**Gesture Recognition Assignment**

As we know there are multiple neural network architecture suitable to solve this problem. Some of the well-known architectures are Conv3D, CNN + LSTM, CNN + GRU, CNN + Bidirectional LSTM, CNN + Bidirectional GRU. To not complicate things at first, we started with Conv3D architecture.

**Conv 3D**

**Experiment 1**

**Generator function:**

We are starting with a base Conv3D model with 30 images per sequence and resizing all the images to 100,100. Normalizing by channel using min\_max normalization.

**Training Sequences:** 100

**Validation Sequences:** 20

**Epochs:** 2

**Batch Size:** 10

**Learning Rate:** 0.01

**Architecture:**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Layer (type) Output Shape Param #

=================================================================

conv3d\_1 (Conv3D) (None, 28, 98, 98, 32) 2624

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

conv3d\_2 (Conv3D) (None, 26, 96, 96, 64) 55360

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

max\_pooling3d\_1 (MaxPooling3 (None, 13, 48, 48, 64) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

conv3d\_3 (Conv3D) (None, 11, 46, 46, 128) 221312

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

max\_pooling3d\_2 (MaxPooling3 (None, 5, 23, 23, 128) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

flatten\_1 (Flatten) (None, 338560) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_1 (Dense) (None, 32) 10833952

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_2 (Dense) (None, 5) 165

=================================================================

Total params: 11,113,413

Trainable params: 11,113,413

**Results:**

Epoch 2/2

30/30 [==============================] - 63s 2s/step - loss: 1.6106 - categorical\_accuracy: 0.2133 - val\_loss: 1.6016 - val\_categorical\_accuracy: 0.2000

**Description:**

For the base Conv3D model we got only 20% of Validation accuracy. Let’s try the same model with all the image sequences.

**Experiment 2**

**Generator function:**

We are starting with a base Conv3D model with 30 images per sequence and resizing all the images to 100,100. Normalizing by channel using min\_max normalization.

**Training Sequences:** 663

**Validation Sequences:** 100

**Epochs:** 2

**Batch Size:** 10

**Learning Rate:** 0.01

**Architecture:**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Layer (type) Output Shape Param #

=================================================================

conv3d\_1 (Conv3D) (None, 28, 98, 98, 32) 2624

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

conv3d\_2 (Conv3D) (None, 26, 96, 96, 64) 55360

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

max\_pooling3d\_1 (MaxPooling3 (None, 13, 48, 48, 64) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

conv3d\_3 (Conv3D) (None, 11, 46, 46, 128) 221312

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

max\_pooling3d\_2 (MaxPooling3 (None, 5, 23, 23, 128) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

flatten\_1 (Flatten) (None, 338560) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_1 (Dense) (None, 32) 10833952

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_2 (Dense) (None, 5) 165

=================================================================

Total params: 11,113,413

Trainable params: 11,113,413

**Results:**

Epoch 2/2

67/67 [==============================] - 146s 2s/step - loss: 12.9907 - categorical\_accuracy: 0.1940 - val\_loss: 12.5721 - **val\_categorical\_accuracy: 0.2200**

**Description:**

For the base Conv3D model we got only 22% of Validation accuracy. We have got an increase of 2% in the validation accuracy. Increasing the number of epochs to 10, to check if we see any increase in the accuracy rate.

**Experiment 3**

**Generator function:**

**[Remains the same as Experiment 2]**

We are starting with a base Conv3D model with 30 images per sequence and resizing all the images to 100,100. Normalizing by channel using min\_max normalization.

**Training Sequences:** 663

**Validation Sequences:** 100

**Epochs:** 10

**Batch Size:** 10

**Learning Rate:** 0.01

**Architecture:**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Layer (type) Output Shape Param #

=================================================================

conv3d\_1 (Conv3D) (None, 28, 98, 98, 32) 2624

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

conv3d\_2 (Conv3D) (None, 26, 96, 96, 64) 55360

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

max\_pooling3d\_1 (MaxPooling3 (None, 13, 48, 48, 64) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

conv3d\_3 (Conv3D) (None, 11, 46, 46, 128) 221312

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

max\_pooling3d\_2 (MaxPooling3 (None, 5, 23, 23, 128) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

flatten\_1 (Flatten) (None, 338560) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_1 (Dense) (None, 32) 10833952

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_2 (Dense) (None, 5) 165

=================================================================

Total params: 11,113,413

Trainable params: 11,113,413

**Results:**

Epoch 10/10

67/67 [==============================] - 137s 2s/step - loss: 1.5091 - categorical\_accuracy: 0.3657 - val\_loss: 1.4905 - **val\_categorical\_accuracy: 0.3500**

**Description:**

From experiment 2 to experiment 3 after increasing the epochs to 10, we got an accuracy of 35%.

