Comparison of CNN and Contour Algorithm for Number Identification Using Hand Gesture Recognition

Adarsh Pathak
B.Tech,Computer Science Engineering
Medi-Caps University, Indore, India
adarsh24699@gmail.com

Faraz Ahmed Khan
B.Tech,Computer Science Engineering
Medi-Caps University, Indore, India
faraz1299@gmail.com

***Abstract*— This paper compares the performance of two methods for hand gesture recognition for number identification. The image is captured employing a web camera system and undergoes many process stages before recognition of the numbers. Some of these stages include capturing the images, noise elimination, application of the CNN and contour algorithm to predict the number. Once the hand is been placed in the region of interest the CNN algorithm predicts the number and gives output in the frame using deep learning techniques, whereas the contour algorithm creates the boundary of the hand and predict the number using the convexity hull defects algorithm and gives the output in the frame. The proposed methods of CNN and Contour achieved the accuracy of 91.2% and 93.8% respectively.**

Keywords- Contour, Convexity Hull, Convolutional Neural Network (CNN), Hand Gesture Recognition.

I. INTRODUCTION

Gestures are a type of nonverbal communication within which visible bodily actions are accustomed to communicating vital messages, either in situ of a speech or along and in parallel with spoken words. Gestures embrace the movement of the hands, face, or different elements of the body. Physical non-verbal communication like strictly communicative displays, proxemics, or displays of joint attention differs from gestures, that communicate specific messages. With the assistance of those hand gestures, we can determine numbers, characters, etc using numerous algorithms that have numerous applications. With advances in computer vision technology, hand gesture communication is seeing loads of application in human-computer interaction. This paper presents a way to compare the performance of two algorithms for number identification using hand gesture recognition. Some modern systems use a hand gesture, mostly number of fingers raised within the region of Interest to perform various operations such as Play, Pause, seek forward, seek back word in a video player (for instance VLC media player) [1]. Other applications employing a vision system for hand gesture recognition are developed for multimedia system device management [2]. In recent years, several analysis groups have adopted machine-learning strategies to train models for classification, such as support vector machine [3], convolutional neural network (CNN) [4], recurrent neural network [5] and so on.

II. LITERATURE REVIEW

Deepak K. Ray, et al. [1]: explains the use of hand gesture recognition for controlling the VLC media player options such as play, pause etc. with the help of raised fingers detected.

S.G.Rayo [2]: explains the use of hand gesture recognition for managing multimedia devices using machine learning techniques.

T.-N. Nguyen, et al. [3]: shows the performance of support vector machines for classification problems such as hand gesture recognition for character and number identification.

Hung-Yuan Chung, et al. [4]: shows the performance of CNN for hand gesture recognition with the help of tracking of hand by kernelized correlation filters (KCF) algorithm.

Chigozie Enyinna Nwankpa, et al. [6]: explains the performance of CNN on various activation functions by comparing their accuracy and learning rates for the same dataset.

Guifang Lin, Wei Shen [7]: explains the performance of CNN by using the improved ReLU by varying hyperparameters.

Yanan Xu1, et al. [9]: explains the contour and convexity defects algorithm for hand gesture recognition by calculating the ratio of convex hull and contour area.

III. PROPOSED METHOD

The project is divided into two algorithms first, the Convolutional Neural Networks (CNN) and the second is the Contour algorithm. In the first algorithm, the image is captured using the webcam and the dataset is created (Figure 1). After the creation of the dataset for all gestures, the model is trained. After the completion of training, the prediction is performed using CNN and output is displayed on the screen. In the second algorithm the image is taken at runtime using the webcam and contour and convex hull are formed and with the help of convexity defects, the output is predicted and displayed on the screen.

A. Convolutional Neural Networks (CNN)

Convolutional Neural Networks are made of neurons which consists of weights and biases that can be learned. Every neuron receives many inputs, takes a weighted sum, pass it through the activation function and gives an output.

