HUMAN ACTIVITY RECOGNITION SYSTEM USING DEEP LEARNING

Mr. S.A.K. JAINULABUDEEN[1], KUMARA VAIBHAV C[2], SHABARI RUBAN M[3], SHALIN P SUNIL[4]

[1] Assistant Professor, Department of CSE, Panimalar Engineering College

[2] Department of CSE, Panimalar Engineering College

[3] Department of CSE, Panimalar Engineering College

[4] Department of CSE, Panimalar Engineering College

1.Abstract:

Human Activity Recognition using smartphone sensors like accelerometers is one of the hot topics of research. It is one of the time series classification challenges. Many methods are being discussed when it comes to human activity recognition. Normally Video Regression models are used, and one of the most common video regression models used is the Convolutional Neural network(CNN). But when it comes to CNN it consumes large processing power, and they are also considered slow. That is why CNN is mostly used for object detection problems. We can use LRCN (Long-term Recurrent Convolutional Network) to recognize various activities of humans like standing and walking. The LRCN model is a recurrent neural network that can learn order dependency in sequence prediction challenges. This model is employed because it aids in the recall of values over arbitrary periods. LRCN is also considered to be fast and more accurate while compared to CNN when it comes to video regression problems.

2.Introduction:

The difficulty of predicting what a person is doing based on a trail of their movement using sensors is known as Human Activity Recognition (HAR).

Standing, sitting, jumping, and moving up and downstairs are common examples of movements. Sensors, such as those found on a smartphone or a vest, are frequently used to gather accelerometer data in three dimensions (x, y, z). An intelligent computer system can then offer aid after the subject's action is noticed and known. It's a difficult topic to solve because there's no clear analytical approach to link sensor data to specific actions in a broad sense. The high number of sensor data acquired (e.g., tens or hundreds of observations per second) and the traditional use of hand-crafted features and heuristics from this data in constructing predictive models make it technically tough.

Deep learning algorithms have lately found success on HAR problems due to their capacity to learn higher-order features automatically.

[6] Sensor-based activity recognition aims to extract high-level information about human activities from a large number of low-level sensor readings. In recent years, traditional pattern recognition techniques have made remarkable improvement. Those methods, on the other hand, frequently rely on heuristic hand-crafted feature extraction, which may limit their generalizability. Deep learning has recently advanced to the point that it is now able to execute autonomous high-level feature extraction, resulting in promising results in a variety of fields.

The main objective of this work is to build a useful model that can predict activities from a live video frame by frame.

3.Existing System:

The purpose of this study is to see how effective deep learning-based algorithms are at recognising and analysing different elements of human behaviour. However, the majority of our video surveillance systems are still run in the old-fashioned method, with anomalies and evidence gathered solely through offline videos. It's difficult to create real-time alarms, pop-up notifications, and continuously monitor crisis situations.

As a result, real-time human behaviour identification technologies must be developed to reduce security staff workload and increase productivity.

4.Literature Survey:

Human Activity Recognition Based On Convolutional Neural Network[1] - Smartphones are ubiquitous and becoming increasingly sophisticated, with ever-growing sensing powers. Recent years, more and more applications of activity recognition based on sensors are developed for routine behavior monitoring and helping the users form a healthy habit. In this field, finding an efficient method of recognizing the physical activities (e.g., sitting, walking, jogging, etc) becomes the pivotal, core and urgent issue. In this study, we construct a Convolutional Neural Network (CNN) to identify human activities using the data collected from the three-axis accelerometer integrated in users' smartphones. The daily human activities that are chosen to be recognized include walking, jogging, sitting, standing, upstairs and downstairs.

Human Activity Recognition Using Convolutional Neural Networks[2] - Using smartphone sensors to recognize human activity may be advantageous due to the abundant volume of data that can be obtained. In this paper, we propose a sensor data based deep learning approach for recognizing human activity. Our proposed recognition method uses linear accelerometer (LAcc), gyroscope (Gyr), and magnetometer (Mag) sensors to perceive eight transportation and locomotion activities. The eight activities include: Still, Walk, Run, Bike, Bus, Car, Train, and Subway. In this study, the Sussex-Huawei Locomotion (SHL) Dataset of three participants are used to recognize the physical activities of the users. Fast Fourier Transform (FFT) spectrograms generated from the three axes of the LAcc, Gyr, and Mag sensor data are used as input data for our proposed Convolutional Neural Network (CNN) model.

