

Computer Vision - COL780

Assignment 4 : Hand Gesture Detection

Ansh Prakash
2016CS10367

Adithya Anand
2016CS10377

Motivation of the assignment is to easily classify the hand gestures into different classes. The environment can be classified as stop, next, previous and background class. Requirement is to create a model to perform the above task. Classification of hand gestures has its own application.

1. Methodology

Our approach to handle with the above required problem consists usage of convolution neural layers. The architecture of our model consists of five neural layers. However, pre-processing of the dataset has been done in order to overcome the noise of the image.

1.0 Pre-Processing:

We tried Edge Detection but it varied a lot with illumination. So, we dropped that idea.

After that, we did skin detection which improved the accuracy and allowed the neural network to learn as the distribution of images doesn't vary a lot as it gets a binary image and most of the time the irrelevant background gets removed which allows to generalise it to generic images as well.

1.1 Objective function :

Our model basically deals with the cross entropy loss:

Cross Entropy Loss : we are minimizing the negative log likelihood.

Loss : $(-\sum Y(o, c) * \log(po, c))$

As a results, when the loss is minimized the classification accuracy for the classes increases.

1.2 Network Architecture :

The architecture has a basically 2 convnet block which mainly extract the coarse line to the fine level of details in the Image frames.

Layer	Kernels	Size	Stride	Padding
Conv2D	8	5	1	0
BatchNorm2D				
Relu				
MaxPool2D		5	2	
Conv2D	16	5	1	0
BatchNorm2D				
Relu				
MaxPool2D		5	3	
FC1		400 X 128		
FC2		128 X 32		
FC3		32 X 4		

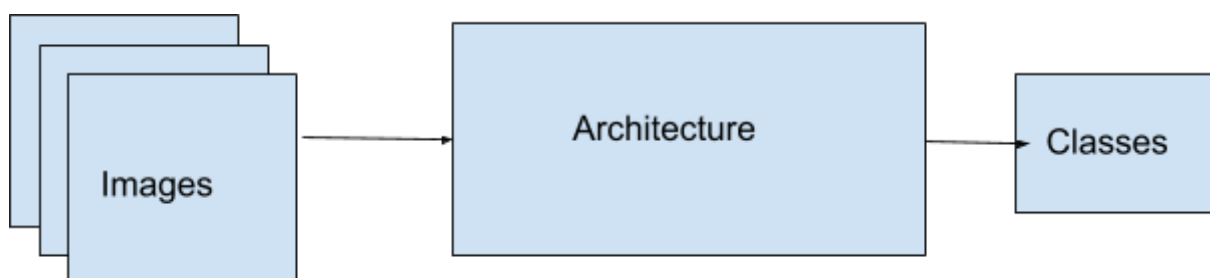


Fig: Block Diagram

1.3 Training :

The training of the proposed model is basically the video of varied sizes. However, the basic form are the images which are frames extracted the video which are further being . Size of the images are 50×50 . No of images from for each class nearly 4000 in count. Training images basically been extracted from the dataset which is self prepared with various environment on the background.

Training has been done on the CPU of 8RAM with the learning rate set as 0.001, the batch size is set to be 64. We have trained the model till 100 epochs.

2. Dataset and Experimental Evaluation

We have prepared two dataset. Basically Dataset 1: is prepared with plain background and Dataset 2: is formed with generic background which consists of recording collected at different locations. The plain background was indeed the white board and the other backgrounds were of hostel room surroundings, outside green bushes, black chartboard and the moving screen of the laptop. The actual training images are the black and white images extracted from the frames of the recordings.



Fig : Few frames of the 50×50 images

3.Results

Accuracy on training set:

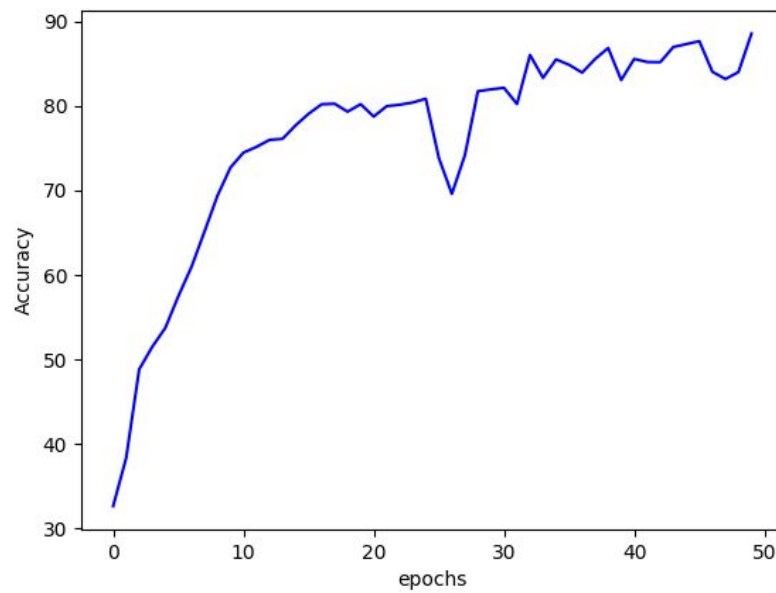


Fig: Accuracy VS Epochs in Training set

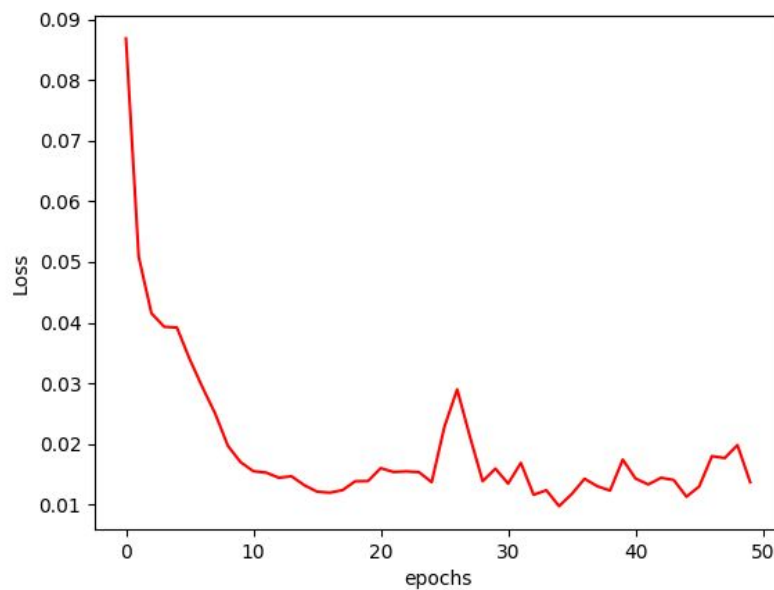


Fig: Loss VS Epochs in Training set

Accuracy in Test Set:

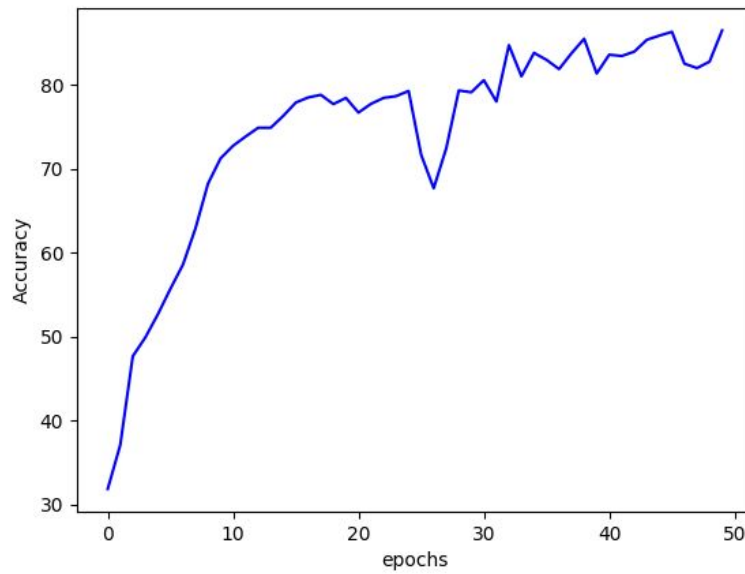


Fig: Loss VS Epochs in Test set

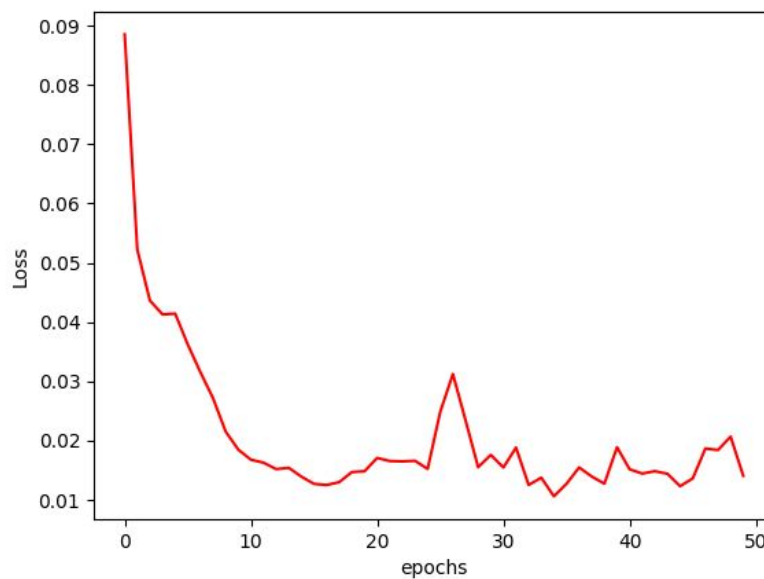


Fig: Loss VS Epochs in Test set

Accuracy at 50 epoch

Training Set(%)	Test Set(%)
-----------------	-------------

82	85
----	----

4.Failure:

Our model has do show failure at some cases when the background and the hand seems indistinguishable. It consider all as a background resulting in misclassification of the image.

We were facing the problem of overfitting however, we corrected it by introducing regularization and dropouts.

5. Acknowledgement:

We did have few discussion with our batchmate Ayush Verma which resulted into valuable improvement of our model