AHMEDABAD UNIVERSITY
School of Engineering and Applied Science

Human Activity Recognition

Mid Semester Presentation

Prepared by: Group 07

Ashna Jain (201501008)
Dhruti Chandarana(201501015)
Charmi Chokshi (201501021)
Hena Ghonia (2015010132)
Janvi Patel (201501072)
Nishi Patel (201501076)

Subject: Computer Vision
Guided by: Dr. Mehul Raval
Date: September 19, 2018

Motivation

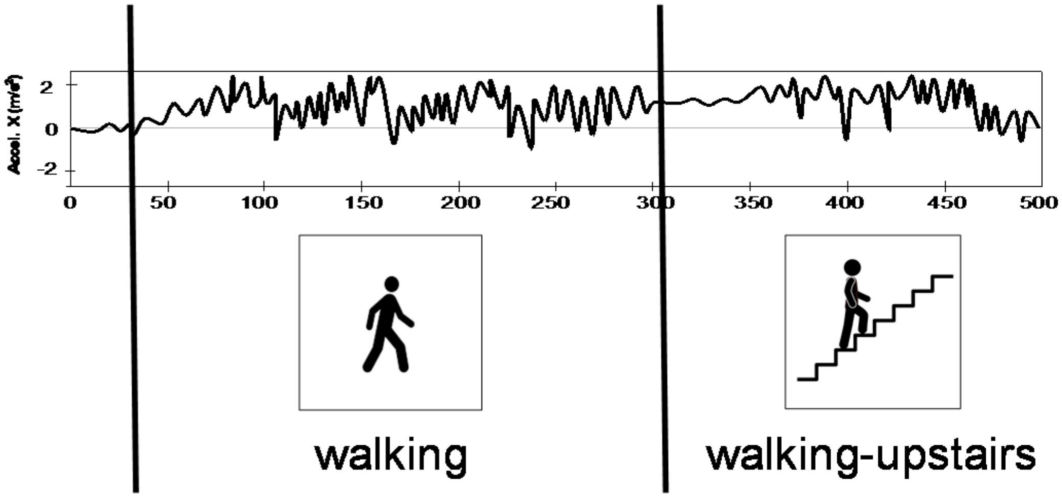
- Human Activity Recognition is one of the active research areas in computer vision for various contexts like
  - security surveillance
  - Healthcare
  - human computer interaction

- Due to rapidly increasing amount of video records, based on automatic video analysis such as visual surveillance, human-machine interfaces, sports video analysis, and video retrieval it is important to make the detection more accurate and robust.

Problem Statement

Deep Learning based Human Activity Recognition

- The objective is to classify activities into one of the activities performed.



Literature Review: 1

- **Two Stream Convolutional Networks for Action Recognition Videos by Karen Simonyan and Andrew Zisserman(2014) [6]**
  - Incorporates spatial and temporal networks.
  - Limited training data - ConvNet trained on multi-frame dense optical flow achieves good performance
  - Multitask learning- To decrease over-fitting, one could consider combining the two datasets into one

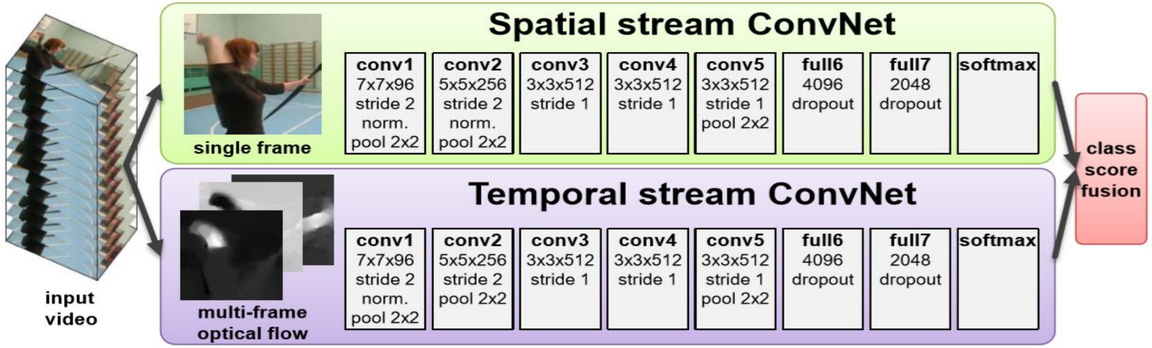
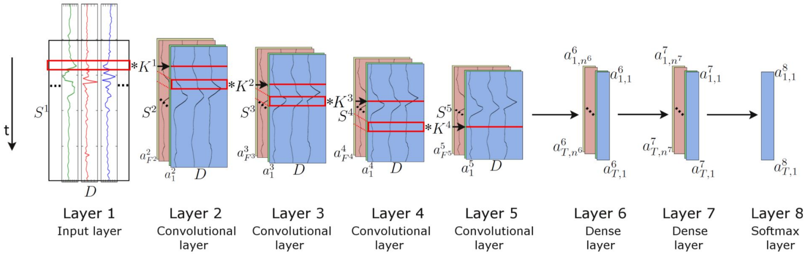


