Computer Vision | Action Recognition Project

Description

This project is an identification of different actions from video clips (a sequence of 2D frames). It involves two main approaches of Single Stream Network.

Data Understanding

Classes

Data is classified into 5 classes {Diving, Jumping, Basketball, Tennis, Walking}.

Train Data

Contains 474 videos.

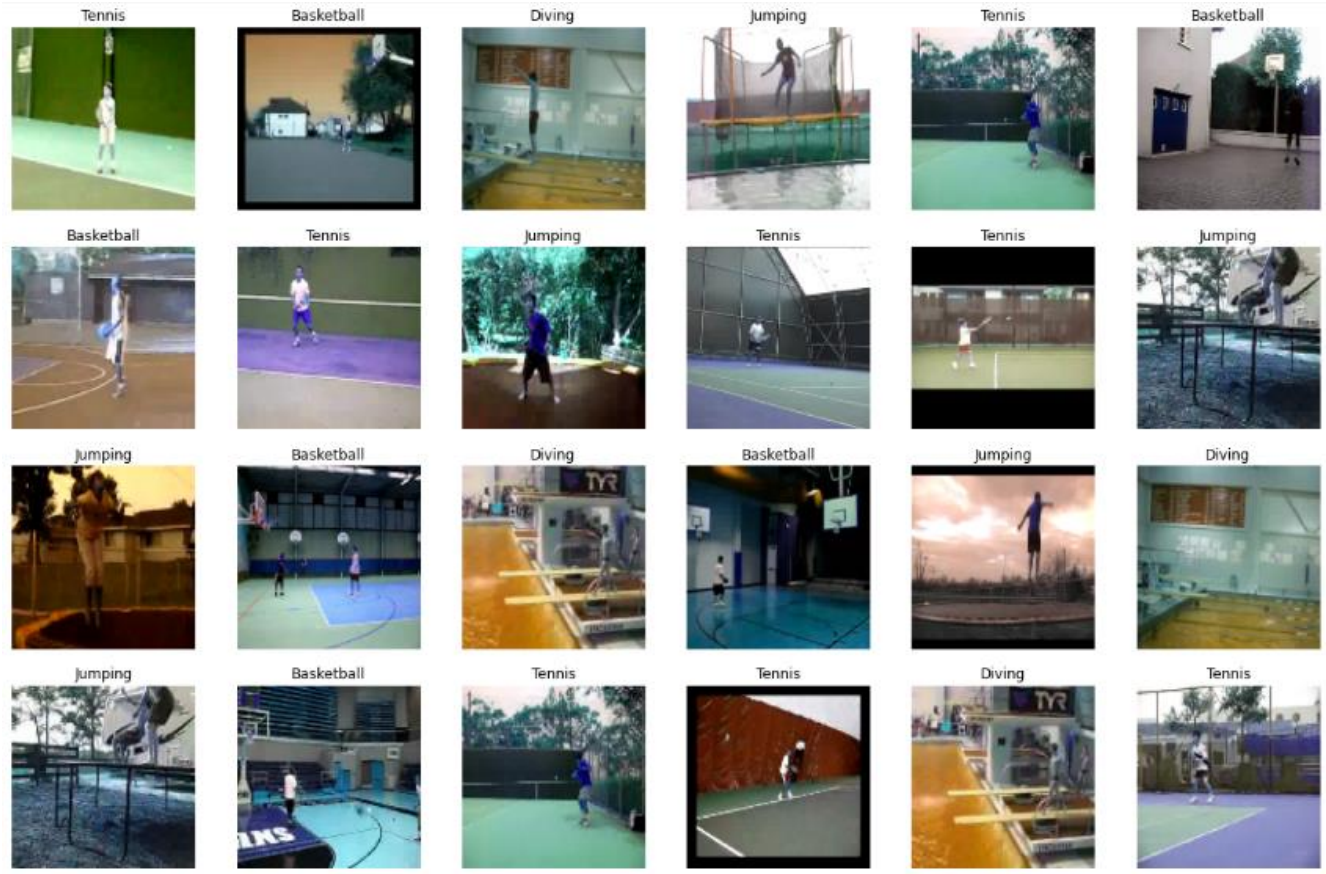
Divided into:

Class	Count
Diving	113
Jumping	100
Basketball	89
Tennis	105
Walking	67

Different Shapes of Videos:

Shape	Count
(320.0, 240.0, 239.0)	44
(320.0, 240.0, 201.0)	32
(320.0, 240.0, 151.0)	15
(320.0, 240.0, 238.0)	12
(320.0, 240.0, 105.0)	9
..	..
(320.0, 240.0, 163.0)	1
(320.0, 240.0, 310.0)	1
(320.0, 240.0, 88.0)	1
(320.0, 240.0, 401.0)	1
(320.0, 240.0, 71.0)	1

Samples of 24 random videos:



Test Data

Contains 126 videos.

