Human Activity Recognition in Smart Home With Deep Learning Approach

Conference Paper · April 2019

DOI: 10.1109/SGCF.2019.8782290

2 authors:

Homay Danaei-mehr
Gazi University

Hüseyin Polat
Gazi University

Human Activity Recognition in Smart Home With Deep Learning Approach

Homay Danaei Mehr
Computer Engineering Department
Gazi University, Faculty of Technology
Ankara, Turkey
h.danai.mehr@gmail.com

Huseyin Polat
Computer Engineering Department
Gazi University, Faculty of Technology
Ankara, Turkey
polath@gazi.edu.tr

***Abstract*— Vision-based human activity recognition in smart homes has become a significant issue in terms of developing the next generation technologies which can improve healthcare and security of smart homes. Recently, deep learning models that aim to automatic extraction of low-level to high-level features of input data instead of using complicated conventional feature extraction methods have achieved significant improvements in the classification of a large amount of data especially vision-based datasets. Therefore, in this study in order to recognize human action of a smart home video dataset (DMLSmartActions) Convolutional Neural Networks (CNNs) architecture as a deep learning model has been proposed. Moreover, the performance of the proposed method has been compared with the previous methods which have used traditional machine learning methods on the same dataset. Experimental results demonstrated that the proposed deep learning model has achieved 82.41% accuracy rate in the classification of human activity.**

Keywords*— *Convolutional neural networks, deep learning, human activity recognition, smart home

I. INTRODUCTION

That the smart home will play a significant role in providing intelligent dwellings in the future is inevitable fact. Not only does smart home technology control the incorporated lightning, heating, electrical and all domestic components, but it also can recognize activity of all home residents. Moreover, in conjunction with recognizing the activity of occupants, by utilizing machine learning techniques it can make decision and prepare sufficient devices and services based on user's need. Therefore, due to increasing demand for human activity recognition in terms of security and health care especially elderly and child care, it has become as a noteworthy issue in recent years [1]. The methods of activity recognition include sensor and vision-based categories. Sensor based methods which contain wearable and ambient sensors seem to be traditional methods of human activity recognition in smart homes. In both sensors based approaches in order to analyze the human motion, data have been collected and conveyed by sensors. Since wearable sensors which have been attached to the body and ambient sensors have been installed all around the home, they can be annoying for residents. In addition, sensors can produce noise and wrong alarms which can lead to inaccurate results [2]. However, they could achieve satisfying results in human activity recognition in smart homes [3],[4]. In recent years, in order to overcome the mentioned deficiency of sensors in collecting accurate and sufficient data, vision-based methods have gained popularity in human activity recognition researches. Moreover, vision-based methods take the advantage of using diverse camera types to provide more accurate and adequate data than sensor based methods [5].

In this regard, in order to perform data classification and detection, most vision-based methods have used traditional pattern recognition and machine learning methods. Since in traditional methods in order to acquire the features of video frames or images, complicated handcraft methods are being used, vision-based recognizing human activity is a complex method. Furthermore, using handcraft methods by some local descriptors such as Histogram of Gradient (HOG) and Scale-Invariant Feature Transform (SIFT) to achieve low-level features can be acceptable for some fixed datasets. However, since handcrafted features are limited to a certain dataset, achieving effective features from a new dataset and adjusting the manually selected low-level features to a new dataset and condition is a challenging task [6]. Nevertheless, there are significant studies that have used traditional pattern recognition and conventional machine learning methods in human action recognition on video datasets. Amiri et. al., have used conventional feature extraction methods, i.e. Harris3D feature detector and STIP feature descriptor [7] to extract spatiotemporal features from DMLSmartActions activity dataset for activity recognition [8]-[11]. In both [9],[10], K-means, Sparse Coding (SC) and Non-Negative Sparse Coding (NNSC) algorithms [12] have been used to build visual words from extracted spatiotemporal features. Different types of Support Vector Machine (SVM), i.e. X2-SVM, Linear SVM and Intersection SVM have been used to classify the dataset. Experimental results in both papers showed that X2-SVM using SC algorithm has achieved the highest accuracy rate compared to the other used methods by 58.20%. Moreover, in [10] this group could improve the performance of Linear SVM using NNSC feature extractor by 2.1% to achieve 59.65% accuracy rate. In [11] they have proposed SVM classifier using Meta classifier (MC) [13] and Naïve Bayesian classifier (NBC) to classify the same human activity dataset. Results showed that their proposed MC method (by 77.19% accuracy rate) has surpassed the NBC method (by 56.32% accuracy rate) in classification of DMLSmartActions dataset. However, the proposed MC method has achieved significant accuracy rate in classification of other datasets as well. Furthermore, this group in another work in order to recognize the human activity have used Speed Up Robust Features (SURF) and STIP for extracting features from images of DMLSmartActions dataset. Moreover, NNSC algorithm has been used to construct the visual words. For measuring the similarity between images or videos kernel function has been used. In this regard they have compared their proposed methods with their previous works. Consequently, their proposed kernel method combined with SVM and NNSC achieved the highest accuracy rate (79.9%). Moreover, their proposed kernel function combined with SVM achieved the second highest accuracy rate compared to their previous same works by 62.6% accuracy rate [14].

