Hand Gesture Recognition using Convolutional Neural Network

Conference Paper · November 2021

DOI: 10.1109/i-PACT52855.2021.9696463

5 authors, including:

Dhanasekaran Parasuraman
VIT University
8 PUBLICATIONS **6** CITATIONS

SEE PROFILE

Sharmila Ashok
VIT University
52 PUBLICATIONS **683** CITATIONS

SEE PROFILE

Hand Gesture Recognition using Convolutional Neural Network

S. Shanmugam
School of Electrical Engineering
Vellore Institute of Technology
Vellore, India
gsshansvn@gmail.com

Lakshmanan S A
Dept of ECE
Amrita School of Engineering,
Amrita Vishwa Vidyapeetham
Chennai, India
sa_lakshman@ch.amrita.edu

P . Dhanasekaran
School of Electrical Engineering
Vellore Institute of Technology
Vellore, India
dhakshay20@gmail.com

P. Mahalakshmi
School of Electrical Engineering
Vellore Institute of Technology
Vellore, India
pmahalakshmi@vit.ac.in

A. Sharmila
School of Electrical Engineering
Vellore Institute of Technology
Vellore, India
asharmila@vit.ac.in

2021 Innovations in Power and Advanced Computing Technologies (i-PACT) | 978-1-6654-2691-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/i-PACT52855.2021.9696463

Abstract— **The implicit message is usually brought to the spectators through activities involving various body parts like hands, face and arms. This is prominently known as Gesture and many such gestures are generally performed through hands in an involuntary manner. Smartness is to keep tabs on these hand gestures and derive purposeful details out of it. Convolutional neural networks (CNN) track these complex movements and help in extracting prime features. In this paper, training and testing were done consecutively with the aid of images to check the effectiveness of CNN and results are presented. .**

Keywords—Convolutional Neural Network, Hand Gesture, Images

I. INTRODUCTION

Convolutional neural network (CNN) gives an edge over the other neural networks by directly classifying patterns from images, text, video and sound thus eliminating the need for any other manual medium. Due to the same reason, it is widely used in the niche of deep learning [1-5]. All they do is hunt for patterns in pictures to detect a face, scene and objects. The motive of this paper is to detect patterns and thereby count the number of fingers using CNN with training and testing phases. The block diagram of the proposed work is shown in Fig.1 The image acquisition toolbox belonging to MATLAB and laptop webcam was used to capture images of hand gestures.

The application of CNN in the domains of feature extraction is hierarchical as it consists of three major layers which are the convolutional, subsampling or pooling and fully connected layers [6-9]. The high-level layers deal with conjectural information whereas the low-level layers articulates the substandard facts which is beneficial to the complicated classification tasks. The correlated points of the input is mapped to the next layer or the feature map by the correlation layer. Subsequently many feature maps are built using various kernels.

II. CONVOLUTIONAL NEURAL NETWORK

One must be aware of the significance of the technology emanated from vision in the world of human-computer-interaction (HCI). As hand gesture is an expression of ideas and thoughts, one cannot ignore it for the technologies desired for HCI. Applications of hand gestures are seen in numerous fields these days which are but not limited to: Applications of hand gestures are seen in numerous fields these days which are but not limited to:

- Virtual Reality communication: Gestures have a huge contribution in advancement of virtual reality communications as they aid in stimulating authentic management of virtual objects by the usage of hands, for 3D display interactions or 2D displays that simulate 3D interactions [10].
- Smart TV and its UI communication: [11] In order to establish a conventional TV remote's control, dynamic hand gestures were used for the purpose of navigation key and with it three hand gesture recognition methods were also implemented.
- Sign language: This is an effective method to help the physically impaired and disabled persons to deal with computers [12].
- Video games: A player's hand or body position could be traced to establish a control in movement. Location of Objects used in games such as cars can also be tracked in a similar manner [13].
- Connect and communicate with Displays in common areas.
- Tablet PC and desktop Applications: Gesture way of communicating with computer appliances is an effective substitute to interconnect with mouse and keyboard

The following are the methods to provide inputs to the hand gesture recognition system.