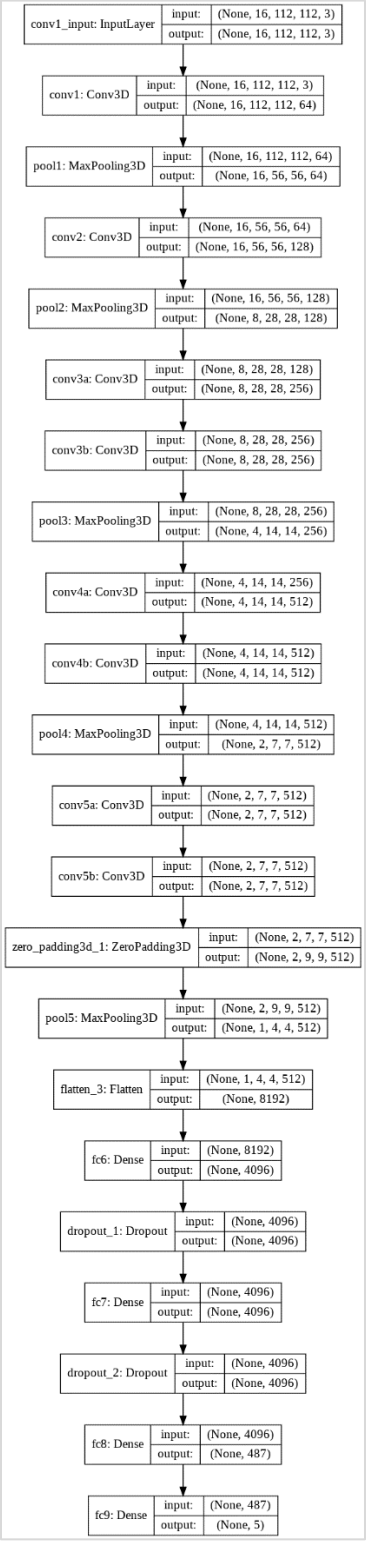
Different Shapes of Videos:

Shape	Count
(320.0, 240.0, 239.0)	11
(320.0, 240.0, 179.0)	4
(320.0, 240.0, 101.0)	4
(320.0, 240.0, 119.0)	4
(320.0, 240.0, 115.0)	3
..	
(320.0, 214.0, 177.0)	1
(320.0, 240.0, 176.0)	1
(320.0, 240.0, 201.0)	1
(320.0, 240.0, 130.0)	1
(320.0, 240.0, 87.0)	1

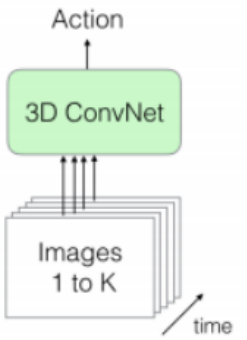
Deep Learning Algorithms

3D - CNN

A Single Stream Network - One Network for Spatial information.
Transfer Learning Model trained on Sports Data.
Input Shape: (16, 112, 112, 3)



