International Journal of Future Generation Communication and Networking
Vol.14, No. 1, (2021), pp. 4206 – 4212

CNNs Framework for Group Activity Recognition

Payal Jape[1] , Sumit Kumar[2], Kumar Satyam[3], S.S.Bhosale[4] ,H.B.Ghorpade[5], C.G.Patil[6]

Department of E&TC Engineering, SKNCOE, SPPU, Pune, India
[1]payaljape12@gmail.com
[2]sumit042014@gmail.com
[3]0000satyamsingh@gmail.com

Abstract

Human group activity recognition has drawn the attention of researchers worldwide because of the significant role it plays in many applications, including video surveillance and public security. To ensure high detection accuracy, current state-of-the-art tracking techniques require human supervision to identify objects of interest before automatic tracking can take place. Activity recognition in humans is one of the active challenges that finds its application in numerous fields such as, medical health care, military, manufacturing, assistive techniques and gaming. Due to the advancements in technologies the usage of smartphones in human lives become inevitable. The sensors in the smartphones help us to measure the essential vital parameters. These measured parameters enable us to monitor the activities of humans, which we call as human activity recognition. In this System, we present a framework based on convolutional neural networks (CNNs) for group activity recognition.

Keywords - *Group Activity Recognition, Convolutional Neural Network*

I.I. INTRODUCTION

Group activity recognition is an important problem in video understanding and has many practical applications, such as surveillance, sports video analysis, and social behavior understanding. To understand the scene of multiple persons, the model needs to not only describe the individual action of each actor in the context, but also infer their collective activity. The ability to accurately capture relevant relation between actors and perform relational reasoning is crucial for understanding group activity of multiple people. However, modeling the relation between actors is challenging, as we only have access to individual action labels and collective activity labels, without knowledge of the underlying interaction information. It is expected to infer relation between actors from other aspects such as appearance similarity and relative location. Therefore, it is required to model these two important cues when we design effective deep models for group activity understanding. In this work, we address the problem of capturing appearance and position relation between actors for group activity recognition. Our basic aim is to model actor relation in a more flexible and efficient way, where the graphical connection between actors could be automatically learned from video data, and inference for group activity recognition could be efficiently performed. Specifically, we propose to model using CNN (Convolution Neural Network) for recognizing group activities. Thus, our model can discover and learn the potential recognizing group activities is a more flexible way. Once trained, our network can not only recognize individual actions and collective activity for group activity understanding. Objectives of this work are to examine activities from video sequences or still images to determine the activities of a person or a group of persons Motivated by this fact, and human activity recognition systems aim to correctly classify input data into its underlying activity category.

ISSN: 2233-7853 IJFGCN

III.LITERATURE SURVEY

A survey of various techniques involved in activity recognition in general and human activity recognition in specific. The different methodologies used are also introduced. Applications reported in brief and challenges identified gives an insight into the area of computer vision based human activity recognition [1] System examined the use of CNN as feature learning method for HAR. Both IMU and audio-based HAR were considered. Results at all stages confirmed that CNNs challenge the state-of-the-art HCF-based approaches, while providing a standardized and automated way to accomplish the feature learning step.[2]

System presented a CNN model for the HAR problem. Focused on a set of activities extracted from a common exercise program for fall prevention, training our model data sampled from different sensors, in order to explore the classification capabilities of each individual unit, as well as groups of units [3].

The System Propose an end-to-end deep architecture, Differential Recurrent Convolutional Neural Networks (DRCNN), for group activity recognition. DRCNN does not require human trajectory or bounding-box information and achieves fully automatic group behavior understanding without sacrificing performance. When no human supervision is utilized for state-of-the-art methods, DRCNN outperforms these techniques by a significant margin [4]. Developed a transfer learning framework using convolutional neural networks to build a personalized activity recognition model with minimal user supervision [5].

Presented a framework based on multi-stream convolutional neural networks (CNNs) for group activity recognition. Streams of CNNs are separately trained on different modalities and their predictions are fused at the end. They evaluate the method on the Volleyball and Collective Activity datasets [6]. Proposed system accurately recognizes six types of physical activities, viz. walk, walk-upstairs, walk-downstairs, sit, stand and lay using convolutional neural network framework [7].

