JCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-6, Issue-3 E-ISSN: 2347-2693

Human Activity Recognition Using LSTM Networks

P.Sharma[1*], S.Chaudhary[2], Komal [3]

[1*]CSE, Amity School of Engineering and Technology, Amity University, Gurugram, India
[2]CSE, Amity School of Engineering and Technology, Amity University, Gurugram, India
[3]CSE, Amity School of Engineering and Technology, Amity University, Gurugram, India

Corresponding Author: sarikacse23@gmail.com, Tel.: +91-9911193360

Available online at: www.ijcseonline.org

Received: 18/Feb//2018, Revised: 24/Feb2018, Accepted: 16/Mar/2018, Published: 30/Mar/2018

Abstract**—** Deep learning has shown great improvements in all the computer vision and image interpretation tasks. In this paper a fully automated deep model for human activity recognition has been proposed which do not include any prior knowledge. In the first step of the proposed method, model automatically learns all the temporal and spatial features for recognition. In the second stage of the method memory network which is recurrent in nature is used to classify the various human actions. The results obtained from the suggested method are compared with all the rage methods. Outcomes show that the suggested method has better accuracy as compared to various alternative techniques available.

Keywords**—**MLP, LSTM, TDR

I. INTRODUCTION

Deep learning represents the largest trend in machine learning over the past decade. Since the beginning of the umbrella term the range of ways it encompasses has magnified rapidly, and can still driven by the resources of both tutorial and industrial interests. Deep learning has become accessible to everybody via machine learning frameworks like Torch7 [1], and has had important impact on a spread of application domains like object recognition, speech recognition, Data analysis and many more [2].

Human activity recognition includes segmenting the time series data captured from Smartphone or body-worn sensors followed by applying manually designed feature extraction techniques and finally applying a classification strategy to classify the videos into different human activities going on. The basic activity flow chart is shown in Figure 1.

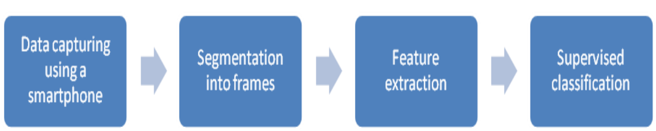

Figure 1. Basic flow of activities in Human activity Recognition

This paper describes a unique method for activity recognition using long short term memory networks. The paper is organized as follows. In the second section all the previous, related work is discussed. In the third section the proposed method is explained. Results are discussed in the forth section and finally the paper is summarized in the last section.

II. RELATED WORK

It is very difficult to find and design handcrafted features for robust recognition of any activity or object in an image or video. Deep learning has given a great platform and excellent performances also in self learning robust features from row data [3].The method has shown great improvements in the area of object classification [4] and speech recognition [5] but, very few work has been done in the area of activity or action recognition from subsequent frames or sequential data. A MLP (Multi-layer perceptron) based method was proposed was proposed to classify each activity in a frame of video [6] but the method was not able to incorporate the temporal dependencies in the data. A gesture recognition system with shallow bidirectional LSTM which included only one forward and one backward hidden layer was proposed to explore temporal dependencies in the data [7]. The accuracies in the system were improved but the system was still not a robust system. Deep recurrent neural network architecture with hand-designed features was used for frame based activity recognition [8]. However, the handcrafted hierarchical features were not taking into consideration the occurrence of different joints.
In this paper a novel LSTM network has been proposed that take into consideration all the temporal information in the data. In earlier proposed LSTM architectures dropout was only applied to feed forward networks not to the recurrent [9]. In this paper, in depth dropout is applied to improve the

overall performance of the network.

III. METHODOLOGY

In this paper we have classified the movement into 6 different categories i.e. walking, walking-upstairs, walking-downstairs, sitting and standing. The various steps of the method are discussed in the next sections.

A. Data Capturing and pre-processing

Various videos capturing different activities were used for the classification purpose [10]. The experiments have been carried out with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities (WALKING,WALKING_UPSTAIRS, WALKING_DOWNSTAIRS, SITTING, STANDING, LAYING) wearing a Smartphone (Samsung Galaxy S II) on the waist The obtained dataset has been randomly partitioned into two sets, where 70% of the volunteers was selected for generating the training data and 30% the test data. The videos captures using different smart phones often contain some disturbances. These disturbances are modelled as noise and needs to be removed for further processing. A simple Gaussian noise removal filter was used to remove such disturbances.