Model Plot



Model Summary

Layer (type)	Output Shape	Param #
conv1_input (InputLayer)	(None, 16, 112, 112, 3)	0
conv1 (Conv3D)	(None, 16, 112, 112, 64)	5248
pool1 (MaxPooling3D)	(None, 16, 56, 56, 64)	0
conv2 (Conv3D)	(None, 16, 56, 56, 128)	221312
pool2 (MaxPooling3D)	(None, 8, 28, 28, 128)	0
conv3a (Conv3D)	(None, 8, 28, 28, 256)	884992
conv3b (Conv3D)	(None, 8, 28, 28, 256)	1769728
pool3 (MaxPooling3D)	(None, 4, 14, 14, 256)	0
conv4a (Conv3D)	(None, 4, 14, 14, 512)	3539456
conv4b (Conv3D)	(None, 4, 14, 14, 512)	7078400
pool4 (MaxPooling3D)	(None, 2, 7, 7, 512)	0
conv5a (Conv3D)	(None, 2, 7, 7, 512)	7078400
conv5b (Conv3D)	(None, 2, 7, 7, 512)	7078400
zero_padding3d_1 (ZeroPaddin	(None, 2, 9, 9, 512)	0
pool5 (MaxPooling3D)	(None, 1, 4, 4, 512)	0
flatten_3 (Flatten)	(None, 8192)	0
fc6 (Dense)	(None, 4096)	33558528
dropout_1 (Dropout)	(None, 4096)	0
fc7 (Dense)	(None, 4096)	16781312
dropout_2 (Dropout)	(None, 4096)	0
fc8 (Dense)	(None, 487)	1995239
fc9 (Dense)	(None, 5)	2440

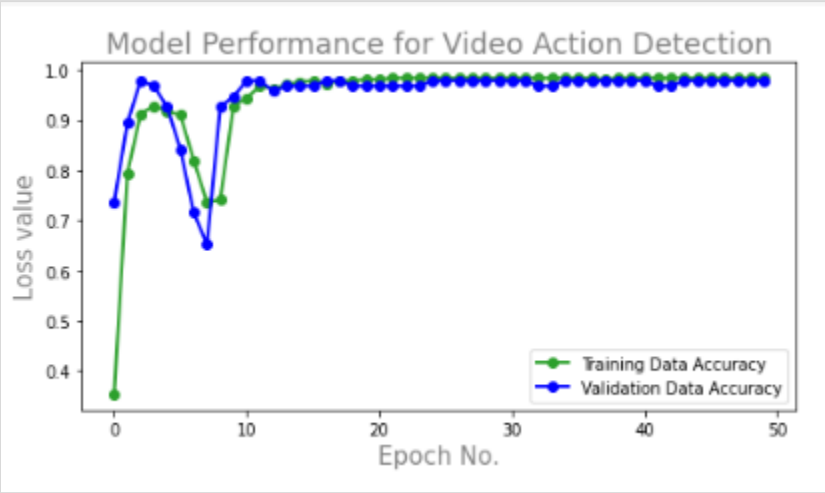
Total params: 79,993,455
Trainable params: 18,778,991
Non-trainable params: 61,214,464