Analysis of Human Activity Recognition using Deep Learning[3] - There will be discussion on mainly two models-2-D Convolutional Neural Network and Long-Short term Memory. In order to maintain the consistency and credibility of the survey, both models are trained using the same dataset containing information collected using wearable sensors which was acquired from a public website. They are compared using their accuracy and confusion matrix to check the true and false positives and later the various aspects and fields, where the two models can separately and together be used in the wider field of Human Activity Recognition using image data have been explained. The experimental results signified that both Convolutional Neural Networks and Long-Short term memory model are equally equipped for different situations, yet Long-Short Term memory model mostly appears to be more consistent than Convolutional Neural Networks.

A Survey on Human Activity Recognition and Classification[4] - Activity Recognition and Classification is one of the most significant issues in the computer vision field. Identifying and recognizing actions or activities that are performed by a person is a primary key goal of intelligent video systems. Human activity is used in a variety of application areas, from human-computer interaction to surveillance, security, and health monitoring systems. Despite ongoing efforts in the field, activity recognition is still a difficult task in an unrestricted environment and faces many challenges. In this paper, we are focusing on some recent research papers on various methods of activity recognition. The work includes three popular methods of recognizing activity, namely vision-based (using pose estimation), wearable devices, and smartphone sensors. We will also discuss some pros and cons of the above technologies and take a view on a brief comparison between their accuracy.

Human Activity Recognition Using Smartphones[5]- Human Activity Recognition(HAR) is classifying activity of a person using responsive sensors that are affected from human movement. Both users and capabilities(sensors) of smartphones increase and users usually carry their smartphone with them. These facts makes HAR more important and popular. This work focuses on recognition of human activity using smartphone sensors using different machine learning classification approaches. Data retrieved from smart phones' accelerometer and gyroscope sensors are classified in order to recognise human activity. Results of the approaches used are compared in terms of efficiency and precision.

5.Proposed System:

The accuracy is high in the proposed system. The proposed system loading speed and execution speed are fast when compared with the existing system. The proposed system is highly efficient and scalable and is also further improved for complex use cases.
With few future enhancements, this proposed system can predict outputs from live video feeds like for example live feeds from a CCTV. As this system predicts action on every frame of a video this system can be used in surveillance programs and supervision programs.

This proposed system makes use of the LRCN algorithm which is a combination of CNN and LSTM. This is a more modern way of handling the problem and it is promised to perform fast with high accuracy and fewer computing resources.

6.Project Module:

There are two modules to this project. The first module discusses how the deep learning models are trained, and the second model explains how the actual process of predicting the action in a video is carried out.