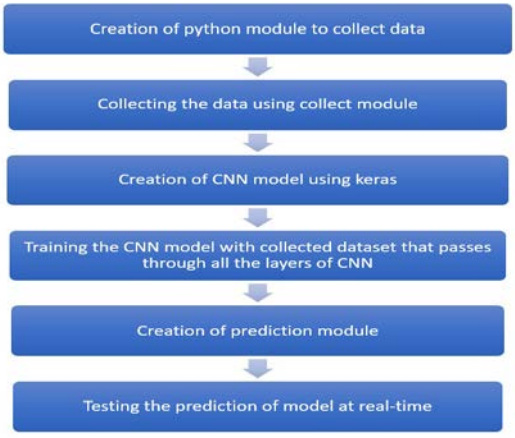


Figure 1: Process Flow of CNN Algorithm

There are four layers in Convolutional Neural Networks:

1. Convolution,
2. Activation Function Layer,
3. Pooling and
4. Full Connectedness (Fully Connected Layer).

1) Convolution Layer
Convolution is used in image processing to classify images. This layer converts the images into a pixel matrix where the pixel is represented in the form of 1's and -1's (1 for image present in pixel and -1 for absence of image in pixel). The calculation in this layer takes place by multiplying the pixel value to the feature pixel and taking the summation and dividing by the total number of pixels in the feature [8].

2) Activation Function Layer
Activation Functions are functions used in Neural Networks to compute the weighted sum of inputs and biases, which decides if a neuron can be fired or not [6]. It manipulates the conferred knowledge through some gradient process typically gradient descent to produce output for the Neural Network, that contains the parameters within the knowledge. In this project the ReLU activation function and adam optimizer is used for image classification these are the most efficient [7].

$$RELU(x) = \begin{cases} 0 \; if \; x < 0 \\ x \; if \; x >= 0 \end{cases}$$

3) Pooling Layer
In this layer, the matrix which is given by the activation function layer is converted into a smaller size. In this layer the calculation takes place by selecting the reduced size of the matrix (generally 2 or 3). This reduced matrix is moved across the original matrix and the maximum value of the matrix is taken [8].

4) Fully Connected Layer
This is the last layer of CNN. The input of this layer is the reduced matrix given by the pooling layer. In this layer, the matrix is converted into a flattened matrix like a list. The pattern of 1's and 0's of the flattened matrix is compared with the flattened matrix of the trained image matrix. The matching of both the matrix is compared with the probability and the output is given for the most probable input match [8].

B. Contour Algorithm

1) Contour
The contour of the hand is defined as when the boundary pixels of the given gestures is connected to form a close figure of the hand space. Once we get the contour, it is used for further calculations in the algorithm for prediction of the gesture [9]. In Figure 2 we can see that the blue line around the hand gesture given through the webcam is the required contour.

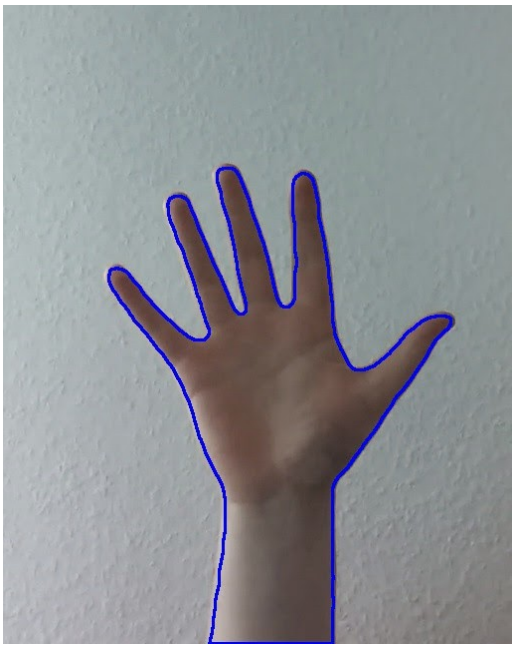


Figure 2: Contour Boundary of the Hand

2) Convex Hull
The convex hull of hand gesture is the polygon formed by joining the end points of contour of the hand gesture [9], as shown in figure 3, the polygon formed by red curve boundary is the convex hull of hand gesture and green curve boundary is contour of the hand.

