

# Online Video Classification: Comparative Analysis of Different Deep Neural Network Architectures

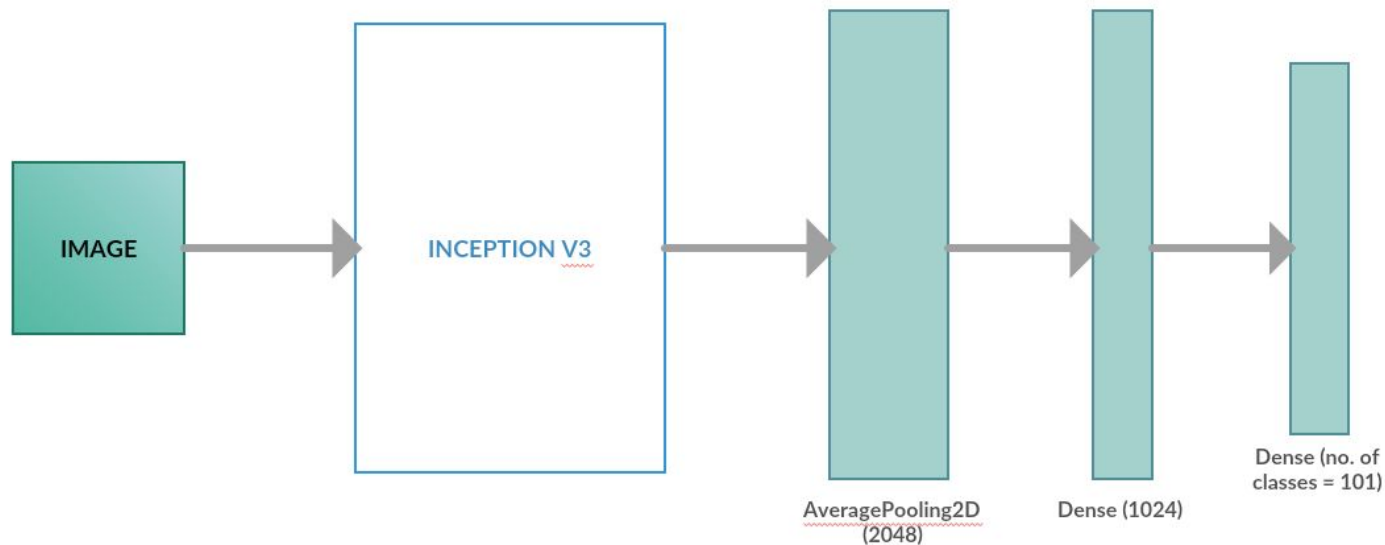
---



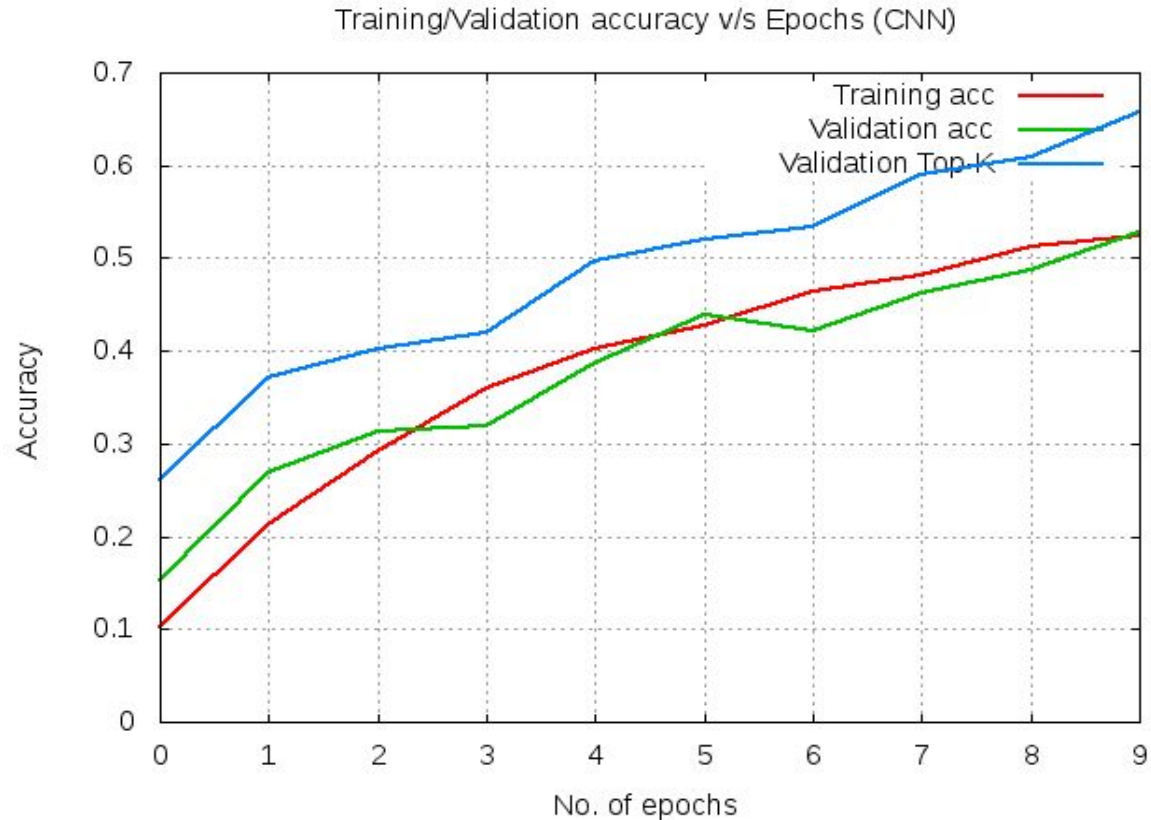
# One Frame at a Time ConvNet

- Classify each video clip based on a single frame
- Temporal component of videos not taken into consideration
- Transfer Learning
- Feature extraction using Inception-v3 pre-trained on ImageNet
- Fine tune the top dense layers to ensure previous learning was retrained
- Retrain the top 2 inception blocks on UCF101 dataset

# One Frame at a Time ConvNet



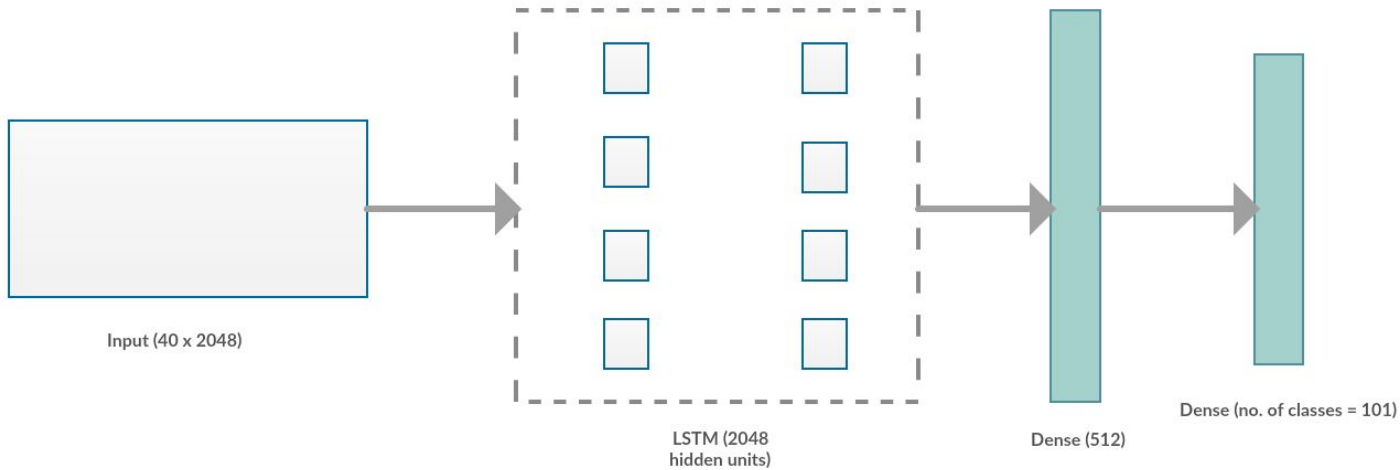
# One Frame at a Time ConvNet



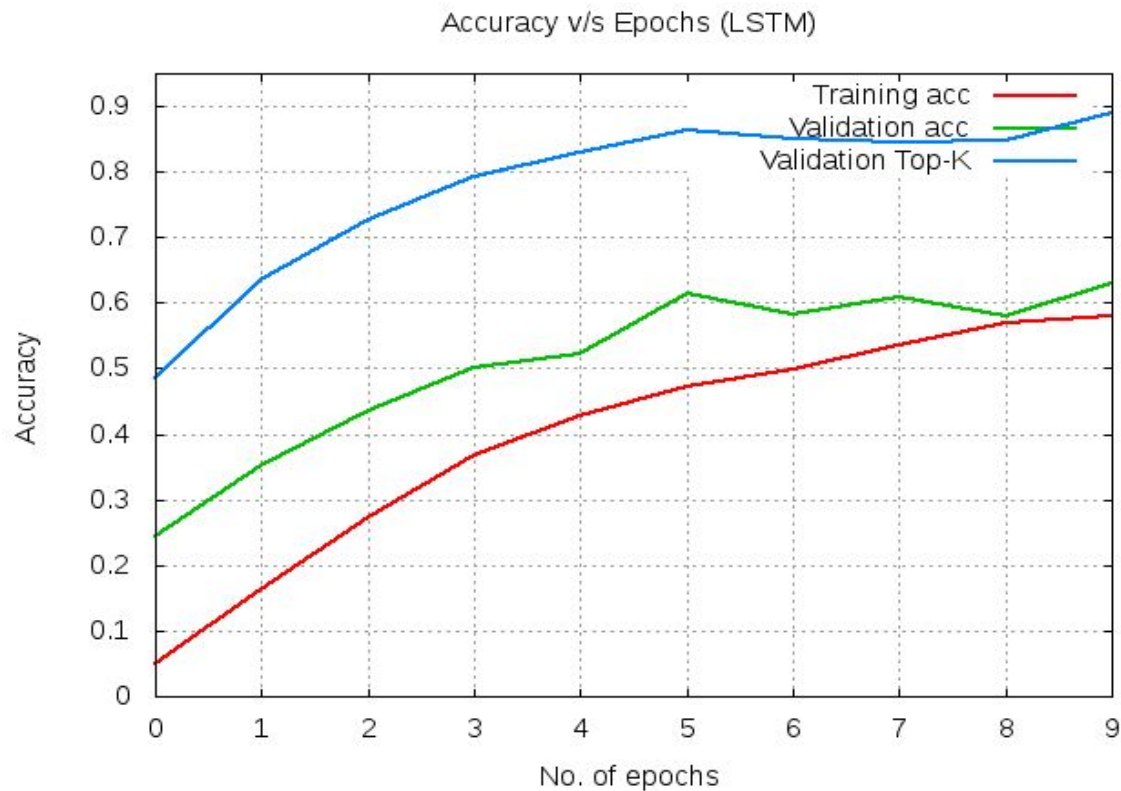
# ConvNet features passed to LSTM RNN

- Ran each frame of the video clips through a pre-trained Inception-v3 model
- Extracted a 2048-d vector of features by cutting the network at the final pool layer
- Converted these features into sequences
- Used a 2048-wide long short term memory (LSTM) layer, followed by a 1024 dense layer, with dropout layers in between for the recurrent neural network.

# ConvNet features passed to LSTM RNN



# ConvNet features passed to LSTM RNN

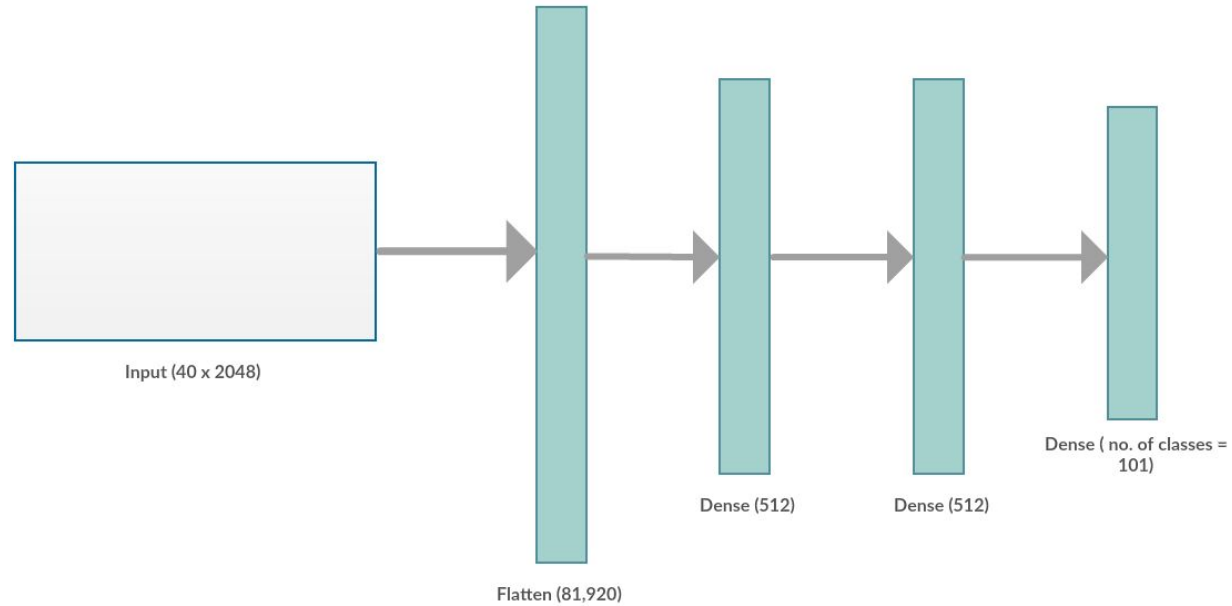




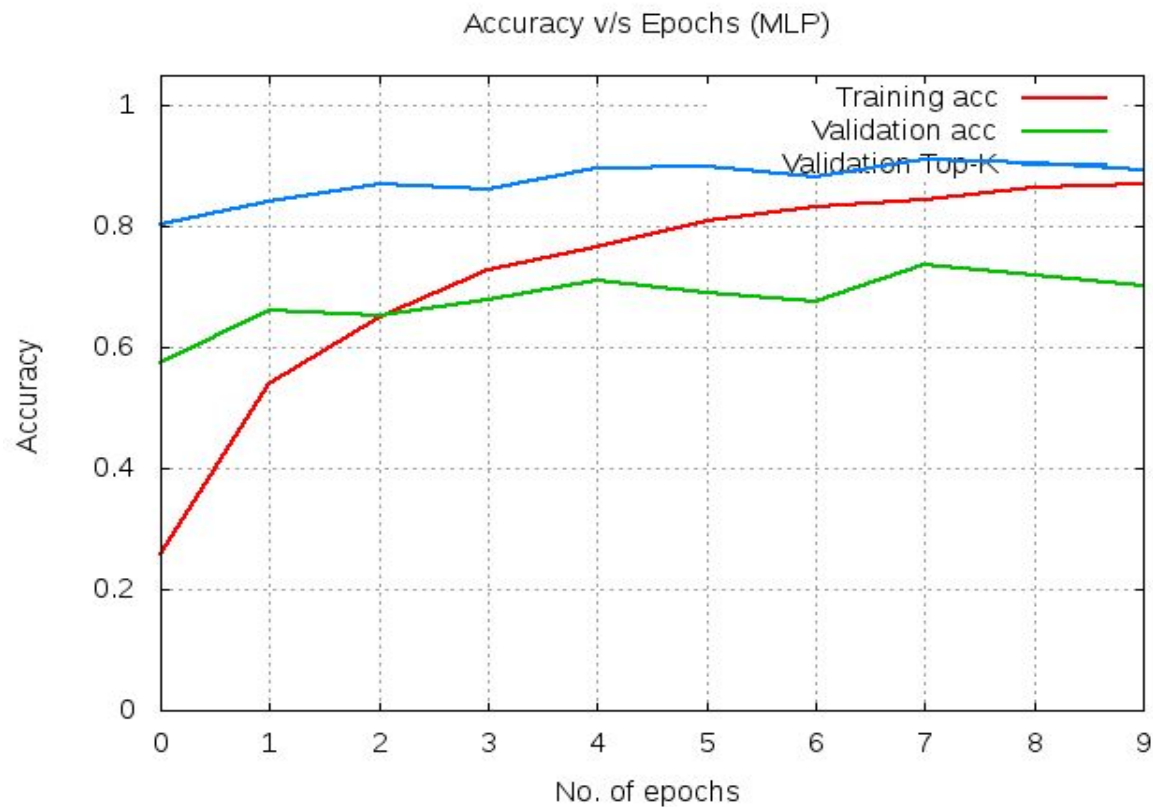
# ConvNet features passed to MLP

- Extracted features from the pre-trained InceptionV3 model with video frames as input
- Flattened these features into a sequence of length 40
- Used this sequence to train a perceptron network of two hidden layers each with 512 units.

# ConvNet features passed to MLP



# ConvNet features passed to MLP



# Thank You!

---

- Darshana Gadre, Mrunmayi Deshpande, Dnyanada Joshi