

Gesture Recognition Deep Learning Model Write up

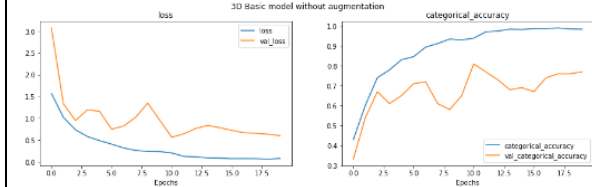
Authors:

Sidhartha Sankar Mahapatro

S S B Phani Pradeep Miriyala

- 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. Cropping the images to have the targeted gestures and remove the noise from the images.
 - Level 2 augmentation: Augmentation with warp affine transformation about 10degrees and rotation of 10degrees.

Table 1 List of models and their metrics

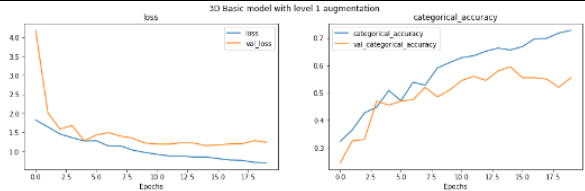
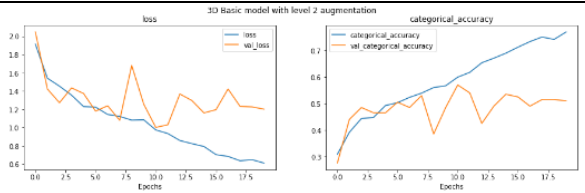
Model Architecture	Parameters	Accuracy [Training, Validation]	Loss [Training, Validation]	Comments								
<div>Model 1 : Basic 3D Model<table><tr><td>Conv3D 16 Features, 3x3 Filter</td></tr><tr><td>Conv3D 32 Features, 3x3 Filter</td></tr><tr><td>Conv3D 64 Features, 3x3 Filter</td></tr><tr><td>Conv3D 128 Features, 3x3 Filter</td></tr><tr><td>Flatten</td></tr><tr><td>FC Dense w/ 128 neurons, 0.5 dropout</td></tr><tr><td>FC Dense w/ 64 neurons, 0.25 dropout</td></tr><tr><td>Softmax Layer</td></tr></table><ul style="list-style-type: none">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.</div>	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	<div>Exp 1:<ul style="list-style-type: none">16 Frames160x160 sized ImageBatch size = 40Epoch size = 25No Augmentation</div>	[98%, 77%]	[0.07, 0.6]	<div></div> <ul style="list-style-type: none">Accuracy difference between training and validation is almost 20% (98 - 77)Validation accuracy is not consistently improving although training accuracy reached maximumThus, this is clearly overfittingThus, this model will not be considered
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												

Gesture Recognition Deep Learning Model Write up

Authors:

Sidhartha Sankar Mahapatro

S S B Phani Pradeep Miriyala

Model Architecture	Parameters	Accuracy [Training, Validation]	Loss [Training, Validation]	Comments								
<div>Model 1 : Basic 3D Model</div> <table><tr><td>Conv3D 16 Features, 3x3 Filter</td></tr><tr><td>Conv3D 32 Features, 3x3 Filter</td></tr><tr><td>Conv3D 64 Features, 3x3 Filter</td></tr><tr><td>Conv3D 128 Features, 3x3 Filter</td></tr><tr><td>Flatten</td></tr><tr><td>FC Dense w/ 128 neurons, 0.5 dropout</td></tr><tr><td>FC Dense w/ 64 neurons, 0.25 dropout</td></tr><tr><td>Softmax Layer</td></tr></table> <div><ul style="list-style-type: none">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.</div>	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	<div>Exp 2:</div> <div><ul style="list-style-type: none">16 Frames120x120 sized ImageBatch size = 40Epoch size = 20Level 1 Augmentation</div>	[72%, 55%]	[0.69, 1.23]	<div></div> <div><ul style="list-style-type: none">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.</div>
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												
<div>Model 1 : Basic 3D Model</div> <table><tr><td>Conv3D 16 Features, 3x3 Filter</td></tr><tr><td>Conv3D 32 Features, 3x3 Filter</td></tr><tr><td>Conv3D 64 Features, 3x3 Filter</td></tr><tr><td>Conv3D 128 Features, 3x3 Filter</td></tr><tr><td>Flatten</td></tr><tr><td>FC Dense w/ 128 neurons, 0.5 dropout</td></tr><tr><td>FC Dense w/ 64 neurons, 0.25 dropout</td></tr></table>	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	<div>Exp 3:</div> <div><ul style="list-style-type: none">16 Frames120x120 sized ImageBatch size = 40Epoch size = 20Level 2 Augmentation</div>	[77%, 51%]	[0.61, 1.20]	<div></div> <div><ul style="list-style-type: none">There is no drastic increase in training accuracy and is always below 80%.</div>	
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												

