**SUMMARY FOR GESTURE RECOGNITION CASE STUDY**

We have experimented with different model architecture for both type CONV 3d and CNN+RNN whose details are given in the below table. We have accomplished the following in the case study:

1. **Generator**:  The custom written generator is able to take a batch of videos as input without any error. Steps like cropping, resizing and normalization has been performed successfully.
2. **Model**: Model is developed that is able to train without any errors and performs well on the total number of parameters and good accuracy achieved. As shown in the table below, Model 8 performs the best for our case study and hence this is the final model.
3. **Write up**: This write up contains the detailed procedure followed in choosing the final model.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| MODEL NAME | NUMBER OF PARAMETERS | AUGMENTATION | TRAINING ACCURACY | VALIDATION ACCURACY | EXPERIMENT | OBSERVATIONS & REASONS |
| MODEL 1 | TOTAL -1117061    TRAINABLE-  1116325 | NO | 93.4% | 47.0% | **Conv 3D (Base Model)**  No Augmentation  Filter: 3x3x3  Batch size:40  Resolution: 160x160  Epochs: 15 | Model is overfitting, so trying with data augmentation and reducing filter size and resolution. |
| MODEL 2 | TOTAL - 1762613    TRAINABLE-  1761109 | YES | 71.4% | 72% | **Conv 3D**  Data Augmentation  Increasing dropouts  Filter: 2x2x2  Batch size:30  Resolution: 120x120  Epochs: 30 | Overfitting solved but the accuracy is poor. Let’s try with more augmentation. |
| MODEL 3 | TOTAL -3638981    TRAINABLE-  3637477 | YES | 86.1% | 83% | **Conv 3D**  More Augmentation  Filter: 3x3x3  Batch size:20  Resolution: 160x160  Epochs: 30 | Accuracy improved, trying with more augmentation & reduced network parameters. |
| MODEL 4 | TOTAL -230949    TRAINABLE-  230453 | YES | 78.6% | 75% | **Conv 3D**  More Augmentation  Reduced Network Parameters  Filter: 3x3x3  Batch size:20  Resolution: 160x160  Epochs: 30 | Stable performance but the accuracy has decreased so let’s try with CNN + RNN type models. |
| MODEL 5 | TOTAL - 1657445    TRAINABLE-  1656453 | YES | 92.38% | 69% | **CNN + LSTM (Base Model)**  Batch size:20  Resolution: 120x120  Epochs: 20 | Model is overfitting, let’s try adding more augmentation and reducing layer |
| MODEL 6 | TOTAL - 2573925    TRAINABLE-  2573445 | YES | 95.85% | 78% | **CNN + LSTM with GRU**  More Augmentation  Reduced Layers  Batch size:20  Resolution: 120x120  Epochs: 20 | More augmentation reduced overfitting, to further improve accuracy, let’s use transfer learning |
| MODEL 7 | TOTAL - 3840453    TRAINABLE-  609541 | YES | 82.81% | 86% | **CNN + LSTM**  Used Transfer Learning  Batch Size: 5  Resolution: 120x120  Epochs: 20 | Validation accuracy is poor as mobilenet weights not trained. Let’s train them also. |
| MODEL 8 | TOTAL -3693253    TRAINABLE-  3669317 | YES | 93.59% | 92% | **CNN + LSTM with GRU**  Used Transfer learning  MobileNet weights trained.  Batch Size: 5  Resolution: 120x120  Epochs: 20 | We get a better accuracy on training mobilenet layer’s weights as well.  Hence, this is our final model. |