HAND GESTURE RECOGNITION SYSTEM

USING AlexNet AND MATLAB

DSP LAB MINI PROJECT REPORT

GROUP MEMBERS:
FARHEENA | B201320EC
IFRAZ MANIKA | B201321EC
HAMDHA ABDUL RASHEED | B200875EC
JITHIN MOHAMMED KHAN | B200946EC

INDEX

ACKNOWLEDGEMENT:

We, the group members of G5 would like to express our sincere appreciation and gratitude to everyone who contributed to the successful completion of this lab project.

First and foremost, we want to express our heartfelt thanks to our Digital Signal Processing Lab in-charge faculties, Dr Sudeep PV and Dr Jobin Francis, who gave us invaluable direction and assistance throughout the project. Their excellent advice and timely suggestions were critical in overcoming the several hurdles that we encountered on the way.

Furthermore, we would like to thank the DSP Lab assistants for their continuous aid and support in providing us with the resources and facilities we needed to complete our work. Their invaluable assistance was critical in guaranteeing the project's smooth completion.

Finally, we would like to thank everyone who has helped us during this effort, especially our friends and family. Their unfailing support and encouragement helped us stay focused and devoted to the endeavor.

Problem Definition:

Hand gesture recognition is a computer vision application that recognizes and classifies human hand gestures in real-time. This system can be used in various applications, such as sign language recognition, human-computer interaction, and robotics.

In this project, we have used the **AlexNet** which is a deep learning model and the **MATLAB** platform to develop a hand gesture recognition system.

Objective:

The objective of this project is to build a hand recognition system that can accurately recognize hand gestures and classify them into different categories.

The system should be able to perform real-time recognition and should have a high accuracy rate.

Methodology:

Our Hand Recognition System using AlexNet and MATLAB consists of three main stages:

- **Data collection**
- **Training the model**
- **Testing the model**.

Data Collection:

For this project, a large dataset of hand images was collected. The dataset consists of images of different hand gestures, including **thumbs up, thumbs down, 0, 1, 2, 3, 4, 5 and none**.

For this hand recognition system we have **9 classes** of images.

When the code is run, 300 images are clicked automatically by showing the hand gesture in the processing area. This is done for each class of image.

The program used for data collection is as follows:

```
clc
clear all
close all
warning off
c=webcam;
x=0;
y=0;
height=500;
width=500;
bboxes=[x y height width];
temp=0;
while temp<=300
    e=c.snapshot;
    IFaces = insertObjectAnnotation(e,'rectangle',bboxes,'Processing Area');
imshow(IFaces);
filename=strcat(num2str(temp),'.bmp');
es=imcrop(e,bboxes);
es=imresize(es,[227 227]);
    imwrite(es,filename);
    temp=temp+1;
    drawnow;
end
clear c;
```

The code initializes a webcam using the command "webcam", and sets up the initial bounding box parameters for capturing images of hand gestures. The height and width are set to 500 pixels each, and the starting position is set to (0, 0).

The code then starts a loop that takes **300 images** of the hand gesture using the webcam. Each image is cropped and resized to **227x227 pixels**, and saved in a file with a unique name.

The file name is created using the "num2str" function to convert the loop index to a string, which is then concatenated with the ".bmp" file extension.

The cropped image is displayed using the "imshow" function with an added annotation that displays the processing area in a rectangle.

After 300 images have been captured, the webcam is cleared using the "clear" command to free up system resources.

This data collection process is an important step in training the CNN model using AlexNet, as it provides a large dataset of hand gesture images to train the model.

The use of the webcam and the MATLAB script allows for efficient and standardized data collection, which is crucial for achieving high accuracy in the final hand recognition system.

Training the Model:

The next step was to train the CNN model using the AlexNet architecture.

AlexNet is a widely used CNN architecture that consists of eight layers, including five convolutional layers and three fully connected layers.

The model was trained using the preprocessed hand image dataset taken with the help of the data collection program.

The training was done on a single CPU and was run for **20 epochs**.

The program used is as follows:

```
clc
clear all
close all
warning off
g=alexnet;
layers=g.Layers;
layers(23) = fullyConnectedLayer(9);
layers(25) = classificationLayer;
allImages=imageDatastore('Training','IncludeSubfolders',true,'LabelSource','foldernames');
opts=trainingOptions('sgdm','InitialLearnRate',0.001,'MaxEPochs',20,'MiniBatchSize',64);
myNet1=trainNetwork(allImages,layers,opts);
save myNet1;
```

The code first initializes the AlexNet model using the command "alexnet". The layers of the model are then extracted using the "Layers" property and stored in a variable called "layers".

Next, the fully connected layer and classification layer of the model are modified to suit the requirements of the hand recognition system. The fully connected layer is modified to have 9 output neurons, one for each hand gesture category.

The classification layer is then updated to reflect the new number of output neurons.