Train and Validation Accuracy

Parameters:
batch_size = 50
no_epochs = 50
learning_rate = 0.0001
validation_split = 0.2
verbosity = 1

```
Train on 379 samples, validate on 95 samples.
```

```
Epoch 28/50
379/379 [==============================] - 18s 46ms/step - loss: 1.4968 - accuracy: 0.9842 - val_loss: 1.4995 - val_accuracy: 0.9789
Epoch 29/50
379/379 [==============================] - 17s 46ms/step - loss: 1.4954 - accuracy: 0.9842 - val_loss: 1.4981 - val_accuracy: 0.9789
Epoch 30/50
379/379 [==============================] - 17s 46ms/step - loss: 1.4940 - accuracy: 0.9842 - val_loss: 1.4968 - val_accuracy: 0.9789
Epoch 31/50
379/379 [==============================] - 18s 46ms/step - loss: 1.4927 - accuracy: 0.9842 - val_loss: 1.4955 - val_accuracy: 0.9789
Epoch 32/50
379/379 [==============================] - 18s 46ms/step - loss: 1.4913 - accuracy: 0.9842 - val_loss: 1.4943 - val_accuracy: 0.9789
Epoch 33/50
379/379 [==============================] - 18s 46ms/step - loss: 1.4900 - accuracy: 0.9842 - val_loss: 1.4930 - val_accuracy: 0.9684
Epoch 34/50
379/379 [==============================] - 17s 46ms/step - loss: 1.4887 - accuracy: 0.9842 - val_loss: 1.4917 - val_accuracy: 0.9684
Epoch 35/50
379/379 [==============================] - 17s 46ms/step - loss: 1.4873 - accuracy: 0.9842 - val_loss: 1.4903 - val_accuracy: 0.9789
Epoch 36/50
379/379 [==============================] - 18s 46ms/step - loss: 1.4860 - accuracy: 0.9842 - val_loss: 1.4890 - val_accuracy: 0.9789
Epoch 37/50
379/379 [==============================] - 17s 46ms/step - loss: 1.4847 - accuracy: 0.9842 - val_loss: 1.4877 - val_accuracy: 0.9789
Epoch 38/50
379/379 [==============================] - 18s 46ms/step - loss: 1.4836 - accuracy: 0.9815 - val_loss: 1.4861 - val_accuracy: 0.9789
Epoch 39/50
379/379 [==============================] - 17s 46ms/step - loss: 1.4821 - accuracy: 0.9842 - val_loss: 1.4847 - val_accuracy: 0.9789
Epoch 40/50
379/379 [==============================] - 18s 46ms/step - loss: 1.4808 - accuracy: 0.9842 - val_loss: 1.4836 - val_accuracy: 0.9789
Epoch 41/50
379/379 [==============================] - 18s 46ms/step - loss: 1.4795 - accuracy: 0.9842 - val_loss: 1.4825 - val_accuracy: 0.9789
Epoch 42/50
379/379 [==============================] - 17s 46ms/step - loss: 1.4782 - accuracy: 0.9842 - val_loss: 1.4813 - val_accuracy: 0.9684
Epoch 43/50
379/379 [==============================] - 17s 46ms/step - loss: 1.4769 - accuracy: 0.9842 - val_loss: 1.4800 - val_accuracy: 0.9684
Epoch 44/50
379/379 [==============================] - 17s 46ms/step - loss: 1.4756 - accuracy: 0.9842 - val_loss: 1.4787 - val_accuracy: 0.9789
Epoch 45/50
379/379 [==============================] - 18s 46ms/step - loss: 1.4743 - accuracy: 0.9842 - val_loss: 1.4774 - val_accuracy: 0.9789
Epoch 46/50
379/379 [==============================] - 17s 46ms/step - loss: 1.4731 - accuracy: 0.9842 - val_loss: 1.4762 - val_accuracy: 0.9789
Epoch 47/50
379/379 [==============================] - 17s 46ms/step - loss: 1.4718 - accuracy: 0.9842 - val_loss: 1.4749 - val_accuracy: 0.9789
Epoch 48/50
379/379 [==============================] - 17s 46ms/step - loss: 1.4705 - accuracy: 0.9842 - val_loss: 1.4737 - val_accuracy: 0.9789
Epoch 49/50
379/379 [==============================] - 17s 46ms/step - loss: 1.4692 - accuracy: 0.9842 - val_loss: 1.4724 - val_accuracy: 0.9789
Epoch 50/50
379/379 [==============================] - 17s 46ms/step - loss: 1.4680 - accuracy: 0.9842 - val_loss: 1.4711 - val_accuracy: 0.9789
```

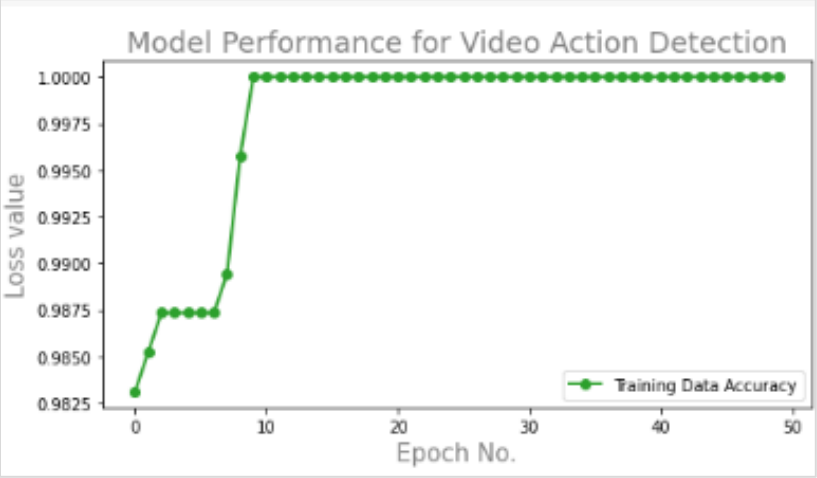


Train All Data Accuracy

Parameters:
batch_size = 50
no_epochs = 50
learning_rate = 0.0001
validation_split = 0
verbosity = 1