Gesture Recognition Deep Learning Model Write up

Authors:

Sidhartha Sankar Mahapatro

S S B Phani Pradeep Miriyala

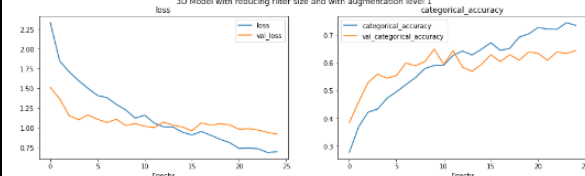
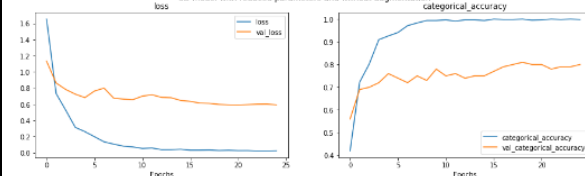
Model Architecture	Parameters	Accuracy [Training, Validation]	Loss [Training, Validation]	Comments
<div>Softmax Layer</div> <ul style="list-style-type: none"> Each Convolution layer is followed by a batch normalization and 3D max pooling. Each fully connected layer is followed by a batch normalization. All activations except last layer are done using ReLU. 				<ul style="list-style-type: none"> 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 <div> <div>Conv3D 16 Features, 2x2 Filter</div> <div>Conv3D 32 Features, 2x2 Filter</div> <div>Conv3D 64 Features, 2x2 Filter</div> <div>Conv3D 128 Features, 2x2 Filter</div> <div>Flatten</div> <div>FC Dense layer and dropout</div> <div>FC Dense layer and dropout</div> <div>Softmax Layer</div> </div> <ul style="list-style-type: none"> 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: <ul style="list-style-type: none"> 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]	<div> </div> <ul style="list-style-type: none"> 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.

Gesture Recognition Deep Learning Model Write up

Authors:

Sidhartha Sankar Mahapatro

S S B Phani Pradeep Miriyala

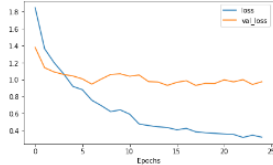
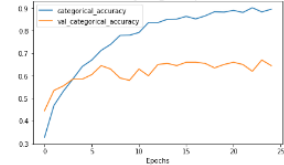
Model Architecture	Parameters	Accuracy [Training, Validation]	Loss [Training, Validation]	Comments								
Model 2 : 3D Model with multiple parameters <table><tr><td>Conv3D 16 Features, 2x2 Filter</td></tr><tr><td>Conv3D 32 Features, 2x2 Filter</td></tr><tr><td>Conv3D 64 Features, 2x2 Filter</td></tr><tr><td>Conv3D 128 Features, 2x2 Filter</td></tr><tr><td>Flatten</td></tr><tr><td>FC Dense layer and dropout</td></tr><tr><td>FC Dense layer and dropout</td></tr><tr><td>Softmax Layer</td></tr></table> <ul style="list-style-type: none">Each Convolution layer is followed by a batch normalization and 3D max pooling.Each fully connected layer is followed by a batch normalization. <p>All activations except last layer are done using ReLU.</p>	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	Exp 5: <ul style="list-style-type: none">16 Frames120x120 sized ImageBatch size = 30Epoch size = 25Level 1 AugmentationLearning Rate 0.0002256 Dense Neurons0.5 Dropout	[73%, 64%]	[0.68, 0.91]	<div></div> <ul style="list-style-type: none">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.
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												
Model 3 : 3D Model with Reduced Parameters <table><tr><td>Conv3D 16 Features, 3x3 Filter</td></tr><tr><td>Conv3D 32 Features, 2x2 Filter</td></tr><tr><td>Conv3D 64 Features, 2x2 Filter</td></tr><tr><td>Conv3D 128 Features, 2x2 Filter</td></tr><tr><td>Flatten</td></tr><tr><td>FC Dense layer and dropout</td></tr><tr><td>FC Dense layer and dropout</td></tr><tr><td>Softmax Layer</td></tr></table>	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	Exp 6: <ul style="list-style-type: none">16 Frames160x160 sized ImageBatch size = 20Epoch size = 25No AugmentationLearning Rate 0.0002128 Dense Neurons	[99%, 80%]	[0.03, 0.59]	<div></div> <ul style="list-style-type: none">Accuracy difference between training dataset and validation dataset is high at 20% (99.85 - 80)
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												

Gesture Recognition Deep Learning Model Write up

Authors:

Sidhartha Sankar Mahapatro

S S B Phani Pradeep Miriyala

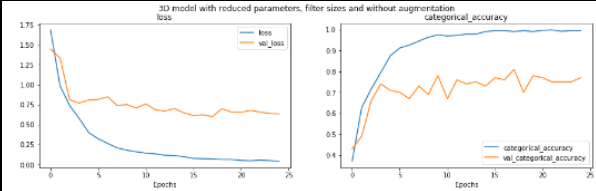
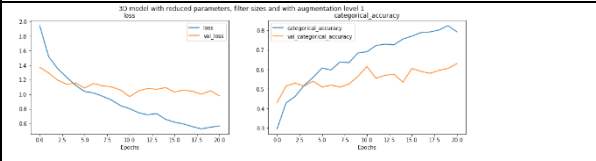
Model Architecture	Parameters	Accuracy [Training, Validation]	Loss [Training, Validation]	Comments								
<ul style="list-style-type: none">Each Convolution layer is followed by a batch normalization and 3D max pooling.Each fully connected layer is followed by a batch normalization.All activations except last layer are done using ReLU.	<ul style="list-style-type: none">0.25 Dropout			<ul style="list-style-type: none">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 <table border="1"><tr><td>Conv3D 16 Features, 3x3 Filter</td></tr><tr><td>Conv3D 32 Features, 2x2 Filter</td></tr><tr><td>Conv3D 64 Features, 2x2 Filter</td></tr><tr><td>Conv3D 128 Features, 2x2 Filter</td></tr><tr><td>Flatten</td></tr><tr><td>FC Dense layer and dropout</td></tr><tr><td>FC Dense layer and dropout</td></tr><tr><td>Softmax Layer</td></tr></table> <ul style="list-style-type: none">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.	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	Exp 7: <ul style="list-style-type: none">16 Frames120x120 sized ImageBatch size = 20Epoch size = 25Level 1 AugmentationLearning Rate 0.0002128 Dense Neurons0.25 Dropout	[89%, 64%]	[0.31, 0.97]	<div><div>3D model with reduced parameters and with level 1 augmentation</div><div><div>loss</div><div>categorical_accuracy</div></div><ul style="list-style-type: none">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.</div>
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												

Gesture Recognition Deep Learning Model Write up

Authors:

Sidhartha Sankar Mahapatro

S S B Phani Pradeep Miriyala

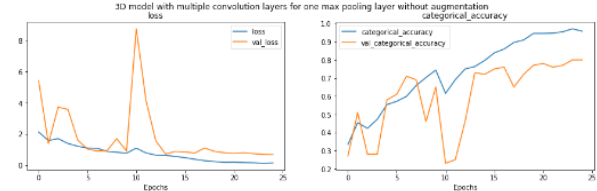
Model Architecture	Parameters	Accuracy [Training, Validation]	Loss [Training, Validation]	Comments								
<div>Model 4 : 3D Model with Reduced Parameters and Changing Filter sizes</div> <table><tr><td>Conv3D 16 Features, 3x3 Filter</td></tr><tr><td>Conv3D 32 Features, 3x3 Filter</td></tr><tr><td>Conv3D 64 Features, 2x2 Filter</td></tr><tr><td>Conv3D 128 Features, 2x2 Filter</td></tr><tr><td>Flatten</td></tr><tr><td>FC Dense layer and dropout</td></tr><tr><td>FC Dense layer and dropout</td></tr><tr><td>Softmax Layer</td></tr></table> <div><ul style="list-style-type: none">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.</div>	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	<div>Exp 8:</div> <div><ul style="list-style-type: none">16 Frames160x160 sized ImageBatch size = 20Epoch size = 25No AugmentationLearning Rate 0.000264 Dense Neurons0.25 Dropout</div>	[99%, 71%]	[0.04, 0.63]	<div></div> <div><ul style="list-style-type: none">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.</div>
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												
<div>Model 4 : 3D Model with Reduced Parameters and Changing Filter sizes</div> <table><tr><td>Conv3D 16 Features, 3x3 Filter</td></tr><tr><td>Conv3D 32 Features, 3x3 Filter</td></tr><tr><td>Conv3D 64 Features, 2x2 Filter</td></tr><tr><td>Conv3D 128 Features, 2x2 Filter</td></tr><tr><td>Flatten</td></tr><tr><td>FC Dense layer and dropout</td></tr></table>	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	<div>Exp 9:</div> <div><ul style="list-style-type: none">16 Frames120x120 sized ImageBatch size = 20Epoch size = 25Level 1 Augmentation</div>	[79%, 63%]	[0.56, 0.97]	<div></div> <div><ul style="list-style-type: none">Training Accuracy is always below 80%Validation accuracy is always below 65%</div>		
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												

Gesture Recognition Deep Learning Model Write up

Authors:

Sidhartha Sankar Mahapatro

S S B Phani Pradeep Miriyala

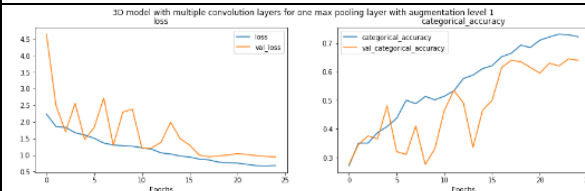
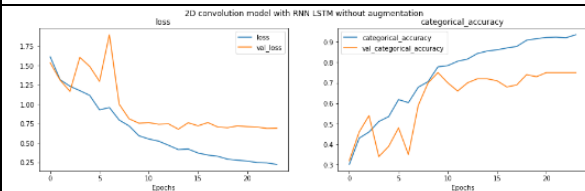
Model Architecture	Parameters	Accuracy [Training, Validation]	Loss [Training, Validation]	Comments
<div>FC Dense layer and dropout</div> <div>Softmax Layer</div> <ul style="list-style-type: none"> Each Convolution layer is followed by a batch normalization and 3D max pooling. Each fully connected layer is followed by a batch normalization. <p>All activations except last layer are done using ReLU.</p>	<ul style="list-style-type: none"> Learning Rate 0.0002 64 Dense Neurons 0.25 Dropout 			<ul style="list-style-type: none"> 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 <div>2 X Conv3D 16 Features, 3x3 Filter</div> <div>2 X Conv3D 32 Features, 3x3 Filter</div> <div>2 X Conv3D 64 Features, 3x3 Filter</div> <div>2 X Conv3D 128 Features, 3x3 Filter</div> <div>Flatten</div> <div>FC Dense layer and dropout</div> <div>FC Dense layer and dropout</div> <div>Softmax Layer</div> <ul style="list-style-type: none"> Each Convolution layer is followed by a batch normalization and 3D max pooling. Each fully connected layer is followed by a batch normalization. 	<p>Exp 10:</p> <ul style="list-style-type: none"> 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]	 <ul style="list-style-type: none"> 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.

