

Gesture Recognition Case Study

Team Members: Vagdevi Pedalenka
Arjun Ashok Kumar

Problem statement: Recognize five different gestures performed by the user which will help users control the RV without using a remote.

The gestures are continuously monitored by the webcam mounted on the TV. Each gesture corresponds to a specific command.

1. Thumbs up: Increase the volume
2. Thumbs down: Decrease the volume
3. Left swipe: 'Jump' backwards 10 seconds
4. Right swipe: 'Jump' forward 10 seconds
5. Stop: Pause the movie

We tried multiple models - Conv3D, CNN+RNN and also tried with transfer learning. The following is a list of models we trained with the results and explanation.

Experiment Number	Model	Result
1	Conv3D	Model template; not trained
2	Conv3D	Categorical accuracy = 0.9110 Validation accuracy = 0.73 (epoch 19) Model overfits.
3	Conv3D	Categorical accuracy = 0.6833 Validation accuracy = 0.33 (epoch 11) Hight degree of overfit.
4	Conv3D	Categorical accuracy = 0.908 Validation accuracy = 0.22 (epoch 24) Improved training accuracy but Very low validation accuracy(overfitting).
5	Conv2D + GRU	Categorical accuracy = 0.7768 Validation accuracy = 0.36 Model overfits

6	Conv2D + GRU	Categorical accuracy = 0.9351 Validation accuracy = 0.50 (epoch 37) Model overfits.
7	Conv2D + LSTM	Categorical accuracy = 0.7376 Validation accuracy = 0.39 (epoch 20) Model overfits.
8	Transfer Learning: MobileNet(without training) + LSTM	Categorical accuracy = 0.9864 Validation accuracy = 0.91 (epoch 28) This model performs very well.
9 (Final model)	Transfer Learning: MobileNet(with training) + GRU	Categorical accuracy = 0.9819 Validation accuracy = 0.96 (epoch 27) This is the best model we got so far.

The model trained using mobileNet + GRU performed the best – good train and validation accuracy and no signs of overfitting. Hence, we can use this as the final model.