**Experiment 4**

**Generator function:**

Changing the image resizing to 150\*150, because there are quite a few frames which are 360\*240 in the data. Also, taking only 28 frames for each sequence, so some videos have only 29 frames (images). Leaving the rest of the parameters same,

**Training Sequences:** 663

**Validation Sequences:** 100

**Epochs:** 2

**Batch Size:** 10

**Learning Rate:** 0.01

**Architecture:**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Layer (type) Output Shape Param #

=================================================================

conv3d\_16 (Conv3D) (None, 26, 148, 148, 32) 2624

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

conv3d\_17 (Conv3D) (None, 24, 146, 146, 64) 55360

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

max\_pooling3d\_11 (MaxPooling (None, 12, 73, 73, 64) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

conv3d\_18 (Conv3D) (None, 10, 71, 71, 128) 221312

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

max\_pooling3d\_12 (MaxPooling (None, 5, 35, 35, 128) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

flatten\_6 (Flatten) (None, 784000) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_11 (Dense) (None, 32) 25088032

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_12 (Dense) (None, 5) 165

=================================================================

Total params: **25,367,493 🡺 There is a significant increase in the parameters.**

Trainable params: 25,367,493

Non-trainable params: 0

**Results:**

67/67 [==============================] - 261s 4s/step - loss: 12.7163 - categorical\_accuracy: 0.2060 - val\_loss: 13.2168 - val\_categorical\_accuracy: 0.1800

Epoch 2/2

67/67 [==============================] - 243s 4s/step - loss: 12.7742 - categorical\_accuracy: 0.2075 - val\_loss: 13.2168 - **val\_categorical\_accuracy: 0.1800**

**Description:**

This model has 28+million parameters because of the increase in input size, so we ran this with only 2 epochs to run the model faster and compare it with earlier experiment Epoch 2 to see if there any increase in the validation and training accuracy. However, there is comparable increase for 2 epochs at all. Trying CNN + LSTM model in the next experiment.

**CNN + LSTM**

**Experiment 5**

**Generator function:**

In this CNN + LSTM Experiment, we have tried with the initial generator function of 100\*100 image resizing on a smaller dataset with 30 frames per sequence.

**Training Sequences:** 300

**Validation Sequences:** 100

**Batch size:** 10

**Epochs: 3**

**Learning Rate:** 0.01

**Architecture:**

Layer (type) Output Shape Param #

=================================================================

TimeDist\_Conv2D\_32 (TimeDist (None, 30, 98, 98, 32) 896

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_Conv2D\_64 (TimeDist (None, 30, 96, 96, 64) 18496

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_MaxPool (TimeDistri (None, 30, 48, 48, 64) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_Conv2D\_128 (TimeDis (None, 30, 46, 46, 128) 73856

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_MaxPool\_2 (TimeDist (None, 30, 23, 23, 128) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_Flatten (TimeDistri (None, 30, 67712) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

LSTM\_128 (LSTM) (None, 128) 34734592

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_5 (Dense) (None, 5) 645

=================================================================

Total params: 34,828,485

Trainable params: 34,828,485

**Accuracy results:**

30/30 [==============================] - 181s 6s/step - loss: 1.6893 - categorical\_accuracy: 0.1900 - val\_loss: 1.6297 - val\_categorical\_accuracy: 0.2300

Epoch 2/3

30/30 [==============================] - 68s 2s/step - loss: 1.6380 - categorical\_accuracy: 0.2133 - val\_loss: 1.6054 - val\_categorical\_accuracy: 0.2400

Epoch 3/3

30/30 [==============================] - 63s 2s/step - loss: 1.6106 - categorical\_accuracy: 0.2133 - val\_loss: 1.6016 - val\_categorical\_accuracy: 0.2400

**Description:**

In this experiment we have 38+million parameters because of the 3 layer Conv2D and a layer of LSTM. We have got only 24% accuracy for 3 epochs.