Vision-based: In vision-based techniques [14], a PC camera is the info gadget for noticing the data of hands or fingers. The Vision-Based techniques require just a camera, subsequently understanding characteristic interaction between people and PCs without the utilization of any additional gadgets. These frameworks will in general supplement biological vision by depicting artificial vision frameworks that are carried out in computer or a potential hardware. Unique sensors know as Kinect sensors can

978-1-6654-2691-6/21/$31.00 © 2021 IEEE

likewise recognize and segment the hands vigorously, along these lines it gives a legitimate base to gesture recognition. Notwithstanding numerous new accomplishments in applying the Kinect sensor to verbalized face acknowledgement [14], human body following [13] and human body tracking [14], it is as yet an open issue to utilize Kinect for hand motion acknowledgement.

Glove-based: In glove-based systems [15], data gloves are utilized which can document the precise places of hand gesture as their positions are directly estimated.

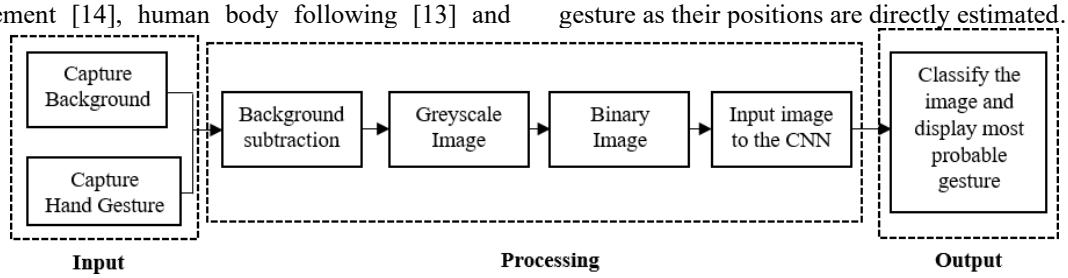

Fig. 1: Block diagram of the procedure followed in this CNN technique

The data-glove based strategies use sensor gadgets for digitizing hand and finger movements into multi-parametric information. The extra sensors make it simple to gather hand configuration and development. Nonetheless, the gadgets are very costly and carry a lot of experience to the users. Glove based gadgets have been used to find a pattern in the gestures for many years. But these devices proved to be tiring and time consuming to the users and have many cables coordinated with the computer. This paved the way to use non-intrusive and methods in which vision is associated for finding gestures. Also, the sensors utilized for the communication through sign language and the motion recognition in the frameworks accessible in the market are very expensive. Therefore, the vision-based sensor was picked to catch the data from the environment.

After the picture is caught from a decent quality camera the undertaking of distinguishing the gesture was to be completed. The techniques for gesture recognition is partitioned into the accompanying classes:

(1) Geometric features dependent on the recognition technique [16-18], the utilization of gestures of the edge qualities and gestures of the regional structure attributes as an recognition includes. The gesture recognition has great versatility and dependability. Yet, the deficiency of the strategy is that learning capacity isn't solid. On account of expanding test size, the recognition rate won't be altogether improved.

(2) Recognition method based on neural network [19]. This statistical method can achieve complex non-linear mapping, and has a classification characteristics and anti-jamming. It is widely used in static gesture recognition, but the feature learning ability of single hidden layer neural network is not strong, and it is easy to fall into the overfitting, which leads to the local optimal situation [20.

(3) Based on the secret Markov model (HMM) recognizable proof technique [21]. This has a solid capacity to depict gesture spatial-temporal gesture. Nonetheless, the HMM model requires to calculate an enormous number of state probability density, the need to appraise the number of parameters more. It makes the recognition pace slower. As of now, the current strategies need further examination on the extraction and expression of gesture strength.

III. HAND GESTURE RECOGNITION

The hand gesture calculation utilizes CNN was discovered to be more valuable than utilizing customary mathematical component-based recognition framework as other body parts, for example, elbow could make the recognition incomprehensible. The work is allowed close by motions, i.e., motions that depend on the movement of hands, yet including palms, wrists and elbows can likewise be incorporated into the picture. The different advances engaged with the improvement of the network are clarified in this part.

