Activity Recognition Using Convolution Neural Network In Smart Home.

P.Rajesh[1], N.Manjunathan[2], S. Gopi[3], A.Suresh[4]

[1,2] Assistant Professor, Department of Computer Science and Engineering,
Vel Tech Rangarajan Dr.Sagunthala R & D Institute of Science and Technology,
Chennai, Tamilnadu, India
[3] Assistant Professor, Department of Computer Science and Engineering, Kingston Engineering College, Vellore, Tamilnadu, gopi.scse@gmail.com
[4] Associate Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India. prisu6esh@yahoo.com

[1]raji.maghudam08@gmail.com, [2]nmanjunathan24@gmail.com

ABSTRACT

Now a days Security and health care activities are the major issues in Smart homes.Vision based action monitoring has been a serious issue in smart homes.Activity recognition supports in various applications like health care,elderly monitoring,Safety,Social networks analysis,monitoring the environment,transportaion monitoring,surveillance systems etc.Hence the most booming technology ,deep learning is used to monitor the human activity .To perform the above experiment a Convolution Neural Network model is used to recognise the human acivity present in a Smart home.The model is tested and trained with a large data set which has large volume of video data collected .From the experiment a significant output is acheived for the inputs provided which ranges from low level to high level quality of inputs.DML smart actions is a popular data set and it used to test our model.The results obtained from the proposed system was around 82.4% in measuring the accuracy rate of the human activity ecognition in smart homes.

I **INTRODUCTION**

The need and cause of human activity recognition are due to various applications like health care,Elderly monitoring,Safety,Social networks analysis,monitoring the environment,transportaion monitoring,surveillance systems etc.The most common need for the experiment is to take care of the elderly people living alone in home.Their activities has to be monitored to provide medical asssistance,ven they stay far away from home.This type of monitoring also is useful to identify if there is any deviations in the normal and routine activities of elderly people.If any abnormalities are found then necessary steps could be taken to help the elderly people.In some cases even hand gestures has to be learnt and identified to know the real meaning of hand movement the people living in a home independently.

Another major application of Human activity recognition is Security and surveillance applications,even though it is beyond the scope of the paper,activity recognition appliaction is the intention of this experiment.Traditional way of surveillance are dine by human.This continous monitoring would lead to stress to the one who performs monitoring.Hence avision based monitoring would be a better choice.The usage of sensors may lead some disadvantages like sensor failures,calibration of sensors etc.So a visual based activity monitoring is used in this analysis in a smart home.We use a Convolution Neural network to classify the activity which was gathered from a smart home.DML smart action data set is used to test and train the input.

Related Works

Smart home will play a vital role in future in providing and performing intelligent tasks.It not only controls machine activities like automatic electronics equipment switching ,sensing smoke ,making alarms,opening and closing doors.It has its own kind of action in human activity monitoring and recognising the occupants through the deep learning techniques.hence by recognising the activities decision can be made in accesing the devices based on the activity performed by the occupant in the smart home.Therefore utilising machine learning models in in applications like health care,elderly monitoring etc has become more significant[1].

These activity recognition can be made through many technical ways like incorporating sensors,deploying monitoring sensors,incorporating wearable sensors to the body of the occupants.These kind of monitoring using vaarious sensors would be annoying to the occupants and might produce disturbnaces like producing noise,may give wrong alarm,breaking of sensors,sensor need to calibrate sensors in regular intervals etc.This may lead to improper and inaccurate results[2].Hence to overcome the deficiencies mentioned in collecting the samples for classification,vision based methods are used to monitor the activities of human in smart home.Vision based methods are more advantageous when compared with the sensor based because of the above said deficiencies.The samples in vision based arecollected using various types of diverse cameras to obtain more accurate and appropriate input than the sensor based sample collection.[5].

While performing input classification in traditional methods used handcraft methods are used with the machine learning models.The obtaine videao frames are considered as Histogram,scale invariant images,which is of low level range and would be accepted by limited dataset only,which would be a challenging task to acheive high accuracy[6].[7] used spatiotemporal features for activity recognition which used the same DML smart action data set.[9] has used Sparse coding have been used to form visual words for extracting spatiotemporal features.Non negative sparse coding method was used in [10].