A novel fast and robust deep convolutional neural network structure (FR-DCNN) for human activity recognition (HAR) using a smartphone. The effectiveness and extends the information of the collected raw data from the inertial measurement unit (IMU) sensors by integrating a series of signal processing algorithms and a signal selection module [8].

IV.IMPLEMENTATION DETAILS OF MODULE

Main goal is to recognize group activity in multi-person scene by explicitly exploiting relation information. We first introduce two widely-adopted datasets and the implementation details of our approach. For each video clip, we track a set of bounding boxes (tracklets) around each person over T = 10 time steps by the object tracker n the Dlib library. To address the problem of person missing in some frames, we adopt the simple strategy used to make up the feature of the missing person by a full-zero matrix. Specifically, System train a CNN model to recognize individuals' actions, and extract the CNN features of individuals on the person's bounding boxes, which are input for group activity recognition. For fair comparison, we employ the pre-trained model to extract the CNN feature of each person on the person's bounding box.

ISSN: 2233-7853 IJFGCN

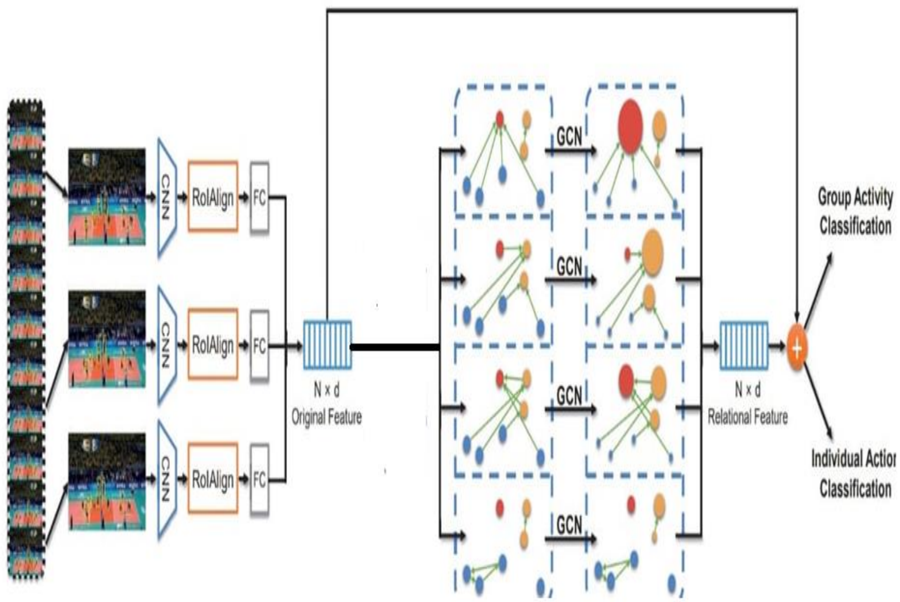

Fig 1:System Architecture

V.EXPERIMENTAL AND RESULT

In order to perform experimental evaluation of the models under consideration, various performance metrics like Accuracy, Precision and Recall is used in this work. Below are some performance confusion metrics are given.

$$\textbf{\textit{Accuracy}} = \frac{True_{possitive}+True_{negative}}{True_{possitive}+True_{negative}+False_{possitive}+False_{negative}} \ldots(1)$$

$$\textbf{\textit{Precision}} = \frac{True_{possitive}}{True_{possitive}+False_{possitive}} \ldots\ldots(2)$$

$$\textbf{\textit{Recall}} = \frac{True_{possitive}}{True_{possitive}+False_{negative}} \ldots\ldots(3)$$

Models are implemented using Python3. All the models are evaluated, where in each split 80% of the data is kept for training purpose (training data) and the rest for testing (testing data). The Convolution and Pooling parts of each of the model may be followed by fully connected layers, with ReLU activation and finally a single-node prediction layer with sigmoid activation function. Here are some screenshots from proposed system where an input video is been given an the output is displayed.

