Activity Recognition from Video

Data:

Videos taken from 7 different activities (3-Similar, 4-Different) :

1. Apply Eye Makeup (Similar)
2. Apply Lipstick (Similar)
3. Brushing Teeth (Similar)
4. Basketball (Different)
5. Diving (Different)
6. Nunchucks (Different)
7. Punch (Different)

Train Data:

8 videos/activity ⇔ 500 frames/video => 4,000 frames/activity
Total frames in training dataset: ~ 28,000

Cross-Validation Data:

10 videos/activity ⇔ 100 frames/video => 1,000 frames/activity
Total frames in CV dataset: ~ 7,000

Test-Data Data:

10 videos/activity ⇔ 100 frames/video => 1,000 frames/activity
Total frames in CV dataset: ~ 7,000
Video resized to size: (224, 224, 3) (RGB images)

Classification	LSTM
Train Shape:	Train Shape:
(29566, 224, 224, 3)	(14783, 16, 224, 224, 3) x 2
(29566, 7)	(14783, 16, 7) x 2
CV Shape:	CV Shape:
(7367, 224, 224, 3)	(7351, 16, 224, 224, 3)
(7367, 7)	(7351, 16, 7)
Test Shape:	Test Shape:
(7320, 224, 224, 3)	(7304, 16, 224, 224, 3)
(7320, 7)	(7304, 16, 7)

Due to memory constraints on Henry Cluster, LSTM is trained in 2 batches each with ~14,000 frames.

Approach 1:

Network:

(https://github.com/LisaAnne/lisa-caffe-public/blob/lstm_video_deploy/examples/LRCN_activity_recognition/train_test_lstm_RGB.prototxt)

Classification Model:

- 5 x Convolutional Layers
- 2 x Max Pooling Layers
- 2 x Batch Normalization
- 2 x Fully connected layers
- Activation: ReLU
- Classification: Softmax
- Loss: Categorical Cross Entropy
- Optimizer: Adam

- Total parameters: 48,327,303
- Total Trainable parameters: 48,326,343

Hyper-parameters:

- Regularization: Dropout (probability: 0.5)
- Batch size: 128
- Learning Rate: 1e-4
- Decay Rate: 1e-2

Sequential Model:

- 5 x Convolutional Layers
- 2 x Max Pooling Layers
- 2 x Batch Normalization
- 2 x Fully connected layers
- 2 x LSTM layers
- Activation: ReLU
- Classification: Softmax
- Loss: Categorical Cross Entropy
- Optimizer: Adam

- Total parameters: 48,504,903
- Total Trainable parameters: 43,174,343

Hyper-parameters:

- Sequence Length: 16
- Regularization: Dropout (probability: 0.5)
- Batch size: 64
- Learning Rate: 1e-4
- Decay Rate: 1e-2
- LSTM cells: 64 cells x 2 layers

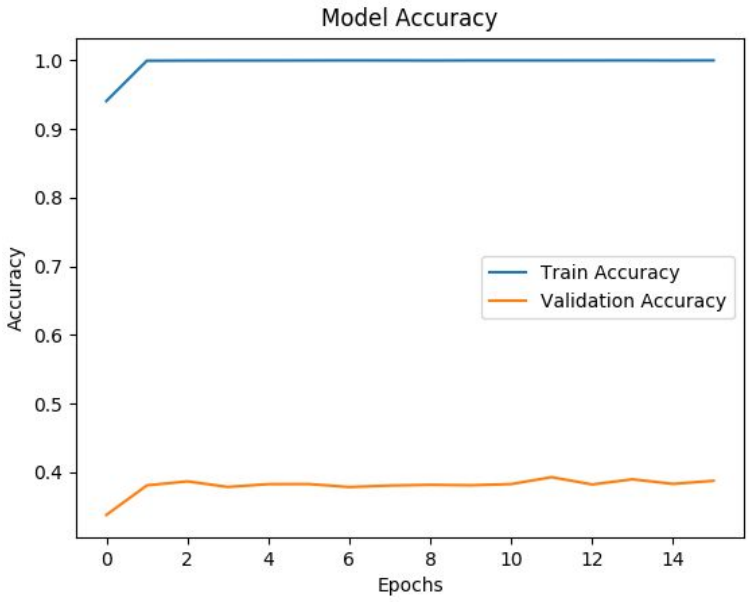
- Due to memory constraints on Henry cluster, batch size is reduced to 64.
- The classification trained weights are used for initialization and made **non-trainable.**

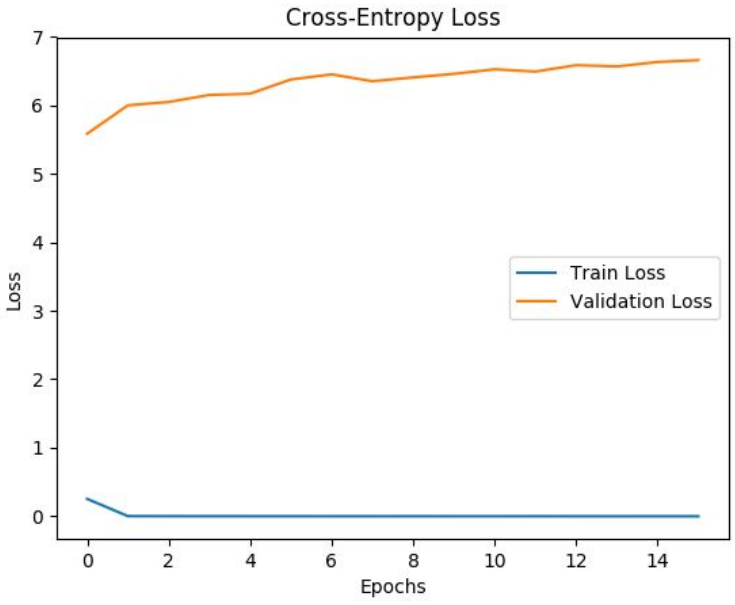
Result:

Classification:

Train Loss	:	0.00029122
Cross Validation Loss	:	6.6589
Test Loss	:	5.7841

Train Accuracy	:	100.00 %
Cross Validation Accuracy	:	38.75 %
Test Accuracy	:	46.95 %

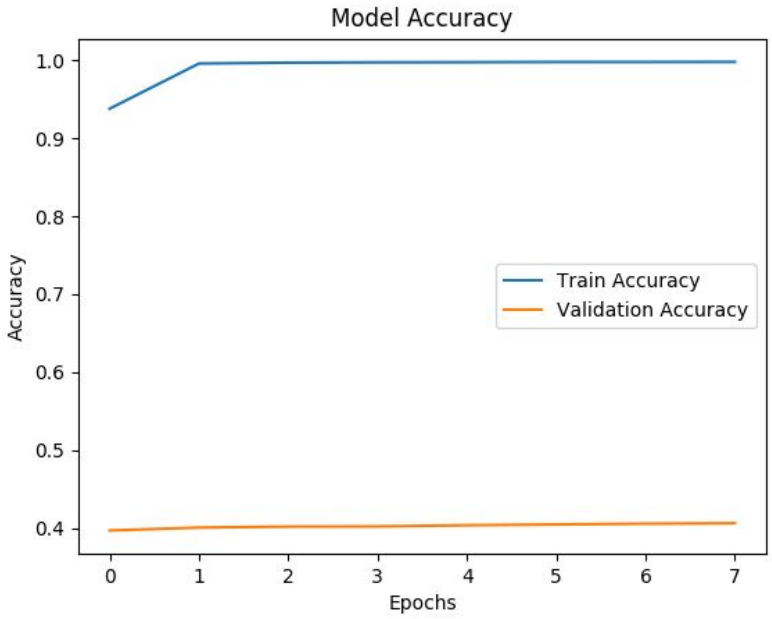


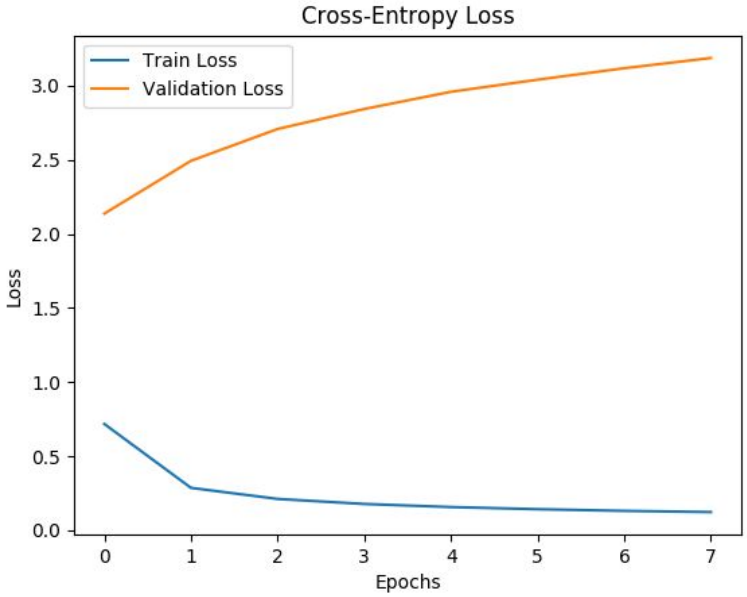


Sequential:

Train Loss : 0.1233
Cross Validation Loss : 3.1875
Test Loss : 2.7415

Train Accuracy : 99.79 %
Cross Validation Accuracy : **40.62** % (Improvement over classification: 38.75 %)
Test Accuracy : **47.72** % (Improvement over classification: 46.95%)





Approach 2: Transfer Learning (VGG-16 pretrained model)

The VGG-16 network is shown as:

https://goo.gl/images/1kAGbi

- The network is pretrained on ImageNet Dataset for 1000 classes
- **Only the convolutional part of the network is taken**

Classification Model:

- The **Convolutional part** is made **non-trainable** so it acts as feature extractor
- A fully connected layer is added after flattening with 256 neurons

- Classification: Softmax
- Loss: Categorical Cross Entropy
- Optimizer: Adam

- Total parameters: 21,139,271
- Total Trainable parameters: 6,424,583

Hyper-parameters:

- Regularization: Dropout (probability: 0.5)
- Batch size: 128
- Learning Rate: 1e-4
- Decay Rate: 1e-2

Sequential Model:

- The **Convolutional part** and the **fully-connected layer with 256 neurons** is made **non-trainable** so it acts as feature extractor

- 2 LSTM layers
- Classification: Softmax
- Loss: Categorical Cross Entropy
- Optimizer: Adam

- Total parameters: 21,183,015
- Total Trainable parameters: 45,543

Hyper-parameters:

- LSTM cells : 32 cells x 2 layers
- Regularization: Dropout (probability: 0.5)
- Batch size: 32
- Learning Rate: 1e-4
- Decay Rate: 1e-2

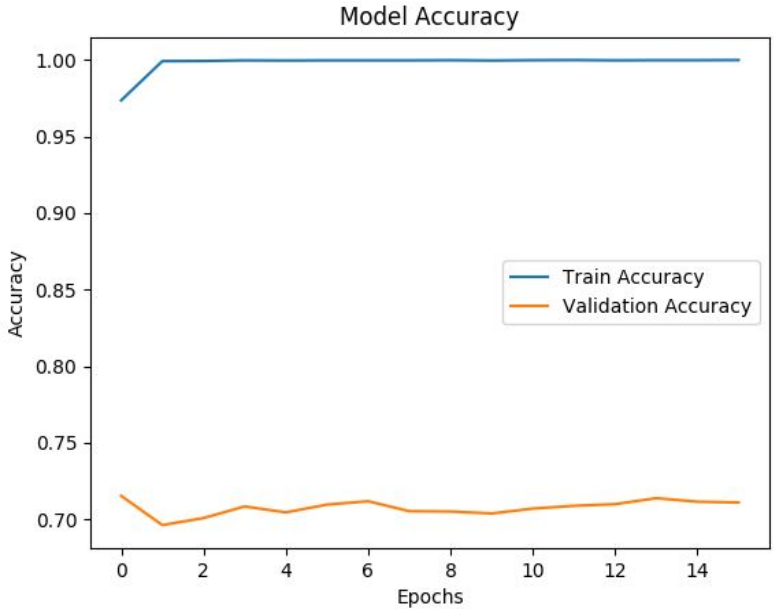
Due to memory constraint on Henry cluster, batch size is reduced to 32.

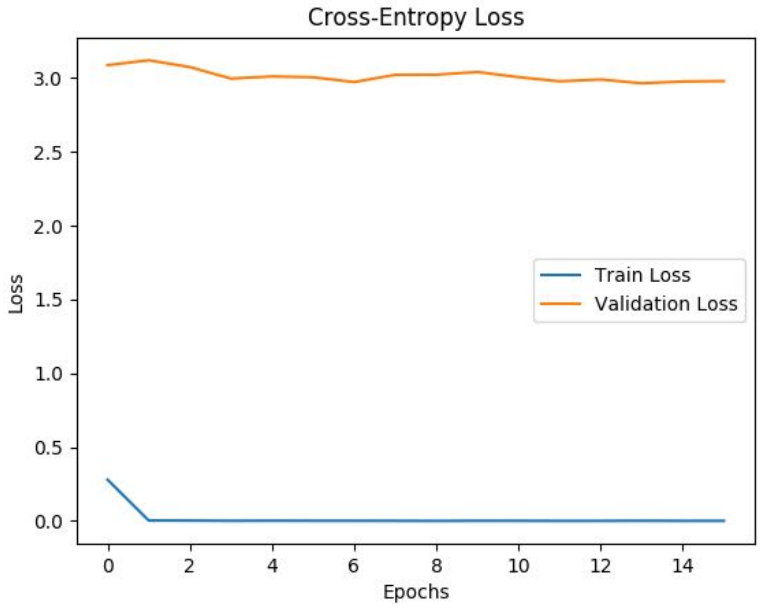
Result:

Classification:

Train Loss : 1.193947e-7
Cross Validation Loss : 2.9791
Test Loss : 4.08048

Train Accuracy : 100.00 %
Cross Validation Accuracy : **71.52** % (Improvement over Approach 1(Sequential) : 40.62%)
Test Accuracy : **60.71** % (Improvement over Approach 1(Sequential) : 47.72%)





Sequential:

Train Loss	:	0.1603
Cross Validation Loss	:	1.6146
Test Loss	:	1.8774

Train Accuracy	:	97.47 %
Cross Validation Accuracy	:	65.86 %
Test Accuracy	:	57.78 %