International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT206466

Deep Learning for Human Action Recognition with Convolution Neural Network

S. Karthickkumar, Dr. K. Kumar M.E., Ph.D.
Computer Science and Engineering, Government College of Technology, Coimbatore, Tamil Nadu, India

Article Info
Volume 6, Issue 4
Page Number: 376-380
Publication Issue :
July-August-2020

Article History
Accepted : 01 Aug 2020
Published : 05 Aug 2020

ABSTRACT

In recent years, deep learning for human action recognition is one of the most popular researches. It has a variety of applications such as surveillance, health care, and consumer behavior analysis, robotics. In this paper to propose a Two-Dimensional (2D) Convolutional Neural Network for recognizing Human Activities. Here the WISDM dataset is used to tarin and test the data. It can have the Activities like sitting, standing and downstairs, upstairs, running. The human activity recognition performance of our 2D-CNN based method which shows 93.17% accuracy.

Keywords : Human activity recognition, Convolutional neural network, 3D accelerometer data.

I. INTRODUCTION

Human Action Recognition (HAR) is a challenging problem in computer vision technology. HAR system used in video surveillance, consumer behavior analysis, health care monitoring systems for the elderly. Smart phone sensor data mining for abnormality detection is proposed in [1]. Here to use the sensors like accelerometers and gyroscopes. It can be implemented by Android platform. Classification of mobile device accelerometer data for unique activity identification is proposed in [2]. Here to use the technique like Principal Component Analysis and Support Vector Machine. The overall success rate will be 92.22%. Activity recognition using cell phone Accelerometer is proposed in [3]. Here to use the sensors like GPS sensors, vision sensors, audio, light sensors, temperature sensors, direction sensors, and acceleration sensors. Convolutional neural network for human action recognition using mobile sensors is proposed in [4]. Here to use the technique like Convolutional Neural Network. The architecture for Feature extraction is one of the key components of activity recognition [Fig.1] is explained in [4].

In this paper Fig.3. shows structure of Convolution Neural Network for Human Action Recognition [4].

Unobtrusive Activity Recognition of Elderly People Living Alone Using Anonymous Binary Sensors and Deep Convolutional Neural Network is proposed in [5].Here to use the technique like Deep Convolutional Neural Network (DCNN). The architecture [5] of the proposed Deep Convolutional

Neural Network classifier is shown in Fig.8. The proposed DCNN accuracy will be 98%. Here to use sensor like anonymous binary sensors that are PIR motion sensors and door sensors.

A deep learning approach for Human Action Recognition is proposed in [6]. Here we use the technique like Convolution Neural Network and Long Short-Term Memory. Here to use the sensor like Inertial Measurement Units (IMUs) such as gyroscopes and accelerometers. Here to use WISDM dataset and it can be released by Fordham University. In this dataset contains list of activities were: walking, jogging, climbing stairs, sitting, standing, typing, brushing teeth, eating soup, eating chips, eating pasta, drinking from a cup, eating a sandwich, kicking a soccer ball, playing catch with a tennis ball, dribbling a basketball, writing, clapping, and folding clothes. The accuracy will be 79.Deep learning algorithms for human activity recognition using mobileand wearable sensor networks: State of the art and researchchallenges is proposed in [7]. It will explain Deep learning algorithms for human activity recognition using mobile and wearable sensor networks.

II. RELATED WORK

In recent years, lot of CNN-based human activities recognition methods have been proposed. Sequential deep learning for human action recognition is proposed in [8]. Hereto use the technique like 3D-Convolutional Neural Network, Recurrent Neural Network, Long Short-term Memory. The limitation will be the adaptation of the training algorithm, especially when calculating the retro-propagated error. 3D-based deep Convolutional neural network for action recognition with depth sequences is proposed in [9]. Here to use the technique like Convolutional neural network, Support Vector Machine. Limitations will be the length of action videos may vary in different dataset; it is difficult to normalize all the videos to one size while preserve enough information of the videos. Depth video-based human activity recognition system using translation and scaling invariant features for life logging at smart home proposed in [10]. Here to use the technique like R-transform, Principle Component Analysis, Hidden Markov Model, Linear Discriminant Analysis. Limitations will be the activities forward fall, backward fall, faint and vomit activities are not clearly discriminated. An effective view and time-invariant action recognition method based on depth videos is proposed in [11]. Here to use the technique like Support Vector Machine. Limitation will be here do not focus on skeleton joints-based action recognition using depth sequences by tracking and combining each joint.

L. Mengyuan et al Enhanced skeleton visualization for view invariant human action recognition is proposed in [12]. Here to use the technique like Multi-stream Convolutional neural network, Hidden Markov Model, Principle Component Analysis. Limitations will be here instead of using hard selection of ten CNN’s in the weighted probability fusion and to use soft probability fusion to provide more flexibility. Abnormal human activity recognition system based on R- transform and kernel discriminant technique for elderly home care is proposed in [13]. Here we use the technique like R-transform, Kernel Discriminant Analysis, Hidden Markov Model and Linear Discriminant Analysis. Limitations will be here discrimination between the body parts which are not observable in binary silhouettes. 3D Convolutional neural networks for human action recognition are proposed in [14]. Here we use the technique like Convolutional Neural Network, Support Vector Machine. Limitation will be Here does not explore the unsupervised training of 3D CNN models machine.

III. DATASETS

Here to use the dataset is WISDM dataset found in WISDM lab.Then we download the WISDM to download activity prediction. It contains user ID, activity and timestamp in MS and x, y and z accelerometer data. The data are not fully structure. You need to do some preprocessing. This dataset contains jogging, running and standing, upstairs, downstairs. Dataset contains 6 columns. Next the timestamp and x, y and z can be converted into numerical data. Here the sampling rate is 20 Hz. Each sample is of the form:

[Participant ID], [timestamp (Unix-based)], [x value], [y value], [z value] as shown below in Figure 2.

	User	Activity	Time	x	y	z
0	33	Jogging	49105962326000	-0.6946377	12.680544	0.50395286
1	33	Jogging	49106062271000	5.012288	11.264028	0.95342433
2	33	Jogging	49106112167000	4.903325	10.882658	-0.08172209
3	33	Jogging	49106222305000	-0.61291564	18.496431	3.0237172
4	33	Jogging	49106332290000	-1.1849703	12.108489	7.205164

Figure 1: Example of accelerometer/gyroscope data

Figure 2: Phone Accel/GyroData when walking

IV. METHODOLOGY AND ARCHITECTURE

More traditional algorithms have been applied to activity recognition. Here to use TensorFlow 2.0. Frame size to predicting the activity 4s i.e. Fs=Fs*4. The training sample will be 425 and testing sample will be 107. Next the data can be reshaping into three- dimensional data. Then to build 2DCNN model. To call sequential layer. In that first layer to add 2D CNN and pass a 16 filters and kernel size will be 2*2 and to use ReLu activation function. Then the dropout will be10% of neurons. Dense layer can have 64 filters and 15% of neurons. For compilation to add Adam optimizers that is Stochastic gradient descent optimizers. Default learning rate will be 0.001.

Our Convolutional Neural Network consists of 5 Conv2D layers, accompanied with Batch Normalizationand Max Pooling layers and lastly a flatten and two dense layers as shown below:

```
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_4 (Conv2D)            (None, 79, 2, 16)         80
_________________________________________________________________
dropout_6 (Dropout)          (None, 79, 2, 16)         0
_________________________________________________________________
conv2d_5 (Conv2D)            (None, 78, 1, 32)         2080
_________________________________________________________________
dropout_7 (Dropout)          (None, 78, 1, 32)         0
_________________________________________________________________
flatten_2 (Flatten)          (None, 2496)              0
_________________________________________________________________
dense_4 (Dense)              (None, 64)                159808
_________________________________________________________________
dropout_8 (Dropout)          (None, 64)                0
_________________________________________________________________
dense_5 (Dense)              (None, 6)                 390
=================================================================
Total params: 162,358
Trainable params: 162,358
Non-trainable params: 0
```

Figure 3: ConvLayer Neural Network Architecture

V. EXPERIMENTS/RESULTS

Our model describes the validation loss is less than training loss. In this we achieved 90% of accuracy. Confusion matrix is to understand how the class can be actually classified like sitting to walking, of all those kinds of things.

