* Each video from training and validation data set has total of 30 frames. In each experiment, a subset of frames is considered for model training and validation. These frames are selected such that they are uniformly spaced across range of 30 frames.
* Three augmentation levels are defined.
  + Level 0 augmentation: No augmentation. Use raw frames as it is.
  + Level 1 augmentation: Augmentation with warp affine transformation about 10degrees.
  + Level 2 augmentation: Augmentation with warp affine transformation about 10degrees and rotation of 10degrees.

Table List of models and their metrics

| **Model Architecture** | **Parameters** | **Accuracy**  **[Training, Validation]** | **Loss**  **[Training, Validation]** | **Comments** |
| --- | --- | --- | --- | --- |
| **Model 1 : Basic 3D Model**   |  | | --- | | Conv3D 16 Features, 3x3 Filter | | Conv3D 32 Features, 3x3 Filter | | Conv3D 64 Features, 3x3 Filter | | Conv3D 128 Features, 3x3 Filter | | Flatten | | FC Dense w/ 128 neurons, 0.5 dropout | | FC Dense w/ 64 neurons, 0.25 dropout | | Softmax Layer |  * Each Convolution layer is followed by a batch normalization and 3D max pooling. * Each fully connected layer is followed by a batch normalization. * All activations except last layer are done using ReLU. | Exp 1:   * 16 Frames * 160x160 sized Image * Batch size = 40 * Epoch size = 25 * No Augmentation | [98%, 77%] | [0.07, 0.6] | * Accuracy difference between training and validation is almost 20% (98 - 77) * Validation accuracy is not consistently improving although training accuracy reached maximum * Thus, this is clearly overfitting * Thus, this model will not be considered |
| **Model 1 : Basic 3D Model**   |  | | --- | | Conv3D 16 Features, 3x3 Filter | | Conv3D 32 Features, 3x3 Filter | | Conv3D 64 Features, 3x3 Filter | | Conv3D 128 Features, 3x3 Filter | | Flatten | | FC Dense w/ 128 neurons, 0.5 dropout | | FC Dense w/ 64 neurons, 0.25 dropout | | Softmax Layer |  * Each Convolution layer is followed by a batch normalization and 3D max pooling. * Each fully connected layer is followed by a batch normalization. * All activations except last layer are done using ReLU. | Exp 2:   * 16 Frames * 120x120 sized Image * Batch size = 40 * Epoch size = 20 * Level 1 Augmentation | [72%, 55%] | [0.69, 1.23] | * There is no drastic increase in training accuracy. Always below 75% * The validation accuracy has increased very slow and is always below 60%. * Validation loss is not consistent when compared to training loss. * The scores are low. Thus, this model is not considered. |
| **Model 1 : Basic 3D Model**   |  | | --- | | Conv3D 16 Features, 3x3 Filter | | Conv3D 32 Features, 3x3 Filter | | Conv3D 64 Features, 3x3 Filter | | Conv3D 128 Features, 3x3 Filter | | Flatten | | FC Dense w/ 128 neurons, 0.5 dropout | | FC Dense w/ 64 neurons, 0.25 dropout | | Softmax Layer |  * Each Convolution layer is followed by a batch normalization and 3D max pooling. * Each fully connected layer is followed by a batch normalization. * All activations except last layer are done using ReLU. | Exp 3:   * 16 Frames * 120x120 sized Image * Batch size = 40 * Epoch size = 20 * Level 2 Augmentation | [77%, 51%] | [0.61, 1.20] | * There is no drastic increase in training accuracy and is always below 80%. * The validation accuracy has increased very slow and is always below 50% and flat after epoch 12. * Validation loss is not consistent when compared to training loss. * Thus, we will not consider augmentation level 2 going forward with other models. |