ISSN: 2233-7853 IJFGCN

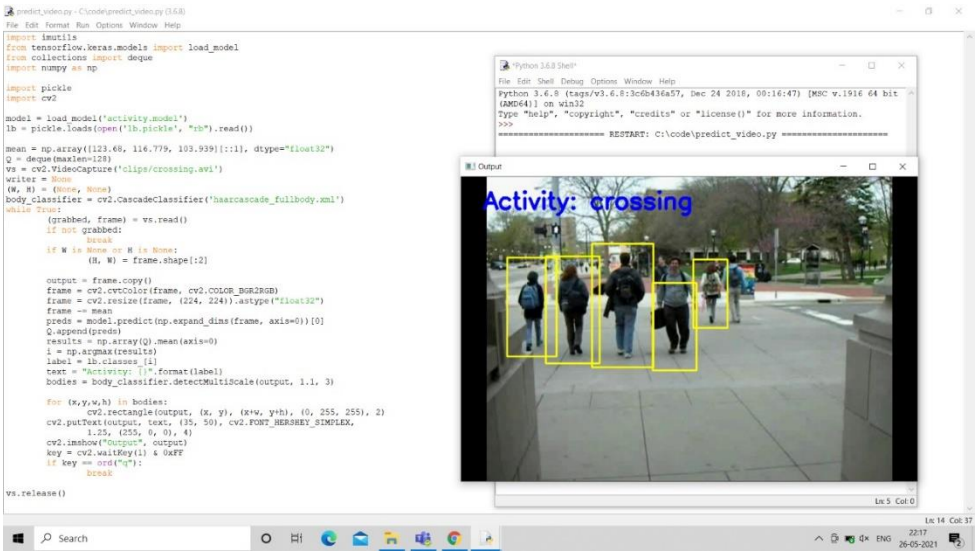

Fig 2: Crossing Activity Detection

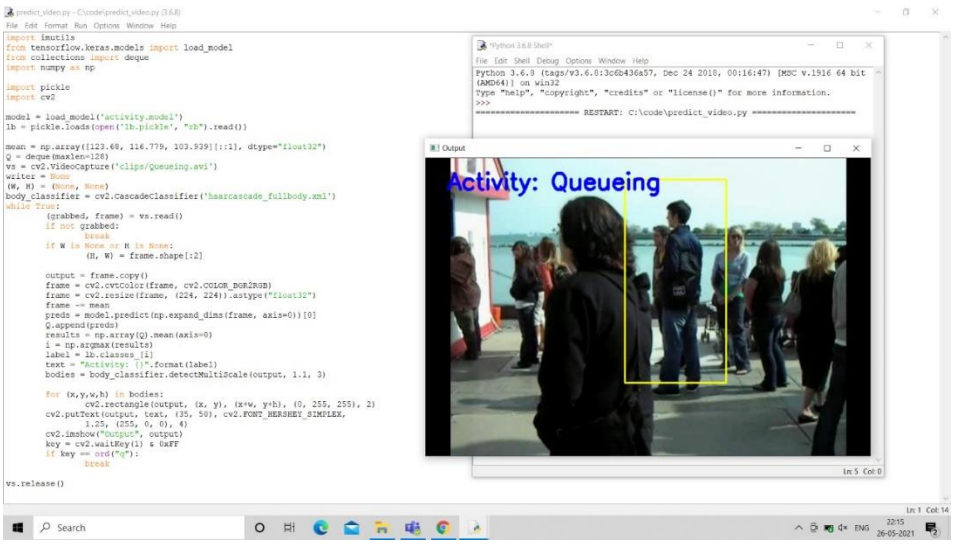

Fig 3: Queuing Activity Detection

ISSN: 2233-7853 IJFGCN

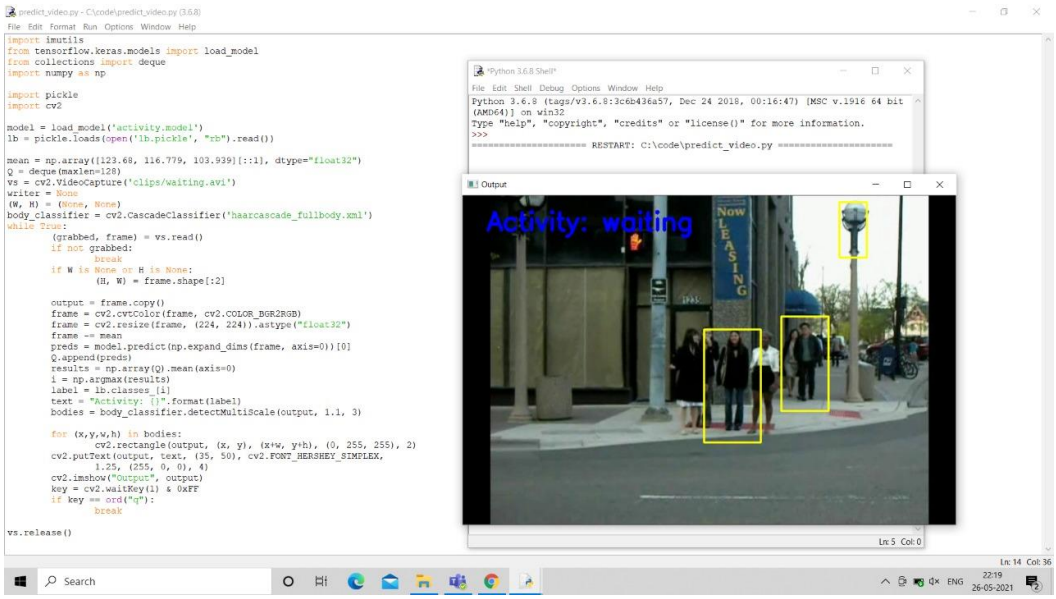

Fig 4: Waiting Activity Detection

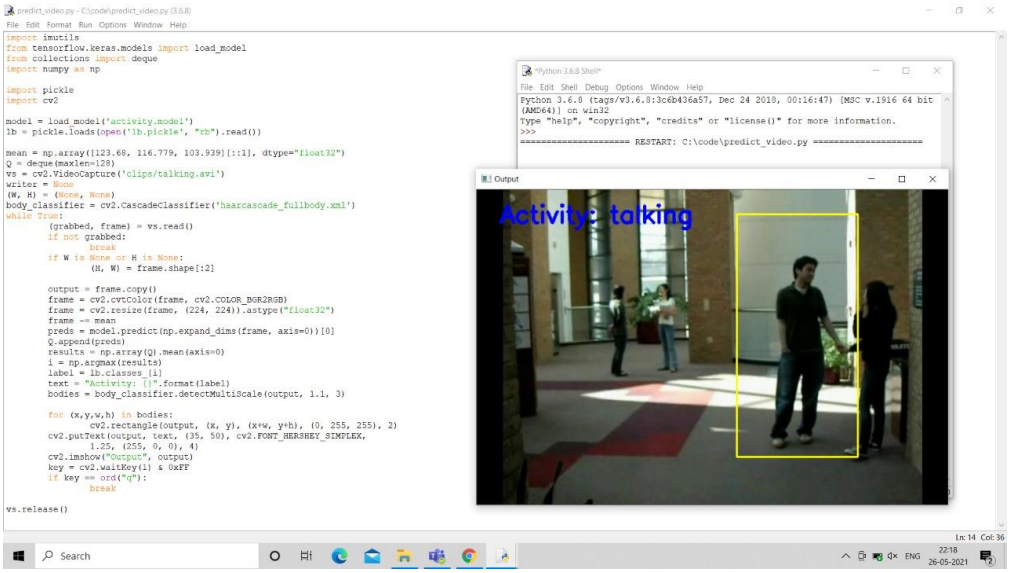

Fig 5: Talking Activity Detection

ISSN: 2233-7853 IJFGCN

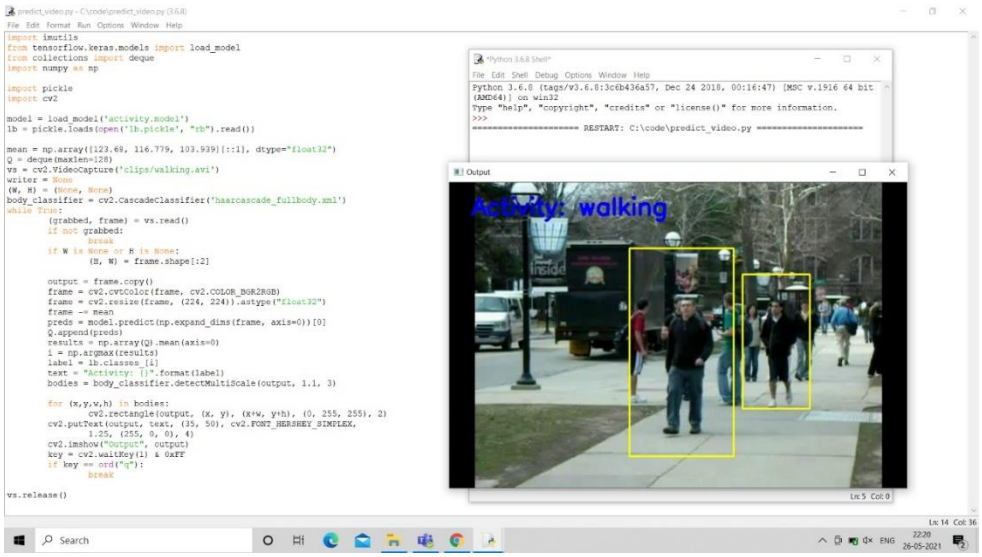

Fig 6:- Walking Activity Detection

VI.CONCLUSION

The system proposed here detects group activity of persons in a given video as input and is converted into frames before reorganization using Convolutional Neural Network. CNN detects the patterns of emotion as we trained the model with large amount of data. We also evaluate the proposed model on datasets and establish new state-of-the-art results.

VII. REFERENCES

