Hand Gesture Recognition Using Deep Convolution Neural Network

Abstract—**Hand gesture recognition is a critical component that offers a natural and intuitive means of communication with machines. In this paper, we are presenting a novel approach to automated hand gesture recognition utilizing a deep convolutional neural network model. This model is designed to address the challenges like variations in hand poses, complex backgrounds, and lighting conditions, so that it can works on real-world applications.**

Keywords ***Index Terms*—CNN (Convolution Neural Network), Opencv**

I. INTRODUCTION

Hand gesture recognition is a component of human-computer interaction. It's applications are in field of robotics ,virtual reality, and sign language interaction. This model Recognizes hand gestures accurately and in real time but variations in hand poses, lighting conditions, and occlusions make it complex. We are focusing on the development of deep learning-based solution, specifically using deep convolutional neural networks (CNNs), to address the challenges associated with hand gesture recognition. we have implemented techniques like background elimination to address those kind of technology. Key challenges within this problem space include:

a)Variability in Hand Gestures: *Hand gestures are of different shapes, sizes, and movements. Recognizing these features and gestures necessitates the ability to handle significant variability.*
b) Complex Backgrounds and Lighting Conditions: *In real-world environments, there are complex backgrounds and varying lighting conditions. So we have to make that our model handles those complexities in such a way that it still provides accurate answers of that particular.*
c) Occlusions: *Occlusions occur when objects or body parts obstruct the view of the hand, making it crucial to develop methods for recognizing partially visible or obscured gestures.*
d) Real-Time Processing: *Real-time processing enables humans to interact with computers in a really quick manner which helps in applications like gesture-based gaming robotics.*

II. METHODOLOGY

A. *System Overview:*

It is an OpenCV and cnn-based hand gesture recognition system. that uses techniques like bounding box and background elimination.

Identify applicable funding agency here. If none, delete this.

B. *Hand Segmentation:*

First, In the input image the hand region must be separated from the background. In recognition of hand gestures, it is one of the most crucial phases. In OpenCV, there are many hand segmentation algorithms like skin color segmentation and contour detection.

C. *Feature Extraction:*

The process of extracting features from a segmented hand region is called feature extraction. There are many characteristics in feature extraction that may be derived from the hand's form, feel, and movement. These characteristics include:

c.1 Shape Features: captures the overall shape of hand such as perimeter, area and aspect ratio.
c.2 Texture Features: captures the texture of hand such as wrinkles and skin color
c.3 Motion Features: Captures motion of hand

D. *Gesture Classification:*

A CNN model is used to classify the hand motion after the features have been extracted. Because CNN models can learn intricate correlations between the features and the gesture classes, they are a good fit for this purpose.
Dataset preparation: *Classes for the final predictions are "Fist", "Five", "None", "Okay", "Peace", "rad" and "Thumb".*

E. *Training the CNN Model:*

A dataset of photos of hand gestures is needed to train the CNN model. Images of the various hand gestures that the system must be able to identify should be included in the collection. The matching gesture class should be labeled on every image. The CNN model can be trained using a variety of deep-learning frameworks once the dataset has been gathered. The tagged photos are sent to the model throughout the training phase, and it is allowed to discover the connections between the features and the gesture classes.

F. *Video capture:*

The user's hand is captured in a video stream using a camera. The hand needs to remain in the square window that is visible on the screen.

G. *Image Processing:*

To accomplish the image augmentation, we employed Keras' ImageGenerator Class. We have used various techniques, including flipping, zooming, shearing, and rescaling. It aids in extrapolating previously unobserved data.

