Following is the summary of experiments and observations for the project.

* We did many more experiments but listing down the main ones along with the explanations.
* In the model column itself, we have tried to give the summary of model.
* Note that dropouts (usually it was set to 0.25/0.5 in dense connections) and batch normalization play an important role for regularization but not explicitly mentioned them in the model column.
* We tried batch size from 20 to 32 for all the models.
* We zeroed down to experimenting with either 15 images OR 18 images per video.
* We zeroed down to experimenting with either 120 by 120 image size OR 84 by 84 image size.
* Final model is model using transfer learning (MobileNet + GRU) number of params – 3.7M
* Data transformations/augmentation didn’t seem to be helping much in improving accuracy or loss so we started avoiding them in the final constructive modeling.

|  |  |  |  |
| --- | --- | --- | --- |
| **Experiment Number** | **Model** | **Result** | **Decision + Explanation** |
| **1.Use starter code generator and model**  **30 Epochs**  **Batch size – 64**  **Frames – 18,**  **Image – 84\*84** | **Conv3D**  **64+128+256+256**  **Optimizer - SGD** | **Catagorical accuracy ~ 22%**  **Validation accuracy ~ 28%** | **Model is underfitting.**  **Might be the number of layers is not enough to produce it.**  **Will increase image size and try.** |
| **2. Added more layers 256 conv3d with dropouts post flattening.**  **Changed activation function to RELU**  **Increased Image Size– 120\*120** | **Conv3D**  **64+128+256+256**  **Optimizer – ADAM**  **Loss – Catagorical Cross entropy.** | **Catagorical accuracy ~ 20%**  **Validation accuracy ~ 17%** | **Model is still underfitting.**  **Dropouts can cause it, layers are not enough to produce results. Model architecture is not enough. And batch size is high. We need to first overfit the model with lower batch size.** |
| **3. Reduce filter size to 2,2,2.**  **Changed activation function to RELU**  **Image – 120\*120** | **Conv3D-**  **16+32+64+128**  **Optimizer – ADAM**  **Loss – Catagorical Cross entropy.** | **Catagorical accuracy ~ 78%**  **Validation accuracy ~ 74%** | **Model is performing well.**  **Reduced filter size (2,2,2).**  **MaxPool- (2,2,2) with same padding. Will try to add more layers to improve accuracy.** |
| **5. adding more layers** | **Conv3D**  **16+32+64+128+256**  **Optimizer – ADAM**  **Loss – Catagorical Cross entropy.** | **Catagorical accuracy ~ 78%**  **Validation accuracy ~ 22%** | **Adding more layers is not performing well.**  **Validation accuracy has decreased a lot.** |
| **4. increased filter size to 3,3,3.**  **conv3d with dropout’s post flattening.**  **Changed activation function to RELU.**  **Optimizer – ADAM**  **Decreased batch size to 20.**  **Image – 120\*120** | **Conv3D- 16+32+128**  **Dense – 256 +128**  **Dropouts – 25%**  **Optimizer – ADAM**  **Loss – Catagorical Cross entropy.** | **Catagorical accuracy ~ 98%**  **Validation accuracy ~ 87%** | **Model is performing well now.**  **Decreased batch size to 20, and lower conv3d neurons to capture high level details in first convolution. Post flattening, we have added 25% dropouts.** |
| **5. Increased dropouts after Dense layer to 50%. To reduce overfitting.**  **Increased batch size to 32.** | **Conv3D**  **16+32+128**  **Dense – 256 +128**  **Dropouts – 50%**  **Optimizer – ADAM**  **Loss – Catagorical Cross entropy.**  **LR – 0.002 with decay factor 0.5** | **Catagorical accuracy ~ 77%**  **Validation accuracy ~ 73%** | **Model is having good validation and train accuracy will small difference.**  **Model is working well after adding dropouts, we can reduce dropouts and try more.** |
| **6. ConvLSTM model.** | **16+32+64+128 Time distributed layer.**  **+128 GRU.**  **Filter Size – 3,3** | **Catagorical accuracy ~ 96%**  **Validation accuracy ~ 21%** | **Model is overfitting.**  **Might be LSTM models are good with image processing.** |
| **7. GRU model.**  **Combined with mobilenet model.**  **BatchSize -64**  **Epochs- 30** | **Mnet +GRU**  **Mnet- first four layer.**  **GRU – 512**  **Optimizer – ADAM** | **Catagorical accuracy ~ 90%**  **Validation accuracy ~ 83%** | **Model is performing well on validation set as well.**  **Mnet with GRU is performing very well on validation set and accuracy difference is also less. Num of trainable parameters – 18.3M.**  **Will decrease model complexity by reducing number of GRU units.** |
|  |  |  |  |
| **Final Model: Model 8**  **Batch Size -32**  **Epochs- 25**  **Initial LR- 0.001**  **Decay factor 0.5**  **Frames per video – 18**  **Image Size- 120\*120** | **Mnet +GRU**  **Mobile net (retrain all weights) + GRU (128 cells) + 128 dense nodes.**  **Optimizer – ADAM**  **Initial LR- 0.001** | **Catagorical accuracy ~ 99%**  **Validation accuracy ~ 97%** | **Model is performing well on validation set as well.**  **This model has less trainable parameter than previous one ~ 3.6M** |
|  |  |  |  |

Conclusion:

1. Conv3D model is performing well after hyperparameter tuning.
2. ConvLSTM +GRU is not performing adequately on image/video classification.
3. **Mobilenet + GRU is performing very well with validation accuracy of 97%. Image size, learning rate, Optimizer, number of GRU units, frames per video, Epochs, Dropouts are affecting the model training and accuracy.**