**Experiment 6**

**Generator function:**

**[Remains same as Experiment 5] except we take all the sequences in this experiment.**

**Training Sequences:** 663

**Validation Sequences:** 100

**Batch size:** 10

**Epochs:** 3

**Learning Rate:** 0.01

**Architecture:**

Layer (type) Output Shape Param #

=================================================================

TimeDist\_Conv2D\_32 (TimeDist (None, 30, 98, 98, 32) 896

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_Conv2D\_64 (TimeDist (None, 30, 96, 96, 64) 18496

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_MaxPool (TimeDistri (None, 30, 48, 48, 64) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_Conv2D\_128 (TimeDis (None, 30, 46, 46, 128) 73856

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_MaxPool\_2 (TimeDist (None, 30, 23, 23, 128) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_Flatten (TimeDistri (None, 30, 67712) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

LSTM\_128 (LSTM) (None, 128) 34734592

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_5 (Dense) (None, 5) 645

=================================================================

Total params: 34,828,485

Trainable params: 34,828,485

**Accuracy results:**

67/67 [==============================] - 201s 3s/step - loss: 1.6668 - categorical\_accuracy: 0.2119 - val\_loss: 1.6256 - val\_categorical\_accuracy: 0.1600

Epoch 2/3

67/67 [==============================] - 136s 2s/step - loss: 1.6171 - categorical\_accuracy: 0.1925 - val\_loss: 1.6303 - val\_categorical\_accuracy: 0.1500

Epoch 3/3

67/67 [==============================] - 140s 2s/step - loss: 1.6201 - categorical\_accuracy: 0.1910 - val\_loss: 1.6106 - **val\_categorical\_accuracy: 0.2100**

**Description:**

There is no increase in the accuracy rate, working on modifying the generator function in the next experiment to better the training data.

**Experiment 7**

**Generator function:**

Modified the Generator function to crop the images to the centre square and then resizing the image to 100\*100, this is to make sure that the important part of the image hand gesture if prominent and also make sure that all the images are in the square shape.

**Training Sequences:** 663

**Validation Sequences:** 100

**Batch size:** 10

**Epochs:** 3

**Learning Rate:** 0.01

**Architecture:**

Layer (type) Output Shape Param #

=================================================================

TimeDist\_Conv2D\_32 (TimeDist (None, 30, 98, 98, 32) 896

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_Conv2D\_64 (TimeDist (None, 30, 96, 96, 64) 18496

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_MaxPool (TimeDistri (None, 30, 48, 48, 64) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_Conv2D\_128 (TimeDis (None, 30, 46, 46, 128) 73856

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_MaxPool\_2 (TimeDist (None, 30, 23, 23, 128) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_Flatten (TimeDistri (None, 30, 67712) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

LSTM\_128 (LSTM) (None, 128) 34734592

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_5 (Dense) (None, 5) 645

=================================================================

Total params: 34,828,485

Trainable params: 34,828,485

**Accuracy results:**

Epoch 3/3

67/67 [==============================] - 129s 2s/step - loss: 1.4183 - categorical\_accuracy: 0.4075 - val\_loss: 1.3123 - **val\_categorical\_accuracy: 0.5300**

**Description:**

The architecture is the same as earlier experiment but this time we have got an accuracy increase up to 53% just by cropping the images before resizing. Trying the same model with more epochs.

**Experiment 8**

**Generator function:**

Modified the Generator function to crop the images to the centre square and then resizing the image to 100\*100, this is to make sure that the important part of the image hand gesture if prominent and also make sure that all the images are in the square shape.

**Training Sequences:** 663

**Validation Sequences:** 100

**Batch size:** 10

**Epochs:** 15

**Learning Rate:** 0.01

**Architecture:**

Layer (type) Output Shape Param #

=================================================================

TimeDist\_Conv2D\_32 (TimeDist (None, 30, 98, 98, 32) 896

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_Conv2D\_64 (TimeDist (None, 30, 96, 96, 64) 18496

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_MaxPool (TimeDistri (None, 30, 48, 48, 64) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_Conv2D\_128 (TimeDis (None, 30, 46, 46, 128) 73856

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_MaxPool\_2 (TimeDist (None, 30, 23, 23, 128) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_Flatten (TimeDistri (None, 30, 67712) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

LSTM\_128 (LSTM) (None, 128) 34734592

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_5 (Dense) (None, 5) 645

=================================================================

Total params: 34,828,485

Trainable params: 34,828,485

**Accuracy results:**

Epoch 15/15

67/67 [==============================] - 133s 2s/step - loss: 0.8023 - **categorical\_accuracy: 0.7358 - val\_loss: 1.1004 - val\_categorical\_accuracy: 0.6000**

**Description:**

Increasing the epochs from 3 to 15 has resulted in an increase of 7% in accuracy. Now the accuracy is at 60%.