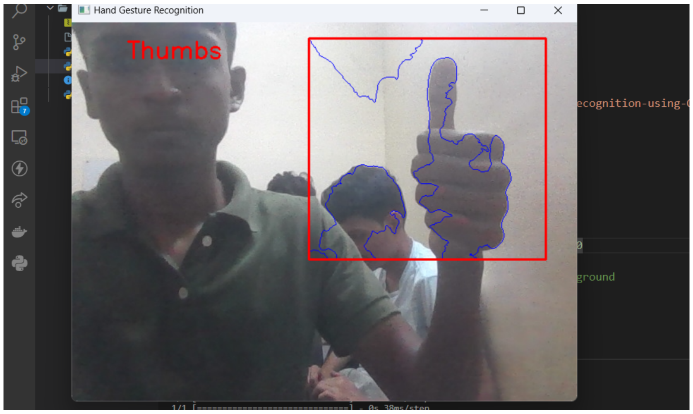

Fig. 1. Demo image

III. Dataset Description

The dataset used has 7 hand gestures:
0: Fist, 1: Five, 2: None, 3: Okay, 4: Peace, 5: Rad, 6: Straight, 7: Thumbs.

A. Architechture

Input method: This method uses the input image, where 640 and 480 are the width and height of the image, and 3 is the number of channels (RGB).
Convolutional layer 1: This layer contains 32 filters of size 3x3. It learns to extract features such as edges, corners, and surfaces from input images.
Max pooling layer 1: This layer reduces the sample of the feature map by a factor of 2, making the network more invariant to translation and scaling.
Convolutional layer 2: This layer contains 64 filters of size 3x3. It learns to extract various features, such as the shape and texture of the hand, from the input image.
Maximum Pooling Layer 2: This technique downsamples the specified map by a factor of 2, making the network more resilient to noise and noise. Changes to the input image.
Flattening layer: This layer converts the 3D feature maps required for all layers into 1D vectors.
Total connections 1: There are 128 neurons in this layer. It learns to combine features extracted from layers into a more interesting representation of the input image.
Full connection method 2: In this system, there are 7 neurons corresponding to 7 descriptive categories that the system can recognize.
Output layer: This layer uses the softmax function to output the result of 7 moves. CNN was trained using the Adam optimizer and categorical cross-entropy loss function. The model is trained 10 times.

The training dataset comprises 7999 images showcasing these hand gestures, distributed among the seven classes. Additionally, there are 4000 images in the test set, all belonging to the same seven classes. These images serve as the basis for training and evaluating models to recognize and classify hand gestures based on the provided labels.

IV. Experimental Settings

Hardware:
A computer with a CPU or GPU capable of running deep learning models.
A webcam or other source of hand gesture images.
Software: *Python 3.6 or higher TensorFlow and Keras libraries, OpenCV library, ImageDataGenerator library*

V. Results and Analysis

CNNs have been shown to be very effective for hand gesture recognition. During one training epoch, the model processed 7999 batches of data. The model achieved a training accuracy of 98.69% with a corresponding loss of 0.0401. However, on the validation dataset, it attained an accuracy of 97.62% with a higher loss of 2.0493. The model demonstrated strong performance on the training data, the drop in accuracy and increase in loss on the validation data suggest a need for further evaluation.

Conclusion

For identifying a wide range of hand movements the CNN-based hand gesture recognition is an easy-to-use tool. Different hand motions can be recognized by the system. This system is an appropriate option for creating gesture recognition for a range of applications because it is easy to install and use.

Possible applications of the system include: Sign language interpretation, Medical diagnosis, Virtual reality, Controlling robots, Playing games, Interacting with computers, and augmented reality. The system can be further improved by training it on a larger dataset of hand gesture images, and by using more advanced CNN architectures.

References

[1] Dhall, I., Vashisth, S. and Aggarwal, G., 2020, January. Automated hand gesture recognition using a deep convolutional neural network model. In 2020 10th International Conference on Cloud Computing, Data Science Engineering (Confluence) (pp. 811-816). IEEE.
[2] Niloy, E., Meghna, J. and Shahriar, M., 2021, July. Hand gesture-based character recognition using opencv and deep learning. In 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI) (pp. 1-5). IEEE.
[3] Bhavana, D., Kumar, K.K., Chandra, M.B., Bhargav, P.S.K., Sanjanaa, D.J. and Gopi, G.M., 2021. Hand sign recognition using CNN. International Journal of Performability Engineering, 17(3), p.314.