Gesture Recognition Case Study

The below experiments were conducted to train the Neural Network Model to recognize 5 hand gestures using Conv3D, Conv2D + LSTM and LSTM with Transfer Learning models.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Exp No | Model | Result | Decision + Explanation | Parameters |
| Test Model | * 3 conv3D Layers * Frames: 16 * Batch: 10 * Epochs: 4 * Ablation: 25 | Accuracy: 0.44  Val Accuracy:0.24 | * The model is working. * Let’s **try overfitting** the model | 2,438,533 |
| Overfit Model | * 3 conv3D Layers * Frames: 16 * Batch: 10 * Epochs: 15 * Ablation: 25 | Accuracy: 1.0  Val Accuracy:0.24 | * The model is overfitting as expected. * Let’s **increase Epoch** and **run the model on full data** set. * **Add dropouts** to reduce overfitting | 2,438,533 |
| 1 | * 3 conv3D Layers with dropouts * Frames: 16 * Batch: 16 * Epochs: 20 | Accuracy: 0.85  Val Accuracy:0.87 | * Accuracy has increased. * Let’s see how the model behaves with **SGD optimizer** | 2,438,533 |
| 2 | * 3 conv3D Layers with dropouts * Frames: 16 * Batch: 16 * Epochs: 20 * Optimizer: SGD | Accuracy: 0.52  Val Accuracy:0.61 | * Accuracy has decreased. * Let’s switch back to **Adam optimizer** * **Add one additional Conv3D layer** | 2,438,533 |
| 3 | * 4 conv3D Layers with dropouts * Frames: 16 * Batch: 16 * Epochs: 20 | Accuracy: 0.79  Val\_accuracy: 0.88 | * Accuracy has Increased * Let’s try **doubling each conv3D layer** | 890,373 |
| 4 | * 4 conv3D Layers with each layer repeated twice and dropouts * Frames: 16 * Batch: 16 * Epochs: 20 | Accuracy: 0.19  Val\_accuracy: 0.23 | * Accuracy has decreased drastically. * Let’s stick to **4 conv3D layers** without repetition * Let’s try with **LR as 0.002** | 1,478,133 |
| 5 | * 4 conv3D Layers with dropouts * Frames: 16 * Batch: 16 * Epochs: 20 * Learning Rate: 0.002 | Accuracy: 0.83  Val\_accuracy:0.84 | * Val\_Accuracy has not improved compared to model 3 * Let’s stick with **default LR i.e., 0.001** * Let’s try **adding BatchNormalization** | 890,373 |
| 6 | * 4 conv3D Layers with dropouts and BatchNormalization * Frames: 16 * Batch: 16 * Epochs: 20 | Accuracy: 0.24  Val\_accuracy: 0.32 | * Accuracy has decreased. * Execution stopped early due to **early stopping** * **Loosing lot of information** when using both Dropout and BatchNormalization * Let’s stick with simple 4 conv3D layers with dropouts * Let’s **increase Epochs** | 890,853 |
| 7 | * 4 conv3D Layers with dropouts * Frames: 16 * Batch: 16 * Epochs: 30 | Accuracy: 0.91  Val\_accuracy: 0.90 | * Accuracy has increased * Let’s **increase Batch** | 890,373 |
| 8 | * 4 conv3D Layers with dropouts * Frames: 16 * Batch: 32 * Epochs: 30 | Accuracy: 0.89 Val\_accuracy: 0.93 | * Val\_Accuracy has increased with higher batch size * Let’s **increase Batch further** | 890,373 |
| 9 | * 4 conv3D Layers with dropouts * Frames: 16 * Batch: 64 * Epochs: 30 | Accuracy: 0.87  Val\_accuracy: 0.90 | * Accuracy has decreased with large batch size * Let’s **increase Frames and reduce Batch size** | 890,373 |
| 10 | * 4 conv3D Layers with dropouts * Frames: 20 * Batch: 16 * Epochs: 30 | Accuracy: 0.86 Val\_accuracy: 0.85 | * Accuracy has not improved much. * Let’s **increase Frames further** | 890,373 |
| 11 | * 4 conv3D Layers with dropouts * Frames: 25 * Batch: 16 * Epochs: 30 | Accuracy: 0.87  Val\_accuracy: 0.89 | * Accuracy has increased * Let’s **increase Batch** | 890,373 |
| 12 | * 4 conv3D Layers with dropouts * Frames: 25 * Batch: 40 * Epochs: 30 | Accuracy: 0.93  Val\_accuracy: 0.92 | * **Accuracy has increase to 93%** * Let’s stop the conv3D model experimentation here and try with **conv2D+LSTM** models | 890,373 |
| 13 | * 4 conv2D + LSTM Layers with dropouts * Frames: 20 * Batch: 32 * Epochs: 30 | Accuracy: 0.91  Val\_accuracy: 0.75 | * **Slightly Overfitting** as val\_accuracy is less * Let’s **decrease the Batch** | 2,539,941 |
| 14 | * 4 conv2D + LSTM Layers with dropouts * Frames: 20 * Batch: 10 * Epochs: 30 | Accuracy: 0.60  Val\_accuracy: 0.70 | * **Accuracy has decreased** * Let’s try using few **Transfer Learning** models | 2,539,941 |
| 15 | * LSTM with Transfer Learning (**Inception V3**) * Frames: 20 * Batch: 32 * Epochs: 30 | Accuracy: 0.74 Val\_accuracy: 0.57 | * **Slightly Overfitting** as val\_accuracy is less * Let’s try with other Transfer Learning model |  |
| 16 | * LSTM with Transfer Learning (**Mobilenet**) * Frames: 20 * Batch: 20 * Epochs: 30 | Accuracy: 0.88 Val\_accuracy: 0.75 | * Has Accuracy of 88% |  |

After conducting experiments with Conv3D, Conv2D + LSTM and LSTM with Transfer Learning models we conclude the Best Model we have achieved is Model 12 which has training accuracy of 0.93% and val\_accuracy of 0.92%

* 4 conv3D Layers with dropouts
* Image size 100 X 100
* Frames: 25
* Batch: 40
* Epochs: 30
* Trainable Parameters: 890,373