

Gesture Recognition Study

Problem: Solving the feature which is developed in smart-TV that can recognise five different gestures performed by user without using remote. The gestures are

- Increase the volume (Thumbs up)
- Decrease the volume (Thumbs down)
- Jump backwards 10 seconds (Left swipe)
- Jump forward 10 seconds (Right swipe)
- Pause the movie (Stop)

We build the model and experiment the various hyper parameters and layers, then captured the results below.

Experiment Number	Model	Result	Decision + Explanation
1	Conv3D Image size: 200 x 200	Throws error	Got an error due to different size between different input set.
2	Conv3D Image size: 64x64 Epochs: 20	Model is not learned and trainable because Training and Validation accuracy/ loss is poor	Introduce Dropout layer to learn effectively
3	Conv3D	Training accuracy: 0.97, Validation accuracy: 0.32	Though training accuracy is improved, validation performance is poor. Rather using flatten module, we can introduce Global Average pooling
4	Conv3D Image size: 64x64 Epochs: 40	Training accuracy: 0.96, Validation accuracy: 0.47	Though validation accuracy is improved, still the result prediction gives poor performance. Rather using Conv3D, we can try to use Conv2D and GRU Neural network

5	Conv2D + GRU Image size: 64x64 Epochs: 40	Training accuracy: 0.94, Validation accuracy: 0.67	When we use Con2D and gated mechanism i.e., Time Distributed GRU Neural network, the model performance is improved drastically i.e., getting good validation result. Going check with LSTM neural network.
5	Conv2D + LSTM Image size: 64x64 Epochs: 40	Training accuracy: 0.94, Validation accuracy: 0.96	When we use Con2D and Time distributed LSTM, we got the best result i.e., Training and validation accuracy is greater than 94% though it consumes more training time.

Conclusion: *When we use Convolution 2D and Time Distributed LSTM neural networks for sequential data as in Gesture recognition, we are getting the better results compared to other models and getting good training and validation accuracy and minimum loss.*