Assignment 4

October 23, 2019

In this assignment, we will do the basics of a camera based music player control.

1 Task Description

The task in this assignment is to identify the direction of the gesture made by a user in front of the camera. We will limit the gestures to next, previous and stop signs as shown in the figure below. So, the model would predict one of four classes - Previous, Next, Stop and Background/Others.

Figure 1: Previous, Next and Stop gestures

1.1 Plain Background

1. In front of a plain background, make the gestures with your hands. Record a video with variation in pose, brightness and hands (at least your own and your partner's hands should work).

2. Split the video in train and test splits. Train a model which detects the above gestures. Keep in mind the model limitations as mentioned in Section 2. Keep the initialization seed fixed so that the output can be reproduced.

3. Include the training loss and validation loss graphs over iterations in your report. Justify the hyperparameters used.

4. For the demo, take input from the webcam and display text on the image itself. Since the inference may take time, you may skip the frames in between.

1.2 Generic Background

1. In front of a generic background, take similar images (with a smaller scale of hand) and use the images to train your model. Some examples are as follows:

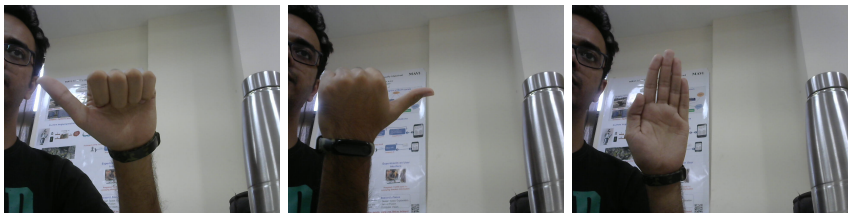

Figure 2: Previous, Next and Stop gestures (Cluttered)

2. Split the video in train and validation splits. Train a model which detects the above gestures. Keep in mind the model limitations as mentioned in Section 2. Keep the initialization seed fixed so that the output can be reproduced.
3. Include the training loss and validation loss graphs over iterations in your report. Justify the hyperparameters used.

2 Model Specifications:

1. Only a total of 5 layers can be used.
2. The starting image dimensions have to be 50x50.
3. A maximum of 4000 images per class can be used for training.
4. Using any pre-trained weights is not allowed.
5. Any pre-processing can be done on the images. For example, you may use grayscale images/edges as input or use background subtraction rather than using the RGB image.

3 Submission Instructions

Create a zip file with your python script and weights/model file. The default run mode should be webcam (device 0). During the demo, you should be able to start the training and show plots/loss values as the training progresses. You may use Keras/Pytorch for implementation.