B. Segmentation into frames

The video data obtained after pre-processing is then sampled in fixed-length sliding windows often called as frames of 2 seconds each with a 50 % overlap so that there is no distortion in the data due to segmentation. Some of the pre-processed frames are shown in figure below.

Figure 2. L-R: sitting position , Laying position , Walking position

C. Feature learning using LSTM Networks

Recurrent neural networks are very successful in modelling sequential information. These networks take as input a time series of feature vectors and convert them into a probability vector for the classification purpose. Dependence of the time series data is maintained as output at each stage is calculated based on the output at previous stage using the equation below.

$$o_t = \theta (W_{ih} x_t + b) \qquad (1)$$

Where denotes output vector , is the weight matrix from input layer to next hidden layer conversion, is the input vector, is the weight matrix from hidden layer to next hidden layer conversion, is the output at previous stage and b is the bias vector.

LSTM is basically an improved RNN as shown in fig 3. It is more complex but easier to train and give better results as it contains an input gate , a forget gate , a cell , an output gate and an output response. The input gate and forget gate are responsible for governing the flow of information into and out of the cell. The output gate controls how much information from the cell is passed to the output. The memory cell has a self connected recurrent edge of weight 1, ensuring that the gradient can pass across many time steps without vanishing or exploding. As a result it avoids the problem associated with simple RNN i.e. Vanishing Gradient Problem.

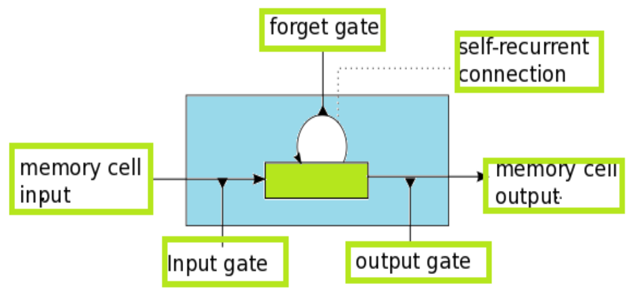

Figure 3. LSTM Network

IV. RESULTS AND DISCUSSION

The trained network showed a very high accuracy or 92 %. The confusion matrix designed for various positions is shown in figure 4.

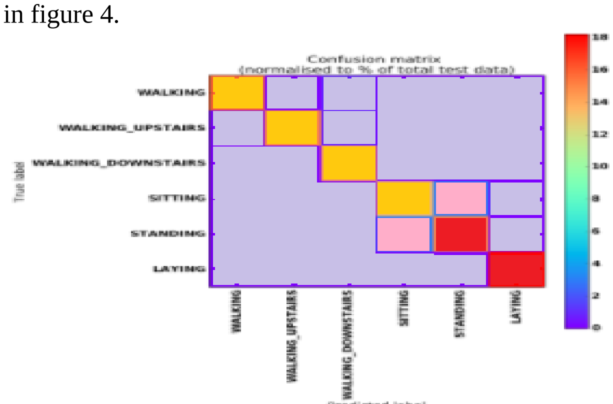

Figure 4. Confusion matrix

The method was also compared with other state of the art methods and a graphical comparison is shown in figure 5.

For detecting the performance of proposed method a metric names as True Detection rate is calculated as :

$$TR = \frac{Correct\ classified\ positions}{Total\ number\ of\ positions}$$

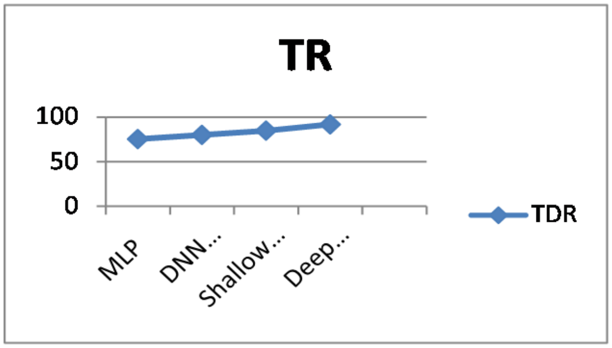

Figure 5. Comparative analysis

V. Conclusion

This paper presented and evaluated a unique LSTM network to identify various human activities captured in the form of videos using a smart phone. It was observed that the method was giving better results as compared to other state of the art methods. The method can be improved further by adding more activities and training the network accordingly.

