iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IX **Month of publication:** September 2023

DOI: https://doi.org/10.22214/ijraset.2023.55704

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IX Sep 2023- Available at www.ijraset.com

Computer Interaction Using Hand Gestures

Mr. Vishwanath M K[1], Prajwal S D[2], Karthik K[3], Achyuth J Shankar[4], Suraj U Mogali[5]
Dept of ECE The NIE Mysore, India

Abstract: Computer interaction using hand gestures is theprocess of automatically recognizing hand gestures that enablespeople to interact with the computer. Usually, interaction with a computer is done using input devices like keyboards, mice, etc., and the need to develop contactless human-computer interaction methods is the most desired feature during these pandemic days. Gesture recognition is a significant step towardsachieving contactless new types of human-computer interaction. A deep learning approach, convolution neural network, is used to recognize the hand gestures and mapped operation to the respective gesture is performed.

I. INTRODUCTION

With the recent expansion of computer sciencetechnologies like smartphones and the internet, thehuman-computer interaction field got a new meaning, the link between user activity andmultiple computers - where computers became anydevice that runs programs [1]. We use computers indifferent forms for several applications. Interactionof humans with the computers is a great field of interest and is known as HCI - Human Computer Interaction. We argue that HCI has emerged as a design-oriented field of research, directed at large towards innovation, design, and construction ofnew kinds of information and interaction technology. But the understanding of such an attitude to research in terms of philosophical, theoretical, and methodological underpinnings seems however relatively poor within the field [2]. Different input and output devices are designed over the years with the resolution of enabling the communication between computers and humans, the two most known are the keyboard and mouse.

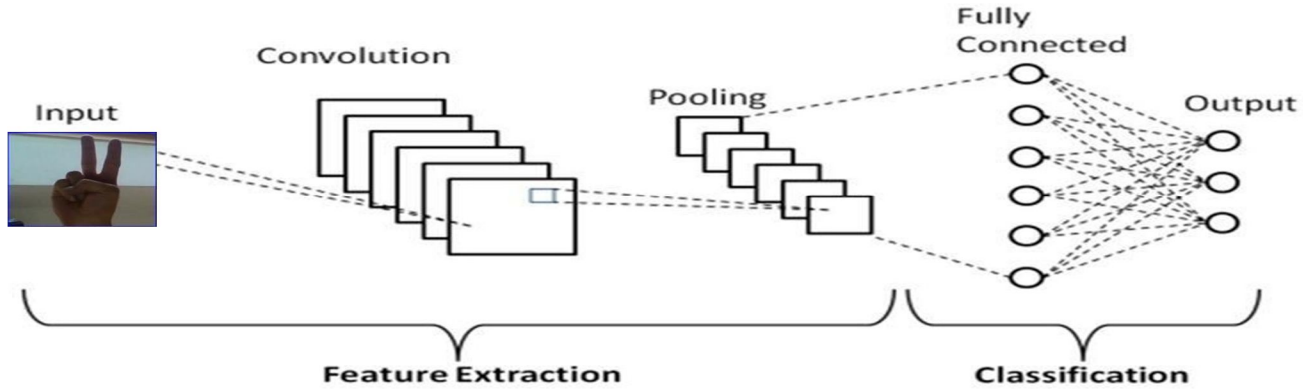
These to have been used normally for interaction between computers. But in this new age of Intelligence there is need of developing new ways of interaction between Humans and Computers forthe ease of communication of Humans withcomputers.

Deep learning is a subclass of machine learning, which is essentially a neural network with three or more layers. These neural networks attempt to simulate the actions of the human brain, although far from matching its ability, allowing it to learn from huge amounts of data. While a neural networkwith a single layer can still make estimated predictions, additional hidden layers can help to optimize and improve the accuracy. We have used convolutional neural network to recognize the handgesture.

II. METHODOLOGY

A. Dataset

The data set is a collection of images of alphabets from the American Sign Language, separated in 29folders which represent the various classes [3]. The training dataset contains 87,000 images of height 200 pixels and width 200 pixels. There images are classified into 29 classes, of which 26 are for the letters A-Z and for SPACE, DELETE and NOTHING These 3 classes are very helpful in real-time applications, and classification.



From these 29 classes we have picked 4 classes andmapped to hand gestures as below

Fig. 1 A Simple CNN Architecture

1) F is mapped to SUPER
2) V is mapped to VICTORY
3) L is mapped to LOSER
4) A is mapped to PUNCH

B. CNN Architectures

CNN is a kind of deep learning classification methodology which has many successful records inimage analysis and classifications tasks [4].