**Experiment 9**

**Generator function:**

[Same as experiment 9, but changed the image size to 150\*150 and only 28 frames per sequence].

**Training Sequences:** 663

**Validation Sequences:** 100

**Batch size:** 10

**Epochs:** 15

**Learning Rate:** 0.01

**Architecture:**

**[added another LSTM layer]**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Layer (type) Output Shape Param #

=================================================================

TimeDist\_Conv2D\_32 (TimeDist (None, 28, 148, 148, 32) 320

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_Conv2D\_64 (TimeDist (None, 28, 146, 146, 64) 18496

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_MaxPool (TimeDistri (None, 28, 73, 73, 64) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_Conv2D\_128 (TimeDis (None, 28, 71, 71, 128) 73856

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_MaxPool\_2 (TimeDist (None, 28, 35, 35, 128) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_Conv2D\_64\_2 (TimeDi (None, 28, 33, 33, 64) 73792

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_Flatten (TimeDistri (None, 28, 69696) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

LSTM\_128 (LSTM) (None, 28, 128) 35750400

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

LSTM\_64\_2 (LSTM) (None, 64) 49408

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_2 (Dense) (None, 5) 325

=================================================================

Total params: 35,966,597

Trainable params: 35,966,597

Non-trainable params: 0

**Accuracy results:**

Epoch 15/15

67/67 [==============================] - 168s 3s/step - loss: 0.7221 - categorical\_accuracy: 0.7896 - val\_loss: 1.0699 - val\_categorical\_accuracy: 0.5900

**Description:**

Increasing the image size and reducing the frame to 28, and adding another LSTM layer, has not resulted in much difference in the accuracy from Experiment 8. Although training accuracy went up to 78%. Trying more experiments with the data set in the next one.

**Experiment 10**

**Generator function:**

In this generator function, we have cropped the images to the centre, resized the image to 150\*150, performed min\_max\_normalization on the whole image, changed the image to gray scale and then finally did a binary\_erosion to brighten the dark spots in the image to track the hand movements.

image = crop\_center\_square(image)

image = imresize(image, (150,150))

image = min\_max\_normalization(image)

image = morpho\_transform(image)

**Training Sequences:** 663

**Validation Sequences:** 100

**Batch size:** 10

**Epochs:** 15

**Learning Rate:** 0.01

**Architecture:**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Layer (type) Output Shape Param #

=================================================================

TimeDist\_Conv2D\_32 (TimeDist (None, 28, 148, 148, 32) 320

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_Conv2D\_64 (TimeDist (None, 28, 146, 146, 64) 18496

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_MaxPool (TimeDistri (None, 28, 73, 73, 64) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_Conv2D\_128 (TimeDis (None, 28, 71, 71, 128) 73856

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_MaxPool\_2 (TimeDist (None, 28, 35, 35, 128) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_Conv2D\_64\_2 (TimeDi (None, 28, 33, 33, 64) 73792

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_Flatten (TimeDistri (None, 28, 69696) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

LSTM\_128 (LSTM) (None, 28, 128) 35750400

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

LSTM\_64\_2 (LSTM) (None, 64) 49408

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_1 (Dense) (None, 5) 325

=================================================================

Total params: 35,966,597

Trainable params: 35,966,597

Non-trainable params: 0

**Accuracy results:**

Epoch 14/15

67/67 [==============================] - 159s 2s/step - loss: 1.0491 - categorical\_accuracy: 0.6090 - val\_loss: 1.3524 - val\_categorical\_accuracy: 0.5100

Epoch 00014: saving model to CNNLSTM\_model\_init\_2018-10-2407\_24\_25.876320/model-00014-1.04908-0.60896-1.35242-0.51000.h5

Epoch 15/15

67/67 [==============================] - 156s 2s/step - loss: 1.0527 - categorical\_accuracy: 0.6075 - val\_loss: 1.3524 - val\_categorical\_accuracy: 0.5100

**Description:**

The accuracy we got is only 51 % with this CNN + LSTM model.