The next step is to create an "imageDatastore" object that reads the training images from the "Training" folder and its subfolders. The "LabelSource" property is set to "foldernames" to automatically label each image with its respective folder name.

The "trainingOptions" function is used to specify the training options for the CNN model. In this case, the stochastic gradient descent with momentum (SGDM) optimizer is used with an initial learning rate of 0.001, a maximum of 20 epochs, and a mini-batch size of 64.

The "trainNetwork" function is then used to train the modified AlexNet model on the training images. The layers and options are passed as inputs to this function, along with the training images. The resulting trained network is saved in a file called "myNet1" using the "save" function.

This training process is an essential step in the Hand Recognition System using AlexNet and MATLAB, as it allows the CNN model to learn and classify hand gestures accurately.

The modification of the fully connected layer and classification layer is necessary to reflect the specific hand gesture categories used in the system.

The training options and hyper-parameters can be adjusted to optimize the performance of the model on the training data.

Testing the Model:

The model is then tested by using a live webcam feed where the user shows the hand gesture in the processing area.

The gesture that is identified by the model will be shown as the title of the webcam feed.

The program used for testing is as follows:

```
clc;
close all;
clear all;
warning off
c=webcam;
load myNet1;
x=0;
y=0;
height=500;
width=500;
bboxes=[x y height width];
while true
   e=c.snapshot;
   Ifaces = insertObjectAnnotation(e,'rectangle',bboxes,'Processing Area');
es=imcrop(e,bboxes);
   es=imresize(es,[227 227]);
   label=classify(myNet1,es);
   imshow(Ifaces);
   title(char(label));
   drawnow;
end
```

The code first initializes a webcam using the command "webcam" and loads the previously trained network "myNet1" using the "load" function.

The initial bounding box parameters for capturing images of hand gestures are set up similarly to the data collection process.

The code then starts an infinite loop that captures images of the hand gesture using the webcam. The current image is cropped and resized to 227x227 pixels, and the trained network is used to classify the hand gesture.

The resulting label is displayed as the title of the image using the "title" function.

The cropped image is displayed using the "imshow" function with an added annotation that displays the processing area in a rectangle.

This testing process is an essential step in the Hand Recognition System using AlexNet and MATLAB, as it allows the trained CNN model to be evaluated on real-time input.

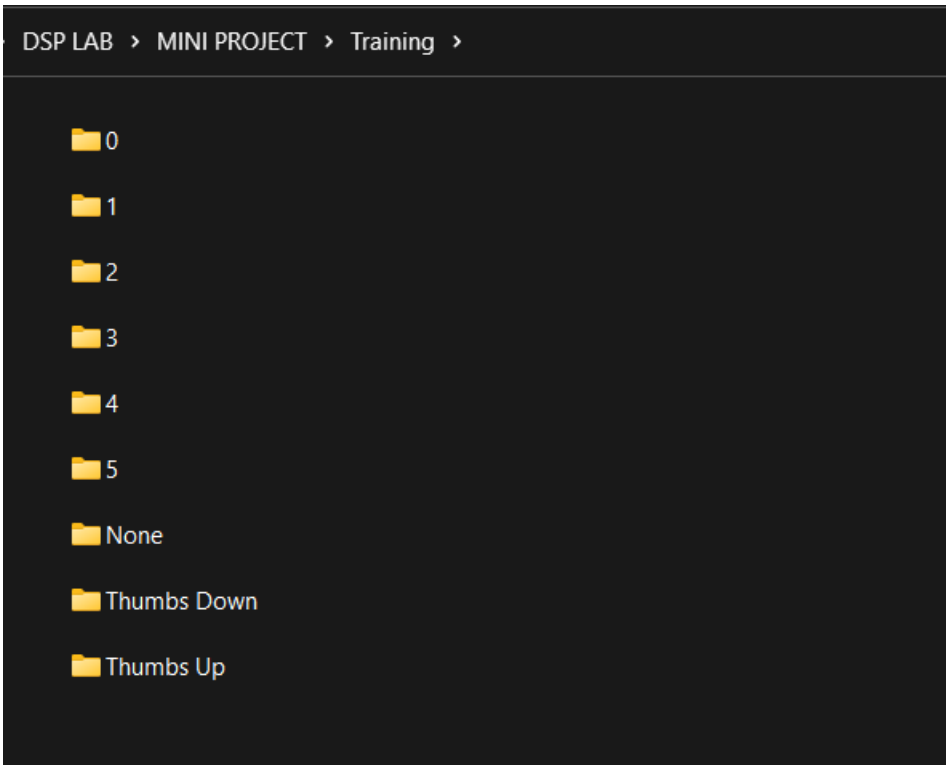
The use of the webcam and MATLAB script allows for efficient and standardized testing, which is crucial for achieving high accuracy in the final hand recognition system.

The ability to display the label of the hand gesture in real-time provides immediate feedback on the performance of the system and allows for quick adjustments and improvements to be made.

Results:

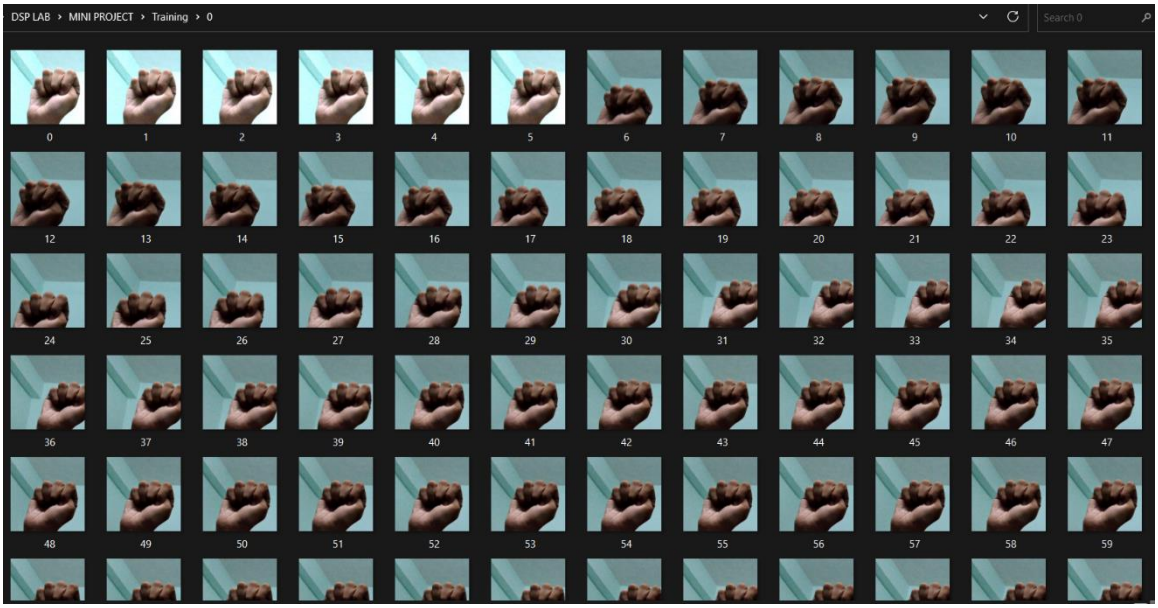
The Hand Recognition System was able to accurately recognize and classify hand gestures in real-time.

The results after data collection are as follows:



Here, these are the folders to create the dataset and the name of each folder indicates the label of each class of image ie. 0 indicates the folder containing images containing the hand-sign images of 0.

The contents of folder ‘0’ are as follows containing 300 images for training the CNN model:



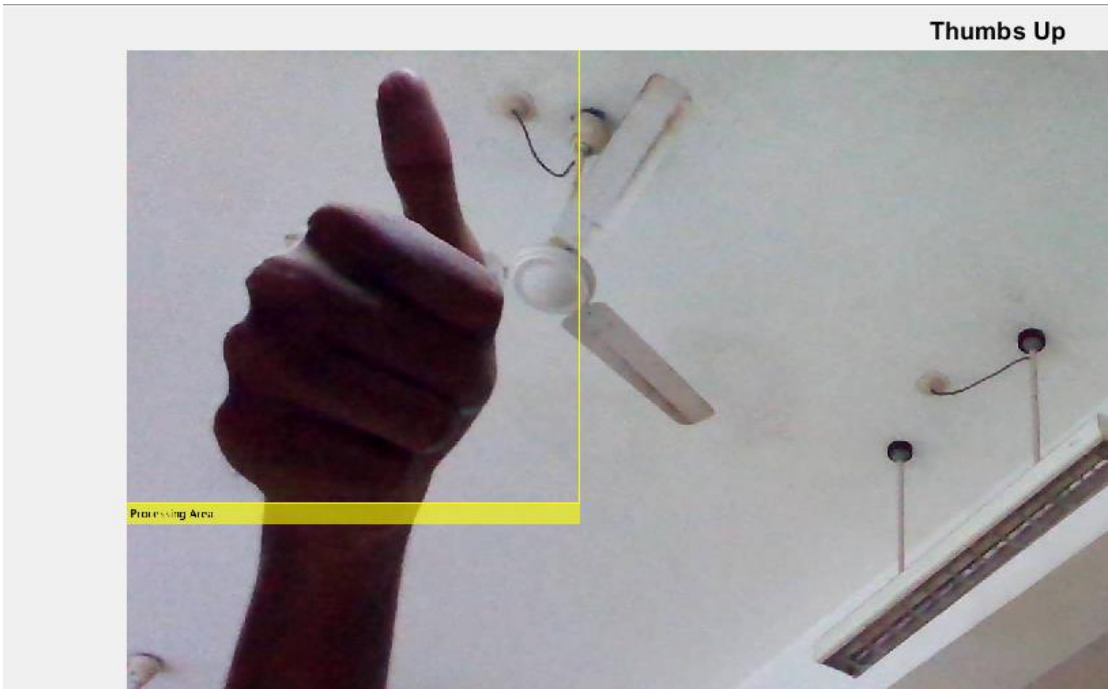
The training results are as follows:

Command Window

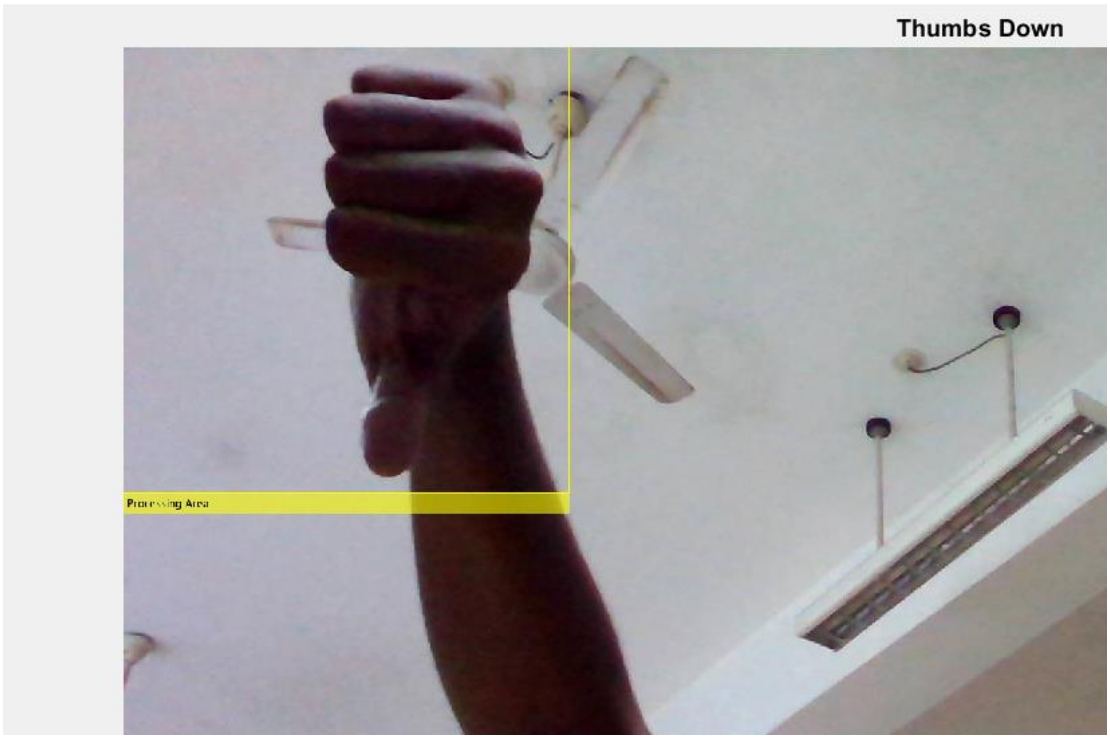
```
Training on single CPU.
Initializing input data normalization.
|=========================================================================================|
|  Epoch  |  Iteration  |  Time Elapsed  |  Mini-batch  |  Mini-batch  |  Base Learning  |
|         |             |   (hh:mm:ss)   |   Accuracy   |     Loss     |      Rate       |
|=========================================================================================|
|       1 |           1 |       00:00:01 |        9.38% |       4.4448 |          0.0010 |
|       2 |          50 |       00:01:24 |       98.44% |       0.0232 |          0.0010 |
|       3 |         100 |       00:02:50 |      100.00% |       0.0005 |          0.0010 |
|       4 |         150 |       00:04:18 |      100.00% |       0.0002 |          0.0010 |
|       5 |         200 |       00:05:46 |      100.00% |       0.0011 |          0.0010 |
|       6 |         250 |       00:07:14 |      100.00% |       0.0007 |          0.0010 |
|       8 |         300 |       00:08:42 |      100.00% |       0.0035 |          0.0010 |
|       9 |         350 |       00:10:11 |      100.00% |       0.0017 |          0.0010 |
|      10 |         400 |       00:11:39 |      100.00% |       0.0001 |          0.0010 |
|      11 |         450 |       00:13:08 |      100.00% |       0.0012 |          0.0010 |
|      12 |         500 |       00:14:36 |      100.00% |       0.0054 |          0.0010 |
|      14 |         550 |       00:16:06 |      100.00% |   5.3893e-05 |          0.0010 |
|      15 |         600 |       00:17:35 |      100.00% |   1.8908e-05 |          0.0010 |
|      16 |         650 |       00:19:03 |      100.00% |       0.0001 |          0.0010 |
|      17 |         700 |       00:20:32 |      100.00% |   6.4015e-06 |          0.0010 |
|      18 |         750 |       00:22:00 |      100.00% |   2.9952e-05 |          0.0010 |
|      20 |         800 |       00:23:29 |      100.00% |   1.9071e-05 |          0.0010 |
|      20 |         840 |       00:24:40 |      100.00% |   8.3554e-06 |          0.0010 |
|=========================================================================================|
Training finished: Max epochs completed.
fx >>
```

The results obtained for each gesture during testing have been shown below:

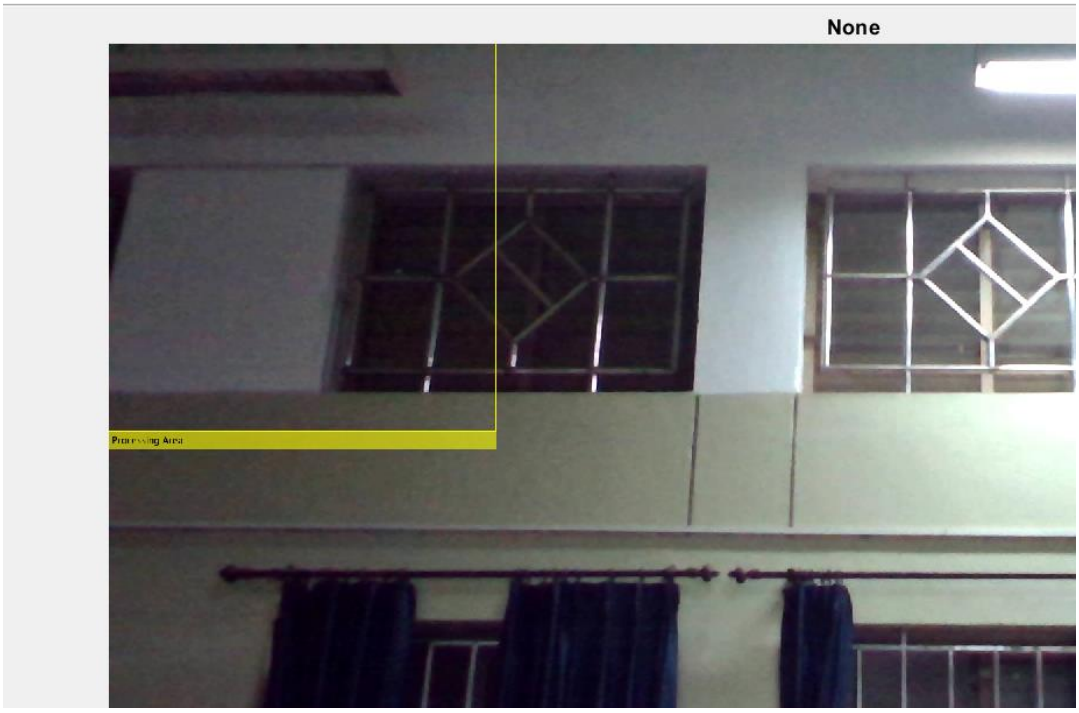
1) **Thumbs Up:**



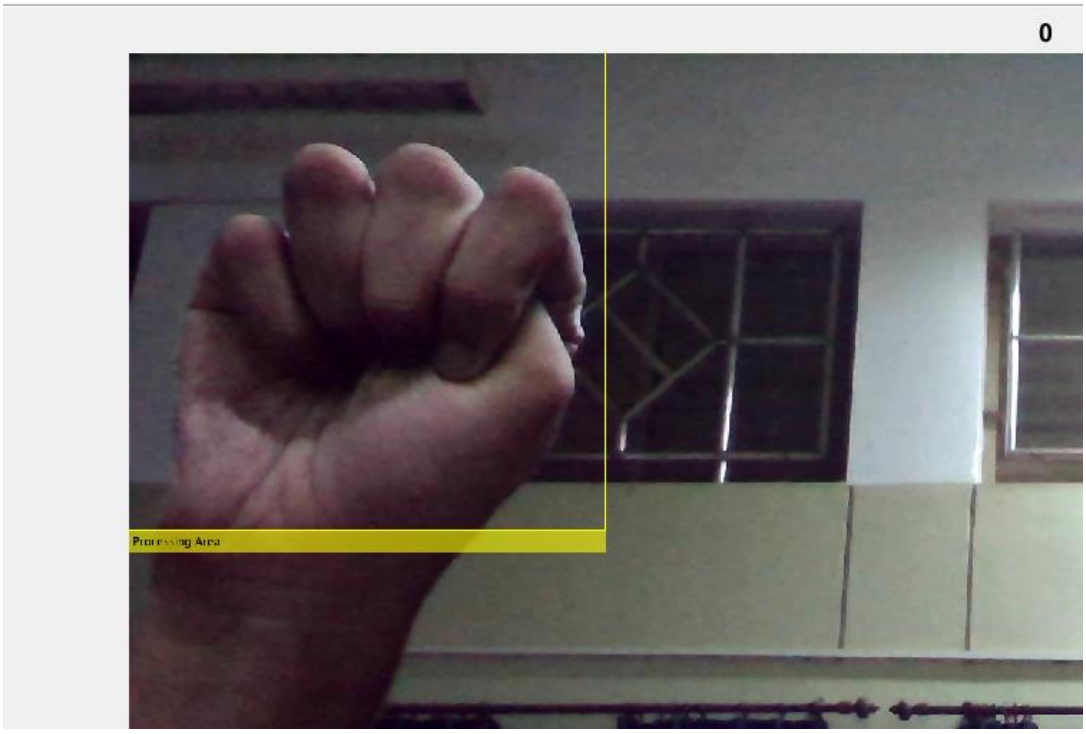
2) **Thumbs Down:**



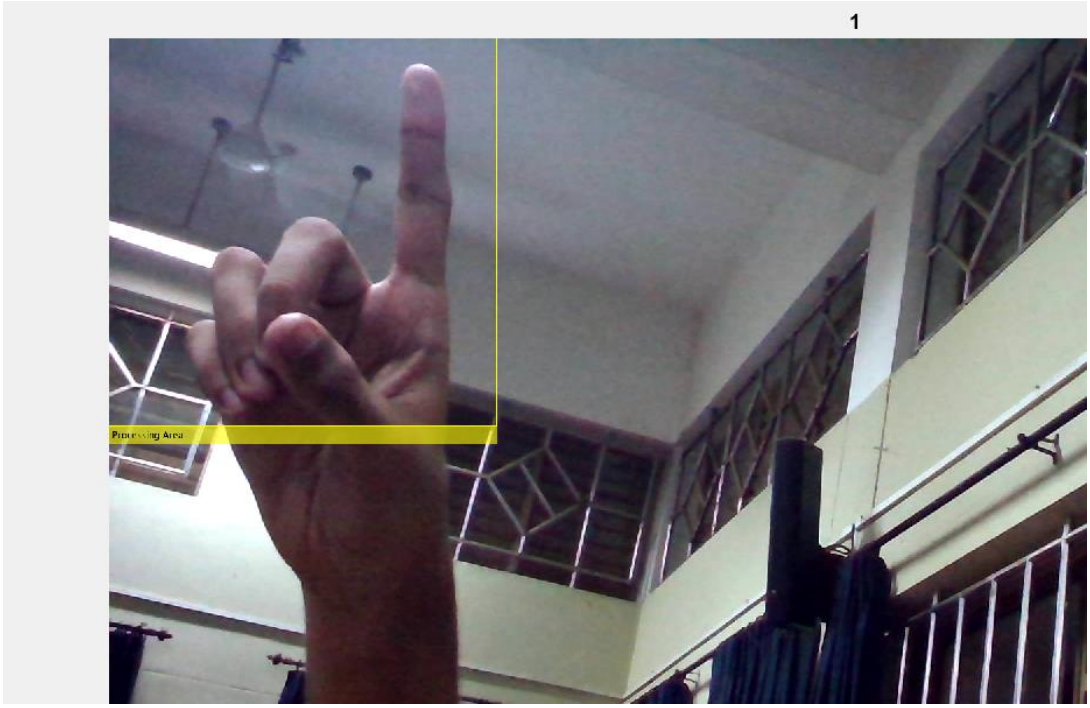
3) **None:**



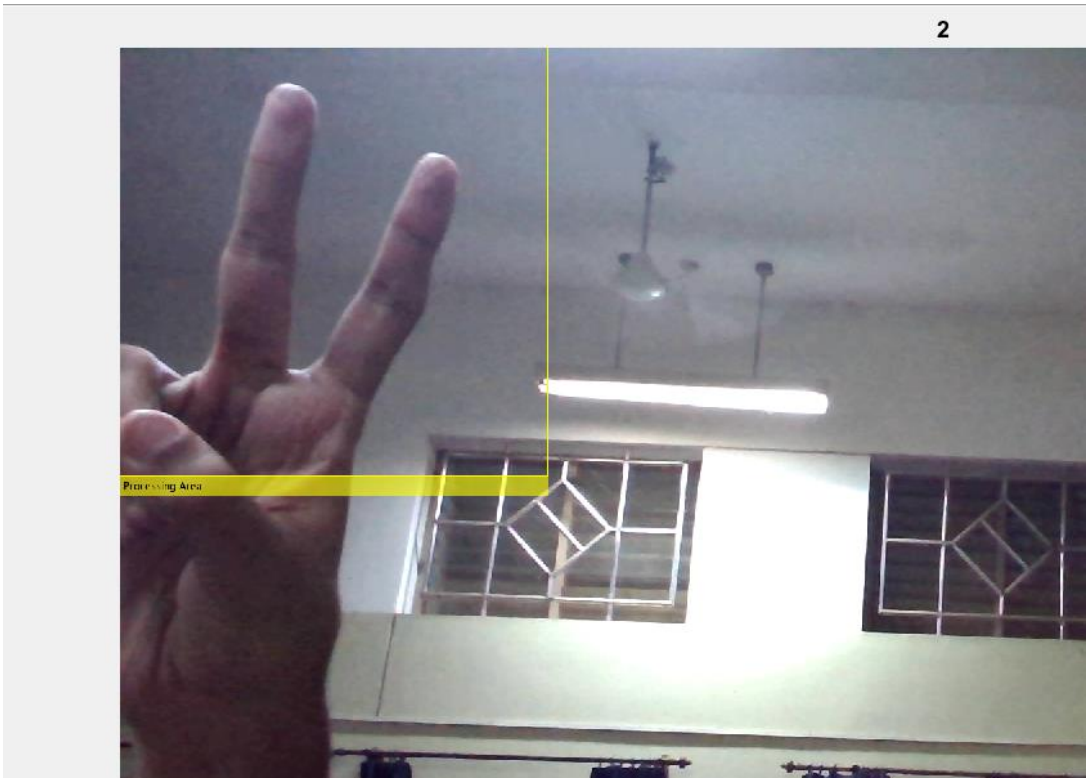
4) **Zero (0):**



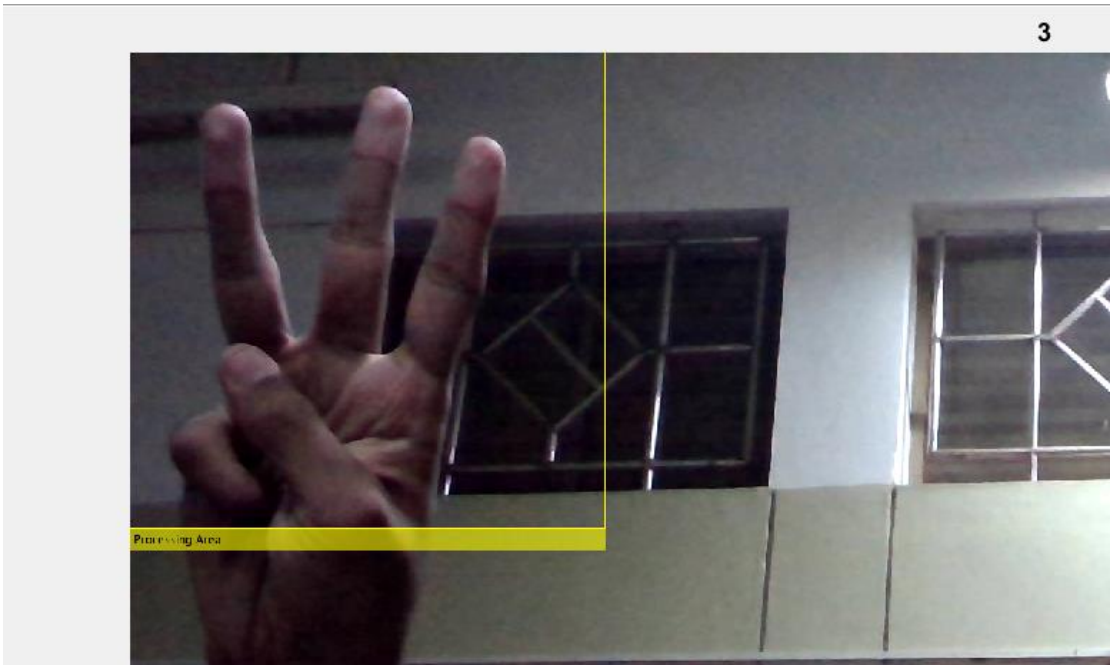
5) **One (1):**



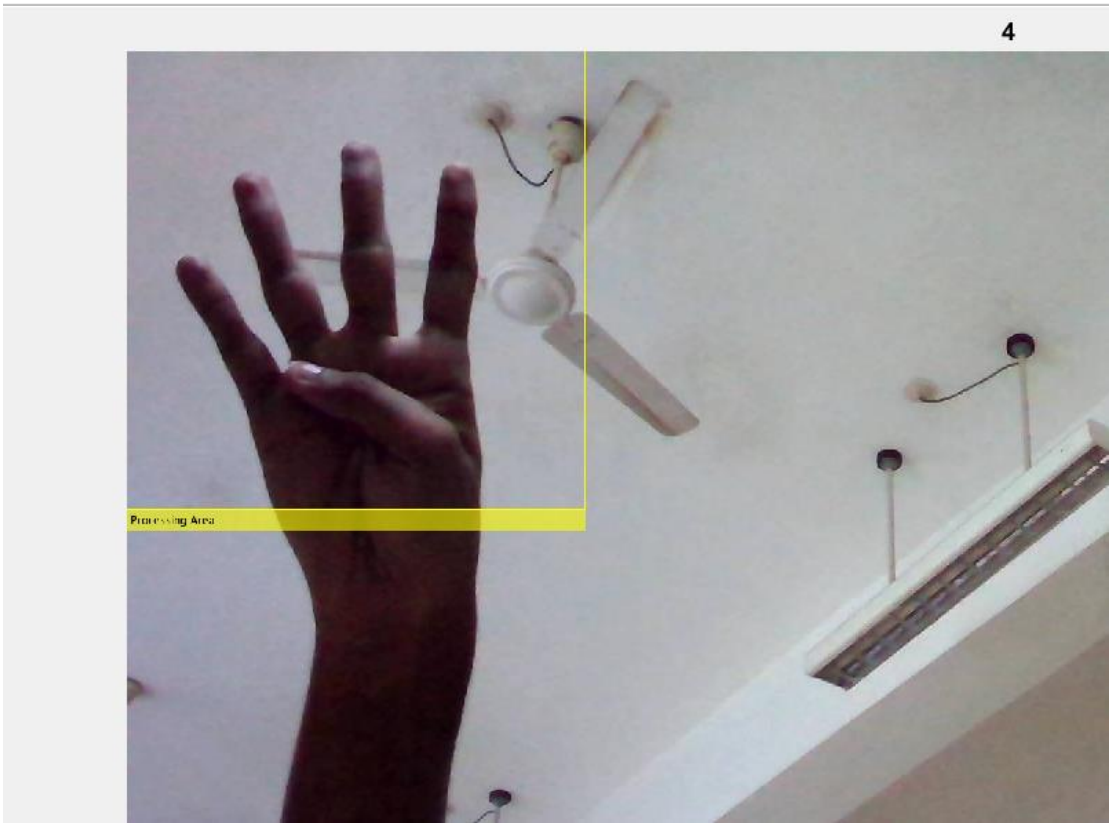
6) **Two (2):**



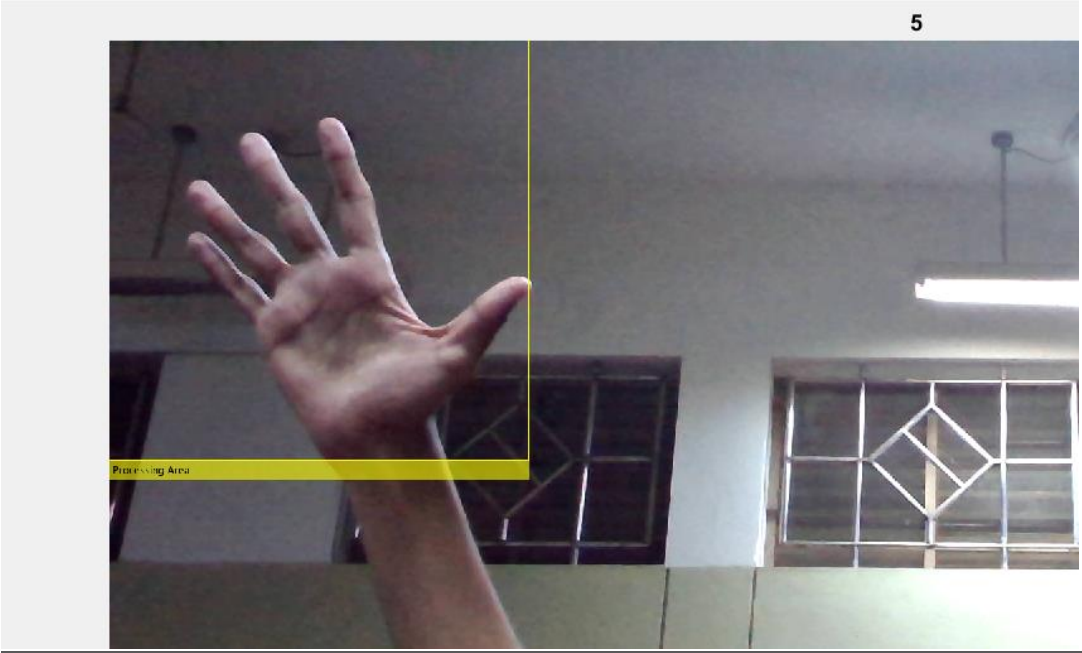
7) **Three (3):**



8) **Four (4):**



9) **Five (5):**



Conclusion:

The Hand Recognition System using AlexNet and MATLAB is an accurate and efficient system for recognizing hand gestures.

The system was successfully able to recognize gestures showing thumbs up, thumbs down, 0, 1, 2, 3, 4, 5 and none through the live webcam feed.

Data collection, testing and training of the model was done and high accuracy results were obtained

The use of AlexNet and MATLAB allowed for a high degree of accuracy and real-time recognition.

This system can be used in various applications, including sign language recognition, gesture-controlled devices, and virtual reality environments.