveyed by the centre of the palm, the position of the fingers and the shape of the hand. Other existing systems ignore the local configuration of the fingers and only consider the global configuration of the body. These systems have been used successfully in some HCI applications with a small number of defined gestures. If computers can understand the movements or gestures of human hands, we can close the gap and the task will become much easier. Human hand recognition is used in image processing, network security, robotics and many mother fields. Gesture recognition has been a very active research topic lately. As the interest and value of gesture detection increases, experiments are performed to validate application and system results. Many countries have their own standards for sign language gestures and interpretation. For example, the alphabet of Korean Sign Language is different from that of Indian Sign Language. It highlights the richness of sign languages, but also their ambiguity. Deep learning requires mastery of gestures to achieve reasonable accuracy. Any technique can be used to identify the logo.

II. RELATED WORK

Hand gestures are a form of nonverbal communication widely used in deaf communication, robotic control, human-computer interaction (HCI), home automation and medical applications. Gesture based research papers use many different techniques, including those based on instrumented sensor technology and computer vision. Gesture recognition can be recognized through human movement. But the automatic recognition of human gestures is a big challenge. This article reviews some previous research. In this article, we focus on creating the most accurate way to communicate or interact between computers and humans. We used a
webcam based approach where user provided gestures are captured, processed, then predicted by the machine, and then disabled for specific gestures. For the construction of the project model, we used the CNN algorithm, but to draw this conclusion, which model to use and why, we studied some previous research papers. In addition, the price of such equipment is quite high. For this investigation, they used a different approach, based on computer vision of gestures.
Human arm motion recognition using deep reinforcement learning in this approach, they used a deep reinforcement algorithm. 2019 Jaya Prakash Sahoo, Samit Ari, Sarat Kumar Patra Gesture recognition using PCA. Deep CNN reduced functions and SVM classifier authors in this model, they used a Support Vector Machine (SVM) model). In 2018 Guillaume Devineau, Wang Xi, and Fabien Moutardel, and Jie Yang used Deep Learning Skeletal Data Gesture Recognition In this model, they use a CNN to create a skeletal model of the hand.

Year	Author	Title	Methodologies
2020	Munir Oudah, Ali Al-Naji and Javaan Chahl	Hand Gesture Recognition Based on Computer Vision	It uses a model on instrumental glove approach.
2018	W.Soek, Y.Kim, C.Park.	Recognition of Human Arm Movement Using Deep Reinforcement Learning	In this approach they have used a deep reinforcement algorithm.
2019	Jaya Prakash Sahoo, Samit Ari, Sarat Kumar Patra	Hand Gesture Recognition using PCA-based Deep CNN Reduced Features and SVM Classifier Author	In this model they have used the support vector machine (SVM) model.
2018	Guillaume Devineau, Wang Xi and, Fabien Moutardel and Jie Yang	Deep Learning for Hand Gesture Recognition on Skeletal Data	In this model they created a hand skeletal model using CNN.

The authors of the article [1] used an approach based on an instrumented glove model, since these wearable sensors can be used to capture hand movement and position.
Additionally, they can easily provide precise coordinates of palm and finger position, orientation, and configuration using sensors on the glove. However, this method requires the user to be physically connected to the computer, which reduces the convenience for the user to interact with the computer. In this investigation, they used a different approach, namely that camera vision sensors are a common, appropriate and applicable technology because it enables contactless communication between humans and computers. The cameras are available in different configurations such as monocular, fisheye, TOF and IR. However, this technique involves several challenges, including lighting changes, background issues, occlusion effects, complex backgrounds, and trade off processing time for resolution and frame rate pictures and show foreground or background objects that have the same skin colour or otherwise appear as hands. The article [2] studies deep reinforcement learning methods. Algorithms using this approach learn patterns from sensors using only reward feedback without class labels. It allows users to control IoT devices and generate desired arm movement patterns without creating beacons. In this paper, the performance of a convolutional neural network (CNN) with a DQN model is compared to that of a LSTM model with DQN.
Reinforcement learning has become very popular since the appearance of AlphaGo, an artificial intelligence Go program developed by Google Deep Mind. Therefore, in this investigation, we found that the model used is a deep reinforcement algorithm, which is too advanced for the system and requires a huge data set. Deep features are extracted from the fully connected layers of the pretrained Alex-Net. The PCA dimensionality reduction technique is then used to reduce redundant features in the feature vectors. The proposed technique does not require any manual segmentation or localization technique. However, in this survey, support vector machines (SVM) are used for image classification, which is not good for large terabyte datasets. In the article [4], they proposed a New 3D gesture recognition method based on deep learning model. In their paper [1], they propose a novel convolutional neural network (CNN) in which sequences of skeletal hand joint positions are processed by parallel convolutions; we then study the performance of this model on the performance of the gesture sequence classification task. The model only uses hand skeletal data, not depth images. Thus, through

all these studies, we have successfully used deep learning to create models through convolutional neural network. N

III. PROBLEM STATEMENT

Suppose you work as a data scientist in a home electronics company that produces modern smart TVs. You want to develop a cool feature in your smart TV that can recognize five different gestures made by the user, which will help the user to control the TV without using a remote control. A webcam mounted on the TV constantly monitors gestures. Each gesture corresponds to a specific command: Thumbs up: Increase volume, Thumbs up: Decrease volume, Swipe left: go back 10 seconds, Swipe right: go forward 10 seconds, Stop put the movie paused.

IV. (1) PROPOSED MODEL

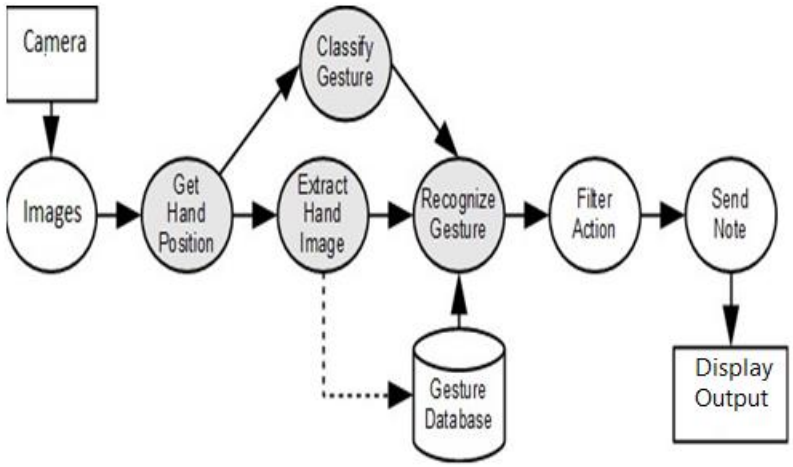

Fig. (1) Proposed System Model

Camera and Image- We are going to capture the images of hand signs with the camera. The images will be in the 3D and also in RGB images. Images will be the inputs for our model. We need to convert these images in the 1D and in grey images to reduce the load on the system.

Get Hand Position- With the help of a camera, a device detects hand or body movements, and a machine learning algorithm segments the image to find hand edges and positions. With our laptops camera we will capture the images. But the camera quality needs to be good for the accuracy of the recognition.

Extract Hand Image- The system distinguishes a hand from the background using colour and depth data. The hand sample is further divided into the arm, wrist, palm, and fingers. The system ignores the arm and wrist since they don't provide gesture information. Also, there should not be any body parts behind your hand or any other person standing next or in the screen of the camera.

Gesture Database- The hand gesture recognition dataset is presented, composed by a set of near infrared images acquired by the Leap Motion Sensor. The database is composed of 12 different hand-gestures and there are total 1,200 images in total. There are various images of only one gesture for the speed and the accuracy of the system.

Recognize Gesture- The system collects all extracted features into a feature vector that represents a gesture. A hand gesture recognition solution, using AI, matches the feature vector with various gestures in the database and recognizes the user's gesture.

Filter Action- Image filtering is used to enhance the edges in images and reduce the noisiness of an image. This technology is used in almost all smartphones. Although improving an image using the image filtering techniques can help in the process of object detection, face recognition and all tasks involved in computer vision.

IV. (2) Proposed System Architecture

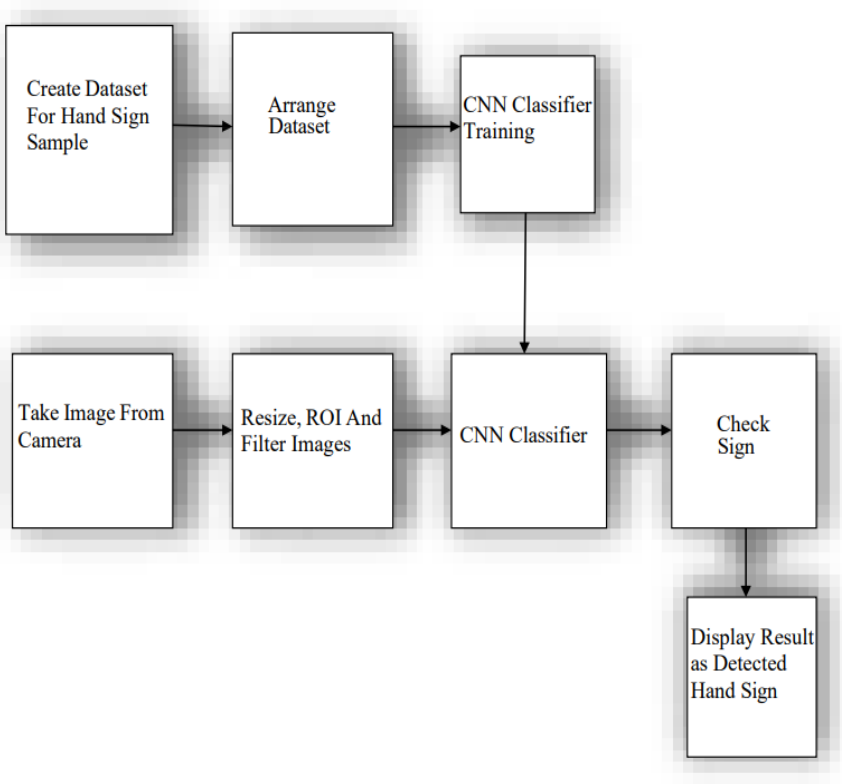

Fig.2 Proposed System Architecture

This dataset contains total 1650 images of 11 different gestures. For training purpose, there are 1650 images in each directory and for testing purpose there are 1650 images in each directory. We use this dataset primarily for hand gesture recognition task.
Now after getting the data set, we need to pre-process the data a bit and provide labels to each of the images given there during training the data set. To do so we can see that name of each image of the training data set so we will use that to our advantage then we use one hot encoder for the machine to understand the labels CNN – it is a machine learning algorithm for machines to understand features picture with foresight and remember the properties to guess whether the name a new image is inserted into the machine. In neural networks, a convolutional neural network (CNN) is one of the main categories for image recognition, images classification. Object detection, face recognition, etc. are some of the areas where CNNs are widely used. After CNN classifier we are going to take the images and after taking the images, we are going to apply following operations on them. We are going to take the images and after taking the images, we are going to apply following operations on them.
When using CNN algorithms with input images, the pre-processing technique of resizing images tend to be an experimental task. Images in complex datasets are often not the same sizes. To resize the images will use the squeeze for this purpose The RoI pooling layer, a Spatial pyramid Pooling (SPP) technique is the main idea behind CNN and the reason that it outperforms R-CNN in accuracy and speed respectively. Filter is a window that scans the image. A filter is used to judge which feature do a pixel belong to. Does it belong to an arc, straight line, diagonal, etc After Resizing, ROI and filtering of the image CNN model predict the output and according to the output it gives the meaning of the sign. After predicting the output, it will show the result in the form of message, order, etc.