**Experiment 11**

**Generator function:**

In this generator function, we have cropped the images to the centre, resized the image to 150\*150, performed min\_max\_normalization on the whole image, changed the image to gray scale and then finally did a binary\_erosion to brighten the dark spots in the image to track the hand movements.

image = crop\_center\_square(image)

image = imresize(image, (150,150))

image = min\_max\_normalization(image)

image = morpho\_transform(image)

**Training Sequences:** 663

**Validation Sequences:** 100

**Batch size:** 10

**Epochs:** 15

**Learning Rate:** 0.01

**Architecture:**

Layer (type) Output Shape Param #

=================================================================

TimeDist\_Conv2D\_32 (TimeDist (None, 28, 148, 148, 32) 320

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_Conv2D\_64 (TimeDist (None, 28, 146, 146, 64) 18496

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

batc\_1 (TimeDistributed) (None, 28, 146, 146, 64) 256

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_MaxPool (TimeDistri (None, 28, 73, 73, 64) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Drop\_out\_1 (TimeDistributed) (None, 28, 73, 73, 64) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_Conv2D\_128 (TimeDis (None, 28, 71, 71, 128) 73856

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_MaxPool\_2 (TimeDist (None, 28, 35, 35, 128) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_Flatten (TimeDistri (None, 28, 156800) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

LSTM\_128 (LSTM) (None, 28, 128) 80347648

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

LSTM\_64\_2 (LSTM) (None, 64) 49408

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

batch\_normalization\_2 (Batch (None, 64) 256

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_1 (Dense) (None, 5) 325

=================================================================

Total params: 80,490,565

Trainable params: 80,490,309

Non-trainable params: 256

**Accuracy results:**

Memory exhaustion error.

**Description:**

Changed the architecture by adding another Conv2D layer and this time the parameters increased to 80million and could run this model because of memory limitation error.

**Experiment 12**

**Conv3D**

**Generator function:**

**[Trying the same generator as Experiment 10 on Conv3D model]**

In this generator function, we have cropped the images to the centre, resized the image to 150\*150, performed min\_max\_normalization on the whole image, changed the image to gray scale and then finally did a binary\_erosion to brighten the dark spots in the image to track the hand movements.

image = crop\_center\_square(image)

image = imresize(image, (150,150))

image = min\_max\_normalization(image)

image = morpho\_transform(image)

**Training Sequences:** 663

**Validation Sequences:** 100

**Batch size:** 10

**Epochs:** 15

**Learning Rate:** 0.01

**Architecture:**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Layer (type) Output Shape Param #

=================================================================

conv3d\_1 (Conv3D) (None, 26, 148, 148, 32) 896

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

conv3d\_2 (Conv3D) (None, 24, 73, 73, 64) 55360

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

max\_pooling3d\_1 (MaxPooling3 (None, 8, 24, 24, 64) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

conv3d\_3 (Conv3D) (None, 6, 22, 22, 128) 221312

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

max\_pooling3d\_2 (MaxPooling3 (None, 2, 7, 7, 128) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

flatten\_1 (Flatten) (None, 12544) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_1 (Dense) (None, 128) 1605760

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_2 (Dense) (None, 5) 645

=================================================================

Total params: 1,883,973

Trainable params: 1,883,973

Non-trainable params: 0\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Accuracy results:**

**Epoch 4/15**

67/67 [==============================] - 146s 2s/step - loss: 0.9463 - categorical\_accuracy: 0.6403 - val\_loss: 0.9361 - **val\_categorical\_accuracy: 0.6900**

**.**

**.**

**.**

**Epoch 15/15**

67/67 [==============================] - 149s 2s/step - loss: 0.2325 - categorical\_accuracy: 0.9433 - val\_loss: 0.9435 - **val\_categorical\_accuracy: 0.6900**

**Description:**

After trying the new generator function with the Conv3D model we got accuracy of up to 69% and this is with just 4 epochs. And the conv3D model has lesser number of parameters ~1.8million compared to the CNN+LSTM model.

**Experiment 13**

**Conv3D**

**Generator function:**

**[Trying the same generator as Experiment 10 on Conv3D model]**

In this generator function, we have cropped the images to the centre, resized the image to 150\*150, performed min\_max\_normalization on the whole image, changed the image to gray scale and then finally did a binary\_erosion to brighten the dark spots in the image to track the hand movements.

image = crop\_center\_square(image)

image = imresize(image, (150,150))

image = min\_max\_normalization(image)

image = morpho\_transform(image)

**Training Sequences:** 663

**Validation Sequences:** 100

**Batch size:** 10

**Epochs:** 15

**Learning Rate:** 0.001

**Architecture:**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Layer (type) Output Shape Param #

=================================================================

conv3d\_1 (Conv3D) (None, 26, 148, 148, 32) 896

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

conv3d\_2 (Conv3D) (None, 24, 73, 73, 64) 55360

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

max\_pooling3d\_1 (MaxPooling3 (None, 8, 24, 24, 64) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

conv3d\_3 (Conv3D) (None, 6, 22, 22, 128) 221312

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

max\_pooling3d\_2 (MaxPooling3 (None, 2, 7, 7, 128) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

flatten\_1 (Flatten) (None, 12544) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_1 (Dense) (None, 128) 1605760

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_2 (Dense) (None, 5) 645

=================================================================

Total params: 1,883,973

Trainable params: 1,883,973

Non-trainable params: 0\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Accuracy results:**

Epoch 00013: saving model to Conv3D\_model\_init\_2018-10-2404\_13\_15.136186/model-00013-1.23534-0.55224-1.31377-0.53000.h5

Epoch 14/15

67/67 [==============================] - 152s 2s/step - loss: 1.2264 - categorical\_accuracy: 0.5358 - val\_loss: 1.3186 - val\_categorical\_accuracy: 0.5200

Epoch 00014: saving model to Conv3D\_model\_init\_2018-10-2404\_13\_15.136186/model-00014-1.22643-0.53582-1.31865-0.52000.h5

Epoch 15/15

**Description:**

Decreasing the learning rate from 0.01 to 0.001 did not help increase the validation accuracy.

**Experiment 14**

**Con2D + GRU**

**Generator function:**

**[Trying the same generator as Experiment 10 on Conv2D + GRU model]**

In this generator function, we have cropped the images to the centre, resized the image to 150\*150, performed min\_max\_normalization on the whole image, changed the image to gray scale and then finally did a binary\_erosion to brighten the dark spots in the image to track the hand movements.

image = crop\_center\_square(image)

image = imresize(image, (150,150))

image = min\_max\_normalization(image)

image = morpho\_transform(image)

**Training Sequences:** 663

**Validation Sequences:** 100

**Batch size:** 10

**Epochs:** 15

**Learning Rate:** 0.001

**Architecture:**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Layer (type) Output Shape Param #

=================================================================

TimeDist\_Conv2D\_32 (TimeDist (None, 28, 148, 148, 32) 320

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_Conv2D\_64 (TimeDist (None, 28, 146, 146, 64) 18496

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_MaxPool (TimeDistri (None, 28, 73, 73, 64) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_Conv2D\_128 (TimeDis (None, 28, 71, 71, 128) 73856

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_MaxPool\_2 (TimeDist (None, 28, 35, 35, 128) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_Conv2D\_64\_2 (TimeDi (None, 28, 33, 33, 64) 73792

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TimeDist\_Flatten (TimeDistri (None, 28, 69696) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

GRU\_128 (GRU) (None, 28, 128) 26812800

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

GRU\_64 (GRU) (None, 64) 37056

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_3 (Dense) (None, 5) 325

=================================================================

Total params: 27,016,645

Trainable params: 27,016,645

Non-trainable params: 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Accuracy results:**

Epoch 15/15

67/67 [==============================] - 157s 2s/step - loss: 0.9456 - categorical\_accuracy: 0.6403 - val\_loss: 1.3506 - val\_categorical\_accuracy: 0.5300

**Description:**

Conv2D + GRU model only has an accuracy of 53%.

**Conclusion:**

Out of all the experiments so far Conv3D model from Experiment 12 is the best both from number of trainable parameters ~1.8million and getting an accuracy of 69% with just 4 epochs.