GESTURE RECOGNITION CAPSTONE PROJECT

Project Description:

In this project, we build a 3D Conv model that was able to predict the 5 gestures correctly for a TV company.

Approach:

Since this project required a lot of computation, we build this project on Kaggle using GPU T4 *2.
The first step was to create a generator function which yields images one by one to the model.
The images were augmented for better training data.

We started with a very basic Conv3D model which was not able to learn and giving very less accuracy (Around 0.15). Then we added layers on top of Conv3D including Batch Normalisation. After 6 Conv3D layers along with Max Pooling and Batch Normalisation, the accuracy went up to ~54%, but was not enough.

Then we used a combination of Time Distributed, Transfer Learning (VGGNet) and GRU. The accuracy was limited to ~75%. We tweaked the images using cropping, affine transformation but still results were not improved.

So we changed the transfer learning to MobileNet. The accuracy improved a little bit about ~78%
Then we added LSTM block to our model which was combination of MobileNet, Time Distributed, GRU and LSTM layers The training accuracy was almost ~100% but validation accuracy was still around 80%. So add the final touch we added 2 dropout layers to our final model.

Observations:

Training Accuracy: 95.45

Validation Accuracy: 97.66

Train Loss: 0.0154

Val Loss: 0.0741