A. Image Capturing

The pictures for the info were acquired from the webcam of the PC on which the MATLAB code was run. Then taking the whole picture caught from the webcam for input, a green box was shown inside which the user needed to give the gesture as shown in Fig. 2.

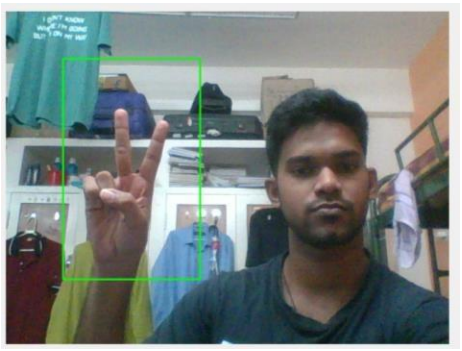

Fig. 2: Only the gesture provided within the green box is considered for gesture recognition

B. Image processing

Every one of the pictures caught by the webcam was handled with the goal that it very well may be effectively be educated by the network. The primary activity performed on the picture was the subtraction of the background in which the captured gesture picture was deducted from the background picture (this picture is first caught when the program begins). After this, the picture is then grey-scaled which lessens the picture size from 80*50*3 to 80*50*1 wherein the RGB layer is decreased to only one layer. This

picture is then changed over into BW picture by utilizing an edge worth of 10. All the above measures did are displayed in Fig.3.

C. Training Set

All the pictures used for preparing the network were independent. Each gesture for example (1, 2, 3, 4, 5) had 80 pictures made which implied an aggregate of 400 pictures were made. The pictures were subsequently made by utilizing a 'for loop'. Each image was either nearer, further or named so that it could assist the network with learning the gesture better and give more precise outcomes when the client gave it some info. A couple of instances of the pictures made are displayed in Fig.4.

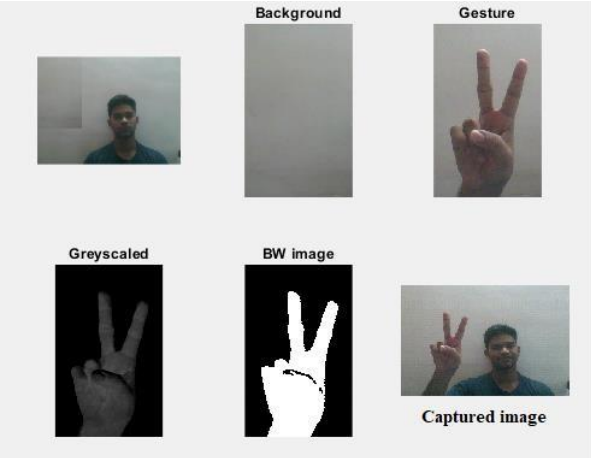

Fig. 3: Steps involved in processing the image that is captured from the webcam.

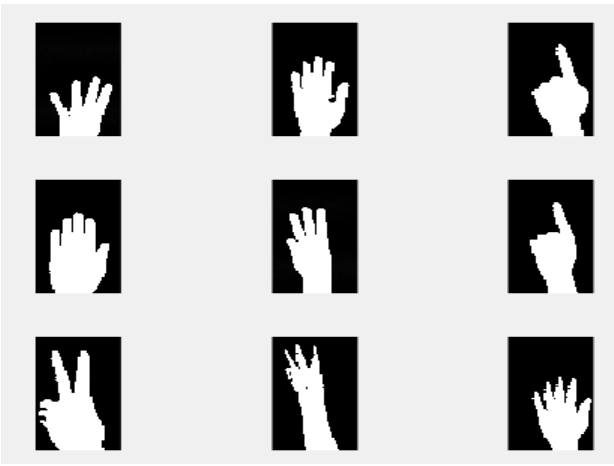

Fig. 4: Random hand gestures developed to train the network

D. CNN architecture

A convolutional neural network (CNN) is quite possibly the most well-known algorithm for profound learning, a kind of AI where a model figures out how to perform a task by undertaking it straightforwardly from pictures, video, text, or sound. CNN's are especially helpful for finding patterns in pictures to perceive objects, faces, and scenes. They gain straightforwardly from picture information, using examples for the classification of images and reducing the requirement for manual feature extraction.