| **Model 2 : 3D Model with multiple parameters**   |  | | --- | | Conv3D 16 Features, 2x2 Filter | | Conv3D 32 Features, 2x2 Filter | | Conv3D 64 Features, 2x2 Filter | | Conv3D 128 Features, 2x2 Filter | | Flatten | | FC Dense layer and dropout | | FC Dense layer and dropout | | Softmax Layer |  * Each Convolution layer is followed by a batch normalization and 3D max pooling. * Each fully connected layer is followed by a batch normalization. * All activations except last layer are done using ReLU. | Exp 4:   * 16 Frames * 180x180 sized Image * Batch size = 30 * Epoch size = 25 * No Augmentation * Learning Rate 0.0002 * 256 Dense Neurons * 0.5 Dropout | [96%, 81%] | [0.11, 0.54] | * Accuracy difference between training and validation is 15% (96 - 81) * Training and validation accuracy are both stable after 10 epochs. * Validation loss and training loss are also stable after 10 epochs. * This model is a **good** candidate. |
| **Model 2 : 3D Model with multiple parameters**   |  | | --- | | Conv3D 16 Features, 2x2 Filter | | Conv3D 32 Features, 2x2 Filter | | Conv3D 64 Features, 2x2 Filter | | Conv3D 128 Features, 2x2 Filter | | Flatten | | FC Dense layer and dropout | | FC Dense layer and dropout | | Softmax Layer |  * Each Convolution layer is followed by a batch normalization and 3D max pooling. * Each fully connected layer is followed by a batch normalization.   All activations except last layer are done using ReLU. | Exp 5:   * 16 Frames * 120x120 sized Image * Batch size = 30 * Epoch size = 25 * Level 1 Augmentation * Learning Rate 0.0002 * 256 Dense Neurons * 0.5 Dropout | [73%, 64%] | [0.68, 0.91] | * Training accuracy is always below 75%. It is also not consistently improving. * Validation accuracy is not consistent and after 10 epochs, the accuracy is not improving. * Training loss is decaying fast although validation loss is not decreasing and flat after 10 epochs. |
| **Model 3 : 3D Model with Reduced Parameters**   |  | | --- | | Conv3D 16 Features, 3x3 Filter | | Conv3D 32 Features, 2x2 Filter | | Conv3D 64 Features, 2x2 Filter | | Conv3D 128 Features, 2x2 Filter | | Flatten | | FC Dense layer and dropout | | FC Dense layer and dropout | | Softmax Layer |  * Each Convolution layer is followed by a batch normalization and 3D max pooling. * Each fully connected layer is followed by a batch normalization. * All activations except last layer are done using ReLU. | Exp 6:   * 16 Frames * 160x160 sized Image * Batch size = 20 * Epoch size = 25 * No Augmentation * Learning Rate 0.0002 * 128 Dense Neurons * 0.25 Dropout | [99%, 80%] | [0.03, 0.59] | * Accuracy difference between training dataset and validation dataset is high at 20% (99.85 - 80) * The model has learned all features from training dataset by end of 6th epoch yet, validation accuracy is not improving consistently. * This model is overfitting on training data. Thus, this model will not be considered. |
| **Model 3 : 3D Model with Reduced Parameters**   |  | | --- | | Conv3D 16 Features, 3x3 Filter | | Conv3D 32 Features, 2x2 Filter | | Conv3D 64 Features, 2x2 Filter | | Conv3D 128 Features, 2x2 Filter | | Flatten | | FC Dense layer and dropout | | FC Dense layer and dropout | | Softmax Layer |  * Each Convolution layer is followed by a batch normalization and 3D max pooling. * Each fully connected layer is followed by a batch normalization. * All activations except last layer are done using ReLU. | Exp 7:   * 16 Frames * 120x120 sized Image * Batch size = 20 * Epoch size = 25 * Level 1 Augmentation * Learning Rate 0.0002 * 128 Dense Neurons * 0.25 Dropout | [89%, 64%] | [0.31, 0.97] | * Validation accuracy is very low and never crossed 70%. * Although, training accuracy is increasing and reaching 90%, validation accuracy is almost flat after 5th epoch. * This model is overfitting and thus this model will not be considered. |
| **Model 4 : 3D Model with Reduced Parameters and Changing Filter sizes**   |  | | --- | | Conv3D 16 Features, 3x3 Filter | | Conv3D 32 Features, 3x3 Filter | | Conv3D 64 Features, 2x2 Filter | | Conv3D 128 Features, 2x2 Filter | | Flatten | | FC Dense layer and dropout | | FC Dense layer and dropout | | Softmax Layer |  * Each Convolution layer is followed by a batch normalization and 3D max pooling. * Each fully connected layer is followed by a batch normalization. * All activations except last layer are done using ReLU. | Exp 8:   * 16 Frames * 160x160 sized Image * Batch size = 20 * Epoch size = 25 * No Augmentation * Learning Rate 0.0002 * 64 Dense Neurons * 0.25 Dropout | [99%, 71%] | [0.04, 0.63] | * Accuracy difference between training and validation is more than 20%. (99.56 - 77) * Model has learnt all features about training data and its validation accuracy is flat after 5th epoch. * This model is overfitting and thus, this model will not be considered. |