The convolutional neural network built takes in input data images of dimensions 200 x 200. It starts with a sequential layer onto which the number of filters, strides, the kernel size of appropriate values are provisioned with. The '**ReLU**' activation function is used, along with batch normalization and max pooling with a pool size of 2 x 2.

we have referred a pre-trained model called VGG-16 and used with custom convolutional neural networks (CNN's). The output layer of the VGG-16model is input to our custom layer. A convolutionallayer with 32 filters, kernel size of (3,3) and defaultstrides of (1,1) is added.

The Image Data Generator module is used for data augmentation of images which is done prior toletting the model work on the images in order to better generalize the model's working on real life images. This technique will help in training the datato improve the accuracy at each epoch.

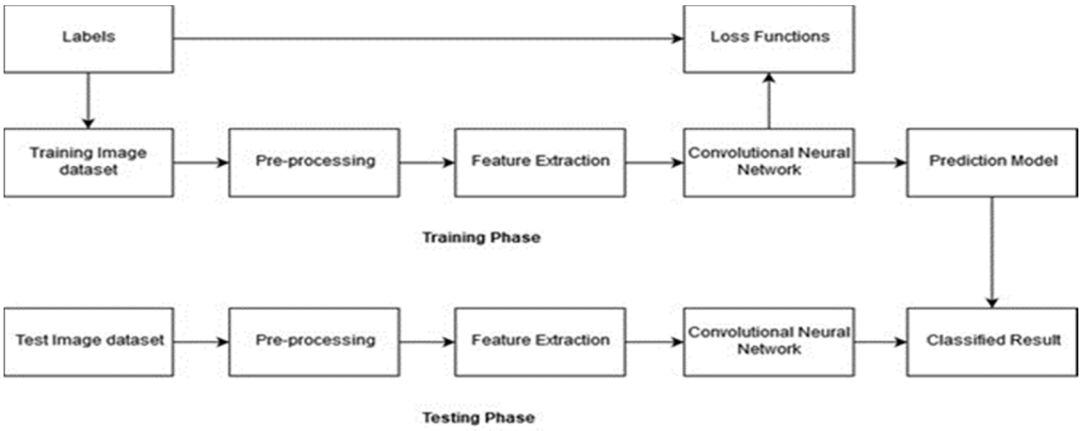


Fig. 2 A Block Diagram Display of the working of a CNN

The 'Adam' optimizer is used to evaluate the accuracy taking into consideration of the 'Categorical cross entropy' loss function. The dataset is split into the training and the validation set and the built model is fit on the training datasetwith validation taken into account.

Convolution is a mathematical operation on two functions to produce a third function that conveys how the shape of one is reformed by the other. Theconvolution layer transforms the input image inorder to extract features from it. In this process, the image is convolved with a kernel which is a small matrix, with its width and height smaller than the image to be convolved, known as convolution matrix.

Followed by convolution layer, pooling layers are used to lower the dimensions of the feature maps. Thus, it lessens the number of parametersto learn, the amount of computation performed and time taken to train the network. The poolinglayer encapsulates the features present in a regionof the feature map produced by a convolution layer. There are many non-linear functions to implement pooling like average and max pooling. The most common pooling technique used is Max pooling, which partitions the input image into a set of rectangles and, for each such sub-region, outputs the maximum

The non-saturating activation function called Rectified Linear Unit (ReLU) is used, whicheffectively eliminates the negative values from an activation map by replacing them with zero. Without affecting the receptive fields of the convolution layers, it introduces nonlinearities tothe overall network. Mathematically, it is expressed as,

Based on the recognised hand gesture, specificoperation is performed using OS module and text is converted to voice using gTTS (Google Text-to-Speech), a Python library and CLI tool to interfacewith Google Translate's text-to-speech API modulein python [5]. The following operations areperformed.

Table 1. List of operations

Hand gesture	Operation
Super	Capture the screen
Loser	Current time in voice
Punch	Opens specific website
Victory	Opens command prompt

Fig. 3 Flow chart of hand gesture operation

Applications can include controlling video games, learning or teaching sign language, or even using it as a substitute for the classic interaction deviceslike keyboard and mouse [6].

$$(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

where x is input to the function.

The final classification is done via fully connected layers which is located after numerous convolutional and max pooling layers. Neurons in a fully connected layer have connections to all units in the previous layer. The "loss function", indicates how training corrects the deviationbetween the predicted output and the true output ofthe network.

Table. 2 Model Summary

Layer Type	Output Shape	Number of Parameter
input_1 (Input Layer)	[(None, 200, 200, 3)]	0
Block1_conv1 (conv2D)	(None, 200, 200, 64)	1792
Block1_conv1 (conv2D)	(None, 200, 200, 64)	36928
Block1_pool (MaxPooling2D)	(None, 100, 100, 64)	0
Block2_conv1 (conv2D)	(None, 100, 100, 128)	73856
Block2_conv2 (conv2D)	(None, 100, 100, 128)	147584
Block2_pool (MaxPooling2D)	(None, 50, 50, 128)	0
Block3_conv1 (conv2D)	(None, 50, 50, 256)	295168
Block3_conv2 (conv2D)	(None, 50, 50, 256)	590080
Block3_conv3 (conv2D)	(None, 50, 50, 256)	590080
Block3_pool (MaxPooing2D)	(None, 25, 25, 256)	0
Block4_conv1 (conv2D)	(None, 25, 25, 512)	1180160
Block4_conv2 (conv2D)	(None, 25, 25, 512)	2359808
Block4_conv3	(None, 25,	2359808

(conv2D)	25, 512)	
Block4_pool (MaxPooing2D)	(None, 12, 12, 512)	0
Block5_conv1 (conv2D)	(None, 12, 12, 512)	2359808
Block5_conv2 (conv2D)	(None, 12, 12, 512)	2359808
Block5_conv3 (conv2D)	(None, 12, 12, 512)	2359808
Block5_pool (MaxPooing2D)	(None, 6, 6, 512)	0
Conv1 (Conv2D)	(None, 4, 4, 32)	147488
Poo11 (MaxPooling2D)	(None, 2, 2, 32)	0
Flatten (Flatten)	(None, 128)	0
FC1 (Dense)	(None, 30)	3870
FC2 (Dense)	(None, 30)	930
Output (Dense)	(None, 5)	155
Total params: 14,867,131 Trainable params: 152,443 Non-trainable params: 14,714,688		

III. RESULTS

The optimal hyperparameter values reported were:No. of filters: 32
Kernel size: 3
Batch size: 128
The model built with these optimal parameter values resulted an accuracy value of 95% in training dataset and 94% in testing dataset. The final predicted output with label is displayed and respective operation is performed.

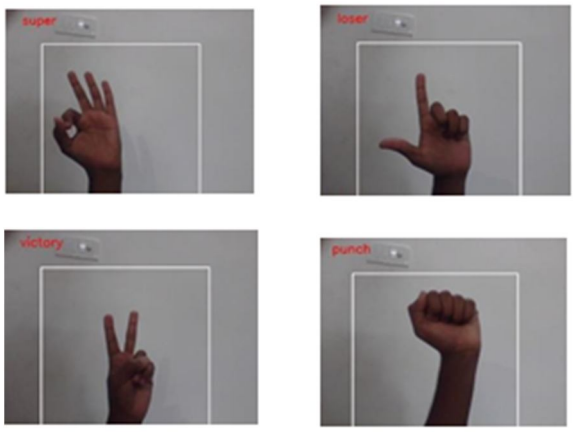


Fig. 4 Model's Prediction

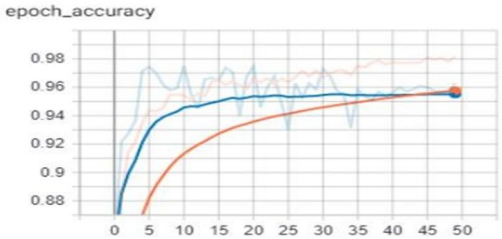


Fig. 5 Model accuracy on training and validation sets for differentepochs.

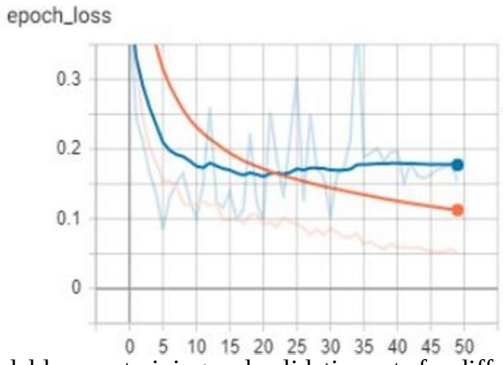


Fig 6. Model loss on training and validation sets for different epochs

IV. CONCLUSION

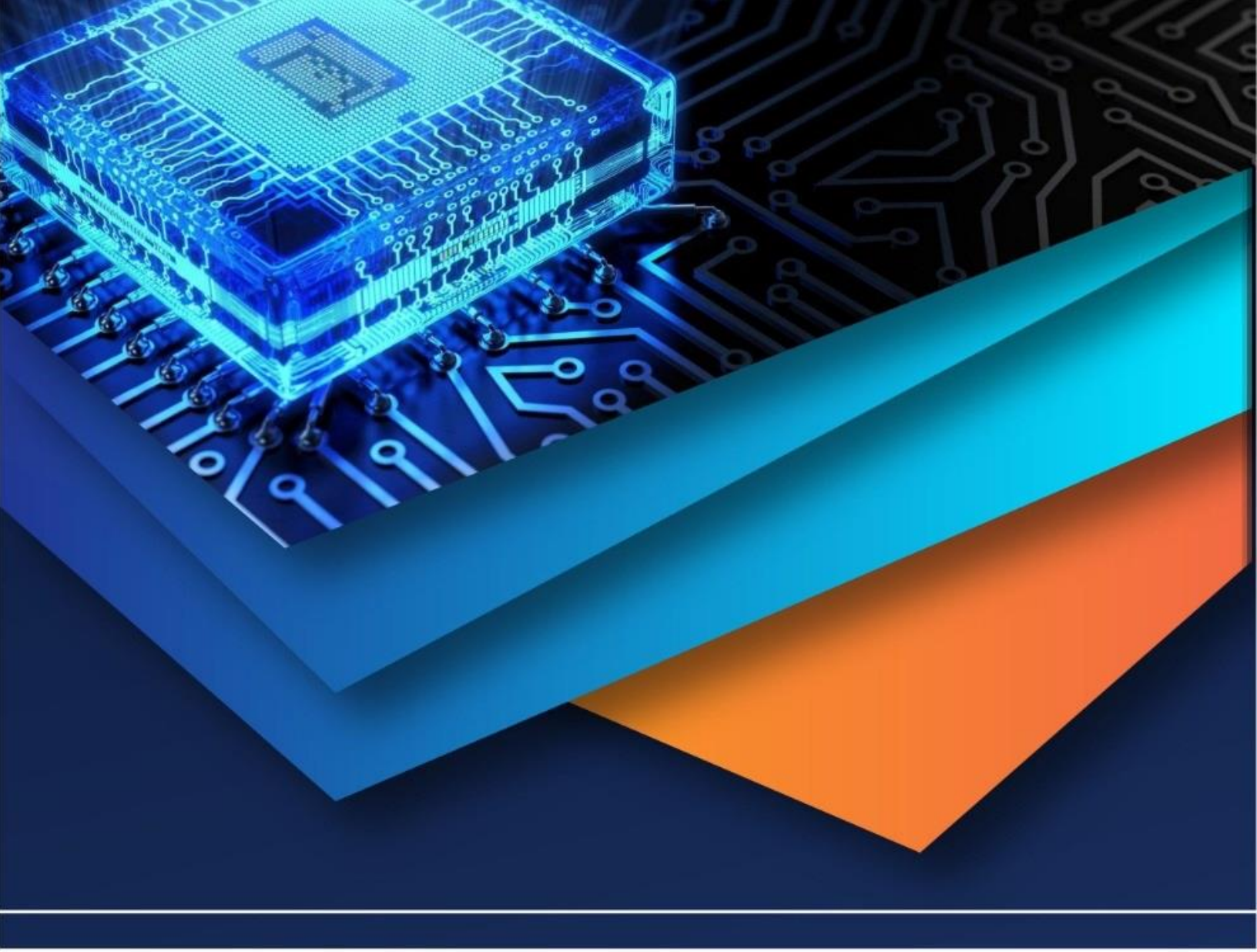
A Convolutional Neural Network is built on 9600 training image samples (having 4 classes) along with hyperparameter tuning (for optimal results) and its performance is evaluated on the test set (consisting of about 2400 image samples). The accuracy values obtained are fairly good with a training accuracy of about 95% and a testing accuracy of about 94%.

The operations which are mapped to respectivehand gestures can be personalised as required and more number of gestures can be added to the existing system.

In future, the human-computer interaction domain will have a rapid growth, making space for innovation and research. The old human-computer interaction devices will become legacy and replaced by new ones.

REFERENCES

[1] J. Lazar, J. H. Feng and H. Hochheiser, Research methods in human-computer interaction, Morgan Kaufmann, 2017.
[2] Fallman, “Design-oriented human-computer interaction,” in SIGCHI conference on Human factors in computing systems, 2003.
[3] “ASL Alphabet,” [Online]. Available: https://www.kaggle.com/datasets/grassknoted/asl-alphabet.
[4] P. S. Neethu, R. Suguna and D. Sathish, “An efficient method for human hand gesture detection and recognition using deep learning convolutional neural networks.,” Soft Comput, vol. 24, March 2020.
[5] “gTTS,” [Online]. Available: https://pypi.org/project/gTTS/.
[6] Bachman, F. Weichert and G. Rinkenauer, “Review of three-dimensional human-computer interaction with focus on the leap motion controller,” Sensors, vol. 18, no. 7, 2018.
[7] Zhan, “Hand Gesture Recognition wuth Convolution Neural Networks,” in IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI), 2019.
[8] H. Aashni, S. Archanasri, A. Nivedhitha, P. Shristi and S. Joythi Nayak, “Hand Gesture Recognition for Human Computer Interaction,” Procedia Coomputer Science, vol. 115, pp. 367-374, 2017.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY