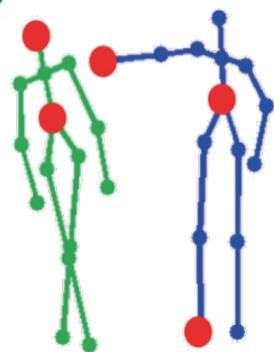
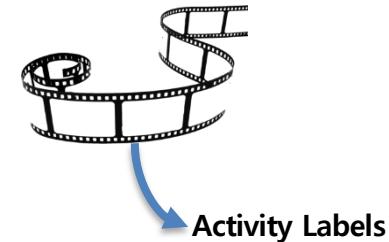


# Human Action and Activity Recognition: Current Trends and Future Directions



# Main Contents

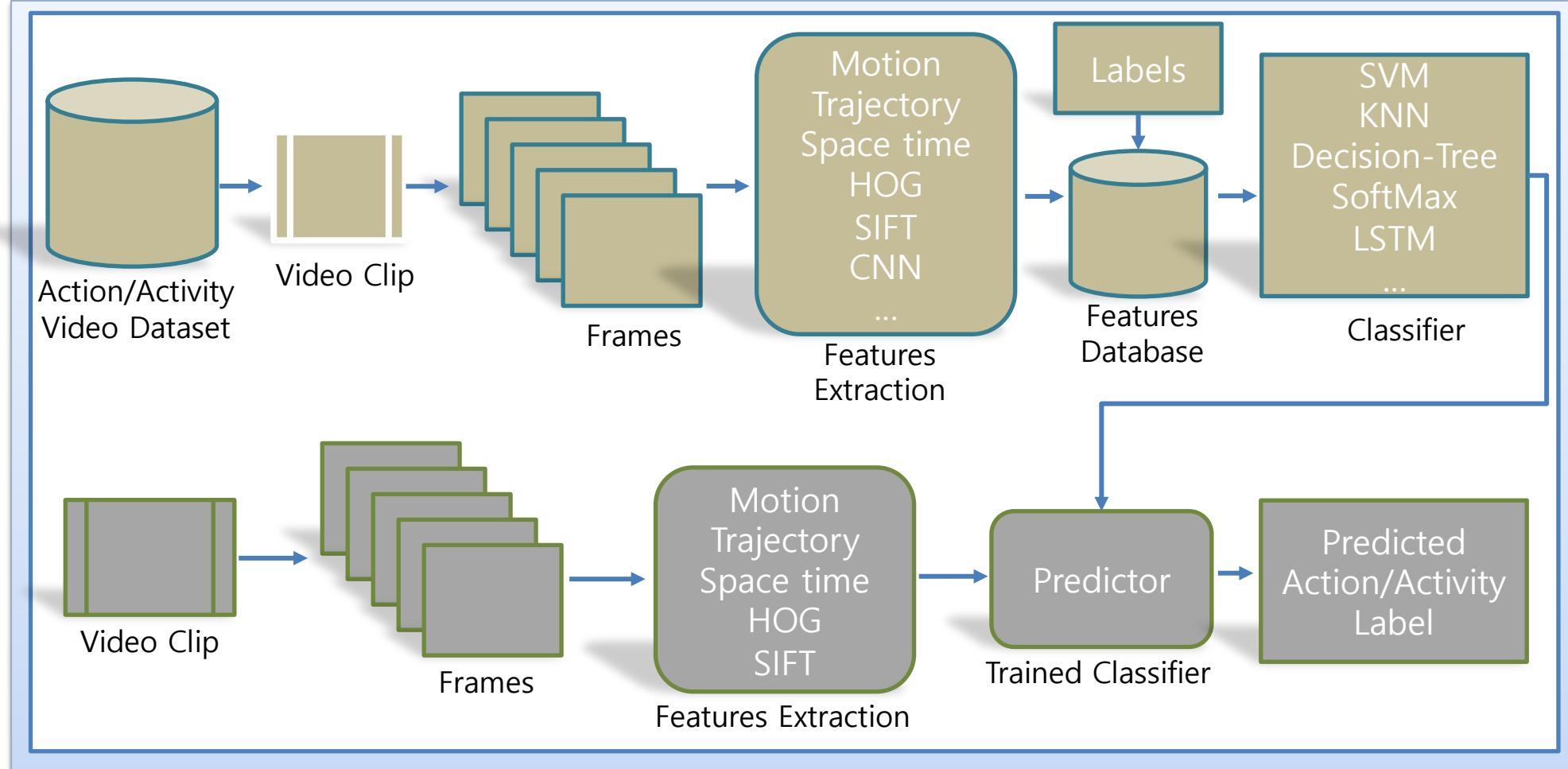


1. Overview
2. Applications
3. Categorization using Available Sensors
4. Literature Survey
5. Benchmark Datasets
6. Action and Activity Recognition Methods
7. Demo Results
8. Major Challenges
9. Future Directions
10. References

# 1. Overview Action and Activity Recognition



- The goal of action and activity recognition is an automated interpretation of ongoing events and their context from video data.



## 2. Applications



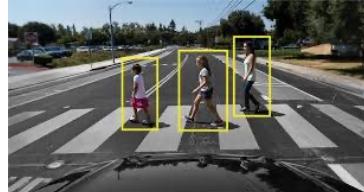
### Smart surveillance



- Suspicious activity recognition in real-time
- Abnormal events at shopping malls, railway station, etc.
- Illegal car parking
- Military activates (aerial images)



### Intelligent driving



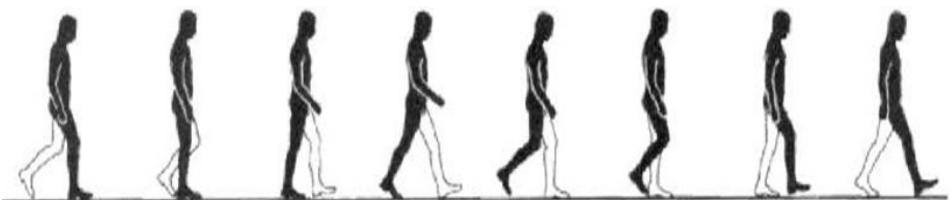
- Driver facial activity recognition
- Real-time driving supporter
- Abnormal movements detection on road



### Behavioral biometrics



- Gait pattern as a biometric
- Gesture and posture recognition and analysis



### Medical

- Elderly health care
- Analysis of orthopedics
- Body postures
- Fitness

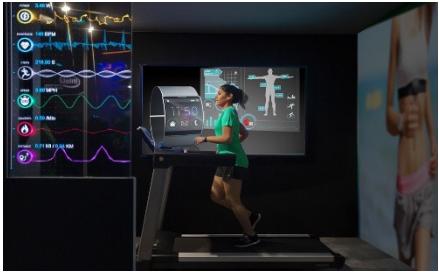


# 2. Applications



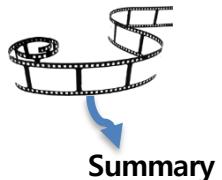
## Sports and exercise

- Analyzing the play and deducing the actions in sports
- Exercise with a feedback system



## Video tagging

- Search-by-video
- Indexing and retrieval
- Video summarization



## Human computer interaction

- Humanoid robot control
- Imitate human motions in a robot
- Virtual and augmented environments

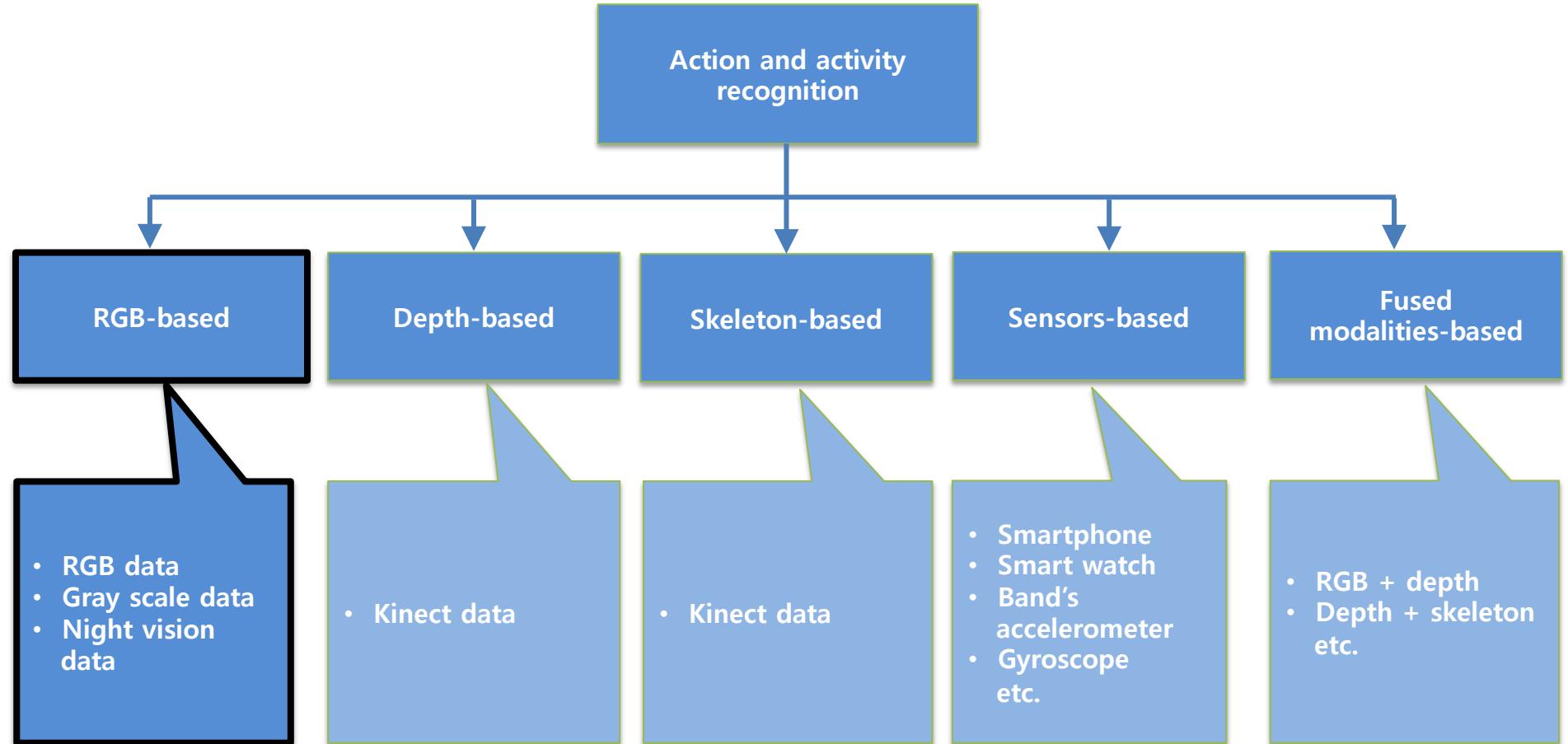


## User-interfaces

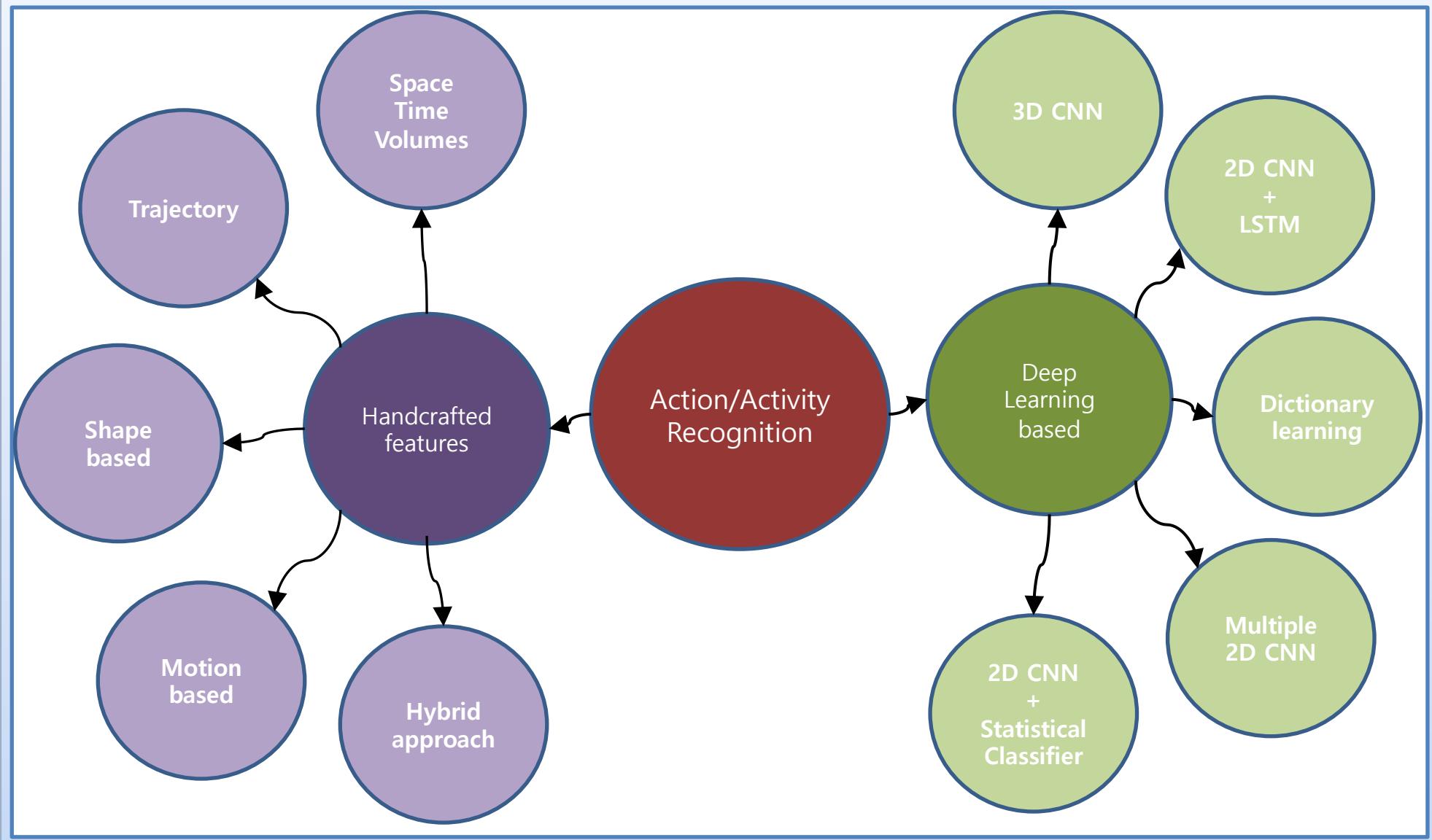
- User-interfaces for blind people to see
- Sharing contextual information with blind people



### 3. Categorization using Available Sensors



### 3. Literature Survey



## 5. Benchmark Datasets



**List of popular action and video datasets used in action recognition research**

Datasets	# Videos	# Actions/ Activities	# Subjects	Modality
UCF101	13,320	101	N/A	RGB
HMDB51	7,000	51	N/A	RGB
Hollywood2	3,669	12	10	RGB
UCF50	50	50	N/A	RGB
ActivityNet	28,000	203	N/A	RGB
Kinetics	500,000	600	N/A	RGB
AVA	57,600	80	N/A	RGB
UCF Sports	150	10	N/A	RGB
YouTube	8,000,000	4,716	25	RGB
Sports-1M	1,133,158	487	N/A	RGB
Moments in Time	1,000,000	339	N/A	RGB

# 5. Datasets



- Sample Action and Activities

- The below figure shows some sample images from different action and activities datasets
- There are many sports activities in the figure such as bowling, diving, soccer penalty along with some other actions like brushing hair, pushing etc

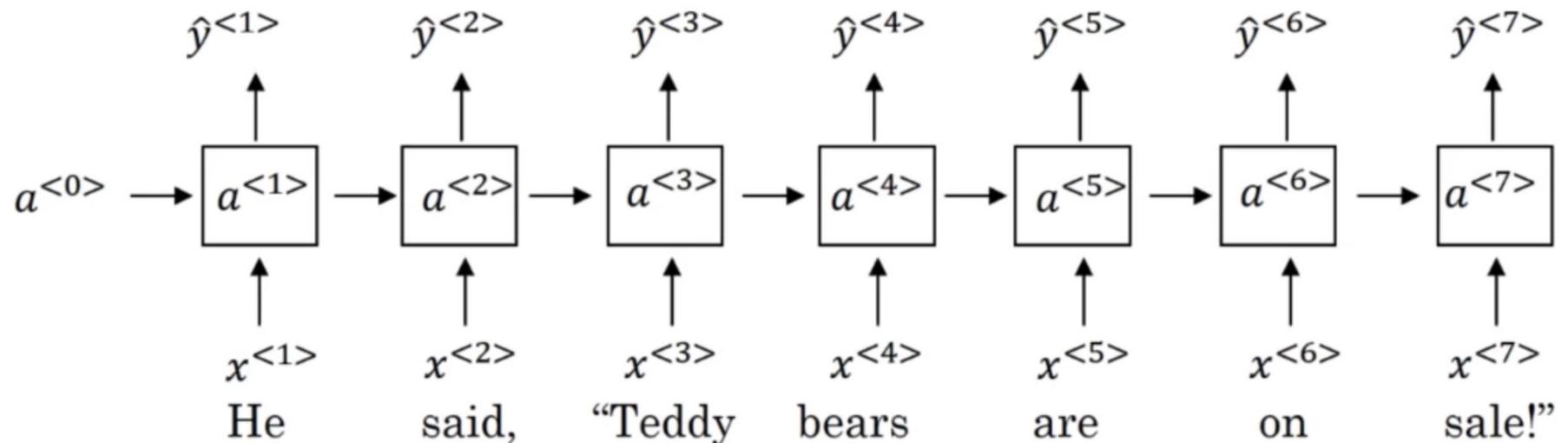




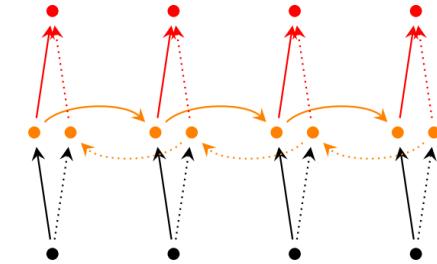
# Getting information from the future

He said, “Teddy bears are on sale!”

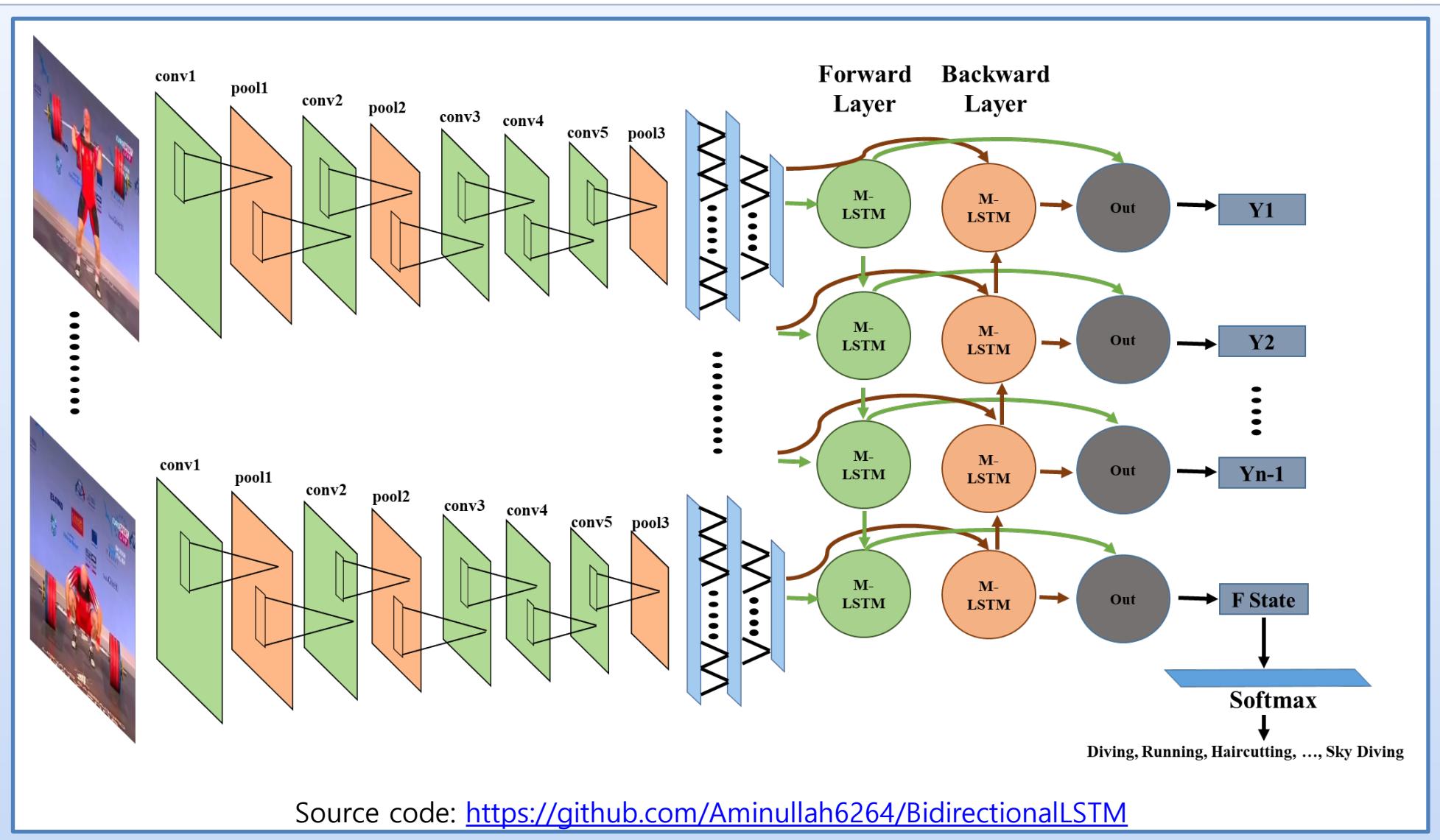
He said, “Teddy Roosevelt was a great President!”



# Bidirectional RNNs



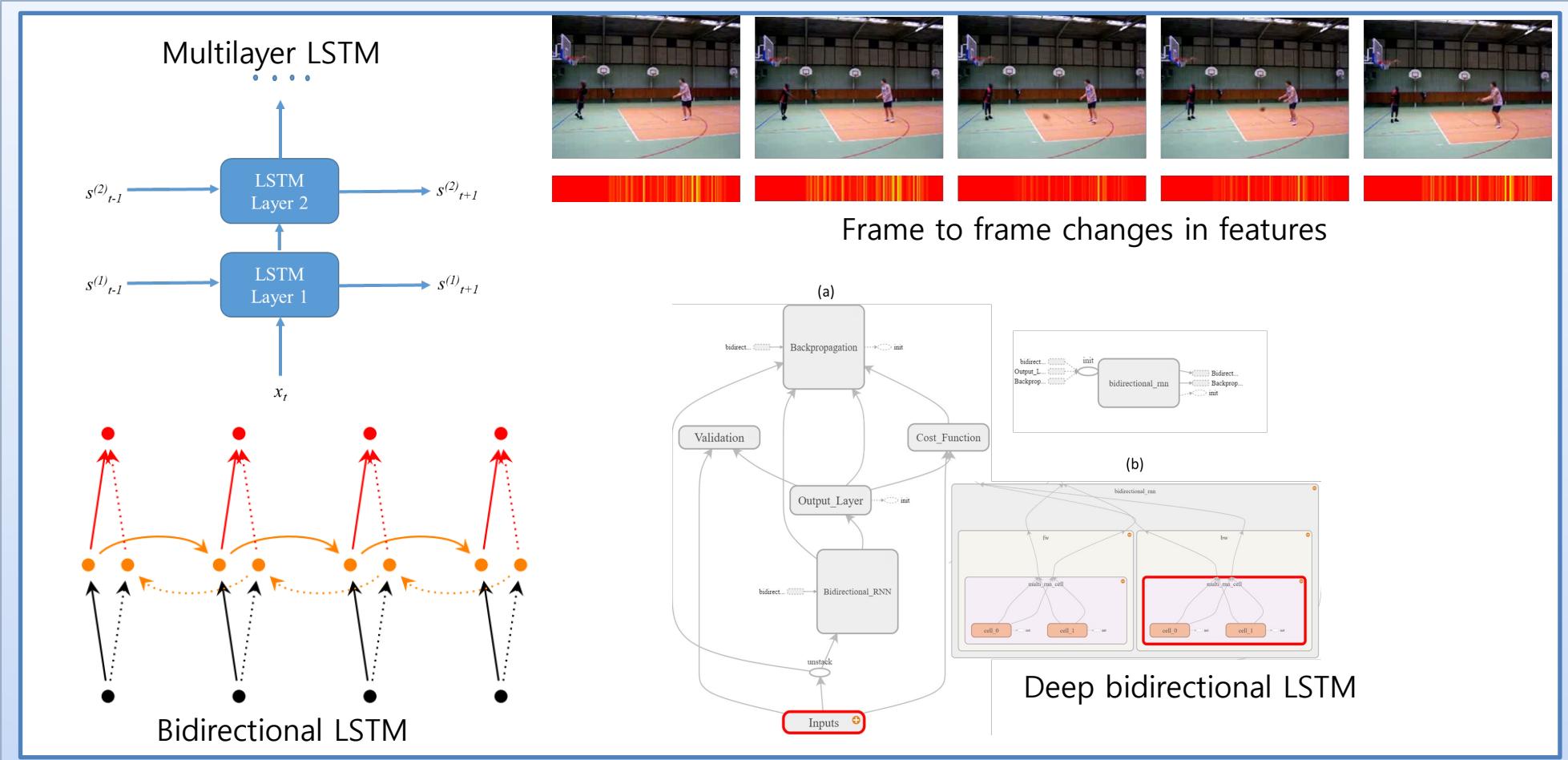
# 5. Action Recognition in Video Sequences using Deep Bi-directional LSTM with CNN Features [1]



# 5. Deep Bidirectional LSTM (DB-LSTM) [1]



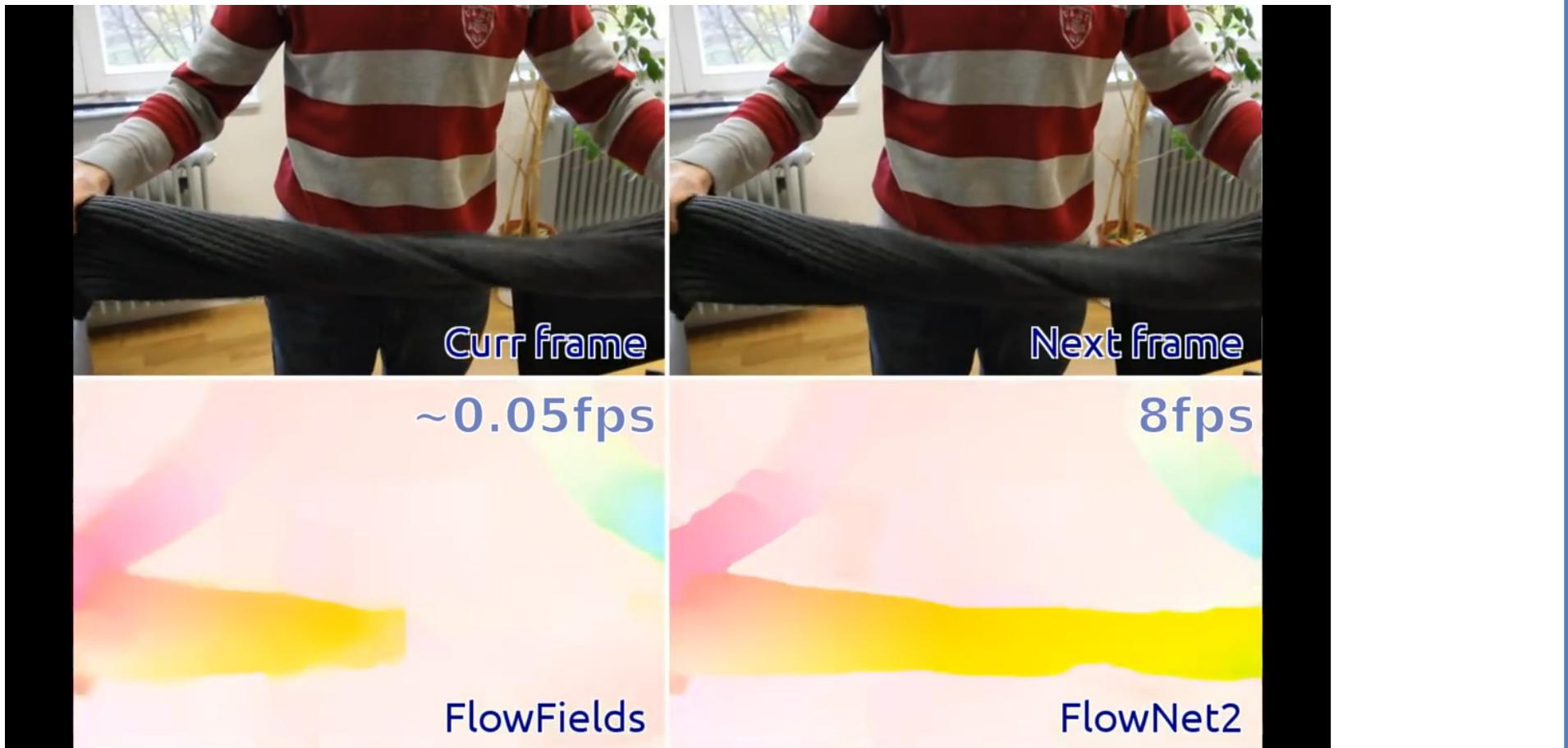
- Bidirectional LSTM:** process information in both directions with two separate hidden layers, which are then fed forwards to the same output layer, providing it with access to the past and future context of every point in the sequence.



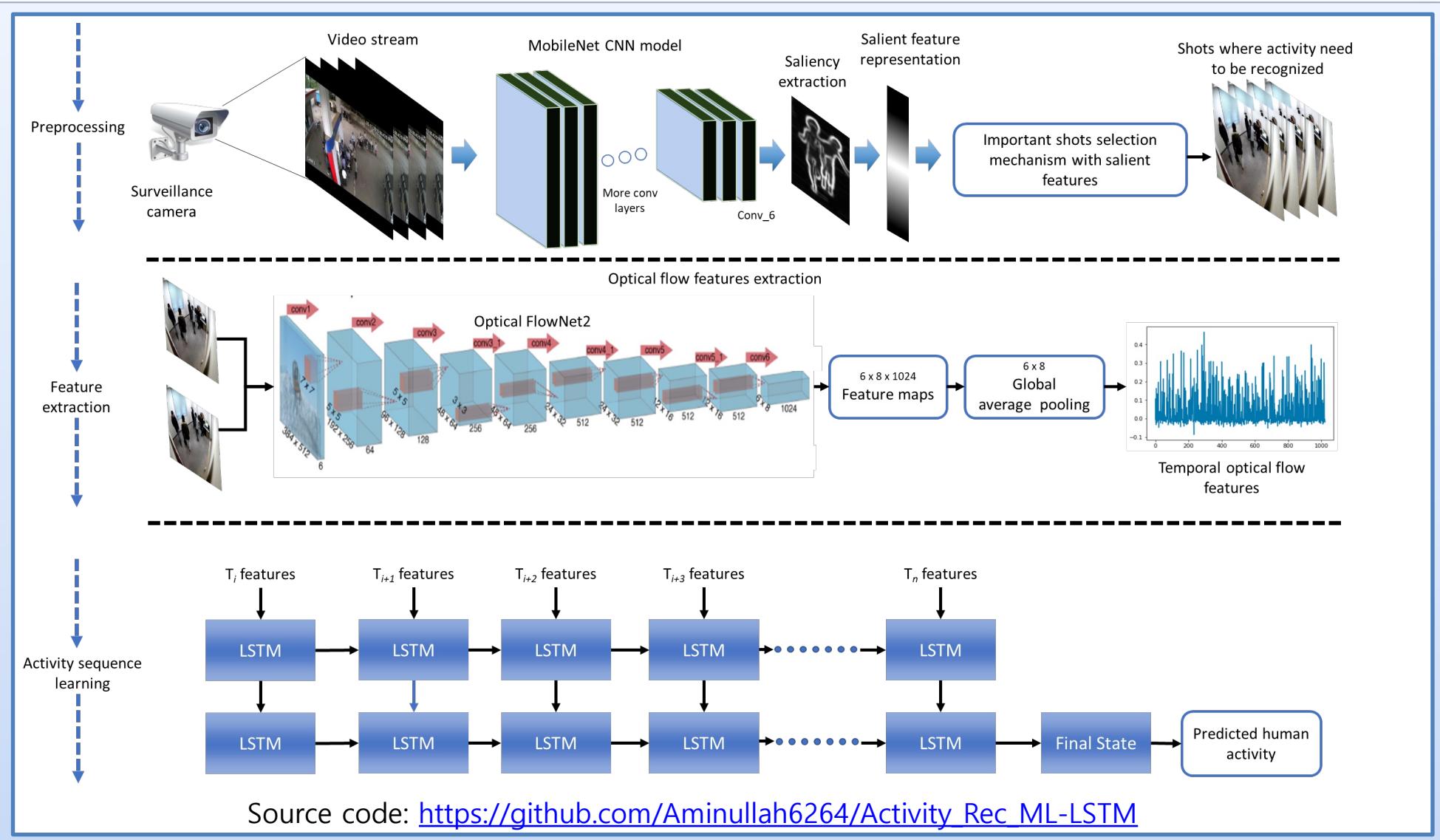
## 5. Activity Recognition using Temporal Optical Flow Convolutional Features and Multi-Layer LSTM [2]



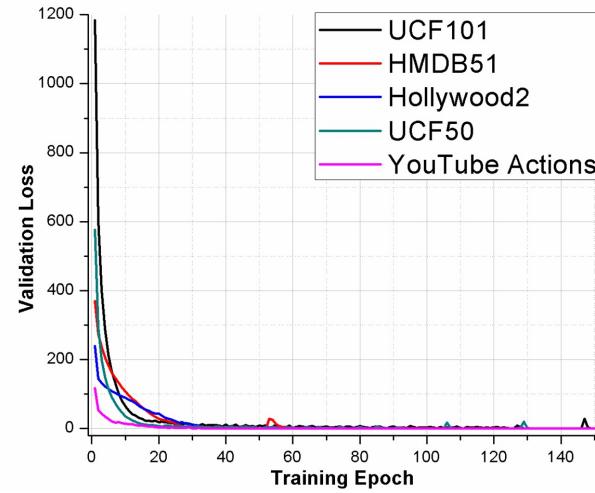
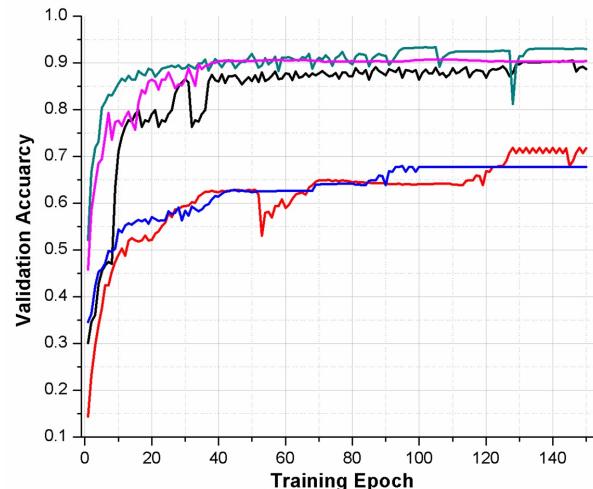
- Optical flow is the pattern of apparent motion of objects, surfaces, and edges in a visual scene caused by the relative motion between an observer and a scene.



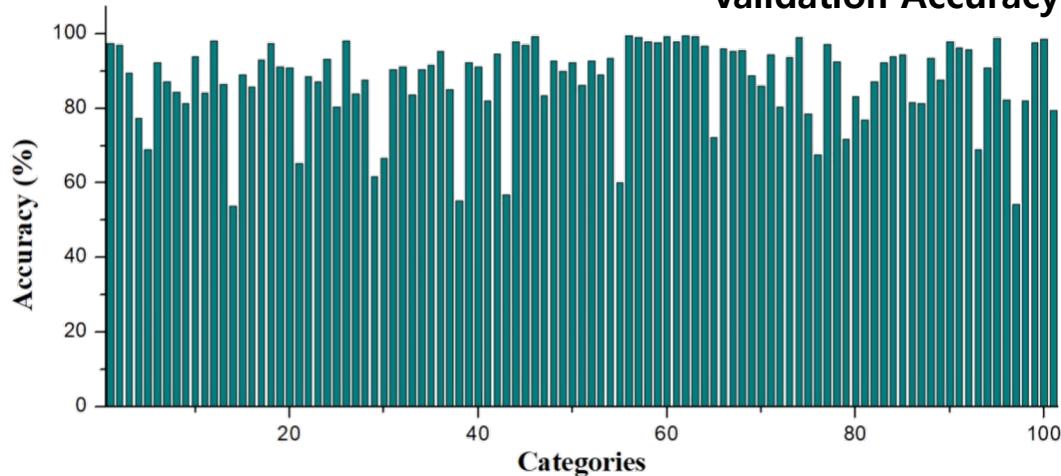
# 5. Activity Recognition using Temporal Optical Flow Convolutional Features and Multi-Layer LSTM [2]



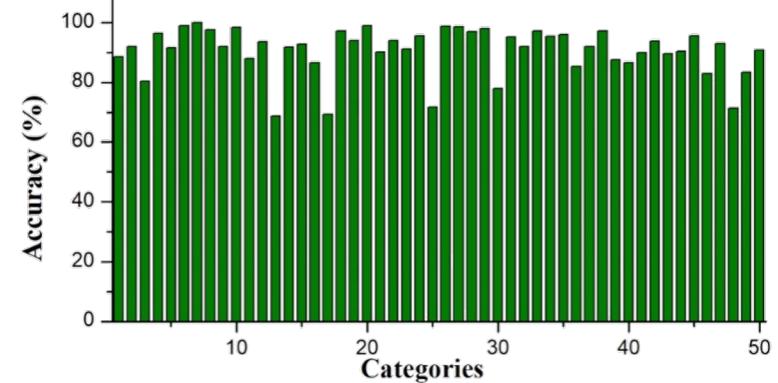
# 5. Activity Recognition using Temporal Optical Flow Convolutional Features and Multi-Layer LSTM [2]



**Validation Accuracy and Loss**

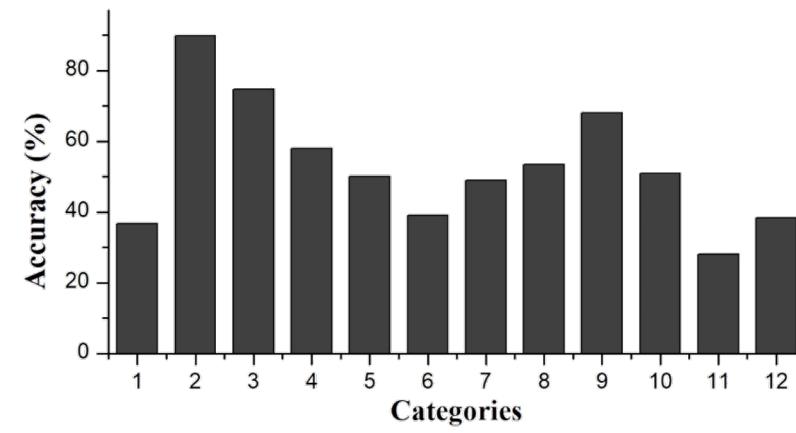
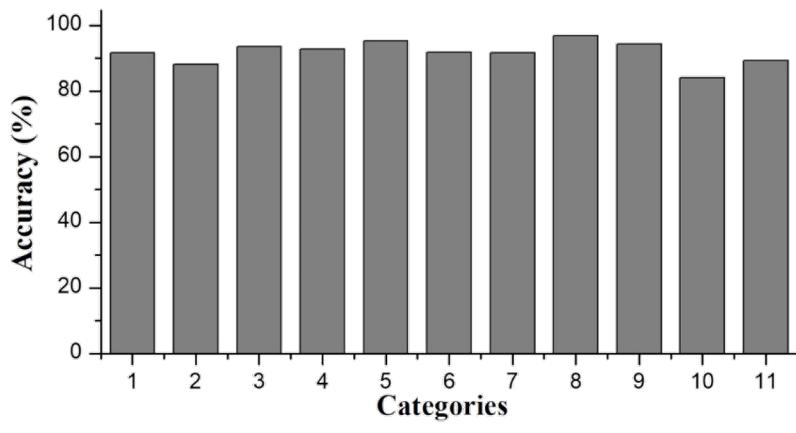
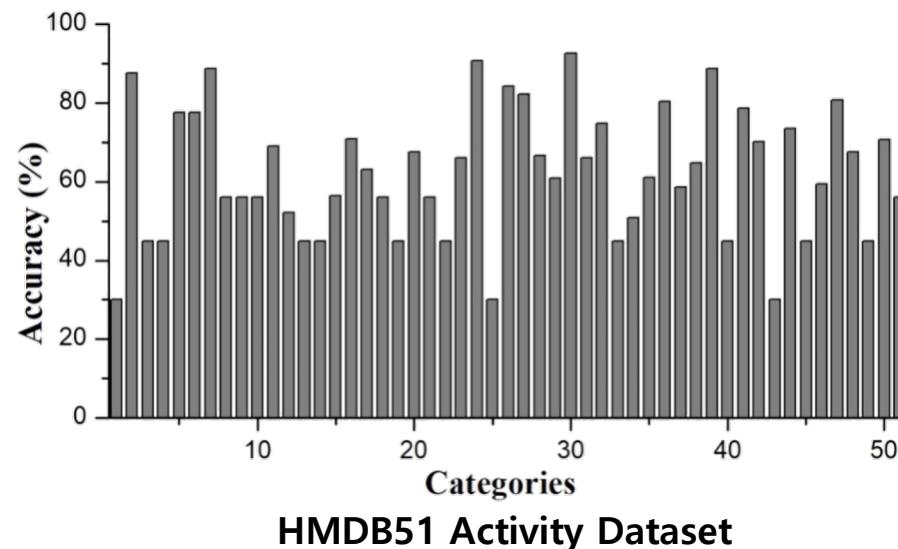


**UCF101 Activity Dataset**

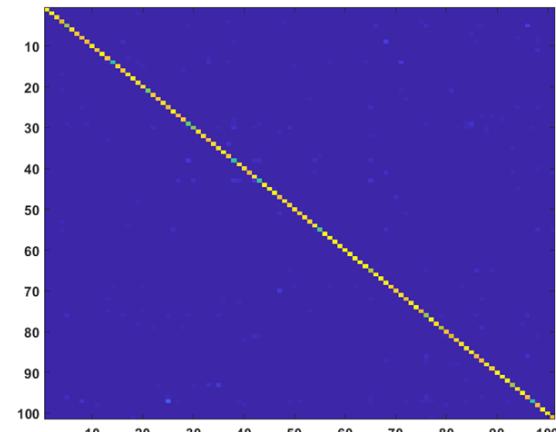


**UCF50 Activity Dataset**

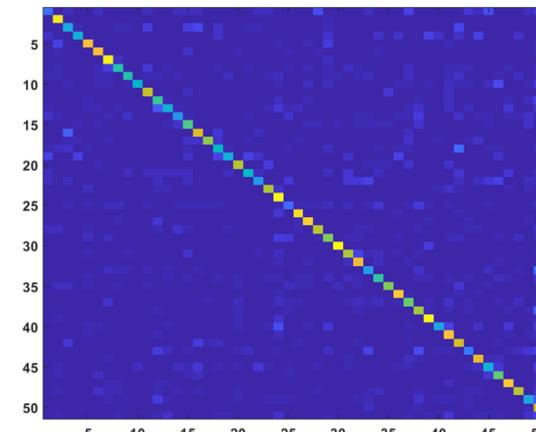
## 5. Activity Recognition using Temporal Optical Flow Convolutional Features and Multi-Layer LSTM [2]



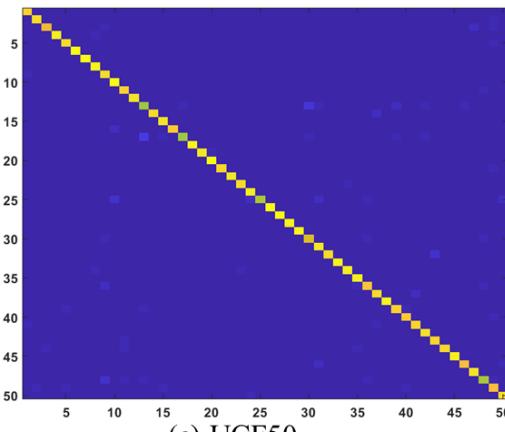
## 5. Activity Recognition using Temporal Optical Flow Convolutional Features and Multi-Layer LSTM



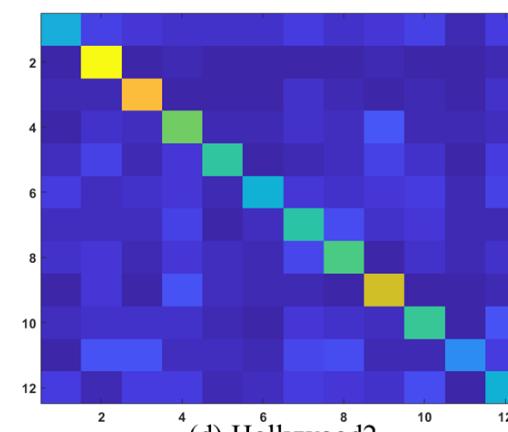
(a) UCF101



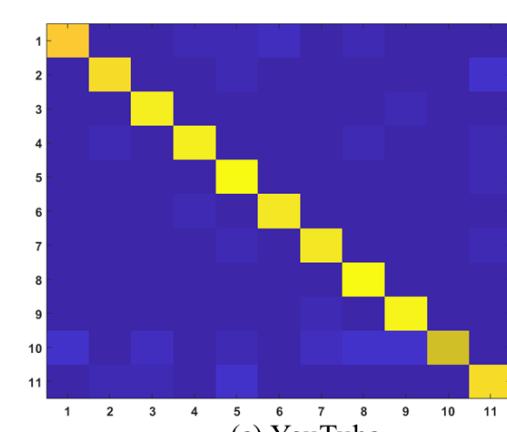
(b) HMDB51



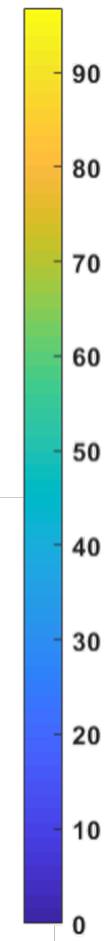
(c) UCF50



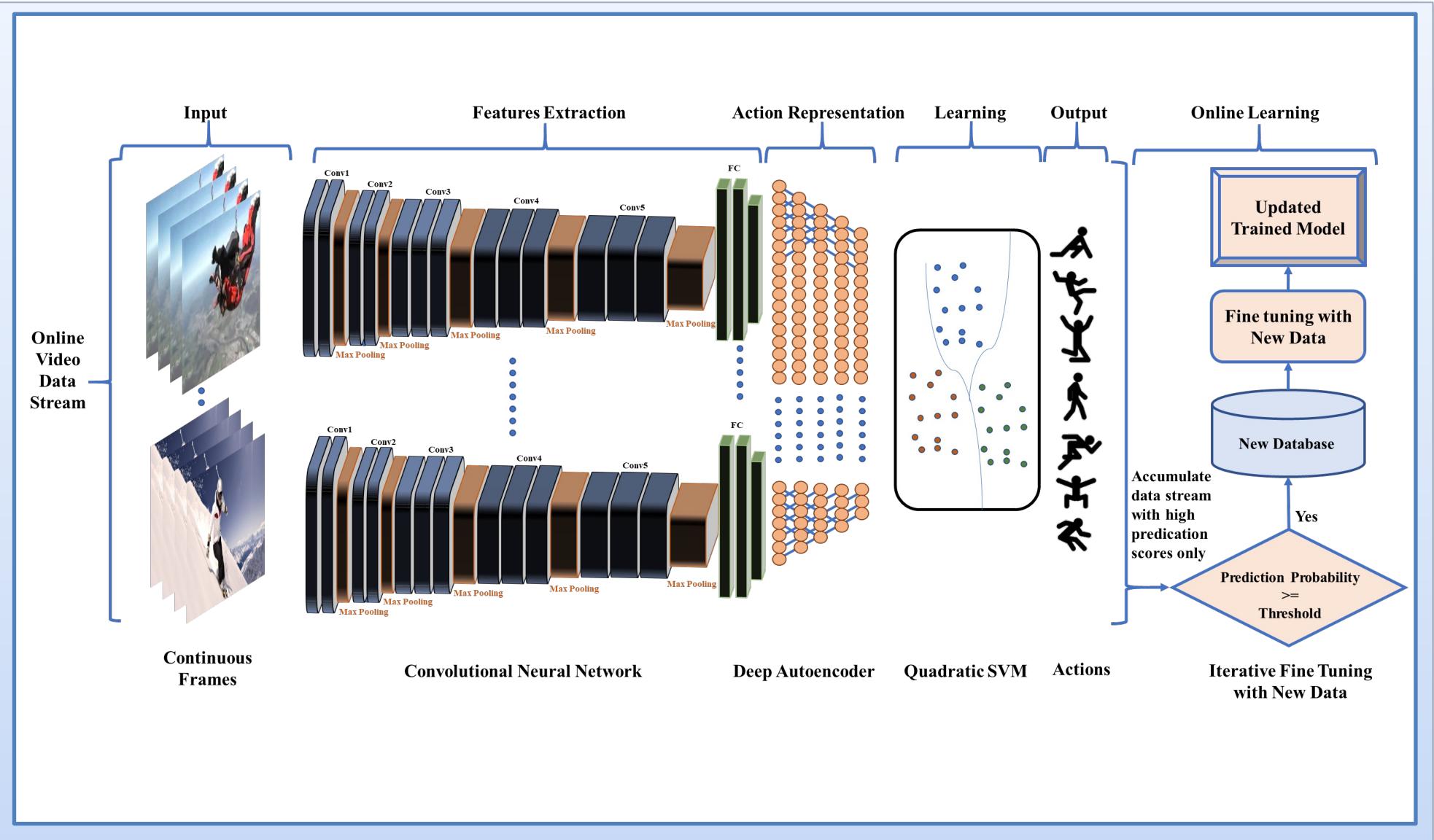
(d) Hollywood2



(e) YouTube

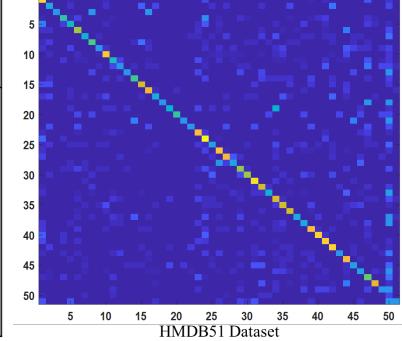
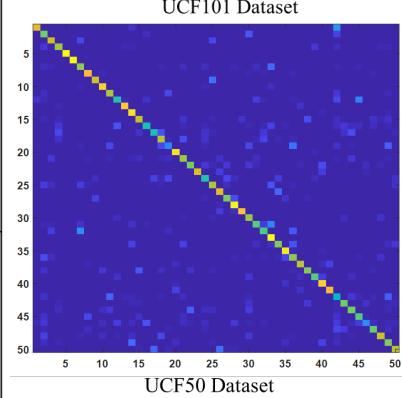
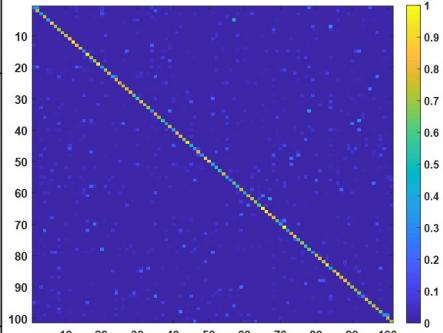


# 5. Action Recognition using Optimized Deep Autoencoder and CNN for Surveillance Data Streams of Non-Stationary Environments [3]



# 5. Action Recognition using Optimized Deep Autoencoder and CNN for Surveillance Data Streams of Non-Stationary Environments [3]

Intervals	Int. 1	Int. 2	Int. 3	Int. 4	Int. 5	Overall Accuracy
<b>Representative frame from each interval</b>						100%
<b>GT/Prediction</b>	Horse Ride / Horse Ride	Horse Ride / Horse Ride				
<b>Confidence score</b>	0.527	0.452	0.485	0.469	0.526	
<b>Representative frame from each interval</b>						100%
<b>GT/Prediction</b>	Baseball / Baseball	Baseball / Baseball	Baseball / Baseball	Baseball / Baseball	Baseball / Baseball	
<b>Confidence score</b>	0.500	0.483	0.333	0.333	0.385	
<b>Representative frame from each interval</b>						80%
<b>GT/Prediction</b>	Breast Stroke / <b>Biking</b>	Breaststroke / Breaststroke	Breaststroke / Breaststroke	Breaststroke / Breaststroke	Breaststroke / Breaststroke	
<b>Confidence score</b>	0.348	0.682	0.717	0.644	0.717	
<b>Representative frame from each interval</b>						80%
<b>GT/Prediction</b>	Dive / <b>Jump</b>	Dive / Dive	Dive / Dive	Dive / Dive	Dive / Dive	
<b>Confidence score</b>	0.312	0.519	0.549	0.509	0.492	



## 7. Demo Results



Class : Basketball

Confidence: 0.548792



## 7. Demo Results



Class : Surfing

Confidence: 0.356943



# 8. Major Challenges



## Actions and human variations and scales



- Human are of different sizes
- Intra- and inter-class variations
- People may appear at different scales in different videos yet perform the same action.



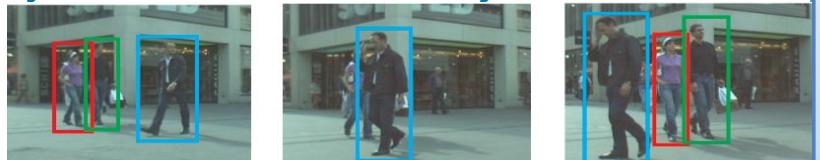
## Camera and objects movements

- The camera may be a handheld camera, and the person holding it can cause it to shake.
- Camera may be mounted on something that moves.
- The subject performing an action (i.e., skating) may be moving with the camera at a similar speed.



## Occlusions and Uneven Predictability

- Action may not be fully visible
- Insufficient annotated data
  - Although Youtube-8M and Sposrts-1M datasets provide millions of action videos, their annotations are generated by retrieval method, and thus may not be accurate



## Background “clutter”

- Other objects/humans present in the video frame.
- Different people perform different actions in different way

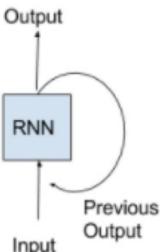
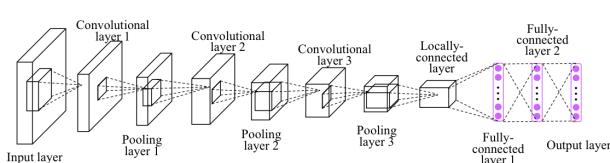


# 9. Future Research Directions



## Benefiting from image models

- Investigating light-weight CNN models for real time action recognition
- Integration with other modalities



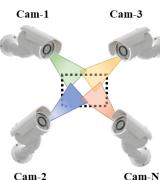
## Utilization of large-scale action dataset

- Real-world data utilization
- Group activity recognition
- Extension of unsupervised deep learning-based methods



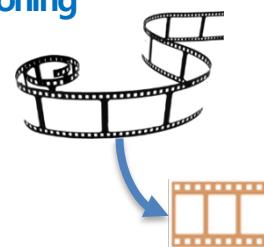
## Multiple actions and activity localization and recognition

- Localize and track activity in sequence of frames
- Recognizing multiple activity in one scene
- Action recognition in multi view data



## Interpretability on temporal extent

- Key frame selection for processing
- Reduced computational requirement and running time
- Intention Reasoning



Real-Time Processing

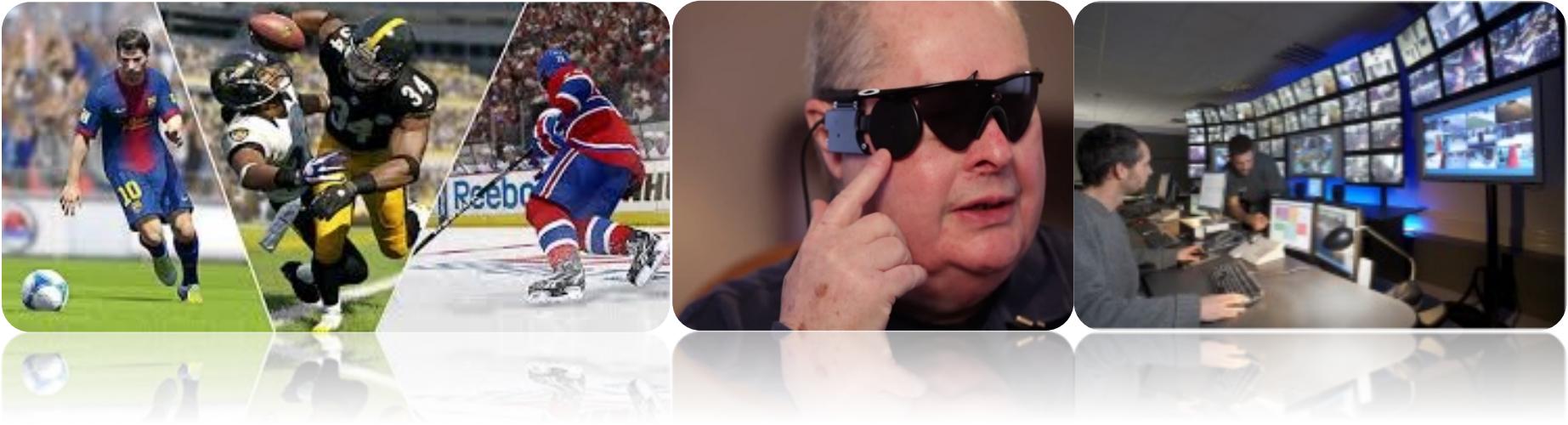
Less Than 1 Sec



## 10. References



1. Amin Ullah, Jamil Ahmad, Khan Muhammad, Muhammad Sajjad, and Sung Wook Baik, Action Recognition in Video Sequences using Deep Bidirectional LSTM with CNN Features, IEEE Access, Vol. 6: 1155 - 1166, 2018. [**IF: 3.244, Q1**].
2. Amin Ullah, Khan Muhammad, and Sung Wook Baik, Activity Recognition using Temporal Optical Flow Convolutional Features and Multi-Layer LSTM, IEEE Transactions on Industrial Electronics, 2018 (In Press). [**IF: 7.05, Q1**]
3. Amin Ullah, Khan Muhammad, Ijaz Ul Haq, and Sung Wook Baik, Action Recognition using Optimized Deep Autoencoder and CNN for Surveillance Data Streams of Non-Stationary Environments, Future Generation Computer Systems, 2018 (**Major Revision**). [**IF: 4.639, Q1**]



# Thank You

## Q&A

