COL780 A4 Report

Shashwat Shivam (2016CS10328)
Atishya Jain (2016CS50393)

November 13, 2019

1 Dataset Generation

We made several videos of hand gestures for each class using both hands of both the partners. We then split the frames to create the dataset. After creating the dataset, we finally preprocess them before feeding them into the network. A part of data was also collected using burst photos from phones to add variety to the dataset.

2 Pre-processing

1. We have 5 channels as input to the model. First 3 channels are R, G and B channels respectively. Fourth channel is the grayscale image and the last 5th channel is the binarised input.
2. From binarising effectively, we threshold the image after smoothening it.
3. After this we reduce the size to 50*50.

Figure 1: Input RGB and Grayscale Channel

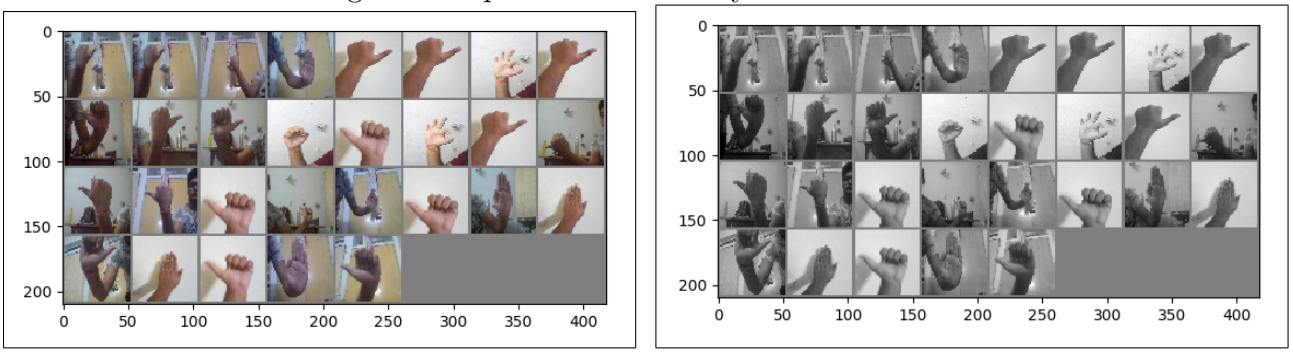
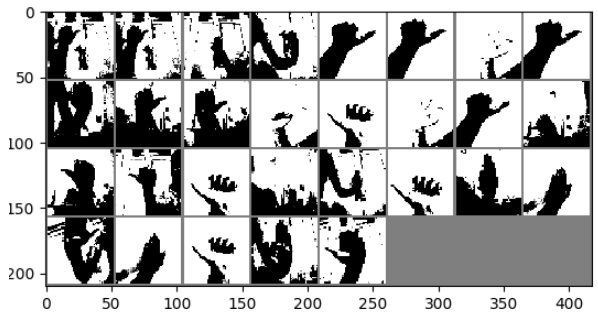


Figure 2: Input Binary Channel



3 Architecture

- Number of Convolution layers: 2
  - Layer1: Conv2d 5 channels to 10 channels followed by batchNorm , ReLu and MaxPooling
  - Layer2: Conv2d 10 channels to 16 channels followed by Dropout, ReLu and Max-Pooling.
- Number of fully connected layers: 3
  - Layer1: 1600 - 256
  - Layer2: 256 - 128
  - Layer3: 128 - 4

4 Training and Validation

We train our model to detect 4 classes (Next, Prev, Others and Pause). Below you can find the attached Training and Validation loss graphs.

Loss is the standard cross entropy loss between in original and predicted classes.

1. Simple Dataset with no background:
   - Final Training F1: 0.99
   - Final Validation F1: 1.00
   - Final Testing F1: 0.917
2. Generic Background Dataset:
   - Final Training F1: 0.969
   - Final Validation F1: 0.982
   - Final Testing F1: 0.974

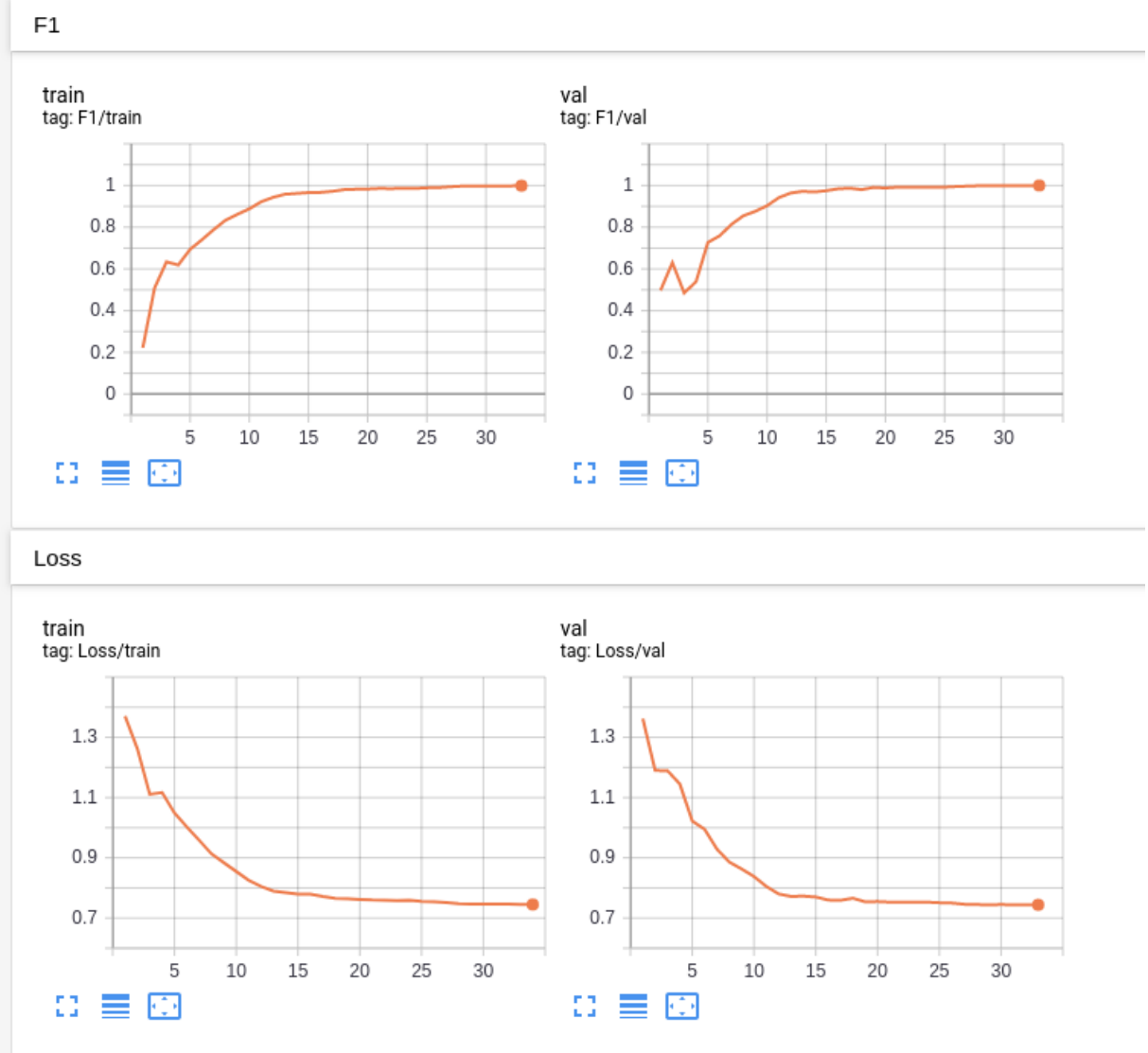


Figure 3: Simple Dataset Plots

5 Hyperparameter Tuning

We tried various learning rates (0.1,0.01,0.001) and different optimizers (SGD, Adam, RMSProp) to get the best training curriculum which was decided based on quality of training curves obtained.

6 Demos

For the Demo, we have a live classification window that labels the hand guestures on the screen.

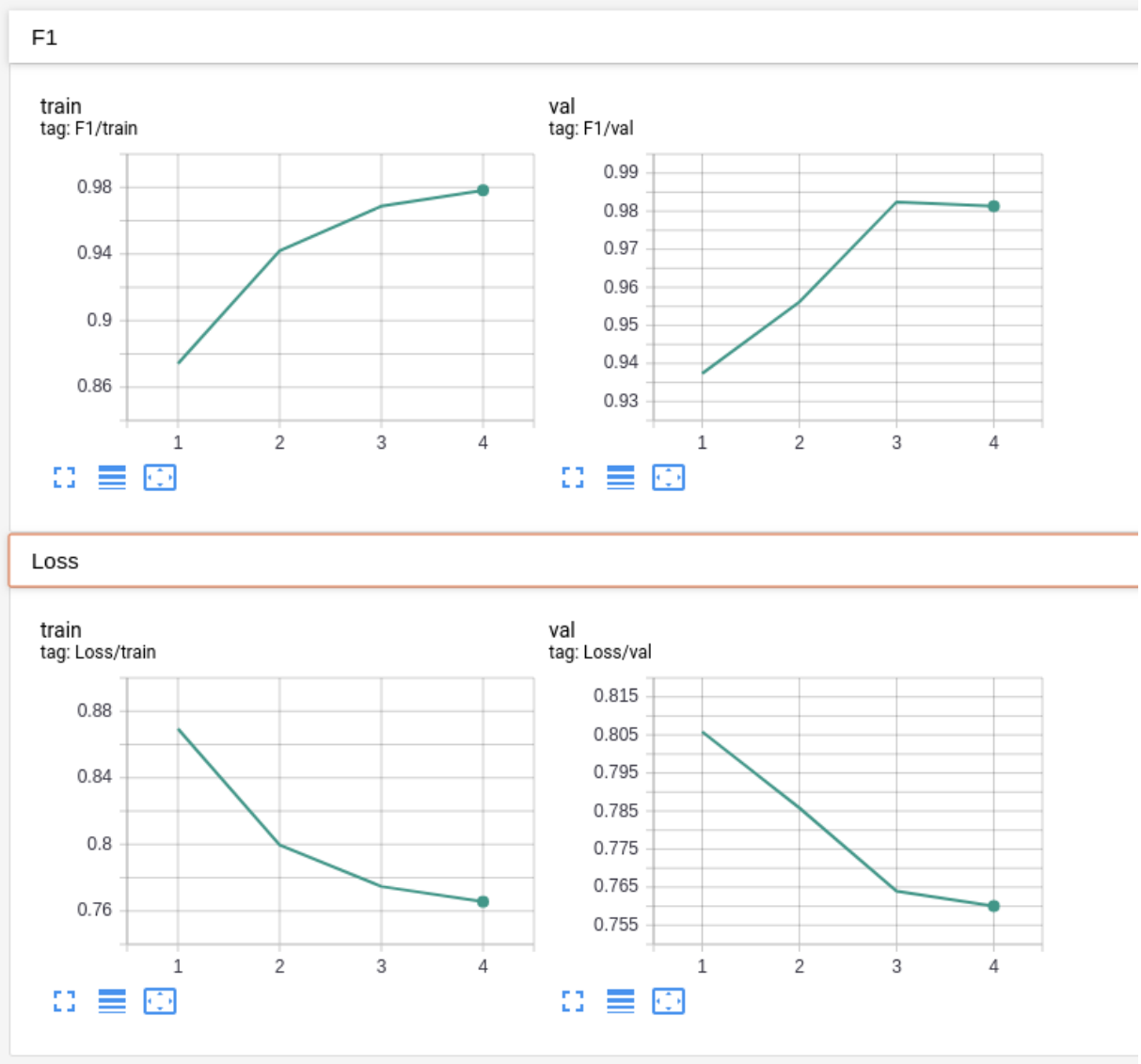


Figure 4: Generic Dataset Plots