| **Model 4 : 3D Model with Reduced Parameters and Changing Filter sizes**   |  | | --- | | Conv3D 16 Features, 3x3 Filter | | Conv3D 32 Features, 3x3 Filter | | Conv3D 64 Features, 2x2 Filter | | Conv3D 128 Features, 2x2 Filter | | Flatten | | FC Dense layer and dropout | | FC Dense layer and dropout | | Softmax Layer |  * Each Convolution layer is followed by a batch normalization and 3D max pooling. * Each fully connected layer is followed by a batch normalization.   All activations except last layer are done using ReLU. | Exp 9:   * 16 Frames * 120x120 sized Image * Batch size = 20 * Epoch size = 25 * Level 1 Augmentation * Learning Rate 0.0002 * 64 Dense Neurons * 0.25 Dropout | [79%, 63%] | [0.56, 0.97] | * Training Accuracy is always below 80% * Validation accuracy is always below 65% * After 10 epochs, Validation accuracy and Validation loss are almost flat and not improving. * This model is not recommended |
| **Model 5 : 3D Model with Multiple Convolution Layers**   |  | | --- | | 2 X Conv3D 16 Features, 3x3 Filter | | 2 X Conv3D 32 Features, 3x3 Filter | | 2 X Conv3D 64 Features, 3x3 Filter | | 2 X Conv3D 128 Features, 3x3 Filter | | Flatten | | FC Dense layer and dropout | | FC Dense layer and dropout | | Softmax Layer |  * Each Convolution layer is followed by a batch normalization and 3D max pooling. * Each fully connected layer is followed by a batch normalization. * All activations except last layer are done using ReLU. | Exp 10:   * 16 Frames * 160x160 sized Image * Batch size = 20 * Epoch size = 25 * No Augmentation * 256 Dense Neurons * 0.5 Dropout | [95%, 80%] | [0.13, 0.67] | * Final accuracies are good yet validation accuracy is not stable. * Validation loss is also not stable. * This model will not be considered due to unstable behaviour. |
| **Model 5 : 3D Model with Multiple Convolution Layers**   |  | | --- | | 2 X Conv3D 16 Features, 3x3 Filter | | 2 X Conv3D 32 Features, 3x3 Filter | | 2 X Conv3D 64 Features, 3x3 Filter | | 2 X Conv3D 128 Features, 3x3 Filter | | Flatten | | FC Dense layer and dropout | | FC Dense layer and dropout | | Softmax Layer |  * Each Convolution layer is followed by a batch normalization and 3D max pooling. * Each fully connected layer is followed by a batch normalization. * All activations except last layer are done using ReLU. | Exp 11:   * 16 Frames * 120x120 sized Image * Batch size = 20 * Epoch size = 25 * Level 1 Augmentation * 256 Dense Neurons * 0.5 Dropout | [72%, 64%] | [0.67, 0.94] | * Training accuracy is always below 75%. * Validation accuracy is very low at 60% * Validation accuracy and loss are unstable. * This model will not be considered. |
| **Model 6 : 2D Convolution model with LSTM cells**     |  | | --- | | Conv2D 16 Features, 3x3 Filter | | Conv2D 32 Features, 3x3 Filter | | Conv2D 64 Features, 3x3 Filter | | Conv2D 128 Features, 3x3 Filter | | Conv2D 256 Features, 3x3 Filter | | Flatten | | LSTM with dropout | | FC Dense layer with dropout | | Softmax Layer |  * Each Convolution layer is followed by a batch normalization and 2D max pooling * All activations except last layer are done using ReLU. | Exp 12:   * 18 Frames * 120x120 sized Image * Batch size = 20 * Epoch size = 25 * No Augmentation * 256 LSTM cells * 256 Dense Neurons * 0.5 Dropout | [93%, 75%] | [0.21, 0.69] | * Training accuracy is increasing after 10 epochs although validation accuracy remained flat. * Training loss is decreasing continuously although validation loss remained flat after 10 epochs. * This model will not be considered. |