References

[1] G. Antonini, M. Bierlaire, and M. Weber, "*Discrete choice models of pedestrian walking behaviour*", Transportation Research Part B: Methodological, Vol:40, Issue:8, pp. 667–687, 2006.

[2] J. Azorin-Lopez, M. Saval-Calvo, A. Fuster-Guillo, and A. Oliver-Albert, *" A predictive model for recognizing human behaviour based on trajectory representation"*, Neural Networks (IJCNN), 2014 International Joint Conference, pp. 1494–1501. IEEE, 2014.

[3] A. Krizhevsky, I. Sutskever, and Hinton, G. E.Hinton, *" Image Net classification with deep convolutional neural network"*, Advances in Neural Information Processing Systems, pp.1097–1105, 2012.

[4] A. Graves, A. Mohamed, and G. Hinton, *"Speech recognition with deep recurrent neural networks",* IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 6645–6649,2013.

[5] D.Weinland, R.Ronfard, and Boyerc, E.Boyerc, *"A survey of vision-based methods for action representation, segmentation and recognition",* Computer Vision and Image Understanding Vol:115, Issue:2, pp. 224–241,2011.

[6] K.Cho and Chen, X.Chen, *" Classifying and visualizing motion capture sequences using deep neural networks",* International Conference on Computer Vision Theory and Applications, pp. 122–130,2014.

[7] G. Lefebvre, S.Berlemont, F.Mamalet, and C. Garcia, *"BLSTM-RNN based 3D gesture classification",* Proceedings of the International Conference on Artificial Neural Networks and Machine Learning, pp. 381–388,2013.

[8] Y.Du, W.Wang and L.Wang, *"Hierarchical recurrent neural network for skeleton based action recognition",* IEEE Conference on Computer Vision and Pattern Recognition, 1110–1118,2015.

[9] V. Pham,T. Bluche, C. Kermorvant, and J. Louradour, *"Dropout improves recurrent neural networks for handwriting recognition",* International Conference on Frontiers in Handwriting Recognition, pp. 285–290,2014.

[10] D. Anguita, A. Ghio, L. Oneto, X. Parra and J. L. Reyes-Ortiz, *"A Public Domain Dataset for Human Activity Recognition Using Smartphones"*, 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2013. Bruges, Belgium 24-26 April 2013.

Authors Profile

Ms. P Sharma pursed Bachelor of Technology from Kurukshetra University in 2010 and Master of Technology from Thapar University in 2012.She is currently pursuing Ph.D. and currently working as Assistant Professor in Department of Computer Science and Engineering, Amity University Gurugram, India since 2017. She is a member of IEEE computer society, a life member of IAENG,IACSIT, ISRD. She has published more than 15 research papers in reputed international journals and conferences including IEEE and it's also available online. Her main research work focuses on image processing, Network Security, Cloud Security and Privacy Computational Intelligence based education. She has 5 years of teaching experience and Research Experience.

Ms. S Chaudhary pursed Bachelor of Technology from Kurukshetra University in 2004 and Master of Technology from Banasthali University of Rajasthan, India in year 2007. She is currently pursuing Ph.D. and currently working as Assistant Professor in Department of Computer Science and Engineering, Amity University Gurugram, India since 2012. She is a member of IET U.K, IACSIT,IAENG,UACEE, IRD INDIA, ISRD. She has published more than 14 research papers in reputed international journals and it's also available online. Her main research work focuses on , Software testing, clustering, Big Data Analytics, Data Mining, IoT. She has 9 years of teaching experience and Research Experience.

Ms. Komal pursed Bachelor of Technology and Master of Technology. She is currently pursuing Ph.D. and currently working as Assistant Professor in Department of Computer Science and Engineering, Amity University Gurugram, India since 2013. She is a member of IET U.K, IACSIT,IAENG,UACEE,.She has published more than 14 research papers in reputed international journals and it's also available online. Her main research work focuses on , Network security, Cloud computing and security, IoT. She has 8 years of teaching experience and Research Experience.