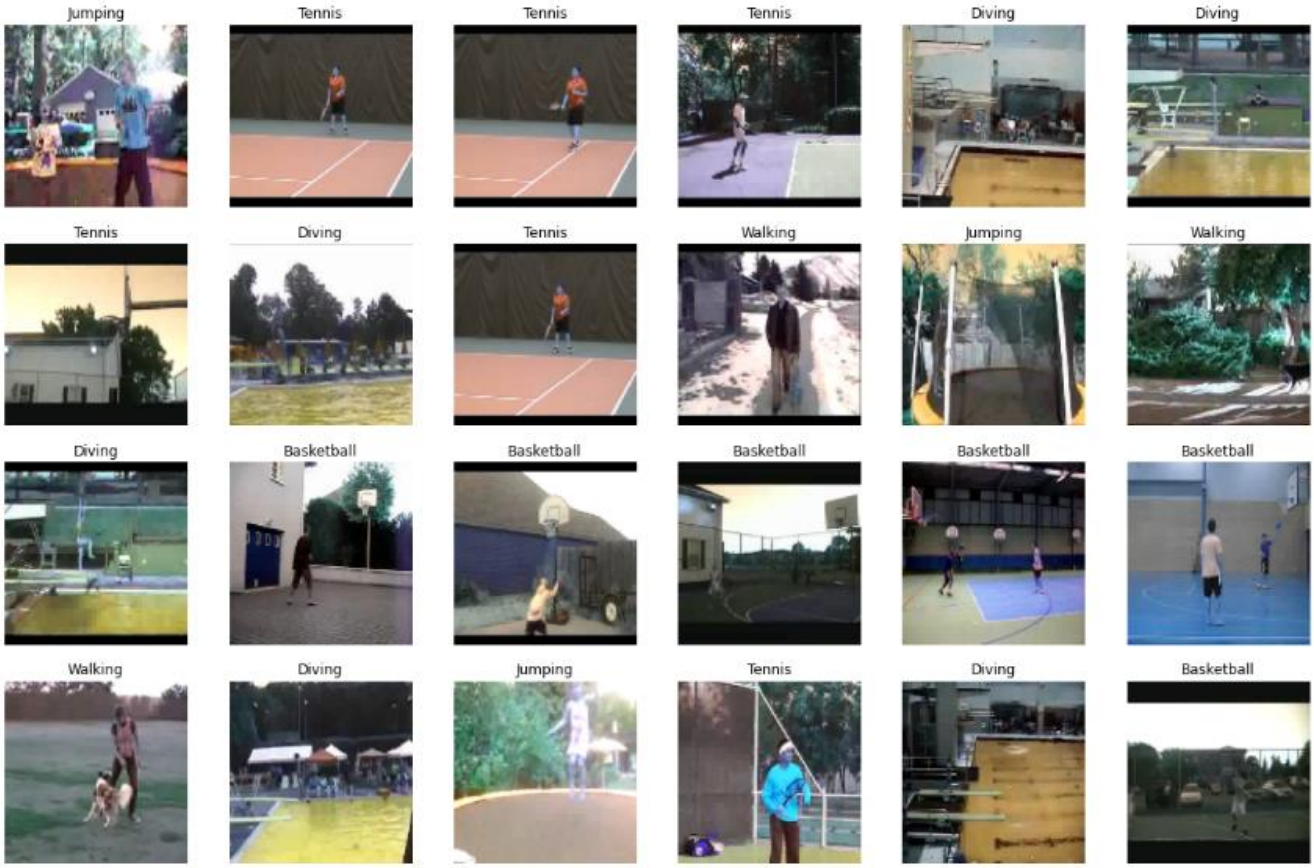
```
Train on all 474 samples.
```

```
Epoch 34/50
474/474 [==============================] - 18s 37ms/step - loss: 1.4128 - accuracy: 1.0000
Epoch 35/50
474/474 [==============================] - 18s 37ms/step - loss: 1.4113 - accuracy: 1.0000
Epoch 36/50
474/474 [==============================] - 18s 37ms/step - loss: 1.4097 - accuracy: 1.0000
Epoch 37/50
474/474 [==============================] - 18s 37ms/step - loss: 1.4082 - accuracy: 1.0000
Epoch 38/50
474/474 [==============================] - 18s 37ms/step - loss: 1.4067 - accuracy: 1.0000
Epoch 39/50
474/474 [==============================] - 18s 37ms/step - loss: 1.4052 - accuracy: 1.0000
Epoch 40/50
474/474 [==============================] - 18s 37ms/step - loss: 1.4036 - accuracy: 1.0000
Epoch 41/50
474/474 [==============================] - 18s 37ms/step - loss: 1.4021 - accuracy: 1.0000
Epoch 42/50
474/474 [==============================] - 18s 37ms/step - loss: 1.4006 - accuracy: 1.0000
Epoch 43/50
474/474 [==============================] - 18s 37ms/step - loss: 1.3991 - accuracy: 1.0000
Epoch 44/50
474/474 [==============================] - 18s 37ms/step - loss: 1.3976 - accuracy: 1.0000
Epoch 45/50
474/474 [==============================] - 18s 37ms/step - loss: 1.3961 - accuracy: 1.0000
Epoch 46/50
474/474 [==============================] - 18s 37ms/step - loss: 1.3945 - accuracy: 1.0000
Epoch 47/50
474/474 [==============================] - 18s 37ms/step - loss: 1.3930 - accuracy: 1.0000
Epoch 48/50
474/474 [==============================] - 18s 37ms/step - loss: 1.3915 - accuracy: 1.0000
Epoch 49/50
474/474 [==============================] - 18s 37ms/step - loss: 1.3900 - accuracy: 1.0000
Epoch 50/50
474/474 [==============================] - 18s 37ms/step - loss: 1.3885 - accuracy: 1.0000
```