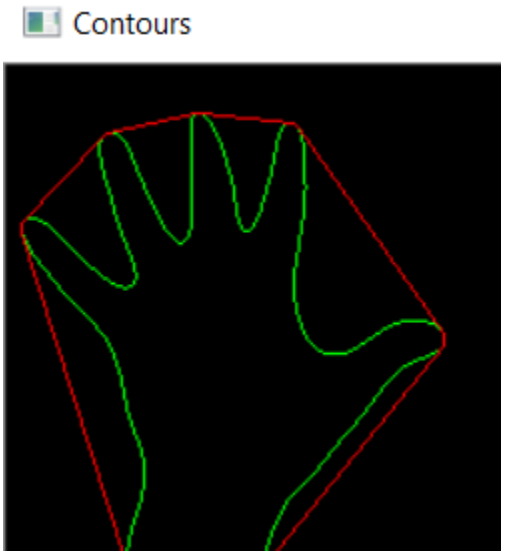


Figure 3: Convex hull Boundary (Red) of Hand

3) Convexity Defects
The convexity defect is defined as the difference in the area between the convex hull and contour. The area which is inscribed in the convex hull but not in the contour is defined as the convexity defect of the hand gesture [9]. In figure 4, the red spots are placed in the convexity defects that lie in between the contour area and hull area. In figure 4 as the number of fingers increases the red spots also increase with the formula of "no. of fingers - 1", therefore if the number of fingers is 5 then the number of convexity defects are 4.

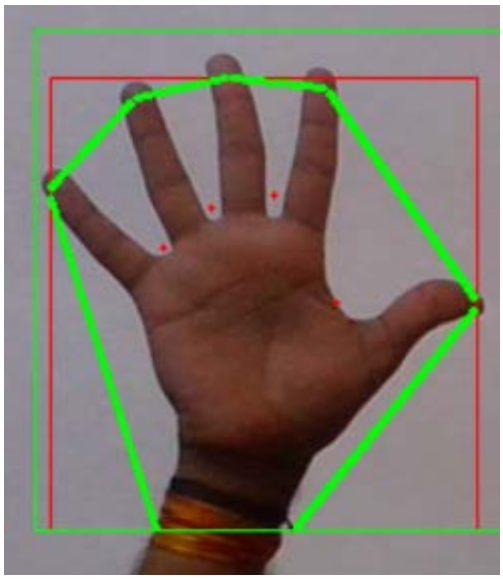


Figure 4: Convexity defects of hand gesture at run time

When the hand is placed in the Region of Interest (ROI), the contour and the convex hull is formed. The fingertips are identified. The angle is calculated between the contour and convex hull of the hand. If the angle is acute Convexity Defect is identified and red spots are placed in the defects (Figure 5). With the help of these red spots, the output is predicted and displayed on the screen (Table 1).

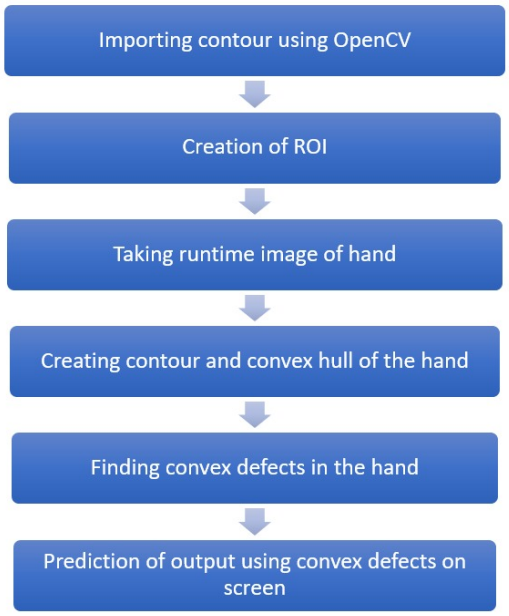


Figure 5: Process Flow of Contour Algorithm

No. of Convexity Defects	0	1	2	3	4
Predicted Output	One	Two	Three	Four	Five

Table 1: Convexity Defects with corresponding output

IV. RESULTS AND DISCUSSION

The web camera captures the images of the hand gesture at the run time in both algorithms. After the hand is placed in the Region of Interest both the algorithms are initiated. In the CNN algorithm, the Region of Interest captures the images at the frames per second(fps) speed of the camera and convert them into an array and processes it through all the four layers of CNN and the hand gesture is detected and the desired number is predicted and displayed on the screen, whereas in Contour algorithm when the hand is put in the region of interest contour and the convex hull of the hand are formed and convexity defects are calculated. With the help of the number of convexity defects, the desired number is predicted and displayed on the screen. Figure 6 shows the predicted number with the ROI using CNN algorithm whereas Figure 7 shows the predicted number with the ROI using the Contour algorithm.

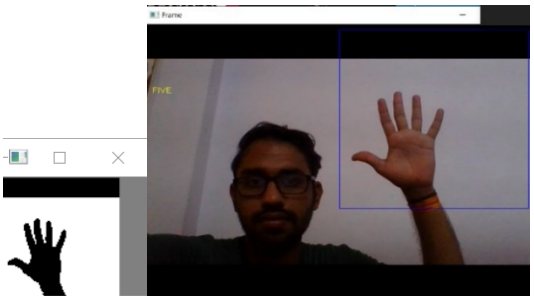
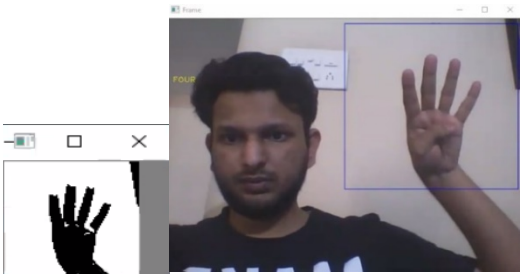
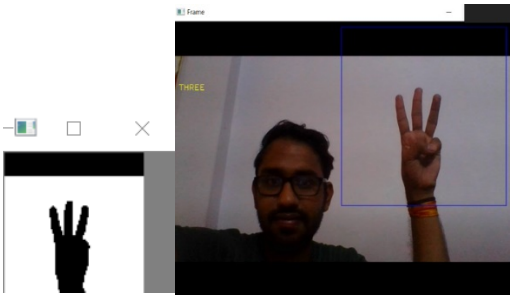


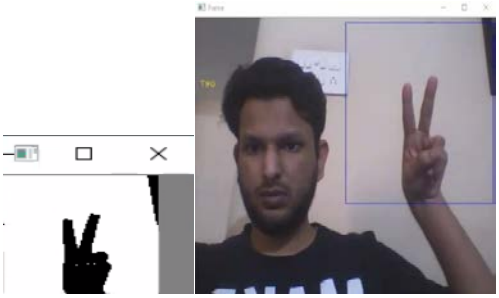
Figure 6: (a) Five



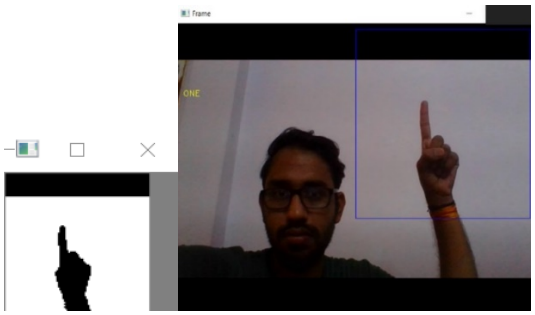
(b) Four



(c) Three



(d) Two



(e) One

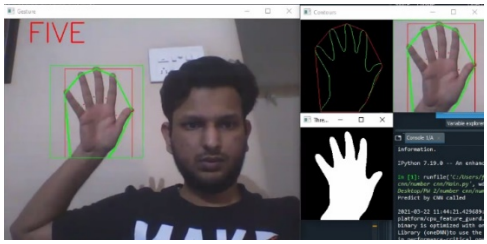
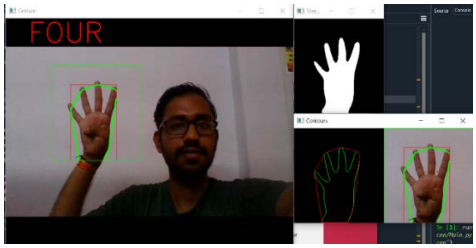
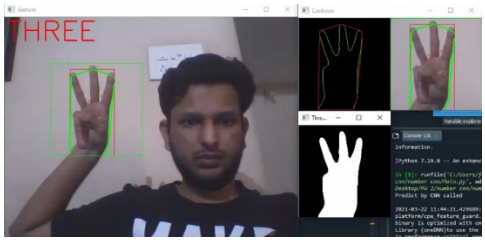


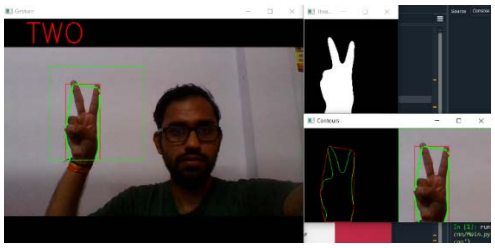
Figure 7: (a) Five



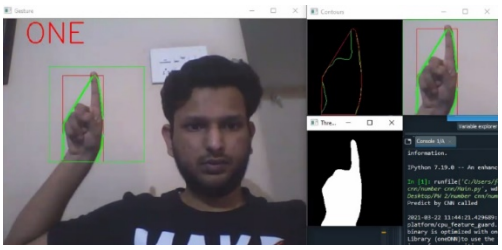
(b) Four



(c) Three



(d) Two



(e) One

Figure 8 and Table 2 shows the correct predictions of CNN and Contour algorithms over 100 images. The accuracy of CNN comes out to be around 91.2% and that of Contour is 93.8% which was calculated manually for every gesture by:

Accuracy = (no. of correct predictions) / (no. of times (100) prediction done)

	One	Two	Three	Four	Five
Correct output by CNN out of 100 images	91	90	92	87	93
Correct output by Contour out of 100 images	94	97	96	89	98

Table 2: Comparison of CNN and Contour over 100 images

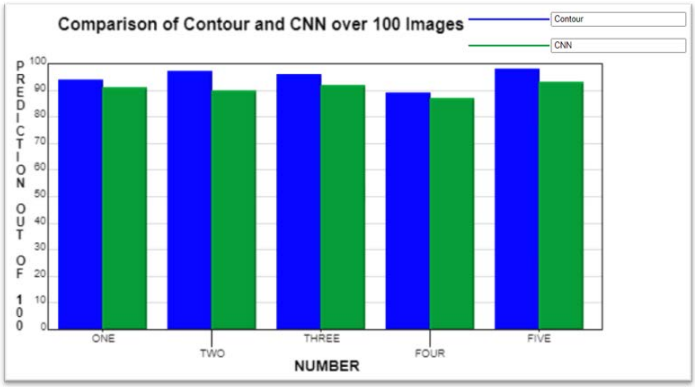


Figure 8: Comparison of correct predictions of CNN and Contour over 100 images

V. CONCLUSION AND FUTURE WORK

The comparison presented in this paper between CNN and Contour algorithm shows that in less number of gestures the Contour algorithm has better accuracy than CNN whereas in more number of gestures the CNN has much better accuracy than Contour. The accuracy of CNN for less number of gestures can be improved by increasing the size of dataset. This project is working on the image currently, further development can lead to detecting the motion of the video sequence and assigning it to a meaningful sentence with TTS assistance. This project can also be further compared with various machine learning algorithms.

REFERENCES

[1] Deepak K. Ray, Mayank Soni, Prabhav Johri, Abhishek Gupta, "*Hand Gesture Recognition using Python*", International Journal on Future Revolution in Computer Science & Communication Engineering ISSN: 2454-4248 Volume: 4 Issue: 6.

[2] S.G.Rayo, "*Hand Gestures and Hand Movement Recognition for Multimedia Player Control*", February 2008, unpublished thesis.

[3] T.-N. Nguyen, D.-H. Vo, H.-H. Huynh, and J. Meunier, "*Geometry-based static hand gesture recognition using support vector machine*," In Proc. The 13th International Conference on Control Automation Robotics & Vision, Singapore, December 10–12, 2014.

[4] Hung-Yuan Chung, Yao-Liang Chung, Wei-Feng Tsai, "*An Efficient Hand Gesture Recognition System Based on Deep CNN*," 2019 IEEE International Conference on Industrial Technology (ICIT).

[5] T. Koizumi, M. Mori, S. Taniguchi, and M. Maruya, "*Recurrent neural networks for phoneme recognition*," In Proc. The Fourth International Conference on Spoken Language Processing, PA, USA, October 3–6, 1996.

[6] Chigozie Enyinna Nwankpa, Winifred Ijomah, Anthony Gachagan, and Stephen Marshall, "*Activation Functions: Comparison of Trends in Practice and Research for Deep Learning*", arXiv:1811.03378v1 [cs.LG] 8 Nov 2018.

[7] Guifang Lin, Wei Shen, "*Research on convolutional neural network based on improved Relu*", 8th International Congress of Information and Communication Technology.

[8] https://www.edureka.co/blog/convolutional-neural-network/ (Date of access: 23rd March 2021)

[9] Yanan Xu1, Dong-Won Park and GouChol Pok "*Hand Gesture Recognition Based on Convex Defect Detection*", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 18 (2017) pp. 7075-7079.