Computer Vision Assignment 4

By
Siddhant Mago(2017CS50419)
Lakshay Saggi(2017CS50412)

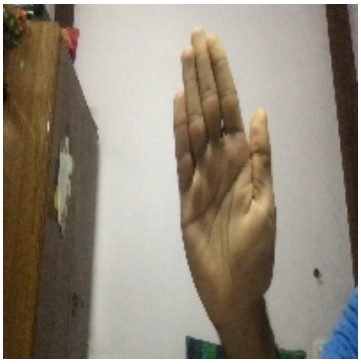
Method:

For both the problems in the assignment statement we have trained a single classifier, our classifier works on the principle of completely removing the background.
The approach adopted by us is as follows:

- We first ask the user to place the webcam at a stable location.
- After that, the user is supposed to press the button 'b'. What this does is assumes that the frame captured when 'b' was pressed is the background and assumes that this background remains fixed and stationary for the remaining execution.
- After that we take each frames and process it so as to subtract the given background from it.
- We then pass the processed image through our neural network, which infers the gesture.

Our neural network has the following architecture:
CONV (7X7X32) => MAXPOOL (2,2) =>
CONV (5X5X32) => MAXPOOL (2,2) =>
CONV (3X3X64) => MAXPOOL (2,2) =>
CONV (3X3X64) => MAXPOOL (2,2) =>
DENSE (with 4 output)

We have used 'reLu' activation for each of the Conv layer and softmax activation for dense layer.
We used batch size = 16, ran our code for 50 epochs and used a learning rate of 0.05. We trained our classifier on around 600-700 images in total. Through, good background subtraction we were able to reduce our problem a lot and thus didn't require a lot of training data.
We chose to start with a 7X7 layer because as we were only allowed 5 layers and we observed that a 3X3 filter was not able to detect properly. Also, our hand covered a major portion of the input image thus 3X3 would not have been sufficient. We used 'reLu' because it is a standard method and gives good results. We used a relatively large number of epochs because we had a small training data.
The gestures supported by our classifier are as follows:

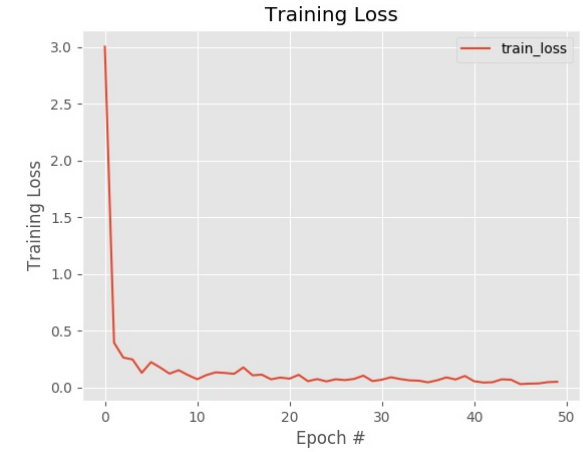



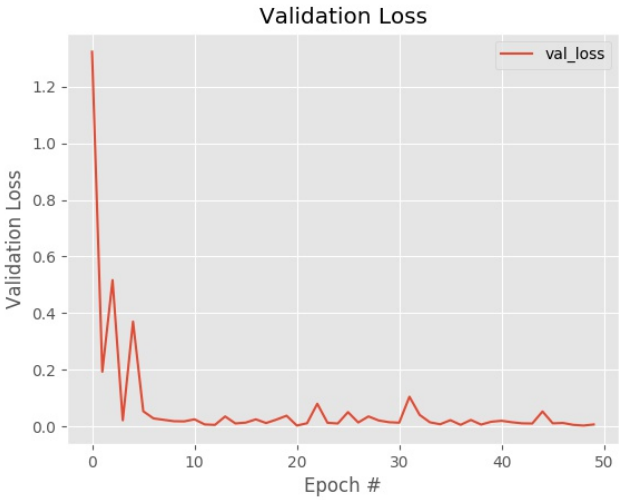

After pre-processing and background subtraction we got the following results:



These images were given as inputs to our neural network. As it is evident from the above images, our classifier doesn't depend on the background as long as it stays fixed. That is we thought it would be better to submit a single classifier for both the problems of the assignment statement.

The graph of validation loss and training loss are as follows:





FUTURE SCOPE:

It is also possible to allow the user to change the stationary background after the code starts executing and also support a lot of new gestures with different poses.