Gesture Recognition Deep Learning Model Write up

Authors:

Sidhartha Sankar Mahapatro

S S B Phani Pradeep Miriyala

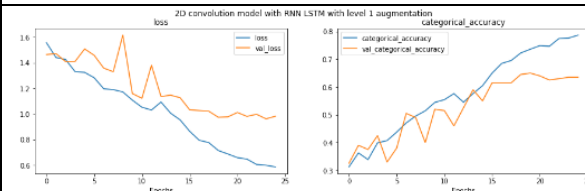
Model Architecture	Parameters	Accuracy [Training, Validation]	Loss [Training, Validation]	Comments								
<ul style="list-style-type: none">All activations except last layer are done using ReLU.												
<div>Model 5 : 3D Model with Multiple Convolution Layers</div> <table><tr><td>2 X Conv3D 16 Features, 3x3 Filter</td></tr><tr><td>2 X Conv3D 32 Features, 3x3 Filter</td></tr><tr><td>2 X Conv3D 64 Features, 3x3 Filter</td></tr><tr><td>2 X Conv3D 128 Features, 3x3 Filter</td></tr><tr><td>Flatten</td></tr><tr><td>FC Dense layer and dropout</td></tr><tr><td>FC Dense layer and dropout</td></tr><tr><td>Softmax Layer</td></tr></table> <ul style="list-style-type: none">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.	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	<div>Exp 11:</div> <ul style="list-style-type: none">16 Frames120x120 sized ImageBatch size = 20Epoch size = 25Level 1 Augmentation256 Dense Neurons0.5 Dropout	[72%, 64%]	[0.67, 0.94]	<div><div>3D model with multiple convolution layers for one max pooling layer with augmentation level 1</div><div></div></div> <ul style="list-style-type: none">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.
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												
<div>Model 6 : 2D Convolution model with LSTM cells</div> <table><tr><td>Conv2D 16 Features, 3x3 Filter</td></tr><tr><td>Conv2D 32 Features, 3x3 Filter</td></tr><tr><td>Conv2D 64 Features, 3x3 Filter</td></tr></table>	Conv2D 16 Features, 3x3 Filter	Conv2D 32 Features, 3x3 Filter	Conv2D 64 Features, 3x3 Filter	<div>Exp 12:</div> <ul style="list-style-type: none">18 Frames120x120 sized ImageBatch size = 20Epoch size = 25	[93%, 75%]	[0.21, 0.69]	<div><div>2D convolution model with RNN : LSTM without augmentation</div><div></div></div>					
Conv2D 16 Features, 3x3 Filter												
Conv2D 32 Features, 3x3 Filter												
Conv2D 64 Features, 3x3 Filter												

Gesture Recognition Deep Learning Model Write up

Authors:

Sidhartha Sankar Mahapatro

S S B Phani Pradeep Miriyala

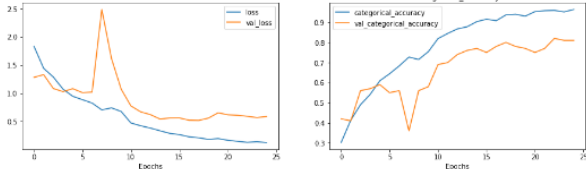
Model Architecture		Parameters	Accuracy [Training, Validation]	Loss [Training, Validation]	Comments
<div><div>Conv2D 128 Features, 3x3 Filter</div><div>Conv2D 256 Features, 3x3 Filter</div><div>Flatten</div><div>LSTM with dropout</div><div>FC Dense layer with dropout</div><div>Softmax Layer</div></div> <div><ul style="list-style-type: none">Each Convolution layer is followed by a batch normalization and 2D max poolingAll activations except last layer are done using ReLU.</div>	<ul style="list-style-type: none">No Augmentation256 LSTM cells256 Dense Neurons0.5 Dropout			<ul style="list-style-type: none">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.	
<div><div>Conv2D 16 Features, 3x3 Filter</div><div>Conv2D 32 Features, 3x3 Filter</div><div>Conv2D 64 Features, 3x3 Filter</div><div>Conv2D 128 Features, 3x3 Filter</div><div>Conv2D 256 Features, 3x3 Filter</div><div>Flatten</div><div>LSTM with dropout</div><div>FC Dense layer with dropout</div><div>Softmax Layer</div></div>	<div>Exp 13:</div> <ul style="list-style-type: none">18 Frames120x120 sized ImageBatch size = 20Epoch size = 25Level 1 Augmentation128 LSTM cells128 Dense Neurons0.25 Dropout	[79%, 63%]	[0.57, 0.98]	<div><div><div>2D convolution model with RNN LSTM with level 1 augmentation</div><div></div></div><div><ul style="list-style-type: none">Validation accuracy and loss are unstableValidation accuracy and loss tend to flat after 15 epochsTraining Accuracy and Validation accuracy are low (below 80% and 60%)This model will not be considered.</div></div>	

Gesture Recognition Deep Learning Model Write up

Authors:

Sidhartha Sankar Mahapatro

S S B Phani Pradeep Miriyala

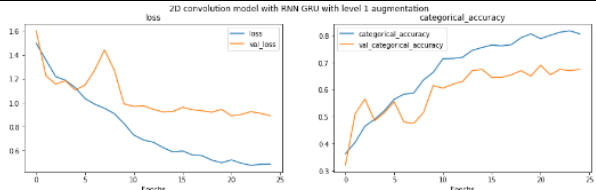
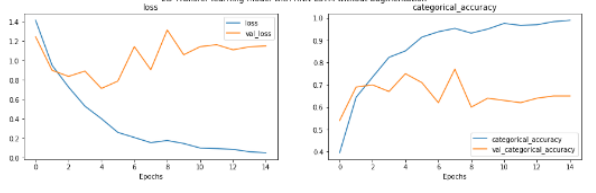
Model Architecture	Parameters	Accuracy [Training, Validation]	Loss [Training, Validation]	Comments									
<ul style="list-style-type: none">Each Convolution layer is followed by a batch normalization and 2D max poolingAll activations except last layer are done using ReLU.													
<div><div>Model 7 : 2D Convolution model with GRU cells</div><table><tr><td>Conv2D 16 Features, 3x3 Filter</td></tr><tr><td>Conv2D 32 Features, 3x3 Filter</td></tr><tr><td>Conv2D 64 Features, 3x3 Filter</td></tr><tr><td>Conv2D 128 Features, 3x3 Filter</td></tr><tr><td>Conv2D 256 Features, 3x3 Filter</td></tr><tr><td>Flatten</td></tr><tr><td>GRU with dropout</td></tr><tr><td>FC Dense layer with dropout</td></tr><tr><td>Softmax Layer</td></tr></table><ul style="list-style-type: none">Each Convolution layer is followed by a batch normalization and 2D max poolingAll activations except last layer are done using ReLU.</div>	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	<div>Exp 14:<ul style="list-style-type: none">18 Frames120x120 sized ImageBatch size = 20Epoch size = 25No Augmentation256 GRU cells256 Dense Neurons0.5 Dropout</div>	[96%, 81%]	[0.12, 0.58]	<div><div><div>2D convolution model with RNN GRU without augmentation</div><div></div></div><ul style="list-style-type: none">Training accuracy and validation accuracy are increasing consistentlyDifference 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</div>
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													

Gesture Recognition Deep Learning Model Write up

Authors:

Sidhartha Sankar Mahapatro

S S B Phani Pradeep Miriyala

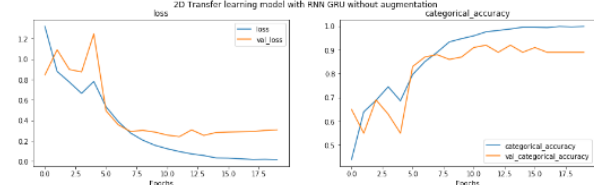
Model Architecture	Parameters	Accuracy [Training, Validation]	Loss [Training, Validation]	Comments									
<div>Model 7 : 2D Convolution model with LSTM cells</div> <div><table><tr><td>Conv2D 16 Features, 3x3 Filter</td></tr><tr><td>Conv2D 32 Features, 3x3 Filter</td></tr><tr><td>Conv2D 64 Features, 3x3 Filter</td></tr><tr><td>Conv2D 128 Features, 3x3 Filter</td></tr><tr><td>Conv2D 256 Features, 3x3 Filter</td></tr><tr><td>Flatten</td></tr><tr><td>GRU with dropout</td></tr><tr><td>FC Dense layer with dropout</td></tr><tr><td>Softmax Layer</td></tr></table><ul style="list-style-type: none">Each Convolution layer is followed by a batch normalization and 2D max poolingAll activations except last layer are done using ReLU.</div>	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	<div>Exp 15:</div> <ul style="list-style-type: none">18 Frames120x120 sized ImageBatch size = 20Epoch size = 25Level 1 Augmentation128 GRU cells128 Dense Neurons0.25 Dropout	[80%, 67%]	[0.48, 0.89]	<div></div> <ul style="list-style-type: none">Validation accuracy and loss tend to flat after 10 epochsTraining Accuracy and Validation accuracy are low (below 80% and 65%)This model will not be considered.
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													
<div>Model 8 – 2D Transfer Learning with LSTM cells</div> <div><table><tr><td>Mobilenet Model with imagenet weights and collapse top layer</td></tr><tr><td>LSTM with dropout</td></tr><tr><td>FC Dense layer with dropout</td></tr><tr><td>Softmax Layer</td></tr></table></div>	Mobilenet Model with imagenet weights and collapse top layer	LSTM with dropout	FC Dense layer with dropout	Softmax Layer	<div>Exp 16:</div> <ul style="list-style-type: none">18 Frames160x160 sized ImageBatch size = 5Epoch size = 20No Augmentation128 LSTM cells128 Dense Neurons	[98%, 65%]	[0.05, 1.14]	<div></div> <ul style="list-style-type: none">Validation accuracy is not improving although training accuracy has improved.Validation loss and accuracy are unstable.					
Mobilenet Model with imagenet weights and collapse top layer													
LSTM with dropout													
FC Dense layer with dropout													
Softmax Layer													

Gesture Recognition Deep Learning Model Write up

Authors:

Sidhartha Sankar Mahapatro

S S B Phani Pradeep Miriyala

Model Architecture	Parameters	Accuracy [Training, Validation]	Loss [Training, Validation]	Comments
<ul style="list-style-type: none">Mobilenet transfer model is followed by batch normalization and 2D Max pooling.All activations except last layer are done using ReLU.	<ul style="list-style-type: none">0.25 Dropout			<ul style="list-style-type: none">Validation accuracy is low (~65%) to be considered.This model will not be considered.
<div>Model 9 – 2D Transfer Learning with GRU cells</div> <div><div>Mobilenet Model with imagenet weights and collapse top layer</div><div>GRU with dropout</div><div>FC Dense layer with dropout</div><div>Softmax Layer</div></div> <ul style="list-style-type: none">Mobilenet transfer model is followed by batch normalization and 2D Max pooling.All activations except last layer are done using ReLU.	Exp 17: <ul style="list-style-type: none">18 Frames160x160 sized ImageBatch size = 5Epoch size = 20No Augmentation128 GRU cells128 Dense Neurons0.25 Dropout	[99%, 89%]	[0.01, 0.30]	<div></div> <ul style="list-style-type: none">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.