In recent years, deep learning techniques have gained successful results in many fields such as recognition of visual objects, natural language processing and audio classification [15]. Convolutional Neural Networks (CNN), Deep Belief Network (DBN) Recurrent Neural Networks (RNN) and Long Short Term Memory (LSTM) are the most common deep learning architectures. In fact, deep learning based methods are models which can extract features automatically without using any handcraft methods. Moreover, features hierarchically are being produced from low-level to high-level features [16]. Consequently, in recent years there are diverse studies in human activity recognition in smart homes which achieved reasonable results. Yu et. al., have used Restricted Boltzmann Machine (RBM), DBN and SVM to classify and fall detection of the video frames which were taken by a camera from a person postures at home. Experimental results showed that RMB with 500 hidden layers achieved more classification accuracy rate (86%) than other two algorithms. Moreover, DBN obtained the least value of false detection rate of fall recognition [17]. Wang et. al., have utilized Deep PCANet [18] to feature extraction of video frames of human fall dataset. Furthermore, in order to classify and detect the falling and standing activities SVM classifier have been used. The sensitivity rate and specificity rate of their used method were 78.85% and 88.87% respectively [19]. Wang et. al., have proposed CNNs architecture to classify human actions of their OA1 and OA2 datasets and CAD120 human action dataset. Their proposed method has exceeded their compared previous works on the same datasets. Moreover, their proposed method achieved average accuracy of 81.2% on CAD120 dataset, 60.1% average accuracy rate on OA1 dataset and 45% average accuracy rate on OA2 dataset [20]. Granada et. al., have used GoogleNet architecture of CNN method to recognize human activity of a kitchen video dataset. Furthermore, fusion methods, e.g. Neural Networks (NNs) and SVM have been used as classifiers. Their used GoogleNet with NNs as fusion method has achieved higher accuracy rate (73%) than GoogleNet with other fusion methods [21]. Monteiro et. al., in order to recognize activities of the kitchen video dataset have utilized the combination of three common CNN architectures, i.e. AlexNet, GoogleNet and SqueezNet with different fusion methods to provide output of the CNNs, i.e. SVM and LSTM. Moreover, they have used post processing method [22] to remove noise in frames of videos. Comparison of the used methods showed that the combination of their used CNNs architectures with LSTM and post processing method have surpassed their other used methods by 78.5% accuracy rate [21].

In this regard, according to the advantage of automatic feature extraction and using large scale datasets of deep learning methods, which achieved state of art results in terms of detecting and recognizing human activity in smart homes, an architecture of CNNs has been proposed. Moreover, instead of using commonplace CNNs, a special CNN architecture to recognize the human activity has been designed. In this regard, a real human activity video dataset (DMLSmartActions) of a smart home has been used. Additionally, the performance of the proposed method has been compared with the other previous used methods on the same dataset. Furthermore, this study is the first deep learning investigation on the mentioned dataset. The structure of this study is arranged as follows: The concept of CNN, the used dataset and the performance evaluation measurements are described in section II. Furthermore, section III presents the proposed architecture and related experimental results. Eventually, conclusions are given in section IV.