Linear SVM and Intersection SVM is used to classify the data ste and it acheived 58.20% of accuracy.in [11] a support vector machine classifier was proposed which used meta classifier and Bayesian classifier.The accuracy rate was around 72.3%.For measuring the similarities images and videos a kernel based functionnis used in[14],which is combined with SVM and NNSC algorithm and it acheived highest accuracy rate of 79.9%.

In recent days machine learning models have produced highest percentage of accuracy wherever it is used,more specifically in areas like image classification.Neural networks like Deep learning classifier,Convolution Neural Network,Recurrent Neural Network,Long Short term Memory classifier are the subset of Machine learning techniques.As mentioned above theses models does not need any handcrafted inputs and could segment and normalise raw input images and could produce higher accuracy from low level to high level input images.[16].[17] has used Deep Belief Network network model to classify for fall detection images ,which was obtained from video inputs taken from a smart home.It produced accuracy rate of 79.4%.

Wang et al used CAD120 dataset to classify their two data sets namely OA1 nad OA2 with the derived CNN model designed for human activity recognition.It managed to get accuracy rate of 60.1% and 45.2% for OA1 and OA2 dataset respectively.[20]. [21] has used fusion methods with the kitchen dataset along with Neural networks and SVM which used fusion methods to classify the inout images.It produced accuracy rate of 73.1%.Monterio et al used the same methodologies but with different classifier models like GoogleNet,AlexNet and SqueezNet and produced the output accuracy with the rate of 78.5%[22].

Paper Organisation

To understand and to analyse the Human activity recognition model using CNN classifier the paper is organised as follows.Section II of this paper contains the Experimental set up and data processing used to conduct the investigation.Section III contains proposed architecture and discussion about the results obtained and concluded in Section IV.

II **EXPERIMENTAL SET UP AND DATA PROCESSING**.

This section contains the details of how the experimental set up is made and how the data are processed to conduct the experiment for Human activity recognition in smart home.A low cost video camera is fixed in a home to monitor the activity of the occupant for around 15 days.The visual images are collected and stored in a memory device for evauating and classifying the input images.These images should be fed into the model ,these video images are cropped with the needed specification so as to fit into the CNN model.To acheive higher efficiency Region of Interest is set to the images,which covers only the occupant image and uncovers the background shades.Then these images are cropped and segmented from 600 X 320 X 3 to 220 X 220 X 3 which better suits the CNN model 's input requirement.The images are collected in live mode ,no temporal images are used for classification as done by other classification.Fig.1 shows a simple CNN classifier used for this experiment.

A.CNN Architecture

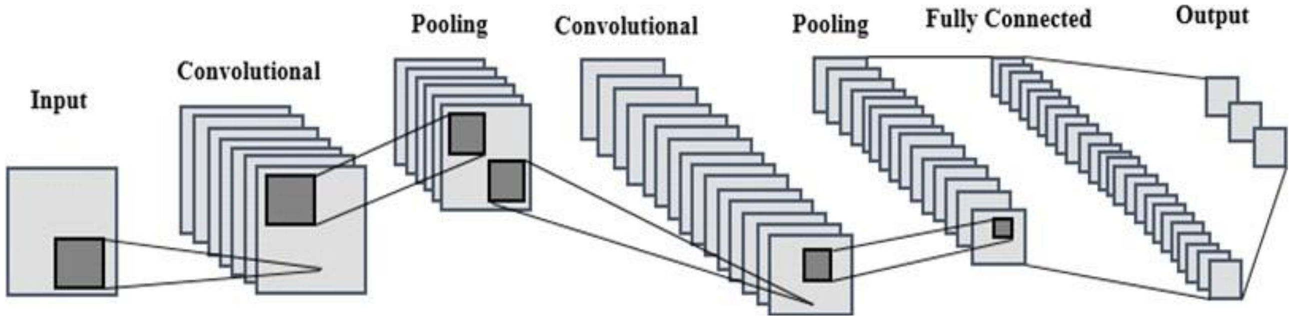


Fig.1 General CNN architecture

The CNN model mimics the biological neural network system ,when used with classifiers like AlexNet,GoogleNet etc.CNN model contains series of common layers called as Convolution layer,pooling layerand fully connected layer.The first layer,Convolution layer are used for calculating the weights.The layer is contains a feature map which is connected with region of previous layer called as Filter banks or kernels.The calculated weight mostly its sum goes through a non linearity function ,for example ReLu.

$$Y_i^l = B_i^l + \sum_{j=1}^{m_l^{(l-1)}} F_{i,j}^{\ l} * W_j^{(l-1)} \qquad (1)$$

The output obtained from the convolution layer is given by (1) where F represents kernel size,m represents feature map,B represents bias,Wj represents filter's weight,Yli is the output,such that j is the feature map of ith layer 1.

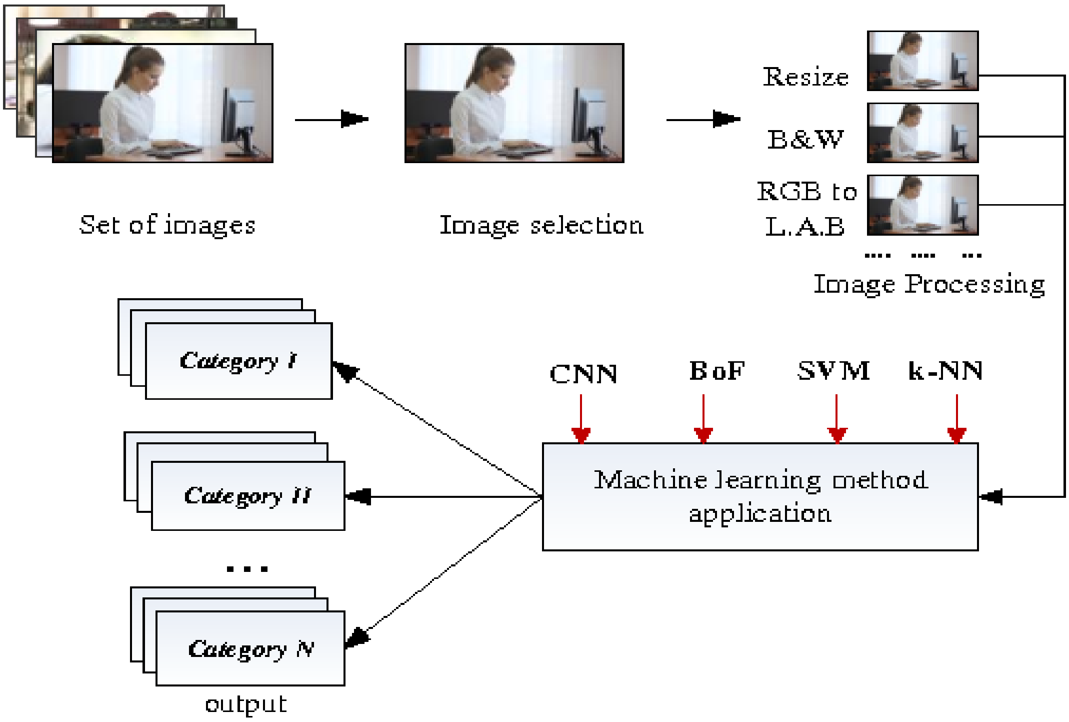


Fig.2.Proposed Architecture

The second layer is the pooling layer of CNN model ,which is used to reduce the dimension of the spatial input of the first layer without changing the depth of the convolution layer.Prevention of overfitting reduction in the training process is the major advantage of the pooling layer.This layer contains max and min operaions in which max pooling layer has produced reasonable results in most areas.The height and the width of the pooling layer is calculated as (2) and (3).W1,H1,D1,F and S are considered as width,height,depth,kernel size and stride size respectively.