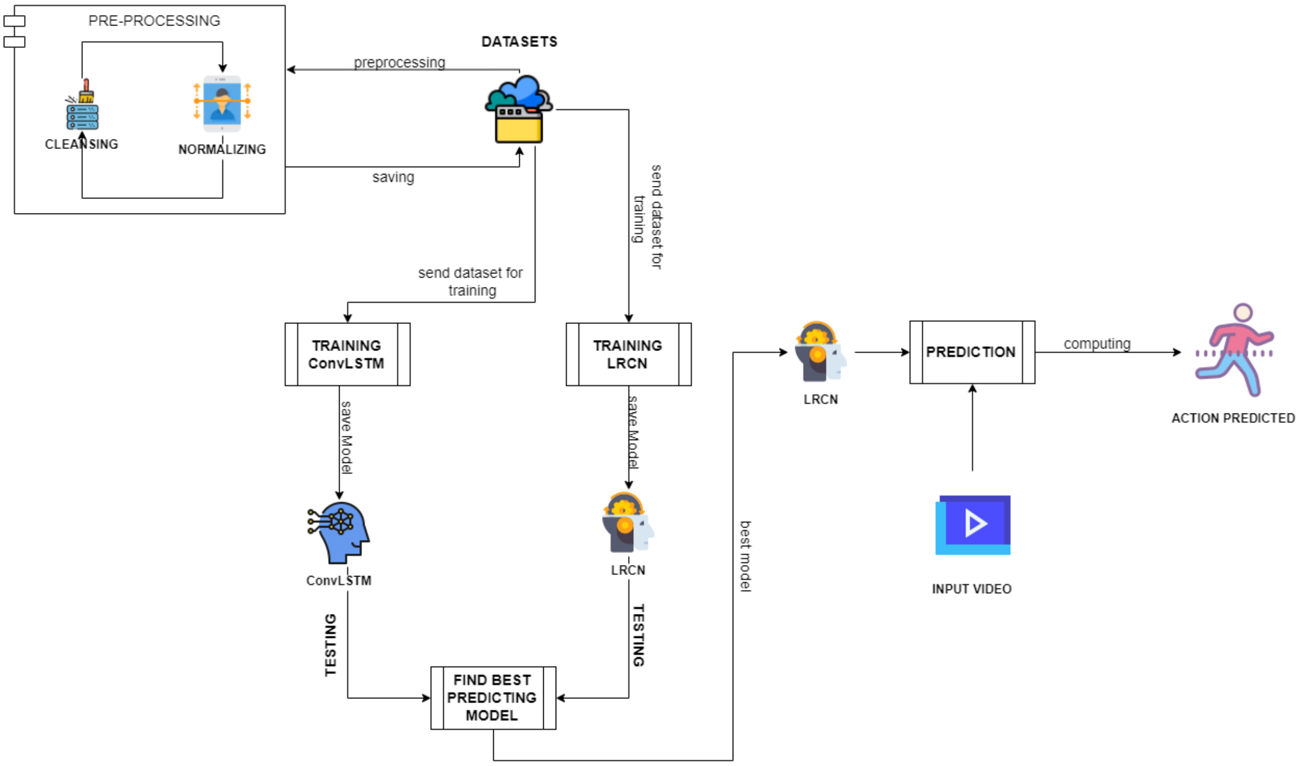

Fig 6.1 – System Architecture

Training Module: The dataset employed in this project is named UCF - 50, and it has 50 action classes, each of which contains clips describing those videos.

The clips are then pre-processed and labeled to serve as training data sets for the two models being used.
Convolutional Neural Network (CNN) and Long Recurrent Convolutional Network (LRCN) are the two models employed here.
The accuracy of these models is tested once they have been trained. The model with the highest accuracy is considered as best performing model.

Activity Prediction Module: The user provides the video input, which is then converted into several frames of images. After that, each frame is input into the best-performing model, which produces an output video that shows the action performed in each frame of the clip.

7.Neural Network Model Implementation:

ConvLSTM Model:

We'll use a combination of ConvLSTM cells to implement the first strategy in this stage. A ConvLSTM cell is a type of LSTM cell that includes convolutional operations in the network. It is an LSTM with convolution included into the architecture, allowing it to recognize spatial characteristics of input while taking into consideration temporal relationships.

This method efficiently captures the spatial relationship between individual frames as well as the temporal relationship between distinct frames when it comes to video classification. The ConvLSTM can take in 3-dimensional input (width, height, num of channels) as a result of this convolution structure, whereas a standard LSTM can only take in 1-dimensional input, making it unsuitable for modelling Spatio-temporal data on its own.

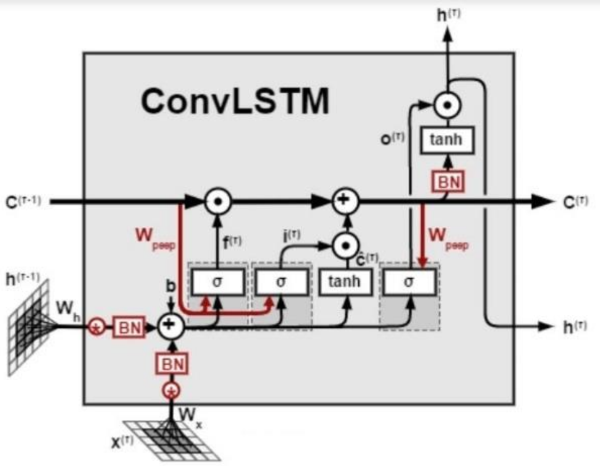

Fig 7.1 – ConvLSTM

LRCN Model:

In this stage, we'll combine Convolution and LSTM layers in a single model to apply the LRCN

Approach. A CNN model and an LSTM model trained separately can be used in a similar way. The CNN model may be used to extract spatial information from video frames, and a pre-trained model that can be fine-tuned for the application can be utilized for this. The CNN features can then be used by the LSTM model to predict the action being performed in the video. But, in this case, we'll use the Long-term Recurrent Convolutional Network (LRCN), which integrates CNN and LSTM layers into a single model. The Convolutional layers are utilized to extract spatial characteristics from the frames, and the retrieved spatial features are given to the LSTM layer(s) for temporal sequence modelling at each time-step. This method allows the network to learn spatiotemporal features in an end-to-end training, resulting in a robust model.

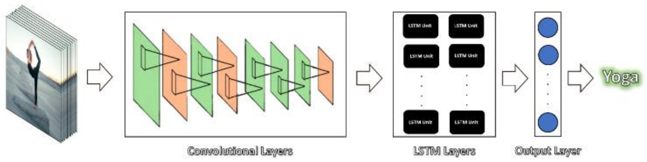

Fig 7.2 - LRCN

We'll also employ a Time Distributed wrapper layer, which allows us to apply the same layer to each frame of the movie separately. As a result, if the layer's input shape was (width, height, num of channels), it may now take input of shape (no of frames, width, height, num of channels), which is highly useful because it allows you to input the entire movie into the model in a single shot.