II. MATERIAL AND METHODS

A. Convolutional Neural Networks(CNNs)

CNNs are one of the most popular architectures of deep learning which simulate biological nervous system like Artificial Neural Networks (ANNs). AlexNet, GoogleNet, SqueezNet and ResNet [23] are the most common architectures of CNN. In comparison with ANNs, CNNs take the advantage of local connections instead of fully connections in all layers except the last layer (Fig. 1).

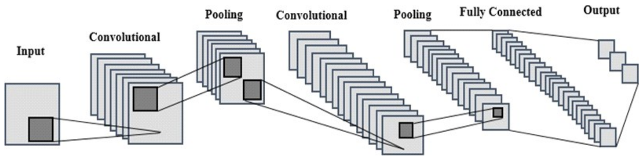

Fig. 1. A general CNN architecture

In this regard, each layer by using the kernels or filter banks connects to the local region of the previous layer. Moreover, CNNs structure contains series of most common layers: The first layer is convolutional layer. In this layer, each region that contains feature maps is connected to the feature maps of a local region in the previous layer by calculating weights that are known as kernels (filter banks). Sum of all local weights goes through a non-linearity function, e.g. Relu [24]. The output of convolutional layer is given by (1) where kernel (filter) size and feature map are defined by F and m respectively. Moreover, bias is defined as B and filter's weight is defined as Wj. Yli is the output that j is the ith feature map of layer l.

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

Pooling layer is the second common layer of CNNs which is used to decrease the spatial dimension of the output of convolutional layer without any change in depth. The advantage of using pooling layer is that by decreasing computational operations it prevents the overfitting in training process. Pooling layer contains min, max and average operations. However max pooling layer has achieved reasonable results in most fields. Height and width of the output of pooling layer can be calculated by (2) and (3); where width, height and depth of input are denoted as W1, H1, D1 respectively. Furthermore, F and S are considered as kernel size and stride size of shifting respectively as well.

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

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

The third common layer is fully connected layer. In this layer each neuron is not only connected to the all neurons of the previous layer but also calculated scores of dataset's classes are given in this layer. Moreover, generally in the last most convolutional layers softmax function [25] is utilized to calculate the probable distribution through the labels of classes. The equation of softmax function is given by (4) where f(z) (softmax) converted the probable values of x between 0 and 1 and, additionally, k is the dimensional value of x [26].

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

B. Dataset

In this study, the DMLSmartActions dataset of the Digital Multimedia Lab of the University of British Columbia has been used [8]. In order to create this dataset, the real and daily actions of seventeen people are captured by static three cameras in two simulated living room environments. Moreover, videos contain twelve different natural activities of people, i.e. using cellphone, walking, writing, reading, sitting down, standing up, putting something, picking something, dropping and picking up, drinking, cleaning table and falling down. The most significant advantage of this dataset is that the large variety of activities of different people with both female and male genders have been considered. Furthermore, some examples of actions are depicted by Fig. 2.

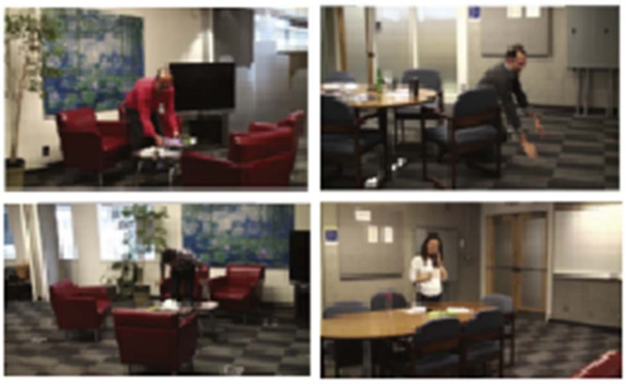

Fig. 2. Example of human activity in the smart home [8].

In this study in order to apply proposed CNN architecture to classify human activity, DMLSmartActions dataset is preprocessed. In this regard, as in the original dataset frame rates of each video streams were set to 30 fps of 640×480 pixels, in this study frames of each video were achieved by its default as well. Consequently, 74120 frames were achieved from 47 videos. In addition, the size of each frame was reduced to 320×256 pixels then 224×224 crops were taken from the resized images. In order to implement the CNN architecture, 75% of dataset was considered as training dataset and the remain 25% of dataset was used for testset.

C. Evaluation Metrics

The multi-class confusion matrix is used to investigate the performance of the classification methods of multi-class datasets. Multi confusion matrix contains two dimensions which indicate the predicted classes by classifier and the actual classes. A multiclass confusion matrix is depicted by Table I which contains A1, A2,…,An classes. Moreover, Nij is determined as the number of samples of class Ai which is predicted as class Aj. Some performance evaluation metrics of this study that have been calculated by confusion matrix are described below [27]. In this study, to clarify the performance measurements action label indicates the selected specific action label for classification between twelve classes. Additionally, no action label is related to the rest of class labels which are not selected as a specific action label for evaluation of classifier.

TABLE I. MULTI CLASS CONFUSION MATRIX [27]

		Predicted		
		A_i	$\ldots A_j \ldots$	A_n
Actual	A_1	N_{11}	N_{1j}	N_{1n}
	A_i	N_{i1}	$\ldots N_{ij} \ldots$	N_{in}
	A_n	N_{n1}	N_{nj}	N_{nn}

- Accuracy: Indicates the proportion of accurate classified samples among all input samples.

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

- Recall (Sensitivity): Demonstrates what proportion of action labels are predicted accurately.

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

- Precision: Demonstrates what proportion of predicted action labels are actually assumed action labels.

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

III. EXPERIMENTAL RESULTS

In this study in order to classify the human activity frames instead of using common CNNs, an architecture of a specific CNN was proposed which contains five convolutional layers, four pooling layers and three fully connected layers. In the last fully connected layer softmax was consider to determine the probability of the 12 classes of activity dataset. Relu activation function was utilized for non-linearity on the output of convolutional and fully connected layers. In the first convolutional layer 96 filters with 3×3 kernel size were applied on the inputs. Moreover, the output of convolutional layer went through the Relu activation function. The max pooling layer was applied to down sample the output of the previous convolutional layer without any change in depth size by applying 2×2 kernel size. After the first pooling layer the second and the third convolutional layers were applied. This architecture continued until the softmax layer.

TABLE II. SUMMARY OF THE ARCHITECTURE OF THE PROPOSED CNN

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

The detail of all layers of proposed CNN is illustrated in Table II. Furthermore, in the training process batch size (number of input samples over one cycle of training) and learning rate were considered as 32 and 0.01.

To evaluate the performance of the proposed deep learning based CNN architecture, results of the used method were compared with the four other studies which have used the DMLSmartActions dataset. Accuracy rates of the proposed CNN for each activity is illustrated by Table III.

TABLE III. ACTIVITY RECOGNITION OF PROPOSED CNN FOR EACH ACTIVITY

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

As it is shown in Table III, the majority of the actions were classified by satisfying accuracy rates. However, some actions are so similar in case of recognizing them such as drop and picking something, reading and writing which results in low accuracy rate. The maximum accuracy rate of classification belongs to walking action (88.71%) and the minimum accuracy rate related to writing action by 69.67%. The overall accuracy rate of human action classifier is 82.41%. It is noticeable that datasets of human activity in smart home are rare and most of available datasets restricted to non-smart indoor environment or do not relate to the real actions of human. Real actions of DMLSmartActions dataset made it to be one of the best dataset in terms of recognizing the human activity in a smart home.

Moreover, the proposed CNN achieved 78.72% and 87.77% for Recall and Precision respectively. In this regard, 78.72% of the assumed action labels among all actions were predicted accurately. However, approximately 21.28% of assumed actions were misclassified. Furthermore, although 12.23% of predicted action labels did not contain the actual labels of assumed labels, 87.77% of predicted action labels contained actual assumed label. By considering accuracy rates of all previous studies on the same dataset (DMLSmartActions) which have used the conventional feature extraction methods and the traditional machine learning methods to classify the human actions it is demonstrated that their proposed SVM classifier combined with kernel and NNSC methods have achieved a higher accuracy rate (79.9%) than other three related works. However, our proposed CNN architecture which takes the advantage of automatic feature extraction of deep learning methods by getting more trainable features achieved the highest accuracy rate compared to all previous works on the same dataset (Table IV).

TABLE IV. ACHIEVED RESULTS OF DIFFERENT ALGORITHMS ON THE DMLSMARTACTIONS DATASET

Algorithms	Accuracy rate (%)	Recall (%)	Precision (%)
χ^2 −SVM using SC [9]	58.20	—	—
SVM with NNSC [10]	59.65	—	—
Proposed MC method [11]	77.19	—	—
SVM with NNSC and proposed kernel [14]	79.9	—	—
Our proposed CNN	82.41	78.72	87.77

IV. CONCLUSION

In this study a CNN architecture as a deep learning method has been proposed to recognize human activity of smart home video dataset (DMLSmartActions). Due to the automatic feature extraction from input data the proposed CNN architecture achieved the highest accuracy rate and, additionally, could improve the performance of the classification of human action compare to the previous studies which have used conventional feature extraction and machine learning methods on the same dataset. Moreover, in this study to achieve accurate results instead of using the commonplace CNN architectures, a specific architecture of CNN has been proposed to adjust completely to the dataset. Therefore, this study is the first study on DMLSmartActions dataset which has been applied by using deep learning models. Indeed, this architecture may not be a sufficient model for any other datasets. However, the performance of the proposed method can be improved by learning the larger amount of data. The future study is planned to be a combination of different deep learning models and classifiers on a huge specific human activity dataset which it will be provided by our team as well.

REFERENCES

[1] Tahmina Zebin, Patricia J. Scully and Krikor B. Ozanyan, "Human Activity Recognition with Inertial Sensors using a Deep Learning Approach," in proc. 2016 IEEE SENSORS, Orlando, FL, USA. Doi: 10.1109/ICSENS.2016.7808590.

[2] Muhammad Mubashir,Ling Shao and LukeSeed, "A survey on fall detection: Principles and approaches," Neurocomputing, vol. 100, p. 144–152, 2013. Doi: 10.1016/j.neucom.2011.09.037.

[3] Parisa Rashidi and Diane J. Cook, "Keeping the Resident in the Loop: Adapting the Smart Home to the User," IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, vol. 39, no. 5, pp. 949-959, Aug. 2009. Doi: 10.1109/TSMCA.2009.2025137.

[4] Xiaohu Fan, Qubo Xie, Xuebin Li, Hao Huang, Jian Wang, Si Chen, Changsheng Xie and Jiajing Chen, "Activity Recognition as a Service for Smart Home: Ambient Assisted Living Application via Sensing Home," in proc. 2017 IEEE International Conference on AI & Mobile Services (AIMS), Honolulu, HI, USA. Doi: 10.1109/AIMS.2017.29.

[5] Liming Chen, Jesse Hoey, Chris D. Nugent, Diane J. Cook and Zhiwen Yu, "Sensor-Based Activity Recognition," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 42, no. 6, pp. 1-19, May 2012. Doi: 10.1109/TSMCC.2012.2198883.

[6] Shuiwang Ji, Wei Xu, Ming Yang and Kai Yu, "3D Convolutional Neural Networks for Human Action Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 1, pp. 221 - 231, March 2013. Doi: 10.1109/TPAMI.2012.59.

[7] Ivan Laptev and Tony Lindeberg, "Space-time interest points," in proc. 2003 Ninth IEEE International Conference on Computer Vision, Nice, France. Doi: 10.1109/ICCV.2003.1238378.

[8] DMLSmartActions dataset: Dataset for human actions in smart environments, (Jan. 2019), University of British Colombia, [Online]. Available: http://dml.ece.ubc.ca/data/smartaction/.

[9] S. Mohsen Amiri, Mahsa T. Pourazad, Panos Nasiopoulos and Victor C.M. Leung, "Non-intrusive human activity monitoring in a smart home environment," in proc. 2013 IEEE 15th International Conference on e-Health Networking, Applications and Services (Healthcom 2013), Lisbon, Portugal. Doi: 10.1109/HealthCom.2013.6720748.

[10] S.M.Amiri, M.T.Pourazad, P.Nasiopoulos and V.C.M.Leung, "Improved human action recognition in a smart home environment setting," IRBM, vol. 35, no. 6, p. 321–328, 2014. Doi: 10.1016/j.irbm.2014.10.005.

[11] S. Mohsen Amiri, Mahsa T. Pourazad, Panos Nasiopoulos and Victor C.M. Leung, "Human action recognition using meta learning for RGB and depth information," in proc. 2014 International Conference on Computing, Networking and Communications (ICNC), Honolulu, HI, USA. Doi: 10.1109/ICCNC.2014.6785361.

[12] P. O. Hoyer, "Non-negative sparse coding," in 2002 Proc. of the 12th IEEE Workshop on Neural Networks for Signal Processing, Martigny, Switzerland. Doi: 10.1109/NNSP.2002.1030067.

[13] M. Buscema, "MetaNet*: The Theory of Independent Judges," Substance Use & Misuse, vol. 33, no. 2, p. 439–461, 1998. Doi: 10.3109/10826089809115875.

[14] S. Mohsen Amiri, Mahsa T. Pourazad, Panos Nasiopoulos and Victor C. M. Leung, "A similarity measure for analyzing human activities using human-object interaction context," in proc. 2014 IEEE International Conference on Image Processing (ICIP), Paris, France. Doi: 10.1109/ICIP.2014.7025480.

[15] Matiur Rahman Minar and Jibon Naher, "Recent Advances in Deep Learning: An Overview," arXiv:1807.08169v1, pp. 1-31, Jul. 2018. Doi: 10.13140/RG.2.2.24831.10403.

[16] G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," Science, vol. 313, no. 5786, pp. 504-507, Jul. 2006. Doi: 10.1126/science.1127647.

[17] Pengming Feng, Miao Yu, Syed Mohsen Naqvi and Jonathon A. Chambers, "Deep learning for posture analysis in fall detection," in proc. 2014 19th International Conference on Digital Signal Processing, Hong Kong, China. Doi: 10.1109/ICDSP.2014.6900806.

[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 Transactions on Image Processing, vol. 24, no. 12, pp. 5017 - 5032, Sep. 2015. Doi: 10.1109/TIP.2015.2475625.

[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, p. 11603–11613, Oct. 2016. Doi: 10.1007/s11042-015-2698-y.

[20] Keze Wang, Xiaolong Wang, Liang Lin, Meng Wang and Wangmeng Zuo, "3D Human Activity Recognition with Reconfigurable Convolutional Neural Networks," in proc. 2014 MM '14 of the 22nd ACM international conference on Multimedia, Orlando, Florida, USA. Doi: 10.1145/2647868.2654912.

[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, Florida, USA.

[22] Shubham Bansal, Shubham Khandelwal, Shubham Gupta and Dushyant Goyal, "Kitchen activity recognition based on scene context," in proc. 2013 IEEE International Conference on Image Processing, Melbourne, VIC, Australia. Doi: 10.1109/ICIP.2013.6738714.

[23] Chunpeng Wu, Wei Wen, Tariq Afzal, Yongmei Zhang, Yiran Chen and Hai Li, "A Compact DNN: Approaching GoogLeNet-Level Accuracy of Classification and Domain Adaptation," in proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA. Doi: 10.1109/CVPR.2017.88.

[24] Yann LeCun, Yoshua Bengio and Geoffrey Hinton, "Deep learning," Nature International Journal of Science, vol. 521, p. 436–444, May 2015.

[25] G. Bouchard, "Efficient Bounds for the Softmax Function and Applications to Approximate Inference in Hybrid models," in the Presentation at The Workshop for Approximate Bayesian Inference in Continuous/Hybrid Systems at Neural Information Processing Systems (NIPS), Meylan, France, May 2007.

[26] Qing Li, Weidong Cai, Xiaogang Wang, Yun Zhou, David Dagan Feng and Mei Chen, "Medical image classification with convolutional neural network," in proc. 2014 13th International Conference on Control Automation Robotics & Vision (ICARCV), Singapore, Singapore. Doi: 10.1109/ICARCV.2014.7064414.

[27] Xinyang Deng, Qi Liu,Yong Deng and Sankaran Mahadevan, "An improved method to construct basic probability assignment based on the confusion matrix for classification problem," Information Sciences, vol. 340–341, pp. 250-261, May 2016. Doi: 10.1016/j.ins.2016.01.033.