Figure 1: **Two-stream architecture for video classification.**

Literature Review: 2

- **Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. [2]**
- DeepConvLSTM and the baseline CNN : topology of the dense layers.
    - MLP: all input and output independent
    - LSTM with RNN: fed back to itself (weight and a unit time delay), provides a memory of past activations: learn temporal dynamics of sequential data.



LSTM recurrent cells: DeepConvLSTM

Non-recurrent and fully connected: baseline CNN

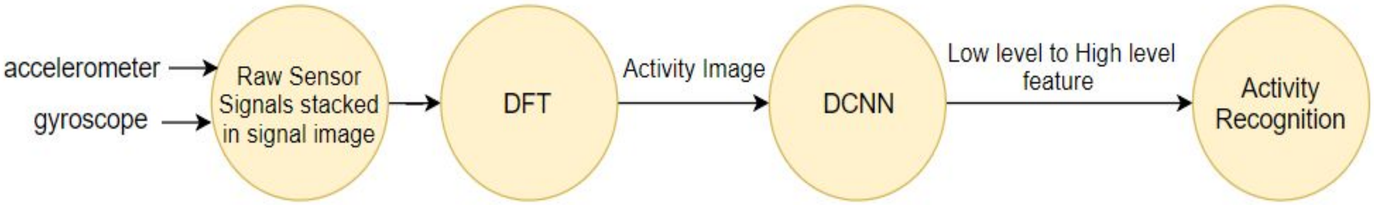
Datasets: OPPORTUNITY, Skoda;

Accuracy :88%, 71.3%

- Feature extraction on sliding windows approach: static and periodic activities.
- Template matching approaches, Hidden Markov modelling: sporadic activities.

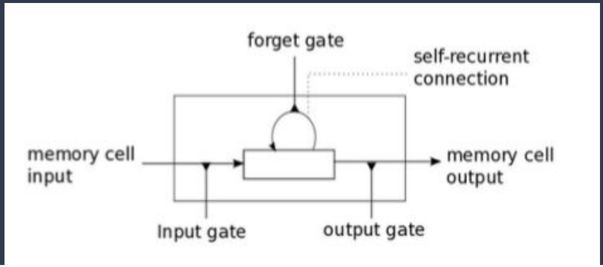
Literature Review: 3

- **Human Activity Recognition using Wearable Sensors by Deep Convolutional Neural Networks [3]**
- Human physical activity recognition based on wearable sensors has applications relevant to our daily life such as health care.
- Implement using wearable devices such as smart phones, smart watches and sport bracelets which embed accelerometers and gyroscopes.



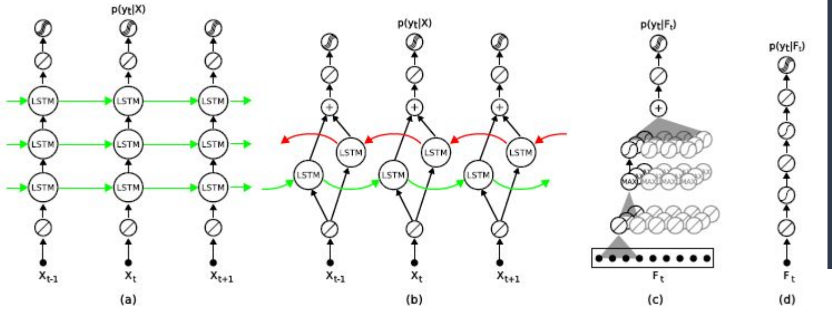
- The proposed DCNN architecture: Input layer - convolution layer1 (5*5 filter) - subsampling layer1 (4*4) - convolution layer2 (5*5) - subsampling layer2 (2*2) - fully connected layer - softmax layer

Literature Review: 4



- **LSTM Networks for Mobile Human Activity Recognition [4]**
- Feature extraction for AR is an important task.Parameters such as mean, standard deviation, entropy, Fourier transform are widely used features in AR.
- Long short-term memory (LSTM) is a recurrent neural network (RNN) architecture which is efficient in predicting results in situations when there are long time lags of unknown size.
- Acceleration of three directions into a three dimensional vector with a sliding window of length N.
- Feature extraction of accelerometer raw data through long short memory network with N timesteps. And finally a new feature vector is generated, which is classified by a multiclassifier.
- Dataset :Wireless Sensor Data Mining(WISDM).It has 1,098,207 records and 6 Class attributes recorded using tri-axial accelerometers.
- Accuracy : 92.1%

Literature Review: 5



- **Deep, Convolutional, and Recurrent Models for Human Activity Recognition Using Wearables[5]**
- Describes 3 different approaches for Human activity recognition (HAR): Deep Feedforward Network, Convolution Neural Network and Recurrent Neural Network.
- It includes comparative studies of suitability and impact of hyper parameters (fANOVA framework).
- CNN approach is suitable for short-term activities, RNN approach is suitable for longer activities.
- DNN is sensible to hyperparameters and require significant efforts to explore parameters.
- Experiments done using 3 different dataset: Opportunity Dataset, PAMP2, Daphent Gait.

Comparative study of Datasets

Table 3. Public HAR datasets (A=accelerometer, G=gyroscope, M=magnetometer, O=object sensor, AM=ambient sensor, ECG=electrocardiograph)

ID	Dataset	Type	#Subject	S. Rate	#Activity	#Sample	Sensor	Reference
D01	OPPORTUNITY	ADL	4	32 Hz	16	701,366	A, G, M, O, AM	(Ordóñez and Roggen, 2016)
D02	Skoda Checkpoint	Factory	1	96 Hz	10	22,000	A	(Plötz et al., 2011)
D03	UCI Smartphone	ADL	30	50 Hz	6	10,299	A, G	(Almaslukh et al., 2017)
D04	PAMAP2	ADL	9	100 Hz	18	2,844,868	A, G, M	(Zheng et al., 2014)
D05	USC-HAD	ADL	14	100 Hz	12	2,520,000	A, G	(Jiang and Yin, 2015)
D06	WISDM	ADL	29	20 Hz	6	1,098,207	A	(Alsheikh et al., 2016)
D07	DSADS	ADL	8	25 Hz	19	1,140,000	A, G, M	(Zhang et al., 2015c)
D08	Ambient kitchen	Food preparation	20	40 Hz	2	55,000	O	(Plötz et al., 2011)
D09	Darmstadt Daily Routines	ADL	1	100 Hz	35	24,000	A	(Plötz et al., 2011)
D10	Actitracker	ADL	36	20 Hz	6	2,980,765	A	(Zeng et al., 2014)
D11	SHO	ADL	10	50 Hz	7	630,000	A, G, M	(Jiang and Yin, 2015)
D12	BIDMC	Heart failure	15	125 Hz	2	>20,000	ECG	(Zheng et al., 2014)
D13	MHEALTH	ADL	10	50 Hz	12	16,740	A, C, G	(Ha and Choi, 2016)
D14	Daphnet Gait	Gait	10	64 Hz	2	1,917,887	A	(Hammerla et al., 2016)
D15	ActiveMiles	ADL	10	50-200 Hz	7	4,390,726	A	(Ravì et al., 2017)
D16	HASC	ADL	1	200 Hz	13	NA	A	(Hayashi et al., 2015)
D17	PAF	PAF	48	128 Hz	2	230,400	EEG	(Pourbabaee et al., 2017)
D18	ActRecTut	Gesture	2	32 Hz	12	102,613	A, G	(Yang et al., 2015)
D19	Heterogeneous	ADL	9	100-200 Hz	6	43,930,257	A, G	(Yao et al., 2017)

Reference: Wang, Jindong, et al. "Deep learning for sensor-based activity recognition: A survey." Pattern Recognition Letters (2018).

Comparison study of Literature review

Article name	Algorithm	Performance metrics	Dataset used
Two Stream Convolutional Networks for Action Recognition Videos [6]	Spatial ConvNet and temporal stream ConvNet -> fusion of two ConvNet by averaging and multiclass SVM.	Maximum accuracy achieved for UCF-101:88% and for HMDB-66.8%	UCF101 and HMDB-51
Deep convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition[2]	DeepConvLSTM for recurrent cells and Baseline CNN for non recurrent cells and fully connected network. MLP but input and output are independent; RNN used for fed back to itself and to provide memory of past activities. LSTM with RNN is used to learn temporal dynamic of sequential data.	Maximum accuracy for OPPORTUNITY - 88% Skoda - 71.3%	OPPORTUNITY (D01) Skoda (D02) mHealth (D13)
Human Activity Recognition using Wearable Sensors by Deep Convolutional Neural Networks [3]	Sensed signals from Accelerometer, gyroscope - From Raw signals stacked image - Discrete Fourier Transform - Activity image - DCNN - Low level to High level features learning - Activity Recognition	Maximum accuracy for UCI - 95.18 % USC - 97.01 % SHO - 99.93 %	UCI, USC, SHO