Training loss and validation loss curves for both models:

ConvLSTM – in this model the total loss went down to almost zero, but still validation loss was much higher and at one point it started to inccrease which indicates the model is overfitting. This means the model is trying to catch more than required data points and in the process it tries to catch noise too, hence the efficency decreases.

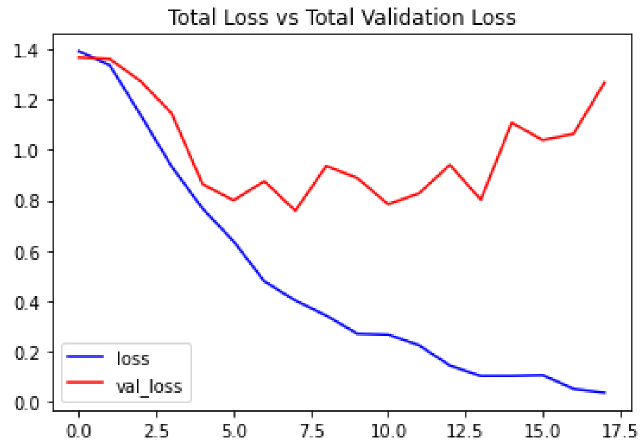

Fig 7.3 – Loss Curve for ConvLSTM

LRCN – in this model the validation loss curve came down with total loss and hence the model was stable and working fine. When both the curves looks similar without any deviation the model with good accuracy and the model is trained well.

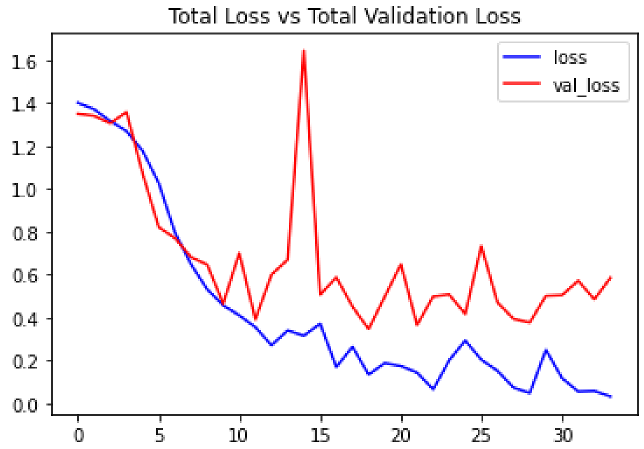

Fig 7.4 – Loss Curve for LRCN

8.Results:

The final step of the proposed system is to detect human activity. Object tracking is described as estimating the trajectory of a target object through a video frame sequence. For testing purpose, we first tested 40% and then 30% of the overall dataset for classification and localization of actions. To accurately evaluate of our proposed approach, we created our own dataset of abnormal patient activities. The final output is predicted which detects human activity like walking, moving.

Predicted outputs:

Fig 8.1 - Walking with dog detection

Fig 8.2 - Horse race detection

Fig 8.3 - Swing

9.Conclusion:

Thus Human activity recognition is successfully implemented. Following extensive testing, comparing results revealed that the model generates exact results with an accuracy of 90.90 percent, which is an important strategy for application in open contexts. Because of its high Frame Per Second, this lightweight model is easy to assess and may be used constantly with the expansion of tiresome computation or picture collapsing.

Since this model can detect activities faster than other models, this system can be used for surveillance and supervision programs. If the GPU resources is improved, this model can be trained to improve its accuracy even more.

Human activity recognition system has so many use cases in the current environment like patient monitoring, traffic monitoring, and also in childcare centers to babysit the kids and in other industries where supervision is done.

10.References:

[1] Wenchao Xu; Yuxin Pang; Yanqin Yang; Yanbo Liu," Human Activity Recognition Based On Convolutional Neural Network", 2018 24th International Conference on Pattern Recognition (ICPR)

[2] Gulustan Dogan; Sinem Sena Ertas; İremnaz Cay," Human Activity Recognition Using Convolutional Neural Networks",2021 IEEE CIBCB

[3] Lamiyah Khattar; Chinmay Kapoor; Garima Aggarwal," Analysis of Human Activity Recognition using Deep Learning",Confluence

[4] Abhay Gupta; Kuldeep Gupta; Kshama Gupta; Kapil Gupta, “A Survey on Human Activity Recognition and Classification”, 2020 IEEE ICCSP

[5] Erhan Bulbul; Aydin Cetin; Ibrahim Alper Dogru,” Human Activity Recognition Using Smartphones”, 2018 IEEE ISMSIT

[6] Jindong Wang, Yiqiang Chen, Shuji Hao, Xiaohui Peng, Lisha Hu, "Deep Learning for Sensor-based Activity Recognition: A Survey", arXiv:1707.03502