$$W_2=\left(\frac{W_1\text{-}F}{S}\right)+1 \qquad (2)$$

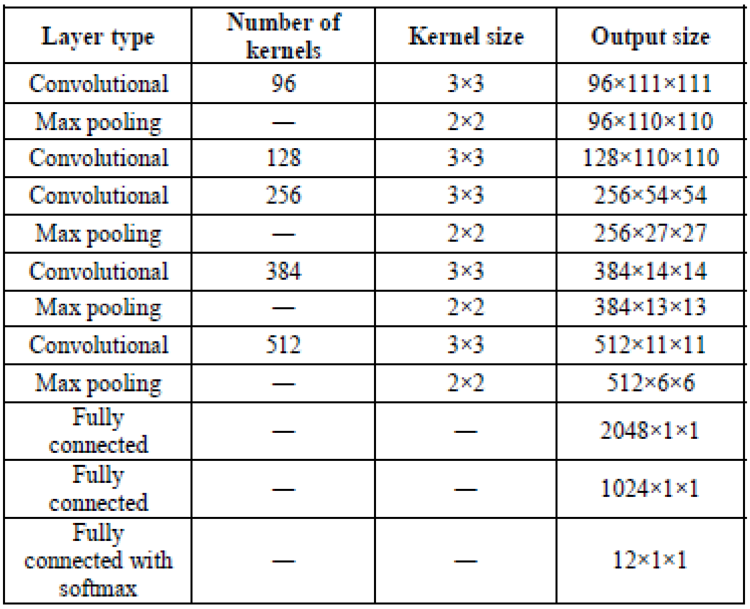
$$H_2=\left(\frac{H_1\text{-}F}{S}\right)+1 \qquad (3)$$

The third layer is the fully connected layer,is used to calculate the scores of the dataset layer,a softmax function is used to calculate the probablity of the inout labels.It is represented by (4).Table I shows the Layer perceptions of the CNN model.

$$f(z)_i=\frac{e^{x_j}}{\sum_{k=1}^{K} e^{x_k}} \qquad \text{For } j=1,\dots,K \qquad (4)$$

It contains various the technical aspects of the various layers present in the investigation.The last column of the table finally predicts the size of the label that are produced as output based on the input images obtained from the data set and also from the real images collected from the smart home.

Table I .Layer Perception of the proposed Model.



Layer type	Number of kernels	Kernel size	Output size
Convolutional	96	3×3	96×111×111
Max pooling	—	2×2	96×110×110
Convolutional	128	3×3	128×110×110
Convolutional	256	3×3	256×54×54
Max pooling	—	2×2	256×27×27
Convolutional	384	3×3	384×14×14
Max pooling	—	2×2	384×13×13
Convolutional	512	3×3	512×11×11
Max pooling	—	2×2	512×6×6
Fully connected	—	—	2048×1×1
Fully connected	—	—	1024×1×1
Fully connected with softmax	—	—	12×1×1

B.Dataset.

We use the DML smart action dataset to complete our investigation.The dataset is formed by coinducting test in wo different smart homes with around 17 people,captured through four cameras.Some examples of the images obtained from the videos is as shown in Fig 3.Various activities are considered for making this investigation