Comparison study of Literature review

Article name	Algorithm	Performance metrics	Dataset used
LSTM Networks for Mobile Human Activity Recognition [4]	LSTM+Multiclass classification	Accuracy : 92.1%	WISDM
Deep, Convolutional, and Recurrent Models for Human Activity Recognition Using Wearables[5]	Finding impact of hyperparameters using fANOVA framework. Three different approaches: CNN, DNN, RNN(bi-directional and forward in time) Dropout and Max-in-Norm regularization	Maximum accuracy for D04 dataset- CNN Approach- 93.7% D01 dataset- RNN(bi-directional) Approach- 74.5% D14 dataset- RNN(forward in time) Approach- 76%	D01, D04, D14

Data Set

- **UCI HAR dataset**
- The dataset was collected from the **in-built accelerometer and gyroscope of a smartphone** worn around the waist of participants.
- Number of subjects: 30
- Number of activities: 06
  - **Walking**
  - **Walking upstairs**
  - **Walking downstairs**
  - **Sitting**
  - **Standing**
  - **Laying**

CNN Network Structure

- **About the inputs**
- That dataset contains 9 channels of the inputs: (acc_body, acc_total and acc_gyro) on x-y-z. So the input channel is 9
- So in the end, we reformatted the inputs from 9 inputs files to 1 file, the shape of that file is [n_sample,128,9], that is, every windows has 9 channels with each channel has length 128

- Convolution + pooling + convolution + pooling + dense + dense + dense
- learning_rate = 0.001
- dropout = 0.8
- training_epoch = 20
- kernel_size = 64 (total 32)

Output

dropout	LR	epoch	train_acc	test_acc
0.7	0.01	100	0.894	0.801
0.7	0.001	10	0.788	0.715
0.8	0.001	10	0.843	0.741

```
Epoch:19,batch:1616,loss:3861.32617188,accuracy:0.81250000
Epoch:19,batch:2416,loss:3103.45263672,accuracy:0.93750000
Epoch:19,batch:3216,loss:0.00000000,accuracy:1.00000000
Epoch:19,batch:4016,loss:2461.26611328,accuracy:0.87500000
Epoch:19,batch:4816,loss:0.00000000,accuracy:1.00000000
Epoch:19,batch:5616,loss:10780.86328125,accuracy:0.87500000
Epoch:19,batch:6416,loss:6024.33496094,accuracy:0.93750000
Epoch:19,batch:7216,loss:0.00000000,accuracy:1.00000000
Epoch:20,batch:016,loss:0.00000000,accuracy:1.00000000
Epoch:20,batch:816,loss:0.00000000,accuracy:1.00000000
Epoch:20,batch:1616,loss:0.00000000,accuracy:1.00000000
Epoch:20,batch:2416,loss:9006.54785156,accuracy:0.68750000
Epoch:20,batch:3216,loss:0.00000000,accuracy:1.00000000
Epoch:20,batch:4016,loss:1838.11120605,accuracy:0.81250000
Epoch:20,batch:4816,loss:4942.12695312,accuracy:0.93750000
Epoch:20,batch:5616,loss:0.00000000,accuracy:1.00000000
Epoch:20,batch:6416,loss:0.00000000,accuracy:1.00000000
Epoch:20,batch:7216,loss:0.00000000,accuracy:1.00000000
Optimization finished!
Accuracy of testing:0.83338988
```

Future Work

- Modification of the current CNN based model for better accuracy
- Implementation of RNN-LSTM based approach
- Create threads by exploiting multiple cores
- Thresholding in the system which restricts assigning label to an unknown activity (outlier detection)

References

[1] Wang, Jindong, et al. "Deep learning for sensor-based activity recognition: A survey." Pattern Recognition Letters (2018).

[2] Ordóñez, Francisco Javier, and Daniel Roggen. "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition." Sensors 16.1 (2016): 115.

[3] Jiang W, Yin Z Proceedings of the 23rd ACM international conference on Multimedia. Human Activity Recognition using Wearable Sensors by Deep Convolutional Neural Networks[C] ACM, 2015:1307-1310.

[4] Yuwen Chen, Kunhua Zhong, Ju Zhang, Qilong Sun and Xueliang Zhao, "LSTM Networks for Mobile Human Activity Recognition" in International Conference on Artificial Intelligence: Technologies and Applications (ICAITA 2016).

[5] Hammerla, Nils Y., Shane Halloran, and Thomas Ploetz. "Deep, convolutional, and recurrent models for human activity recognition using wearables." arXiv preprint arXiv:1604.08880(2016).

[6] Simonyan, Karen, and Andrew Zisserman. "Two-stream convolutional networks for action recognition in videos." *Advances in neural information processing systems*. 2014.

Thank you!