| **Model 6 : 2D Convolution model with LSTM cells**     |  | | --- | | Conv2D 16 Features, 3x3 Filter | | Conv2D 32 Features, 3x3 Filter | | Conv2D 64 Features, 3x3 Filter | | Conv2D 128 Features, 3x3 Filter | | Conv2D 256 Features, 3x3 Filter | | Flatten | | LSTM with dropout | | FC Dense layer with dropout | | Softmax Layer |  * Each Convolution layer is followed by a batch normalization and 2D max pooling * All activations except last layer are done using ReLU. | Exp 13:   * 18 Frames * 120x120 sized Image * Batch size = 20 * Epoch size = 25 * Level 1 Augmentation * 128 LSTM cells * 128 Dense Neurons * 0.25 Dropout | [79%, 63%] | [0.57, 0.98] | * Validation accuracy and loss are unstable * Validation accuracy and loss tend to flat after 15 epochs * Training Accuracy and Validation accuracy are low (below 80% and 60%) * This model will not be considered. |
| **Model 7 : 2D Convolution model with GRU cells**     |  | | --- | | Conv2D 16 Features, 3x3 Filter | | Conv2D 32 Features, 3x3 Filter | | Conv2D 64 Features, 3x3 Filter | | Conv2D 128 Features, 3x3 Filter | | Conv2D 256 Features, 3x3 Filter | | Flatten | | GRU with dropout | | FC Dense layer with dropout | | Softmax Layer |  * Each Convolution layer is followed by a batch normalization and 2D max pooling * All activations except last layer are done using ReLU. | Exp 14:   * 18 Frames * 120x120 sized Image * Batch size = 20 * Epoch size = 25 * No Augmentation * 256 GRU cells * 256 Dense Neurons * 0.5 Dropout | [96%, 81%] | [0.12, 0.58] | * Training accuracy and validation accuracy are increasing consistently * Difference between training and validation accuracies is ~15% (96% - 81%) * This model is a good candidate. * This model may be improved by further tuning hyper parameters |
| **Model 7 : 2D Convolution model with LSTM cells**     |  | | --- | | Conv2D 16 Features, 3x3 Filter | | Conv2D 32 Features, 3x3 Filter | | Conv2D 64 Features, 3x3 Filter | | Conv2D 128 Features, 3x3 Filter | | Conv2D 256 Features, 3x3 Filter | | Flatten | | GRU with dropout | | FC Dense layer with dropout | | Softmax Layer |  * Each Convolution layer is followed by a batch normalization and 2D max pooling * All activations except last layer are done using ReLU. | Exp 15:   * 18 Frames * 120x120 sized Image * Batch size = 20 * Epoch size = 25 * Level 1 Augmentation * 128 GRU cells * 128 Dense Neurons   0.25 Dropout | [80%, 67%] | [0.48, 0.89] | * Validation accuracy and loss tend to flat after 10 epochs * Training Accuracy and Validation accuracy are low (below 80% and 65%) * This model will not be considered. |
| **Model 8 – 2D Transfer Learning with LSTM cells**   |  | | --- | | Mobilenet Model with imagenet weights and collapse top layer | | LSTM with dropout | | FC Dense layer with dropout | | Softmax Layer |  * Mobilenet transfer model is followed by batch normalization and 2D Max pooling. * All activations except last layer are done using ReLU. | Exp 16:   * 18 Frames * 160x160 sized Image * Batch size = 5 * Epoch size = 20 * No Augmentation * 128 LSTM cells * 128 Dense Neurons * 0.25 Dropout | [98%, 65%] | [0.05, 1.14] | * Validation accuracy is not improving although training accuracy has improved. * Validation loss and accuracy are unstable. * Validation accuracy is low (~65%) to be considered. * This model will not be considered. |
| **Model 9 – 2D Transfer Learning with GRU cells**   |  | | --- | | Mobilenet Model with imagenet weights and collapse top layer | | GRU with dropout | | FC Dense layer with dropout | | Softmax Layer |  * Mobilenet transfer model is followed by batch normalization and 2D Max pooling. * All activations except last layer are done using ReLU. | Exp 17:   * 18 Frames * 160x160 sized Image * Batch size = 5 * Epoch size = 20 * No Augmentation * 128 GRU cells * 128 Dense Neurons * 0.25 Dropout | [99%, 89%] | [0.01, 0.30] | * Both training accuracy and validation accuracy are stable. * Validation accuracy is very good at 89%. * Validation loss is also stable. * This model will be considered. |