Person Activity Identification Based on Convolutional Neural Networks

[1]Ms. E. Padma, [2]R. Ruchitha, [3]S. Haritha

[1]Assistant Professor, [2]Graduate Student, [3] Graduate student

Dept. of Computer Science Engineering,

SCSVMV (Deemed to be University), Enathur, India

Abstract- Convolutional neural networks (CNNs) are increasingly being used as a feature learning technique for human activity recognition (HAR). Working independently on a group of people using spotting equipment assisted the Convolution Neural Network established for people's explicit behaviour in public areas. Moreover, a photograph is divided into a visual message with these kinds of elements of person action. Eventually, it has a plan to approach all the photographs employing a strong process called background reduction that tracks changes in image arrangement and aids in the detection of numerous attract. For instance, the training news sets are integrated with a CNN model's framework, and deep learning networks, which are made of random gradient drops, are used to update the model's framework. In the end, different functions involving samples are systematised and known using the obtained system repetition. Area unit will therefore compare the immediate cognitive processes. The findings demonstrate that a convolutional neural network will automatically analyse a person's action model and identify that person's activity without the use of any metadata. The primary foundation of conventional human action detection is the global property of digital figures. Deep Neural Networks (DNNs) have a great ability to recognize any object nowadays due to the growth in computing power and processing capacity, which has successfully brought in a new era of machine learning. It is based on deep learning-based person action spotting applying the CNN model. Convolution Neural Network built for the explicit action of people in public locations was backed by everyone working up on collection of people bearing spotting structure. Importantly, a visual communication with certain human specifics from an action is divided into an image. Eventually, a plan to approach all the photographs using a powerful mechanism called background reduction that tracks changes in image layout and aids in the detection of numerous attract. For instance, CNN's model is outlined in the training news sets, and deep learning networks that are built of random gradient drops are employed to update our model's framework.

Keywords- Human Activity Recognition (HAR), Convolutional Neural Network (CNN), Person Action Identification (DNN).

I. Introduction

A short while ago, person activity identification supported laptop vision earned a lot of attention and support because of the quick development of science, technology, and video signal diverse technologies. Regarding idea of using laptop vision to detect, follow, examine, and keep an eye on human activity. The automation in the field of self-regulating management and pattern recognition is accompanied by a broadcast image. Instead of using the human brain to analyse and be aware of perceptible data, intelligent watching is focused on using the personal computer to convert picture data into a digital signal, it allows intelligent recognition and instinctive analysis of observed images.

In recent years, scientists and researchers have received a lot of attention for their interpretation and approval of human identifications in broadcast police operations. The two categories into which recognition or identification strategies are generally divided are: Considering state-space representation, behavioural recognition and identification techniques supported an identical framework. The recognition behaviour patterns technique supported identical matching alludes to the setup of an honest depiction example series that indicates unchanged prescribed behaviour of the material. Later, it will go with object broadcast image series detection within the example, and if the matching is successful, the behaviour is supposed to possess identical features. If not, the example either resolves recognition but the previously mentioned behaviour is poor.

The method of chance is used to link the reports after the method of behavioural identification supported representation of state-space that describes a specific masquerade report. Because deep learning is used to overcome problems when a data set is dissimilar and correlated, machine learning via neural networks used to be a study that was highly regarded. Pictures are frequently used directly in this instance as a computer file divert to avoid knowledge modification and eliminate the convoluted process within an antiquated identification methodology. Deep learning, a classification method for machine learning, involved numerous algorithms that several analysts found interesting to examine in the discourse of the laptop illusion.

Particularly recent years, deep movement algorithms were used to complete the two-person action. On similar projects, numerous scientists from both here and abroad have done research. For a particular type of semantic network, learning techniques are used. success used in many fields, such as image and speech recognition, and later extended to behavioural identification using analytics.

A. Objective

The goal of this research is to anticipate what a human would do using video as input. Accurately predicting the actions is the major objective. Reading, writing, jogging, using a computer, and playing musical instruments are just a few of the human activities that the model can identify.

In order to analyse a person's behaviour in a real setting, HAR models are designed to provide information on human activities. The goal of human activity recognition (HAR) is to categorise an individual's actions.

These days, gathering this kind of data is not a difficult undertaking. Nearly everyone now has a device that tracks their movements thanks to the development of the Internet of Things. It may be a smartphone, a smartwatch, or even a pulsometer. The deep learning model that is utilised to identify activities is dependent on the devices. Due to the complexity of human actions and the inherent variances between two people, it can be quite difficult to analyse this data.

II. Literature Survey

G. Dogan, S. S. Eras and İ. Cay, "Human Activity Recognition Using Convolutional Neural Networks," 2021 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2021, pp. 1-5, Doi: 10.1109/CIBCB49929.2021. 9562906.The large amount of data that may be collected by using smartphone sensors to detect human activities may be helpful. Here, they suggest a deep learning method based on sensor data for identifying human activities. The suggested recognition system recognizes eight transportation and locomotion activities using sensors from the linear accelerometer (Lack), gyroscope (Gyri), and magnetometer (Mag) families. Still, Walk, Run, Cycle, Bus, Auto, Train, and Subway are among the eight activities. The Sussex-Huawei Locomotion (SHL) Dataset comprising three participants is used to identify the users' physical activity.

Crucian, F., Varekai's, A., Nugent, C. et al. Feature learning for Human Activity Recognition using Convolutional Neural Networks. CCF Trans. Pervasive Comp. Interact. 2, 18–32 (2020). Convolutional neural networks (CNNs) are increasingly being used as a feature learning technique for human activity recognition (HAR). CNNs are capable of automatically extracting features, in contrast to traditional machine learning techniques that demand domain-specific knowledge. CNNs, on the other hand, need a training phase, which makes them vulnerable to the cold-start issue. This work presents a case study where the use of a pre-trained CNN feature extractor is assessed under practical circumstances. The case study is divided into two sections: (1) In order to determine the best candidate models for HAR and produce a pre-trained CNN model, various topologies and parameters are evaluated. The pre-trained model (2) is subsequently used as a feature extractor, and its performance is assessed using a sizable real-world dataset. Both the audio-based HAR and the inertial measurement unit (IMU) CNN applications were taken into consideration

G. Yao, T. Lei, and J. Zhong, "A review of convolutional-neural-network-based action recognition," Pattern Recognition Letters, vol. 118, pp. 14–22, 2019.Video indexing, intelligent surveillance, multimedia understanding, and other industries all use video action recognition extensively. Recently, by combining the learning of deep knowledge using Convolutional Neural Network, it was much enhanced (CNN). After looking at the notable CNN-based action recognition works as a result of this. Since videos are typically seen as 3D spatiotemporal signals and

CNN is primarily developed to extract 2D spatial characteristics from still images, exploiting temporal information is the main challenge when expanding CNN from image to video.

III. PROPOSED METHODOLOGY

A. EXISTING SYSTEM

The prior system, which required a person to sit in front of a monitor to oversee and direct human activity, was manual. It was busy, time-consuming, expensive, and prone to human mistake and carelessness. Also, some systems began to use sensor data to identify human activity, but they required the user to wear them, which restricted the application of activity detection in open environments generally. Moreover, some tasks need the use of a stereo camera. A stereo camera could consist of a pair of cameras positioned next to one another but spaced apart by a very small distance. Despite the fact that these cameras have identical faces, the images they collect have a little spatial variance between them. The existing system contains of the following drawbacks: The greater number of cameras used in this and detecting time period is more by human. Less accuracy. Time consuming. More human effort.

B. Motivation and Problem Statement

Because to its broad potential for application in sectors like video surveillance, human computer interaction, and social video recommendation, behaviour recognition is receiving more and more academic attention. Studies on single and two-person interactions as well as multi-person behaviour recognition have been published in the topic of behaviour recognition research. Many research findings based on the identification of single person behaviour have recently been revealed. When two individuals engage, their conduct is frequently more complicated and includes a wider variety of physical actions than when one person does.

C. Proposed System

By using CNN Algorithms, planned a system that will identify suspicious activity from regular individuals such as walking, fighting with a large group, and terrorist attacks. It will take pictures and issue an alert that will enable the user to flee suspicious activities once it has recorded every suspicious human behaviour and any object that poses a hazard to human safety.

Using phases of still-image processing, CNN model is used to learn visual representations of person activities. CNN uses video as an input or uses real-time activity detection. These frames can be subjected to the CNN algorithm. Features can be extracted from an image after it has been divided into pixels. Here, by recognising potentially dangerous things, spotting hazards to individuals and use the information as a surveillance tool. Threat detection is conceivable by using those dangerous objects and situations to train the CNN algorithm. observing a new hire as they carefully carry out a task (ex., proper steps and procedures when making a pizza, including rolling out the dough, heating oven, putting on sauce, cheese, toppings, etc.). inspecting a food service employee hand for signs of cross-contamination after they use the restroom or handle food. It consists of the following advantages: It is easy to understand , detect and fast to implement. The highest accuracy that predicts images. Helps for better performance of robots in artificial intelligence. Less human effort Less time the CNN algorithm, which predicts images with the maximum accuracy possible. It is simple to learn and quick to implement.