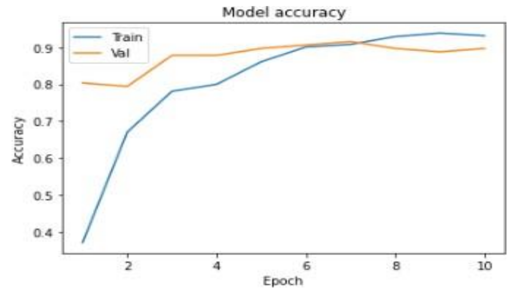

Figure 5: CNN model accuracy

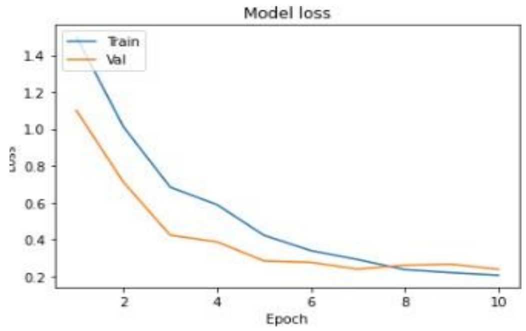

Figure 6: CNN model loss

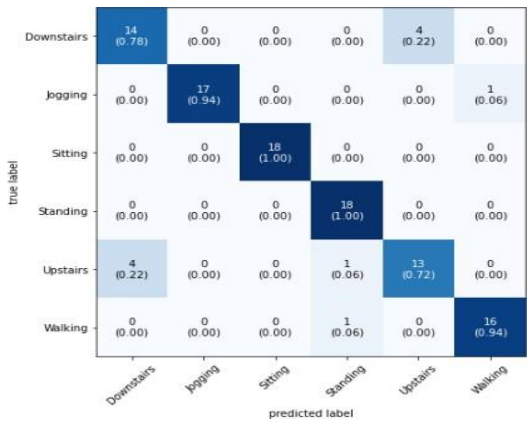

Figure 7: Confusion matrix

VI. CONCLUSION

To improve upon our results going forward on the expanded WISDM dataset that will be released later. Here we achieve the model accuracy of 93.17. Further process the dataset to predict the high accuracy. In future the complex activity can be predicted.

VII.REFERENCES

[1]. Gallagher S. Smartphone Sensor Data Mining for Gait Abnormality Detection. Fordham University, New York. 2014, Feb;2.

[2]. Kuspa, Kai, and Tony Pratkanis. Classification of mobile device accelerometer data for unique activity identification. Stanford Center for Professional Development (2013).

[3]. Kwapisz, Jennifer R., Gary M. Weiss, and Samuel A. Moore. Activity recognition using cell phone accelerometers. ACM SigKDD Explorations Newsletter 12, no. 2 (2011): 74-82.

[4]. Zeng, Ming, Le T. Nguyen, Bo Yu, Ole J. Mengshoel, Jiang Zhu, Pang Wu, and Joy Zhang. "Convolutional neural networks for human activity recognition using mobile sensors." In 6th International Conference on Mobile Computing, Applications and Services, pp. 197-205. IEEE, 2014.

[5]. Gochoo, Munkhjargal, Tan-Hsu Tan, Shing-Hong Liu, Fu-Rong Jean, Fady S. Alnajjar, and Shih-Chia Huang. Unobtrusive activity recognition of elderly people living alone using anonymous binary sensors and DCNN. IEEE journal of biomedical and health informatics 23, no. 2 (2018): 693-702.

[6]. Mathe, Eirini, Apostolos Maniatis, Evaggelos Spyrou, and Phivos Mylonas. A deep learning approach for human action recognition using skeletal information. In GeNeDis 2018, pp. 105-114. Springer, Cham, 2020.

[7]. Nweke, Henry Friday, Ying Wah Teh, Mohammed Ali Al-Garadi, and Uzoma Rita Alo. Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. Expert Systems with Applications 105 (2018): 233-261.

[8]. Baccouche, Moez, Franck Mamalet, Christian Wolf, Christophe Garcia, and Atilla Baskurt. Sequential deep learningfor human action recognition. In International workshop on human behavior understanding, pp. 29-39. Springer, Berlin, Heidelberg, 2011.

[9]. Liu, Zhi, Chenyang Zhang, and Yingli Tian. 3D-based deep convolutional neural network for action recognition with depth sequences. Image and Vision Computing 55 (2016): 93-100

[10].Jalal, Ahmad, Md Zia Uddin, and T-S. Kim. Depth video-based human activity recognition system using translation and scaling invariant features for life logging at smart home. IEEE Transactions on Consumer Electronics 58, no. 3(2012): 863-871.

[11].Liu, Zhi, Xin Feng, and Yingli Tian. An effective view and time-invariant action recognition method based on depthvideos. In 2015 Visual Communications and Image Processing (VCIP), pp. 1-4. IEEE, 2015.

[12].Liu, Mengyuan, Hong Liu, and Chen Chen. Enhanced skeleton visualization for view

invariant human action recognition. Pattern Recognition 68 (2017): 346-362.

[13]. Khan, Zafar A., and Won Sohn. Abnormal human activity recognition system based on R-transform and kernel discriminant technique for elderly home care. IEEE Transactions on Consumer Electronics 57, no. 4 (2011): 1843-1850

[14]. Ji, Shuiwang, Wei Xu, Ming Yang, and Kai Yu.3D convolutional neural networks for human action recognition. IEEE transactions on pattern analysis and machine intelligence 35, no. 1 (2012): 221-231.

Cite this article as :

S. Karthickkumar, Dr. K. Kumar, "Deep Learning for Human Action Recognition with Convolution Neural Network", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 6 Issue 4, pp. 376-380, July-August 2020. Available at
doi : https://doi.org/10.32628/CSEIT206466
Journal URL : http://ijsrcseit.com/CSEIT206466