[1] Prarthana T V, Dr. B G Prasad, “Human Activity Recognition using Computer Vision based Approach – Various Techniques”, International Research Journal of Engineering and Technology (IRJET), Volume: 07 Issue: 06 | June 2020

[2] Federico Cruciani, Ian Cleland, Paul McCullagh, Konstantinos Votis, “Feature learning for Human Activity Recognition using Convolutional Neural Networks ”, Springer, 2020

[3] Antonio Bevilacqua li EL Moussati and Omar Moussaoui , “Human Activity Recognition with Convolutional Neural Networks”, Human Activity Recognition with Convolutional Neural Networks, September 2018

[4] Naifan Zhuang, Jun Ye and Kien A. Hua," Group Activity Recognition with Differential Recurrent Convolutional Neural Networks”,12th IEEE International Conference on Automatic Face and Gesture Recognition,2017

[5] Seyed Ali Rokni and Marjan Nourollahi, "Personalized Human Activity Recognition Using Convolutional Neural Networks“ ,Proceedings of AAAI Conference on AI, 2018

[6] Sina Mokhtrazadeh and Mina Ghadimi Atigh, "A Multi-Stream Convolutional Neural Network Framework for Group Activity Recognition", Computer Vision and Pattern Recognition , Cornell University , 2018

[7] Surya Dhanraj and Dinesh Dash, "Efficient Smartphone-Based Human Activity Recognition Using CNN", International Conference on Information Technology (ICIT), 2019

[8] Wen Qi and Hang Su , "A Fast and Robust Deep Convolutional Neural Networks for Complex Human Activity Recognition Using Smartphone",2019

ISSN: 2233-7853 IJFGCN

[9] Ming Zeng and Bo Yu,“ Convolutional Neural Networks for human activity recognition using mobile sensors”, IEEE
[10] Jiahui Huang and Ning Wang, “TSE-CNN: A Two-Stage End-to-End CNN for Human Activity Recognition”, IEEE

ISSN: 2233-7853 IJFGCN