D. Proposed Architecture

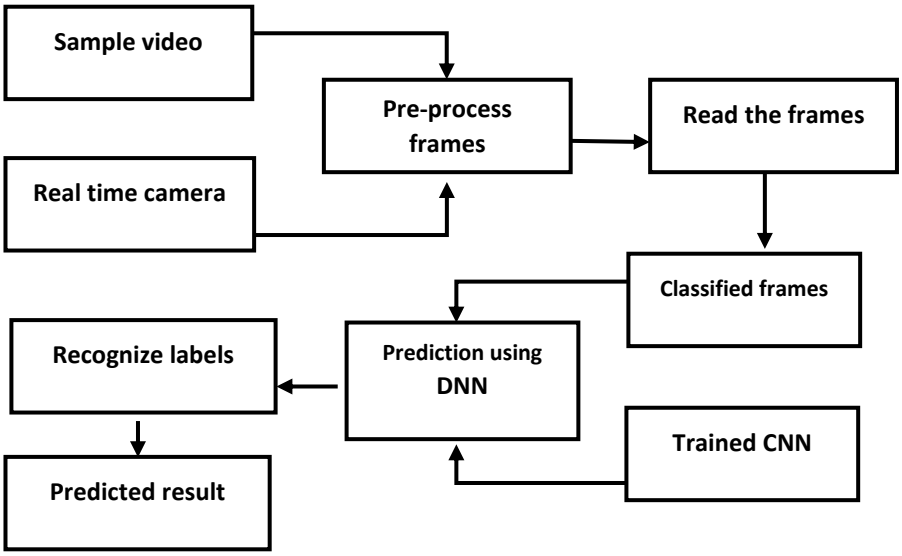

Figure 1: Proposed Architecture

IV. Implementation Process

There are numerous techniques for detecting and identifying human behaviour. A system that is prepared to handle both online and offline video surveillance in order to detect suspicious activity was planned, and it supported a person's conduct. Gather the information in a specific sequence, taking into account human gestures such as walking, sitting, standing, running, etc. Starting with a base set of measured data, feature extraction creates derived values that are meant to be instructive. By identifying crucial details such as the speed of the anatomy during movement, the dynamic properties of limb changes, and so the motion trajectory, image-based human behaviour detection can automatically distinguish between normal human behaviour and intense human behaviour within the context of video surveillance. The use of this technology in sophisticated video surveillance systems is widespread. It is possible to convert the obtained data into pixels. Convolutional networks used have a total of seven layers, including an output layer and a linking layer. First, the input video data is processed according to the second of the processed image's section strategy. Next, adjacent four subsequent images are converted, with the second layer being the largest under-sampling layer. Finally, fully connected layer processing using SoftMax regression model classification follows fully connected layer processing using layer convolution kernel. On the trained dataset, validation and testing can be done to produce the desired results.