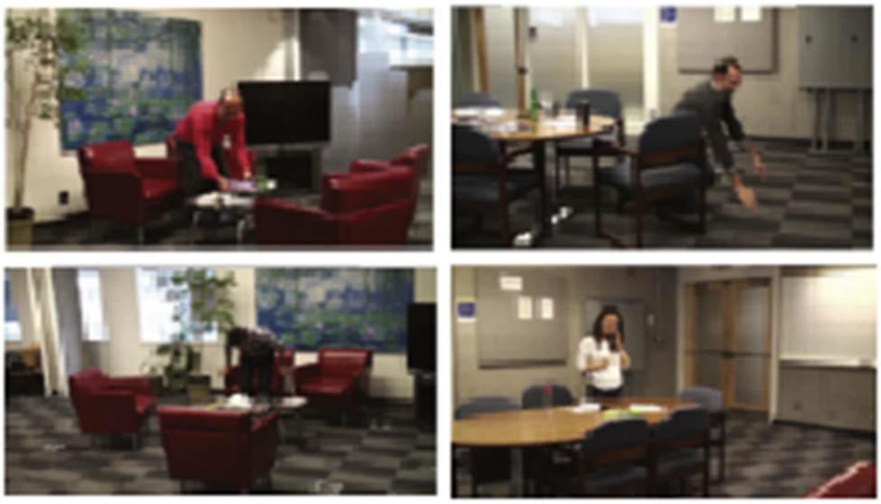
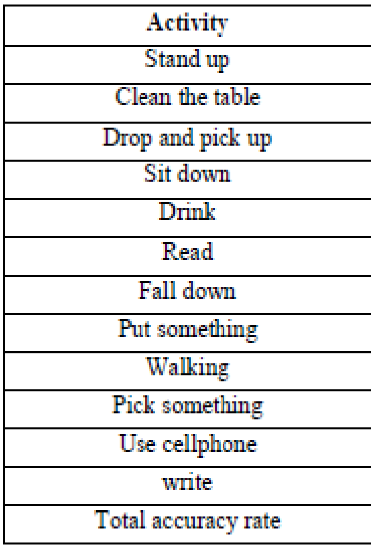


Fig.3 Sample images collected from smart home.

To perform the investigation and to measure the accuracy of the experiment,all possible activities are considered .That is the person in the smart home is allowed to perform all common activities an occupant commonly performs.Some of them are standing,walking,eating, sleeping,siting,walk stair case,using mobile phones,pick ,drop something etc..Table II shows the sample activities considered for analysis.

Table II Sample activities considered for the Investigation.



Activity
Stand up
Clean the table
Drop and pick up
Sit down
Drink
Read
Fall down
Put something
Walking
Pick something
Use cellphone
write
Total accuracy rate

C Metrics used for Evaluation.

The output is evaluated using the metrics ,like Accuracy,Recall or sensitivity,Precision and the method how they are derived are explained as follows.

Accuracy, rate of exactly classified samples within all the inputs.(5).

$$\text{Accuracy}=\frac{\sum_{i=1}^{n} N_{ii}}{\sum_{i=1}^{n}\sum_{j=1}^{n} N_{ij}} \qquad (5)$$

Recall or sensitivity proportion of the accuracy of the live images (6)

$$\text{Recall}_i=\frac{N_{ii}}{\sum_{k=1}^{n} N_{ik}} \qquad (6)$$

Precision proportion of live labels that are actually assumed (7)

$$\text{Precision}_i=\frac{N_{ii}}{\sum_{k=1}^{n} N_{ki}} \qquad (7)$$

III EXECUTION AND ANALYSIS.

The complete experimental set up and the architecture of the investigation is as shown in Fig 2.The video images are collected from the smart home and those videos are converted into images so as to fit into the input layer of the CNN model. These images should be fed into the model ,these video images are cropped with the needed specification so as to fit into the CNN model.To acheive higher efficiency Region of Interest is set to the images,which covers only the occupant image and uncovers the background shades.Then these images are cropped and segmented from 600 X 320 X 3 to 220 X 220 X 3 which better suits the CNN model 's input requirement.The images are collected in live mode ,no temporal images are used for classification as done by other classification.Data recording is done by an Intel Core i7 2.5GHz CPU, and the codes are written in python dependent on the Windows . To imrove the computational efficiency the video images are converted into 640 X 360 X 3 image format. Like other activity recognition models our experiment does not use any temporal images for identifiying the activities.All the images are collected in live mode and tested.By using the evaluation metrics the data obtained are preprocessed.The original data preprocessing is checked for itsaccuracy after the execution of the model with the DML smart action dataset.

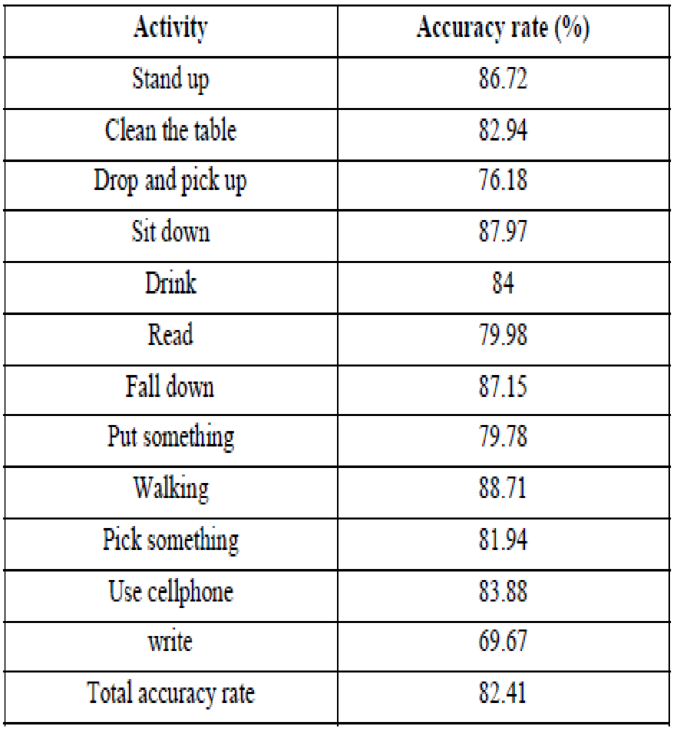
In order to classify the images for activity recognition ,the DML smart action dataset are preprocessed.The original image frames obtained are set to around 32 fps of 600 X 460 pixels.This experiment is conducted through a simple visual camera and around 52 videos samples were collected from where 75430 images were received for the investigation.Also the original size of the image frames are compressed to 330 X 276 pixels then 226 X 226 cropped images are taken for the analysis.75% of the dataset images are used for training and remaining 25% is used for testing.After the analysis the proposed CNN model has outflanked with 82.4% of the accuracy rate.

.

A.**Results and Discussion**

From the results obtained,the Human activity recognition from smart home using CNN model has managed to produce the accuracy of image classification around 82%.The accuracy rate for various activities obtained are shown in Table III.The proposed architecture was successful and it was built using five Convolution layers.Four pooling layers and finally three fully connected layers.A powerful softmax function is used to find the probablity of the inout cases for image classification .In our investigatuon around 12 activities are considered for the classificationThe fully connected layers and convolution layer in some cases may produce non linear output,this non linearity is flattened by a separate ReLu activation function.The inputs are managed by a 3 x 3 kenerl size and the first convolution layer applies 96 filters to the inputs.Down sampling is done by the max pooling layer for getting the output from the previous convolution layer without any difference in the depth size,this could be acheived by applying a 2 X 2 kernel size.As shown in Table III,highest accuracy rate was obtained for walking and writing action could get the lowest accuracy rate.The overall accuracy rate of the human activity recognition is 82.4%.

Table III Accuracy rate obtained for the Activities



Activity	Accuracy rate (%)
Stand up	86.72
Clean the table	82.94
Drop and pick up	76.18
Sit down	87.97
Drink	84
Read	79.98
Fall down	87.15
Put something	79.78
Walking	88.71
Pick something	81.94
Use cellphone	83.88
write	69.67
Total accuracy rate	82.41

IV CONCLUSION AND FUTURE ENHANCEMENT

In this work, a human activity recognition in smart home framework based on the profound CNN model is proposed.The accuracy rate of image classification is compared with various Machine learning models and managed to get around 82% of accuracy which is better than the existing models discussed in related work.The highest accuracy was made possible by the automatic feature extraction from the input provided.Further the results can be further improved by applying more powerful models like AlexNet,GoogleNet etc to acheive more accuracy in the activity recognition.As a future enhancement the video images can be replaced by image sensor for activity recognition.

REFERENCES

[1] Krikor B. Ozanyan et al, "Human Activity Recognition with Inertial Sensors using a Deep Learning Approach," in proc. 2016 IEEE SENSORS.
[2] Mubashir,L Shao et al, "A survey on fall - detection: Principles & approaches," Neurocomputing, vol. 100, p.144–152, 2013.
[3] J. Cook, et al "Keeping the Resident, in the Loop" IEEE Trans on Systems, Man, and Cybernetics - Systems and Humans, vol. 39, no. 52009.
[4] Qubo Xie et al "Activity Recognition as a Service ,for Smart Home " in proc. 2017 IEEE Intl Conference on AI & Mobile Services (AIMS).
[5] Chen et al, "Sensor-Based Activity Recognition," IEEE Trans on Systems, Man, and Cybernetics, Part C vol. 42, no. 6, May 2012..
[6] Wei Xu, Yang , Yu, "3D Convolutional Neural Networks for Human Action Recognition," IEEE Trans on Pattern Analysis and Machine Intelligence, vol. 35, March 2013.
[7] Laptev and Lindeberg, "Space-time interest points," 2003 IEEE Intl Conference on Computer Vision.
8] Dataset for human actions in smart environments, (Jan. 2019), University of British Colombia, [Online].
Available: http://dml.ece.ubc.ca/data/smartaction/.
[9] Victor et al "Non-intrusive activity monitoring in a smart house environment," in proc. 2013 IEEE

conf-Health Networking.

[10] S.M.Amiri, M.T.Pourazad, P.Nasiopoulos and V.C.M.Leung, "Improved human activity recognition in a smart home environment " IRBM, vol. 35, no. 6,2018.

[11] C.M. Leung et al "Human action recognition using meta learning and depth information," in proc. 2014 International Conf on Computing, Networking and Communications (ICNC).

[12] Hoyer, "Non-negative Sparse coding," in 2002 Proc. of the 12th IEEE Workshop on NN for Signal Processing.

[13] Bappaditya, et al. "Towards detection of bus driver fatique based on robust visual analysis of eye state." IEEE Transactions on ITS 2017

[14] T. Pourazad, "The similarity measure for analyzing human activities using human-object interaction context," in proc. 2014 IEEE International Conference on Image Processing (ICIP), Paris, France.

[15] Naher et al "Recent Advances in Deep Learning: An Overview,"June 2018

[16] G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," Science, vol. 313, no. 5786, pp. Jul. 2006.

[17] Syed Mohsen Naqvi and Jonathon A.Chambers, "Deep learning for the posture analysis in fall detection identification," in proc. 2014 19th Intl Conf on Digital Signal Processing,

[18] Tsung-Han Chan, Kui Jia, Shenghua Gao, Jiwen Lu, Zinan Zeng and Yi Ma, "PCANet: A Simple Deep Learning Baseline for Image Classification" IEEE Trans on Image Processing, vol. 24, no.12,

[19] Shengke Wang, Long Chen, Zixi Zhou, Xin Sun and Junyu Dong, "Human fall detection in surveillance video based on PCANet," Multimedia Tools and Applications, vol. 75, no. 19, Oct. 2016.

[20] Wzuo et al , "3D Human Activity Recognition with Reconfigurable CNN," in proc. 2014 MM '14 of the 22nd ACM international conference on Multimedia.

[21] Roger Leitzke Granada, Juarez Monteiro, Rodrigo Coelho Barros and Felipe Rech Meneguzzi, "A Deep Neural Architecture for Kitchen Activity Recognition," in 2017 Proc. of the Thirtieth International Florida Artificial Intelligence Research Society Conference.

[22] Dushyant Goyal, "Kitchen activity recognition based on scene context," in proc. 2013 IEEE International Conference on Image Processing.

[23] Zhang, Chi, Hong Wang, and Rongrong Fu. "Automated detection of fatique based on entropy and complexity measures." IEEE Transactions on Intelligent Transportation Systems 15.1 (2014): 168-177.

.