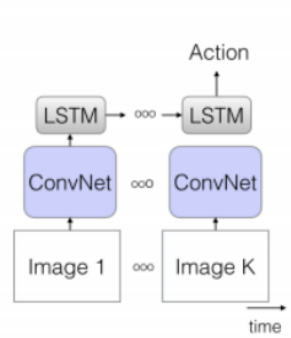
Testing

Test output of 24 random videos:

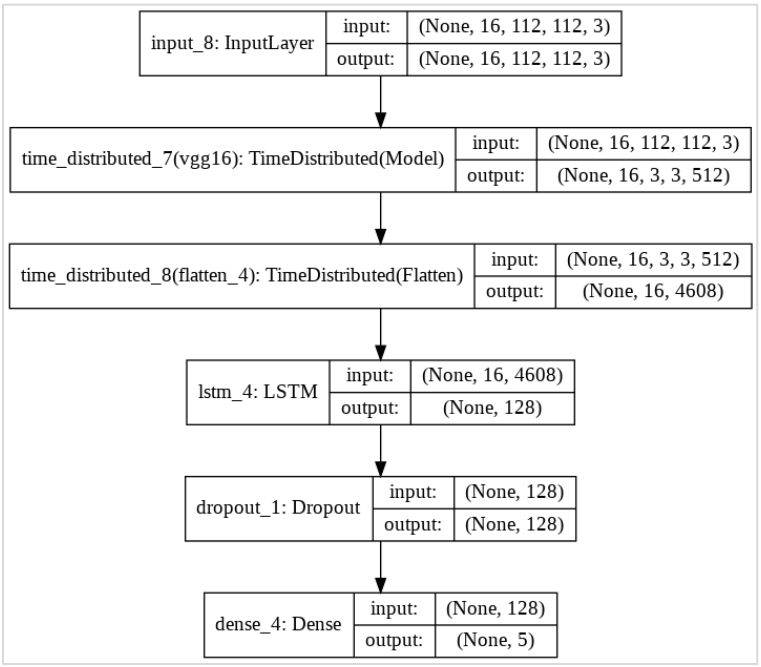


LSTM

A Single Stream Network - One Network for Spatial information.
Transfer Learning Model trained on ImageNet data.
Input Shape: (16, 112, 112, 3)



Model Plot:



Model Summary:

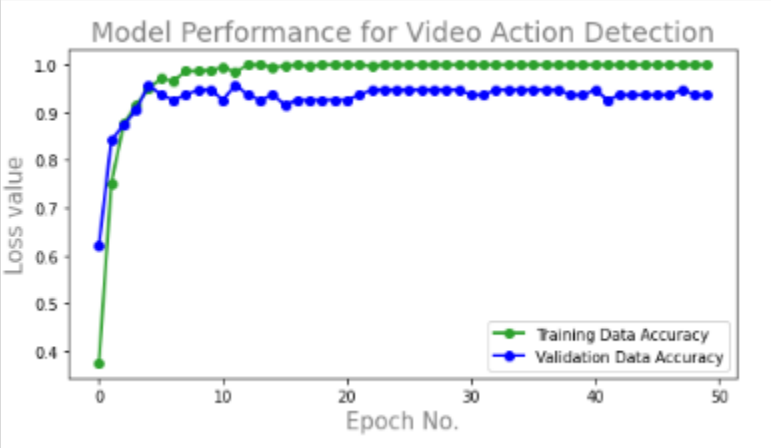
```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_8 (InputLayer)         (None, 16, 112, 112, 3)   0
_________________________________________________________________
time_distributed_7 (TimeDist (None, 16, 3, 3, 512)     14714688
_________________________________________________________________
time_distributed_8 (TimeDist (None, 16, 4608)          0
_________________________________________________________________
lstm_4 (LSTM)                (None, 128)               2425344
_________________________________________________________________
dropout_1 (Dropout)          (None, 128)               0
_________________________________________________________________
dense_4 (Dense)              (None, 5)                 645
=================================================================
Total params: 17,140,677
Trainable params: 2,425,989
Non-trainable params: 14,714,688
```

Train and Validation Accuracy

Parameters:
batch_size = 40
no_epochs = 50
learning_rate = 0.001
validation_split = 0.2
verbosity = 1

```
Train on 379 samples, validate on 95 samples.
```

```
Epoch 34/50
379/379 [==============================] - 18s 46ms/step - loss: 0.0155 - accuracy: 1.0000 - val_loss: 0.1836 - val_accuracy: 0.9579
Epoch 35/50
379/379 [==============================] - 18s 46ms/step - loss: 0.0235 - accuracy: 0.9974 - val_loss: 0.1850 - val_accuracy: 0.9368
Epoch 36/50
379/379 [==============================] - 18s 46ms/step - loss: 0.0163 - accuracy: 0.9974 - val_loss: 0.1907 - val_accuracy: 0.9368
Epoch 37/50
379/379 [==============================] - 17s 46ms/step - loss: 0.0130 - accuracy: 0.9974 - val_loss: 0.1936 - val_accuracy: 0.9368
Epoch 38/50
379/379 [==============================] - 17s 46ms/step - loss: 0.0150 - accuracy: 0.9974 - val_loss: 0.1882 - val_accuracy: 0.9474
Epoch 39/50
379/379 [==============================] - 17s 46ms/step - loss: 0.0160 - accuracy: 0.9974 - val_loss: 0.1979 - val_accuracy: 0.9474
Epoch 40/50
379/379 [==============================] - 18s 46ms/step - loss: 0.0088 - accuracy: 1.0000 - val_loss: 0.2125 - val_accuracy: 0.9368
Epoch 41/50
379/379 [==============================] - 18s 46ms/step - loss: 0.0118 - accuracy: 1.0000 - val_loss: 0.2024 - val_accuracy: 0.9368
Epoch 42/50
379/379 [==============================] - 17s 46ms/step - loss: 0.0097 - accuracy: 1.0000 - val_loss: 0.2091 - val_accuracy: 0.9368
Epoch 43/50
379/379 [==============================] - 18s 46ms/step - loss: 0.0109 - accuracy: 0.9974 - val_loss: 0.2096 - val_accuracy: 0.9368
Epoch 44/50
379/379 [==============================] - 18s 46ms/step - loss: 0.0098 - accuracy: 1.0000 - val_loss: 0.2164 - val_accuracy: 0.9368
Epoch 45/50
379/379 [==============================] - 18s 46ms/step - loss: 0.0120 - accuracy: 1.0000 - val_loss: 0.2137 - val_accuracy: 0.9368
Epoch 46/50
379/379 [==============================] - 18s 46ms/step - loss: 0.0130 - accuracy: 0.9974 - val_loss: 0.2164 - val_accuracy: 0.9368
Epoch 47/50
379/379 [==============================] - 18s 46ms/step - loss: 0.0128 - accuracy: 1.0000 - val_loss: 0.2137 - val_accuracy: 0.9368
Epoch 48/50
379/379 [==============================] - 18s 46ms/step - loss: 0.0116 - accuracy: 0.9974 - val_loss: 0.2135 - val_accuracy: 0.9368
Epoch 49/50
379/379 [==============================] - 17s 46ms/step - loss: 0.0107 - accuracy: 1.0000 - val_loss: 0.2154 - val_accuracy: 0.9368
Epoch 50/50
379/379 [==============================] - 18s 46ms/step - loss: 0.0062 - accuracy: 1.0000 - val_loss: 0.2120 - val_accuracy: 0.9368
```