These days, every field is in need of a system that can recognize objects and computer vision — for example, auto-driving vehicles and face-recognizing applications that depends completely on CNNs. Contingent upon the application, you can fabricate a CNN without any preparation, or utilize a pretrained model with available dataset. A convolutional neural network can have tens or many layers that each figure out how to identify various highlights of a picture. Filters are applied to each preparation picture in many sizes, and the yield of each convolved picture is utilized as the contribution to the following layer. The channels can begin as extremely straightforward features, like brightness and edges, and expansion in intricacy to features that exceptionally point a particular object. CNN is a framework of hidden layers sandwiched between input and output layers just like many other neural networks. These layers perform activities that adjust the information with the aim of learning features explicit to the information. The three layers used in this paper are displayed in the Fig. 5

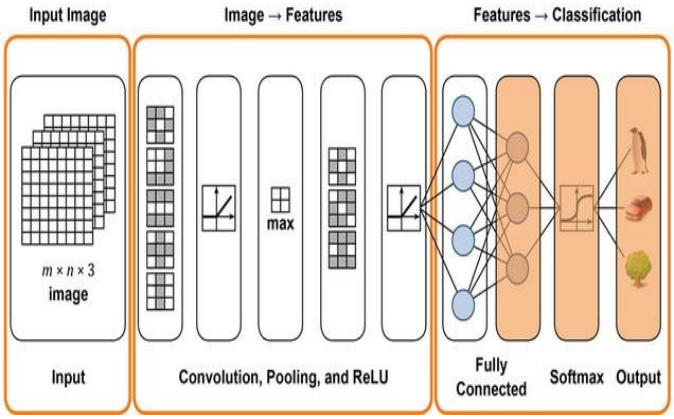

Fig. 5. Various steps involved in training the neural network

and are as per the following:

- Convolution puts the input pictures through a bunch of convolutional channels, every one of which actuates certain features from the pictures.
- Rectified linear unit (ReLU) takes into account quicker and more efficient preparation by planning negative values to zero and keeping up with positive values. This is known as activation, on the grounds that activated filters are carried forward into the following layer.
- The quality of the output can be improved by undergoing down sampling, which is widely known as pooling thus reducing the number of parameters needs to be learned by the network.

Subsequent to learning highlights in numerous layers, the engineering of a CNN shifts to classifying and grouping. The penultimate layer is fully related that yields a vector of K measurements where K speaks the quantity of classes that the network will actually want to anticipate. Pictures belonging to each class is classified by this vector as probabilities. A classification layer is used by the last layer, such as softmax

for providing the result of the classification. The code to prepare this neural network was made on MATLAB by using the deep learning tool kit and image acquisition tool compartment. The network comprises 5 layers of convolution, Pooling and ReLU. A learning pace of 0.01 was utilized for our five-layer convolutional neural network.

E. Training

Subsequent to preparing the network a Validation precision of 91.92% was noticed. The overall time taken to prepare the network was around 7 minutes. Altogether, 100 iterations were run. All the outcomes with respect to the training can be seen in Fig. 6. Out of the absolute 400 pictures made by me 60 for each gesture, for example, absolute 300 pictures were used for training and the excess 100 were utilized for validation of the neural network.

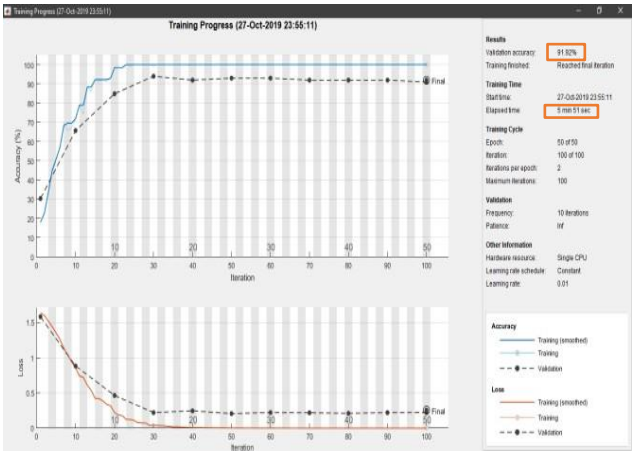

Fig. 6: Convolutional neural network results

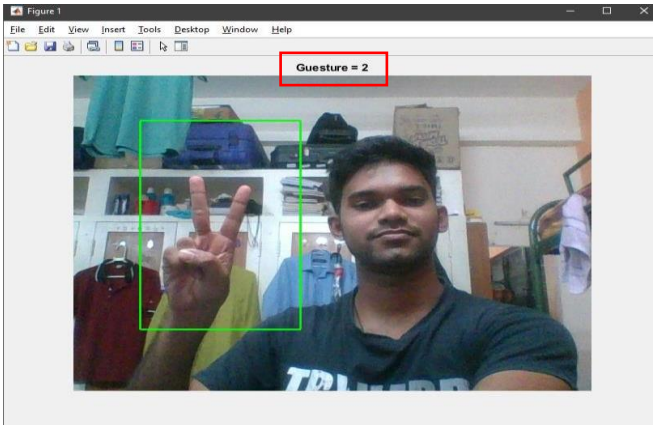

Fig. 7. Detected hand gesture using CNN

IV. RESULTS

The prepared network was tried to identify the hand motion given by the client. The detected hand gesture is shown at the top of the camera film as can be found in Fig. 7. An average gesture recognition time of around 721 milliseconds was noticed yet this outcome anyway is restricted to the equipment of my PC on which the code is run.

V. CONSLUSION

In this paper, the results of hand gesture acknowledgement by making use of Convolutional neural network is presented. This was accomplished with the aid of MATLAB programming. The gesture just inside the green box showed on the screen was considered for gesture acknowledgement. Further, image processing, background deduction, grey-scaling and changing over the picture into a black and white picture was finished. These pictures of around 80 for every gesture for example 400 altogether were made out of which 300 were utilized for preparing and 100 for validation. A general exactness of around 86% was acquired subsequent to testing the trained network. Future work will be focussed on (a) reading gestures without wanting to examine the background first (b) diminish the gesture cognition time from 700 milliseconds to around 100 milliseconds by making use of all the better equipment. (c) read more minute gesture variations like distinguishing between twisted fingers versus straight fingers to convey more gestures.

REFERENCES

[1] G. Li, H. Tang, Y. Sun, J. Kong, G. Jiang, D. Jiang, B. Tao, S. Xu, and H. Liu, “Hand gesture recognition based on convolution neural network,” Cluster Computing, 2017.

[2] M. Elmezain, A. Al-Hamadi, and B. Michaelis, “A robust method for hand gesture segmentation and recognition using forward spotting scheme in conditional random fields,” in Proceedings of the 20th International Conference on Pattern Recognition (ICPR ’10), pp. 3850–3853, August 2010.

[3] Lakshmanan, S.A., Venkateshkumar, M, “Analysis and Design of Lead-Lag Controller and Fuzzy Logic Controller for Boost Converter Applicable to RES” in Proc. IEEE 2nd International Conference on Power, Energy, Control and Transmission Sytems, December 10-11, 2020

[4] A. D. Bagdanov, A. Del Bimbo, L. Seidenari, and L. Usai, “Real- time hand status recognition from RGB-D imagery,” in Proceedings of the 21st International Conference on Pattern Recognition (ICPR ’12), pp. 2456–2459,November 2012.

[5] N. Aloysius, M. Geetha, “A review on deep convolution neural network”, 2017 International Conference on Communication and Signal Processing (ICCSP), IEEE, Chennai, India 2017.

[6] M. R. Malgireddy, J. J. Corso, S. Setlur, V. Govindaraju, and D. Mandalapu, “A framework for hand gesture recognition and spotting using sub-gesture modeling,” in Proceedings of the 20th International Conference on Pattern Recognition (ICPR ’10), pp. 3780–3783, August 2010.

[7] Venkateshkumar, M., Sreedevi, A.G., Lakshmanan, S.A., Yogesh kumar, K.R“Nonlinear Autoregressive Exogenous ANN Algorithm-Based Predicting of COVID-19 Pandemic in Tamil Nadu” Lecture Notes in Networks and Systems, 2022, 209, pp. 545–560

[8] G. R. S. Murthy & R. S. Jadon “A review of vision based hand gestures recognition” International Journal of Information Technology and Knowledge Management July-December 2009, Volume 2, No. 2, pp. 405-410.

[9] N. Aloysius, M. Geetha, “An Ensembled Scale-Space Model of Deep Convolutional Neural Networks for Sign Language Recognition”, Advances in Artificial Intelligence and Data Engineering, Springer Singapore, Singapore, 2020.

[10] Freeman, W., Tanaka, K., Ohta, J. & Kyuma, K. “Computer Vision for Computer Games”. Tech. Rep. and International Conference on Automatic Face and Gesture Recognition, (1996).

[11] Stotts, D., Smith, J. M. & Gyllstrom, K. “Facespace: Endoand Exo-Spatial Hypermedia in the Transparent Video Facetop”. In: Proc. of the Fifteenth ACM Conf. on Hypertext & Hypermedia. ACM Press, (2004) 48–57.

[12] Sharma, R., Huang, T. S., Pavovic, V. I., Zhao, Y., Lo, Z., Chu, S., Schulten, K., Dalke, A., Phillips, J., Zeller, M. & Humphrey, W. “Speech/Gesture Interface to a Visual Computing Environment for Molecular Biologists”. In: Proc. of ICPR’96 II (1996), 964-968.

[13] Feng, Qi & Yang, Cheng & Wu, Xiaoyu & Li, Zhuojia. (2013). A smart TV interaction system based on hand gesture recognition by using RGB-D Sensor. Proceedings - 2013 International Conference on

Mechatronic Sciences, Electric Engineering and Computer, MEC 2013. 1319-1322. 10.1109/MEC.2013.6885271.

[14] Sharmila Ashok, "High level hands-free control of wheel chair-a review, Journal of Medical Engineeirng & Technology, pp. 46-64.

[15] JongShill Lee, YoungJoos Lee, EungHyuk Lee, Seung Hong Hong "Hand region extraction and Gesture recognition from video stream with complex background through entropy analysis" Proceedings of the 26th annual International Conference of the IEEE EMBS San francisco, CA, USA, September 1- 5,2004

[16] Q. Cai, D. Gallup, C. Zhang, and Z. Zhang, "3D deformable face tracking with a commodity depth camera," in Proc. Eur. Conf. Computer Vision, Crete, Greece, 2010, pp. 229–242

[17] J. Wang, Z. Liu, Y. Wu, and J. Yuan, "Mining actionlet ensemble for action recognition with depth cameras," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, Providence, RI, USA, 2012.

[18] S. Mitra, and T. Acharya. (2007). "Gesture Recognition: A Survey" IEEE Transactions on systems, Man and Cybernetics, Part C: Applications and reviews, vol. 37 (3), pp. 311- 324, doi: 10.1109/TSMCC.2007.893280.

[19] Mokhtar M. Hasan, Pramoud K. Misra, (2011). "Brightness Factor Matching For Gesture Recognition System Using Scaled Normalization", International Journal of Computer Science & Information Technology (IJCSIT), Vol. 3(2).

[20] S. Mitra, and T. Acharya. (2007). "Gesture Recognition: A Survey" IEEE Transactions on systems, Man and Cybernetics, Part C: Applications and reviews, vol. 37 (3), pp. 311- 324, doi: 10.1109/TSMCC.2007.893280.

[21] Min B., Yoon, H., Soh, J., Yangc, Y., & Ejima, T. (1997). "Hand Gesture Recognition Using Hidden Markov Models". IEEE International Conference on computational cybernetics and simulation. Vol. 5, Doi: 10.1109/ICSMC.1997.637364