Action Recognition

Using Video Analysis

Introduction

- What is action recognition
- what are different approaches
  - sensor based vs video analysis

Progression on Technology of Video Analysis

a) 2D convolution
b) objection recognition + motion recognition using image processing
c) 3D convolution
d) temporal
    I. 2D convolution + hand crafted features
    II. 2D convolution + RNN
    III. 2D convolution + LSTM

Technology used

- Core Technologies
  - Machine learning
  - Image processing
  - Video processing
  - Deep Learning
  - Transfer Learning

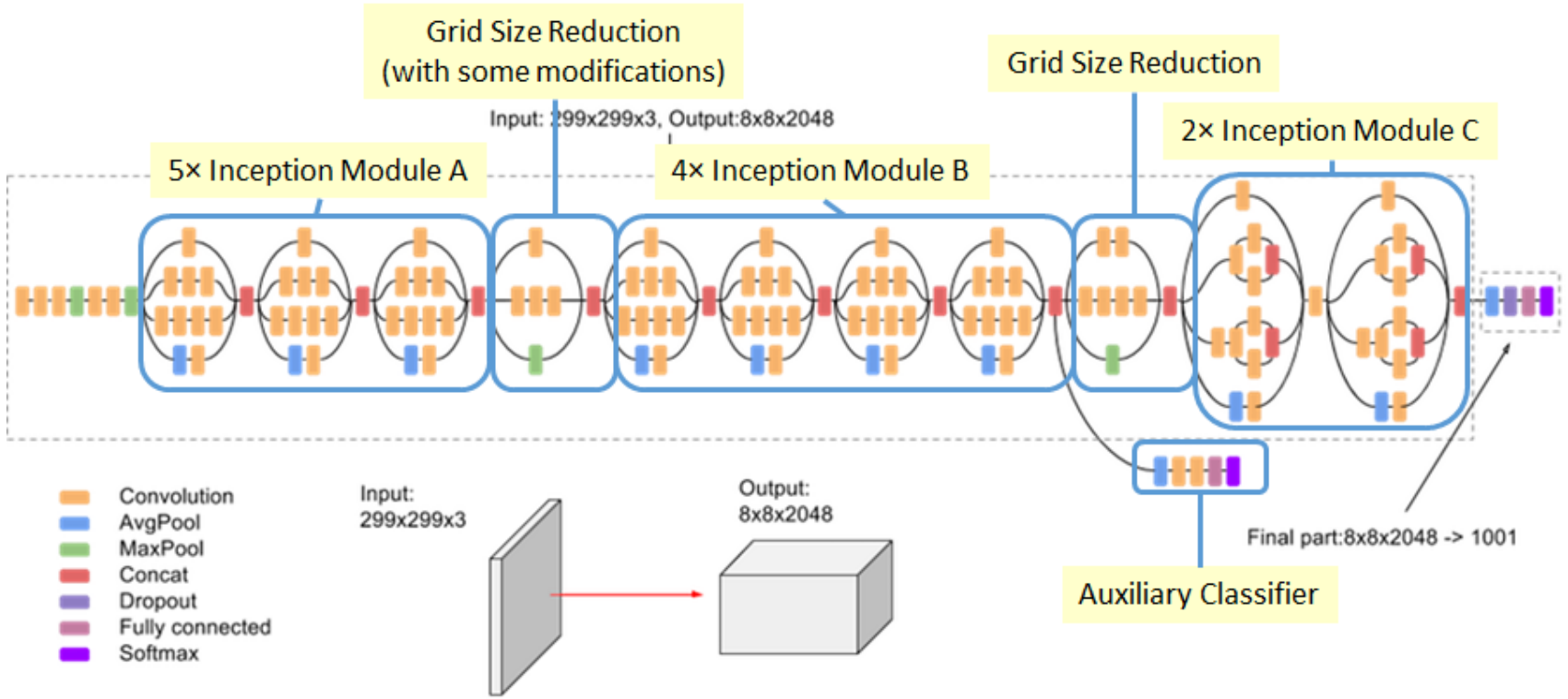
Convolutional Neural Network

- Convolution layer
- RELU
- max-pooling
- fully-connected

Transfer Learning

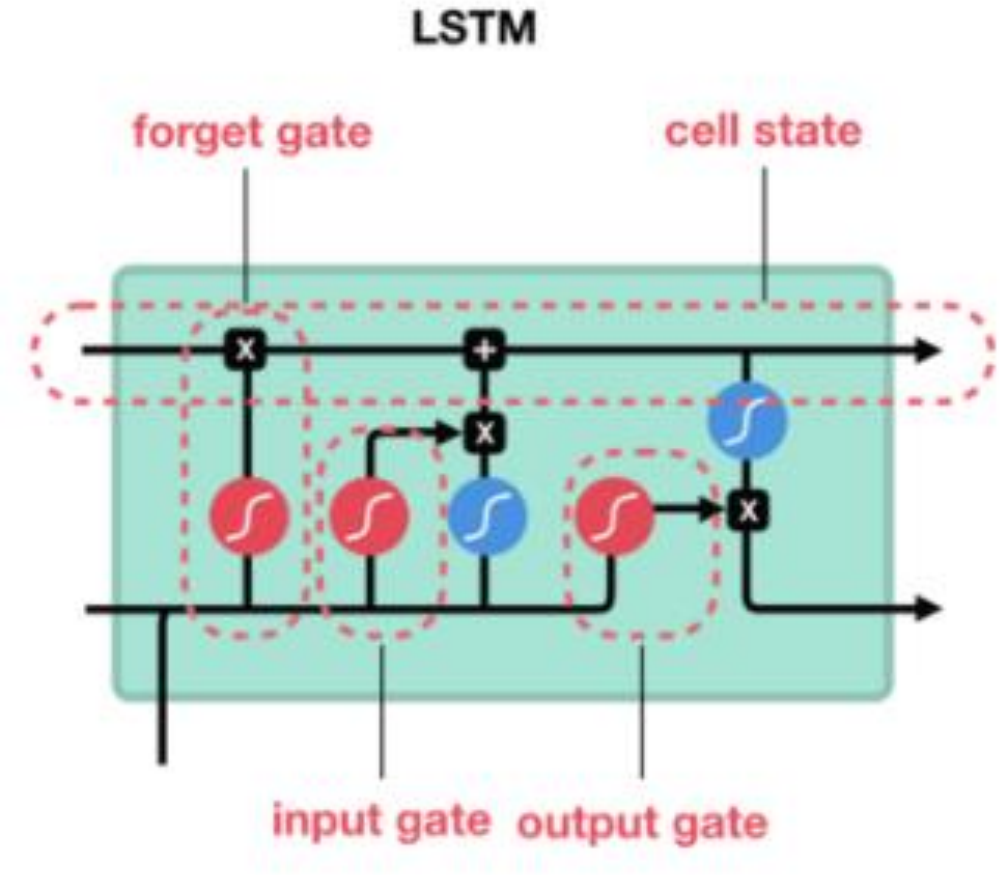
- Two Approaches
  1) develop model approach
  2) pre-trained model approach

Inception Model



Long short-term memory

- RNN vs LSTM



Data Description

- UCF101
- 13320 videos
- 101 classes
- 25groups/class
- 4-7videos/group

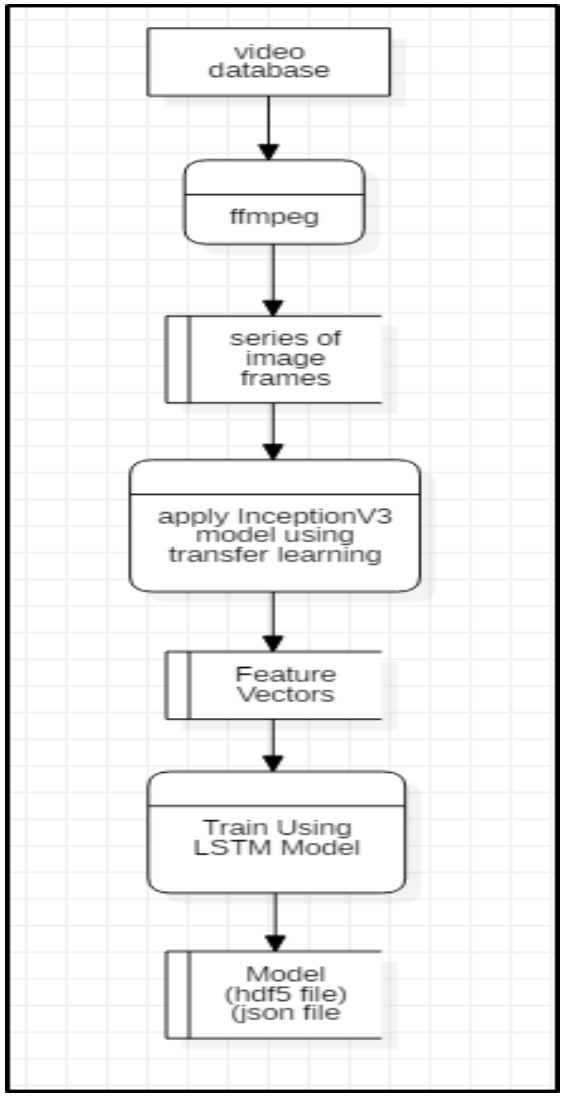
Data Description contd.

- Types of videos in dataset
  A. Human object interaction
  B. Body-motion only
  C. Human-human interaction
  D. Playing musical instruments
  E. Sports

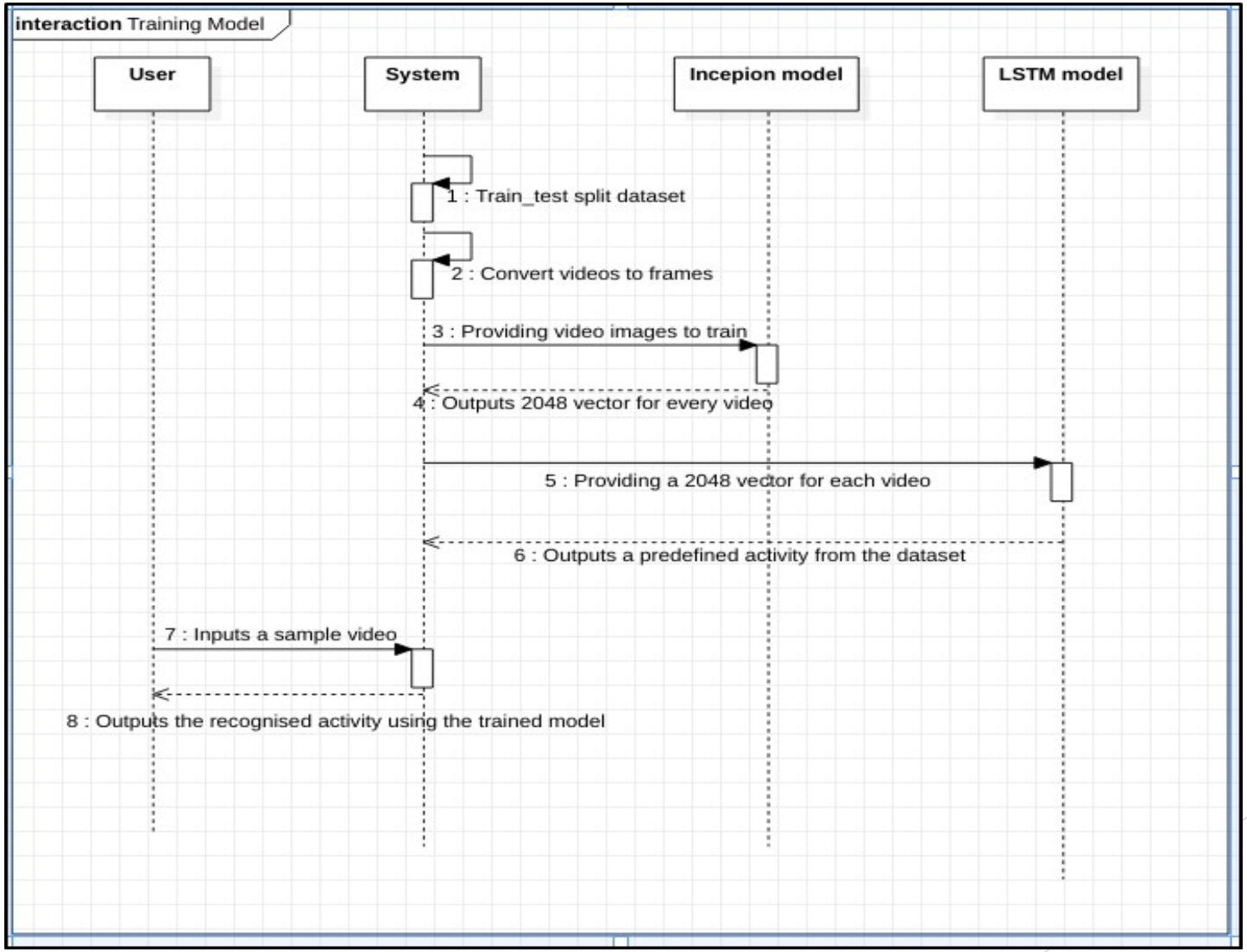
Implementation Details

- Tasks
  - train_test_split
  - extract_frames
  - extract_features
  - generate_textfile
  - model_train
  - predict_class

Data-flow Diagram



Sequence Diagram



Result

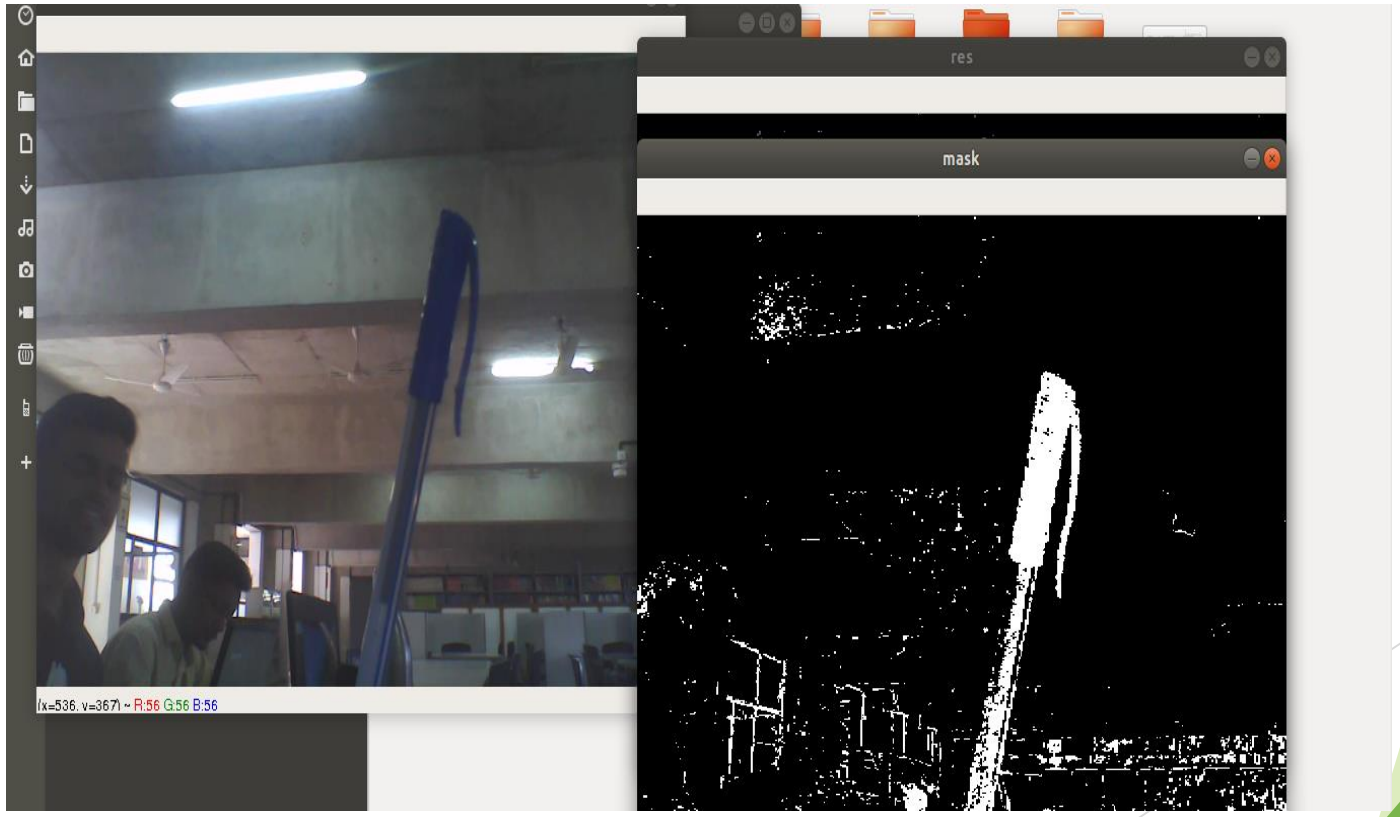
```
331/331 [==============================] - 359s 1s/step - loss: 2.5537 - acc: 0.3780 - val_loss: 2.0013 - val_acc: 0.4876
Epoch 2/15
331/331 [==============================] - 307s 928ms/step - loss: 1.1685 - acc: 0.6690 - val_loss: 0.8741 - val_acc: 0.7501
Epoch 3/15
331/331 [==============================] - 341s 1s/step - loss: 0.7514 - acc: 0.7790 - val_loss: 0.8425 - val_acc: 0.7779
Epoch 4/15
331/331 [==============================] - 393s 1s/step - loss: 0.5422 - acc: 0.8384 - val_loss: 0.5855 - val_acc: 0.8442
Epoch 5/15
331/331 [==============================] - 331s 1000ms/step - loss: 0.4089 - acc: 0.8763 - val_loss: 0.5198 - val_acc: 0.8738
Epoch 6/15
331/331 [==============================] - 262s 791ms/step - loss: 0.3252 - acc: 0.9024 - val_loss: 0.6287 - val_acc: 0.8627
Epoch 7/15
331/331 [==============================] - 227s 686ms/step - loss: 0.2749 - acc: 0.9176 - val_loss: 0.4489 - val_acc: 0.8920
Epoch 8/15
331/331 [==============================] - 211s 639ms/step - loss: 0.2140 - acc: 0.9346 - val_loss: 0.4029 - val_acc: 0.9155
Epoch 9/15
331/331 [==============================] - 214s 646ms/step - loss: 0.1809 - acc: 0.9454 - val_loss: 0.4058 - val_acc: 0.9124
Epoch 10/15
331/331 [==============================] - 215s 649ms/step - loss: 0.1556 - acc: 0.9532 - val_loss: 0.4270 - val_acc: 0.9131
Epoch 11/15
331/331 [==============================] - 218s 657ms/step - loss: 0.1350 - acc: 0.9593 - val_loss: 0.4137 - val_acc: 0.9152
Epoch 12/15
331/331 [==============================] - 219s 661ms/step - loss: 0.1155 - acc: 0.9652 - val_loss: 0.4530 - val_acc: 0.9215
Epoch 13/15
331/331 [==============================] - 217s 657ms/step - loss: 0.1010 - acc: 0.9693 - val_loss: 0.3739 - val_acc: 0.9327
Epoch 14/15
331/331 [==============================] - 207s 625ms/step - loss: 0.0947 - acc: 0.9718 - val_loss: 0.4804 - val_acc: 0.9206
Epoch 15/15
331/331 [==============================] - 222s 670ms/step - loss: 0.0865 - acc: 0.9738 - val_loss: 0.3542 - val_acc: 0.9434

[INFO] : Model Trained.!

[INFO] : Saving Model.!
[INFO] : Model Saved.!
```

Other Works

- Detect pen(blue colour) using OpenCV



Other Works contd.

- YOLO object detection

Future Work

- Building an API for online real-time activity recognition as a service.
- Running and tuning the model for larger and vast dataset Youtube-8M. Which can further improve accuracy in more real-life scenarios.
- Working on latest state-of-the-art 2-stream CNN model for activity recognition.

Thank You