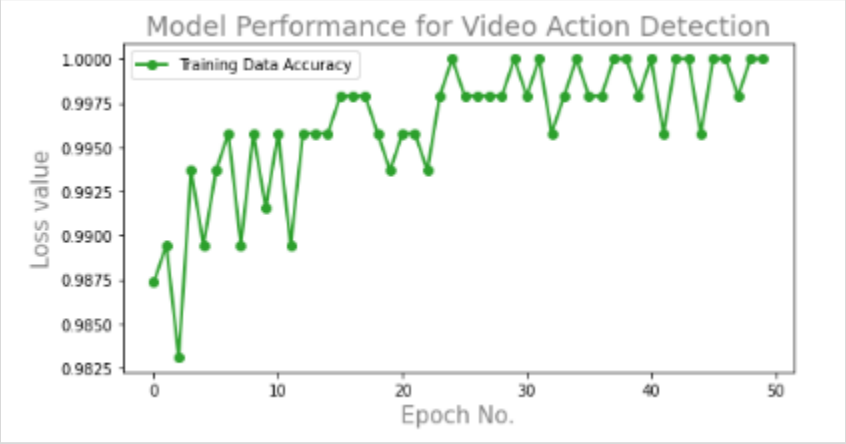


Train All Data Accuracy

Parameters:
batch_size = 40
no_epochs = 50
learning_rate = 0.001
validation_split = 0
verbosity = 1

```
Train on all 474 samples.
```

```
Epoch 35/50
474/474 [==============================] - 18s 37ms/step - loss: 0.0058 - accuracy: 1.0000
Epoch 36/50
474/474 [==============================] - 18s 37ms/step - loss: 0.0097 - accuracy: 0.9979
Epoch 37/50
474/474 [==============================] - 18s 37ms/step - loss: 0.0093 - accuracy: 0.9979
Epoch 38/50
474/474 [==============================] - 18s 37ms/step - loss: 0.0075 - accuracy: 1.0000
Epoch 39/50
474/474 [==============================] - 18s 37ms/step - loss: 0.0085 - accuracy: 1.0000
Epoch 40/50
474/474 [==============================] - 18s 37ms/step - loss: 0.0063 - accuracy: 0.9979
Epoch 41/50
474/474 [==============================] - 18s 37ms/step - loss: 0.0066 - accuracy: 1.0000
Epoch 42/50
474/474 [==============================] - 18s 37ms/step - loss: 0.0107 - accuracy: 0.9958
Epoch 43/50
474/474 [==============================] - 18s 37ms/step - loss: 0.0043 - accuracy: 1.0000
Epoch 44/50
474/474 [==============================] - 18s 37ms/step - loss: 0.0061 - accuracy: 1.0000
Epoch 45/50
474/474 [==============================] - 18s 37ms/step - loss: 0.0083 - accuracy: 0.9958
Epoch 46/50
474/474 [==============================] - 18s 37ms/step - loss: 0.0064 - accuracy: 1.0000
Epoch 47/50
474/474 [==============================] - 18s 37ms/step - loss: 0.0072 - accuracy: 1.0000
Epoch 48/50
474/474 [==============================] - 18s 37ms/step - loss: 0.0073 - accuracy: 0.9979
Epoch 49/50
474/474 [==============================] - 18s 37ms/step - loss: 0.0045 - accuracy: 1.0000
Epoch 50/50
474/474 [==============================] - 18s 37ms/step - loss: 0.0052 - accuracy: 1.0000
```



Testing

Test output of 24 random videos: