Speech Recognition Project

Vishal Mate

IIT Hyderabad

September 26, 2021

Problem statement

This is a short project on Speech Command Recognition. Here you have to generate data using your own recorded commands "Forward","Back","Left","Right","Stop". Set the sampling rate of the recorded voice commands at 16 KHz and generate total 80 utterances of each command.Trim the samples to 1 second. You can use any architecture to recognize the speech commands, you have to fix the hyper-parameters so that the accuracy can be increased. Summarize the model architecture as well as the hyper-parameters. Use 25% of samples for testing and rest for training the model.

Mechanism Involved

1. Generating data using appropriate voice recording software
2. Adding this data to a particular path in drive
3. Loading the data into Colab and Importing required libraries
4. Splitting data into Train and Test
5. Checking Model Summary
6. Fitting the model on to the training and validation data using model.fit()
7. Calculating acccuracy of the model
8. Comparing different plots(Accuracy vs Epochs) obtained by changing hyperparameters

Architecture

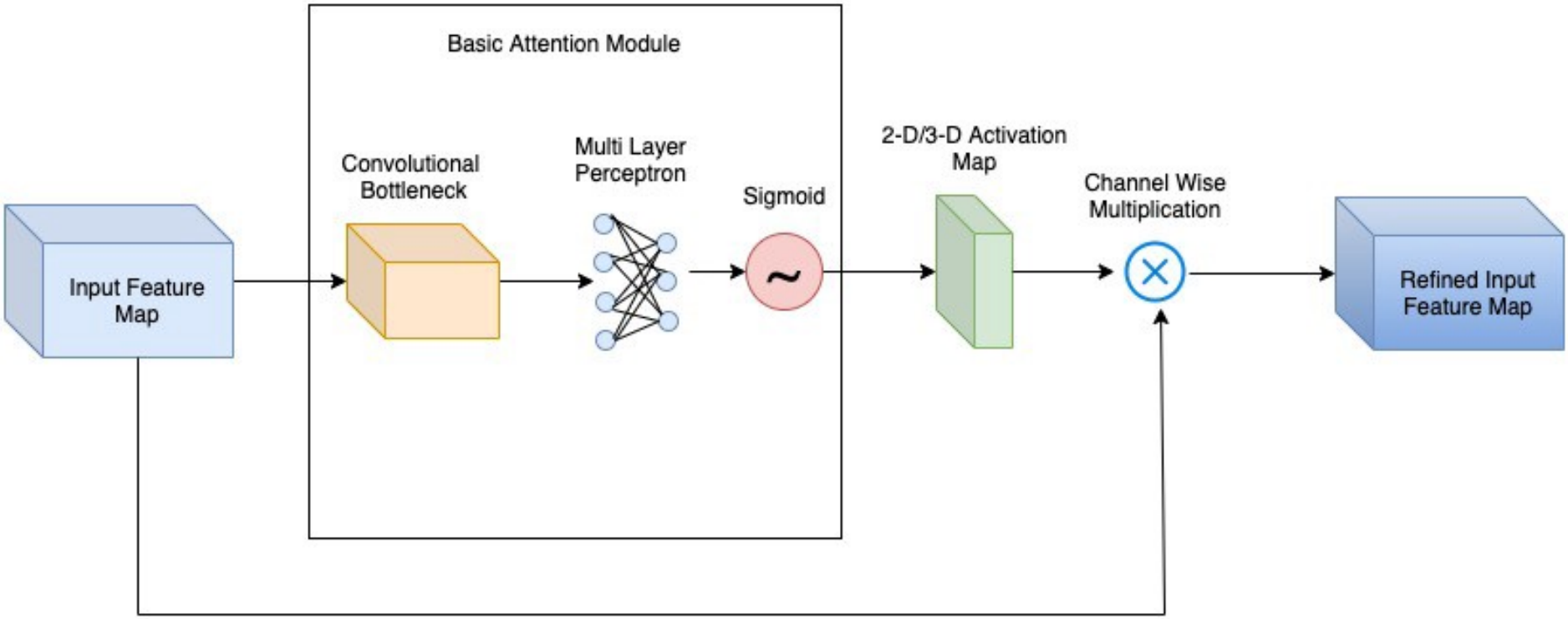
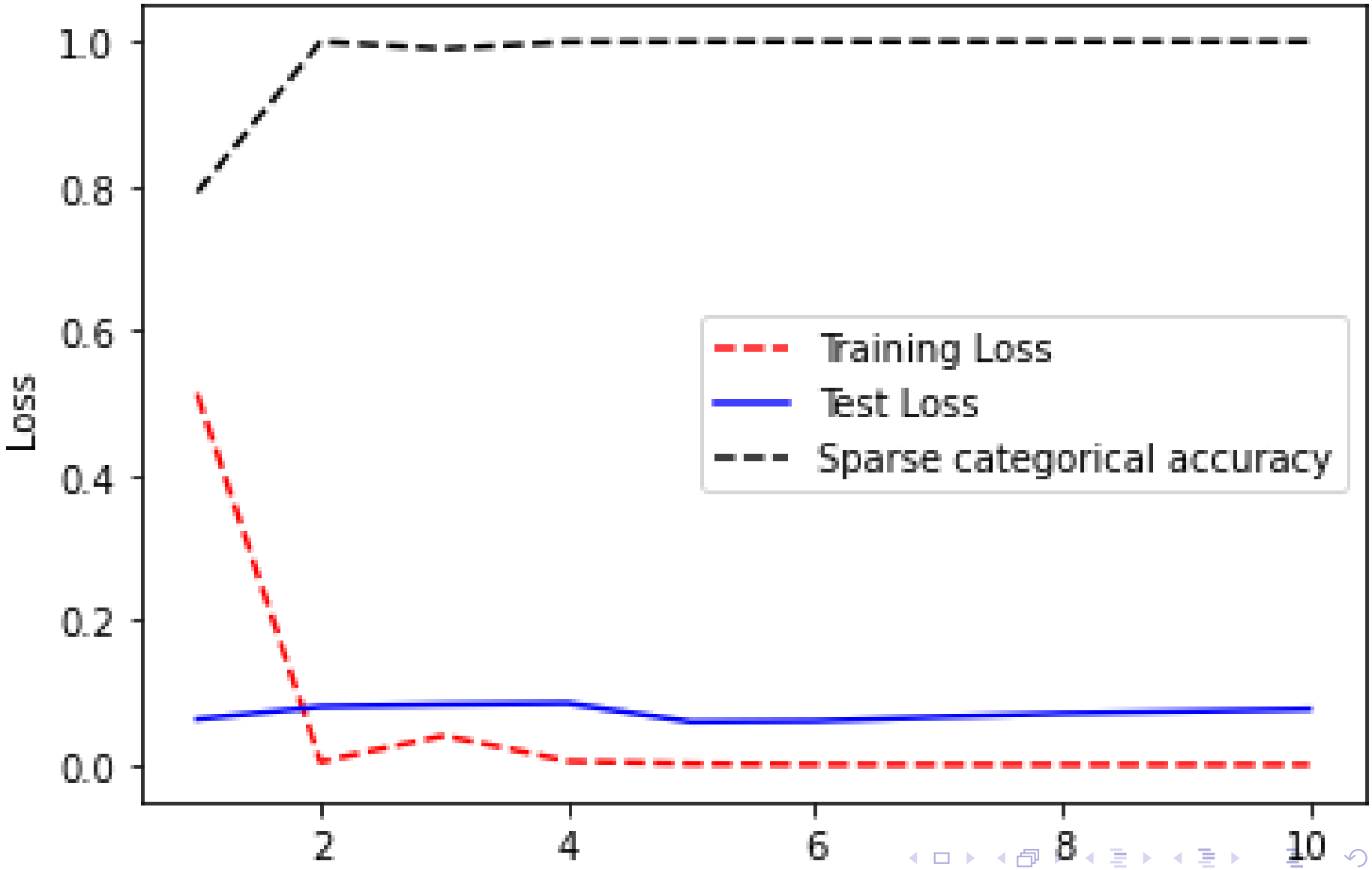
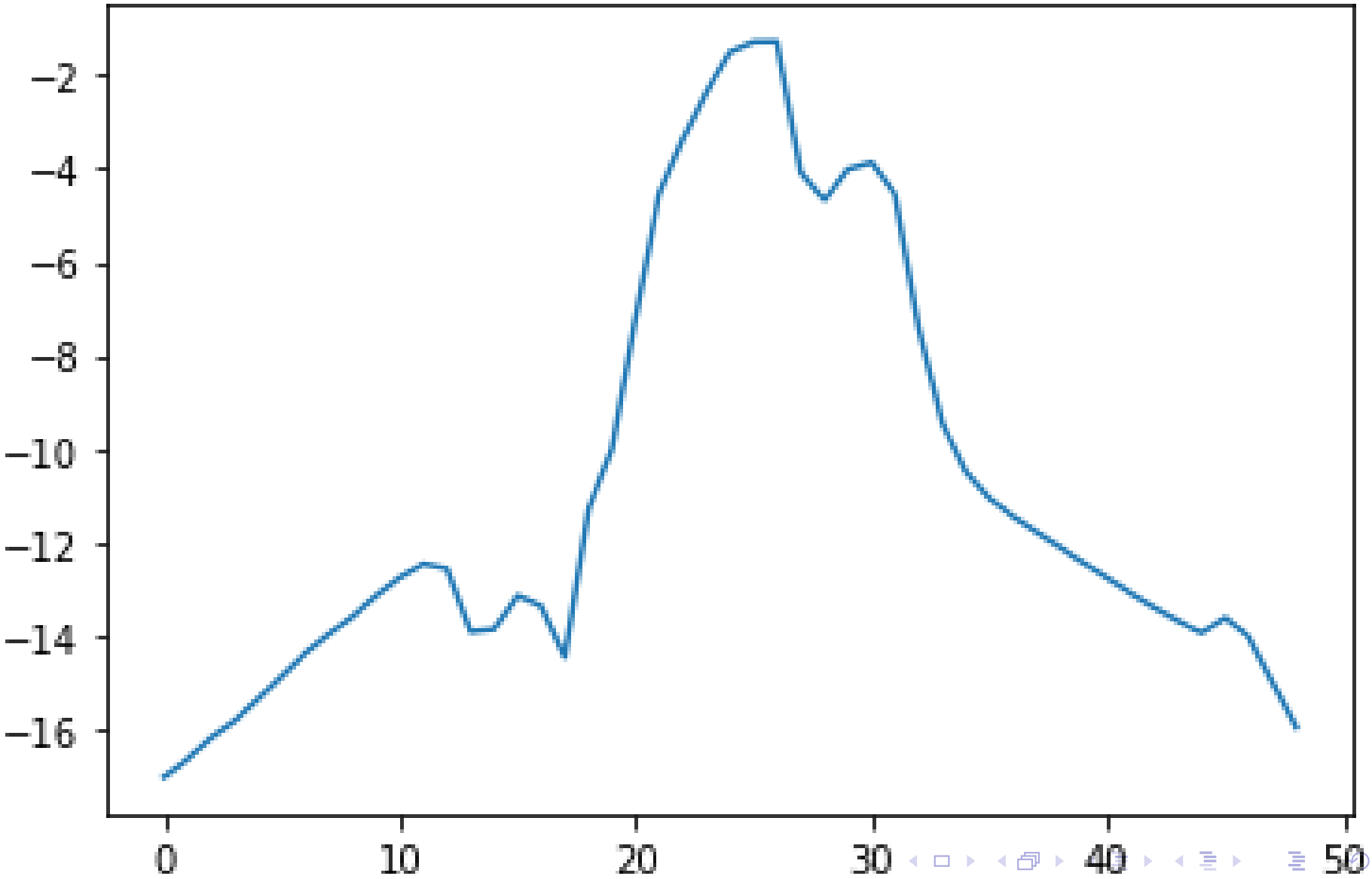


Figure: Plot : Architecture

Results


1.0
0.8
0.6
0.4
0.2
0.0
Loss
2
4
6
8
10
Training Loss
Test Loss
Sparse categorical accuracy

Results



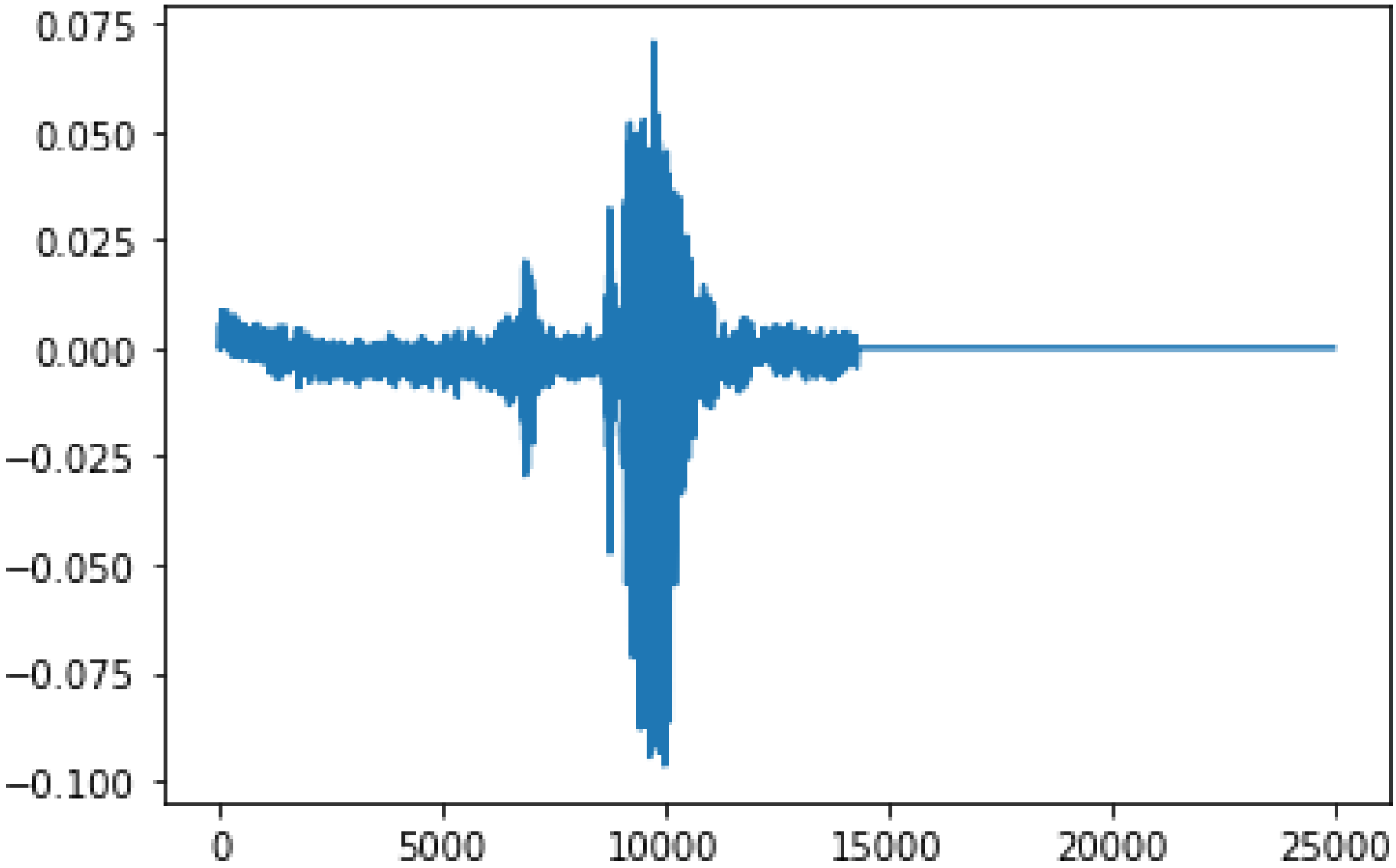


Figure: Plot : Spectogram

Conclusion

1. Accuracy of the model can be increased by changing the hyperparameters such as:
   - Using Double convolutional model increases accuracy
   - Batch size
   - Number of epochs, etc.
2. Training loss is only at start but is decreases drastically at 2 but the test loss is constant.