V. Experimental Analysis

Figure 2: Playing Tennis

Figure 5: Skiing slalom

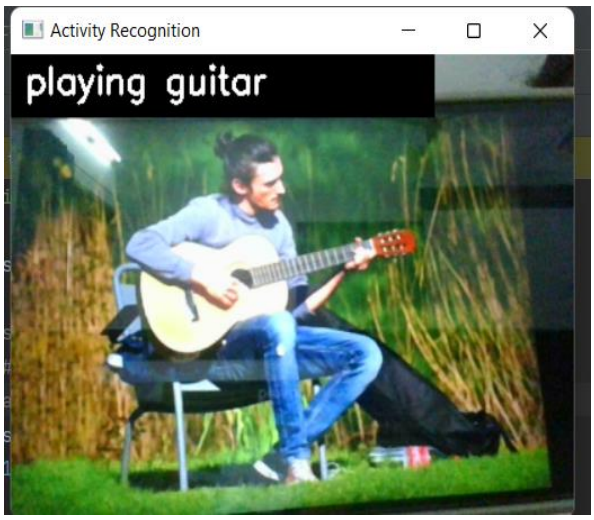

Figure 3: Playing Guitar

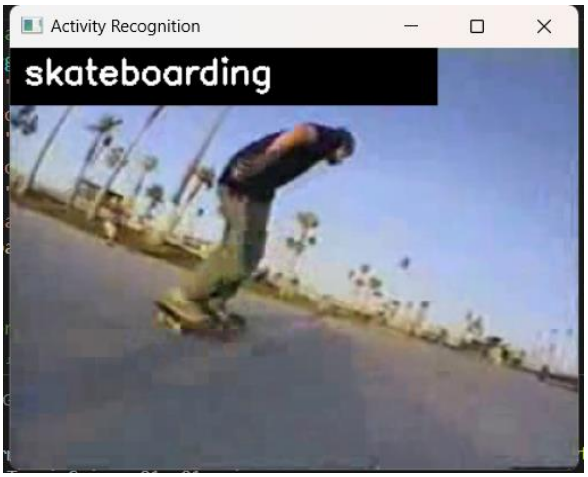

Figure 6: Skateboarding

Figure 4: Salsa Dancing

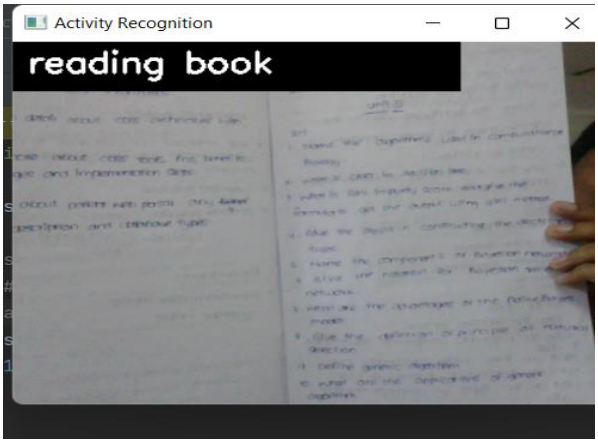

Figure 7: Reading book

VI. Conclusion

The Human Activity Recognition System is being used here. It is a program that recognizes and labels the motions or actions carried out by people. This study begins by conducting a literature review on how to put this application into practice. After doing a requirements and design analysis, providing an input model that, when trained, can identify, and label the activity occurring in the relevant video input. The ResNet architecture of the convolution neural network (CNN) machine learning technique is used to train the model. Overall, it shows how to use OpenCV and Deep Learning to accomplish human activity recognition. Using a human activity recognition pretrained model on the Kinetics dataset, which contains 400–700 human activities and more than 300,000 video clips, will enable completion of this assignment. The number of videos depends on the version of the kinetics dataset you are using. The model makes use of the ResNet architecture, which employs 3D kernels rather than the conventional 2D filters and enables it to incorporate a temporal component for activity identification.

Improvements to current approach, like image uploading along with video and storing the results in a database. For user interaction, creation of a user interface or a website can be done, so that the person who do not know how to use code can also use the project and get benefitted. Planning to increase the number of activities for the experiment in future. Usage of more camera types that will allow us to increase our data sets in other dimensions. As a future enhancement the video images can be replaced by image for activity recognition. Its hidden representation can then be used as input to the recurrent neural network which should treat as time series prediction problem. The model performs admirably on video streams while delivering passable results on image data. This technique serves as the foundation for numerous other activity recognition applications.

VII. References

[1] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection” (2005) In IEEE CVPR, volume 1, pp: 886-893.

[2] D. Geronimo, A. Lpez, D. Ponsa, and A. D. Sappa. “Wavelets and edge orientation histograms for onboard pedestrian detection”,(2007) In Iberian Conference on Pattern Recognition and Image Analysis, volume 2, pp: 418–425.

[3]Shian-Ru ke, Hoang Le Uyen Thuc, Y.-J. Lee, J.-N. Hwang, J.-H. Yoo, and K.-H. Choi, “A review on video-based human activity recognition”,(2013) volume 2, pp: 88–131.

[4] R. Girshick, F. Iandola, T. Darrell, and J. Malik. “Deformable part models are convolutional neural networks”, (2015) In IEEE CVPR,volume 1, pp: 437–446.

[5] K. He, X. Zhang, S. Ren, and J. Sun. “Deep residual learning for image recognition”,(2016) In CVPR,volume 2, pp: 770– 778.

[6] M. Ziaeefard and R. Bergevin, “Semantic human activity recognition and Pattern Recognition”, (2015) volume 48, pp: 2329–2345.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” (2017). Communications of the ACM, Volume 60, pp: 84–90.

[8] S. Zhang, R. Benenson, M. Omran, J. Hosang, and B. Schiele, “How far are we from solving pedestrian detection?” (2016) volume 1, pp: 3488–3496.

[9] R. Girshick, Abhinav Shrivastava, Abhinav Gupta “Training region-based object detectors” (2015), volume 1, pp: 1440–1448.