IV. EXISTING SYSTEM

The paper [1] describes how a wearable glove-based sensor was used to capture hand movement and position. However, this method requires the user to be physically connected to the computer. However, modern glove-based approaches use touchscreen technology, a more promising technology considered for industrial-grade haptics.
The sensor they created based on electronic camera vision is therefore a universally applicable technology as it enables contactless communication between humans and machines.
However, this technique involves several challenges, including lighting changes, background issues, occlusion effects, complex backgrounds, and trade-offs between processing time, resolution, and frame rate. pictures. picture. images, and the appearance of skin tones or hands displayed in the foreground or background. In the second paper, they introduced the deep reinforcement learning method, which uses deep reinforcement learning algorithms to identify human arm movement patterns using IoT sensors. Recent studies have investigated supervised learning, based approaches such as CNNs and RNNs to implement HCI devices.
In paper [2], the performance of a convolutional neural network (CNN) with a DQN model is compared to that of a long-short-term memory (LSTM) model with DQN. The results show that the CNN-based DQN model is more stable than the LSTM-based model, with a classification accuracy of 98.33% in predicting arm motion patterns. In paper [3], they classify gesture poses using a classifier based on a support vector machine (SVM) with a linear kernel. In this analysis, the variance of the sum of deep features is greater than 99.9% are considered redundant features. The final reduction function studies gesture recognition performance. The tabular results show that the performance of the proposed technique is 87.83%. which is better than the average accuracy result of Alex-Net only "FC6" function. They introduce a new 3D gesture recognition method based on deep learning.

V. RESULT AND ANALYSIS

The result of our hand gesture recognition using deep learning reveal that the convolutional neural network (CNN) algorithm has the highest accuracy among the other models. There are various other models in which various algorithms are used like support vector machine (SVM), deep neural network (DNN), deep reinforcement learning algorithm.
By doing the research we found that the convolutional neural network algorithm has the highest accuracy and it is easy and convenient to use for the Pc's and laptops in general use. The CNN model demonstrated its ability to effectively recognise the hand gestures done by the user. This high accuracy indicates the potential of CNN as a reliable tool for recognising hand gestures.
In terms of computing the CNN shows the speed and scalability, allowing for efficient processing of datasets.
To develop and train this model, various libraries are available in Python that specifically used for machine learning and recognizing projects. These libraries provide the inbuilt functions and ability that are used for recognize images in projects. This will enhance the performance of the hand gesture recognition system
While discussing the results of our system it is really important to know and accept the limitations of our model. The result of our recognition system depends on the quality of data. That's why it is important to update dataset for the changing gestures.
Our study points out the work of convolutional neural network algorithm, in recognizing the gestures done by the users. The systems high accuracy and speed can be used for the home appliance objects and also in hospitals. Also, the there are ongoing researches and advancement in field are needed to increase the accuracy and robustness of the system.

VI. CONCLUSION

It discusses different approaches to hand tracking, webcams, wireless technologies, CNNs, OpenCV, deep learning, Tensor Flow.
After analysing five articles on gesture recognition from different authors, we concluded that this article Hand Gesture Recognition Using Convolutional Neural Network is the best. Experimental results show that the proposed system outperforms all other search methods in terms of recognition rate, demonstrating its effectiveness.

VIII. REFERENCES

[1] Guillaume Devineau and Wang Xi and Fabien Moutarde1 and Jie Yang, DEEP Learning for Hand Gesture Recognition on Skeletal Data [2018].

[2] Dhall, Shubham Vashisth, Garima Aggarwal, Automated Hand Gesture Recognition using a Deep Convolutional Neural Network [2018].

[3] W. Seok, Y. Kim, C. Park, Recognition of Human Arm Movement Using Deep Reinforcement Learning [2018].

[4] Jaya Prakash Sahoo, Samit Ari, Sarat KumarPatra,HandGesture Recognition using PCA based Deep CNN Reduced Features and SVM classifier [2019].
[5] Munir Oudah, Ali Al-Naji and Javaan Chahl, Hand Gesture Recognition Based on Computer Vision: A Review of Techniques [2020].