International journal of basic and applied research

www.pragatipublication.com

ISSN 2249-3352 (P) 2278-0505 (E)

Cosmos Impact Factor-5.86

Deep Learning-Based Human Abnormal Activity Detection for Smart Surveillance

[1]Naveen G,[2] Mr. Prasanth Kumar,
[1]M.Tech Scholar, Dept. of CSE (AI&ML), Malla Reddy Technical Campus, Malla Reddy Vishwavidyapeeth, Maisammaguda, Hyderabad, Telangana 500100, India, ngnaveengodi@gmail.com
[2]Assistant Professor, Dept. of CSE, Malla Reddy Technical Campus, Malla Reddy Vishwavidyapeeth, Maisammaguda, Hyderabad, Telangana 500100, India., kundaas.web@gmail.com

Abstract

In the modern world, video surveillance is crucial. After incorporating AI, ML, and deep learning into the system, technology evolved too rapidly. There are a variety of systems in place that use the aforementioned combinations to distinguish between different types of suspicious behavior based on the live monitoring of footages. When it comes to human behavior, the most unexpected factor is how hard it is to tell whether anything is suspicious or not. In a school setting, a deep learning method may distinguish between typical and suspect behavior, and if the former is detected, an alarm will be sent to the appropriate authorities. It is common practice to monitor a video by extracting successive frames from it. There are two sections to the whole structure. The first step is to extract features from video frames; the second is to train a classifier to identify suspicious or normal classes using these characteristics.

Keywords: Convolutional Neural Networks(CNNs), Suspicious Human Behaviors, Proactive Security Management, High False Alarm Rates

Introduction

A problem arises when criminals are not apprehended promptly and appropriate safety measures are not implemented, especially in light of rising crime rates. There are surveillance systems that are always collecting data in most urban and metropolitan locations. The likelihood of suspicious behaviors increasing as monitoring data accumulates at an exponential rate. However, due to the complexity and resource requirements of these jobs, human supervision is necessary for their detection by artificial intelligence. One approach to simplifying a work for automation is to break it down into smaller ones, and then look for the subtasks that might lead to a crime. We try to use our models to identify two primary pathways that might lead to criminal activity.
Nowadays, video surveillance is widely used to keep various locations safe and secure, such as public areas, businesses, healthcare facilities, transit hubs, and schools. Automatic surveillance systems are in high demand because to the exponential growth of metropolitan areas and the complexity of human interactions. People used to be the backbone of traditional video surveillance systems, watching live feeds for signs of suspicious behavior. But these systems aren't perfect; human mistake, exhaustion, and subjective judgment may all cause false positives and inconsistent results.

Thanks to advancements in AI, ML, and DL, surveillance has become a more smarter and automated procedure. These advancements in technology allow computers to filter, analyze, and understand video footage in real-time, which may supplement or even replace human operators when it comes to spotting questionable actions. The nature of human conduct is one of the greatest unknowns when it comes to monitoring. Human activities, in contrast to those of immobile things, are complex, multi-faceted, and subject to a wide range of environmental influences. Because of this, it is quite difficult to tell what constitutes "normal" and "suspicious" behavior.
Keeping people safe is of the utmost importance in any setting where learning takes place, but especially in educational

International journal of basic and applied research

www.pragatipublication.com

ISSN 2249-3352 (P) 2278-0505 (E)

Cosmos Impact Factor-5.86

institutions. Keeping track of all the kids, teachers, and staff members on campus manually isn't going to cut it. As soon as suspicious activity is detected, it should not be left unchecked. This includes acts of aggression, illegal access, test cheating, and other disruptive activities. This highlights the critical need for a smart video surveillance system capable of autonomously distinguishing between benign and malicious human actions.

Literature Survey

The diverse variety of applications of Human Activity Recognition (HAR) in healthcare, smart homes, surveillance, and human-computer interaction has propelled it to the forefront of ubiquitous computing, computer vision, and artificial intelligence. Methodologies like as sensor-based, vision-based, and fusion-based approaches to HAR are systematically reviewed in this paper. Using motion signals like gyroscope and accelerometer data collected by smartphones and other wearables is the mainstay of sensor-based approaches, which provide excellent accuracy in confined spaces but suffer from issues with user compliance and power use. Vision-based approaches, in contrast, estimate human activities using body posture, monitor motion, and model temporal relationships using video data and sophisticated machine learning algorithms. To get over individual constraints and be resilient in different situations, fusion-based methods combine input from vision and sensors. Concerns about privacy, real-time processing limitations, inter-subject variability, and sophisticated activity detection in congested settings are among the major obstacles highlighted by the poll. In addition, it delves into potential uses in security, activity tracking, healthcare, and care for the aged. Lastly, the authors discuss unanswered questions and possible next steps for HAR research, highlighting the importance of multimodal fusion, deep learning, and transfer learning.

The vital uses of Human Activity Recognition (HAR) in smart environments, security, healthcare, and sports analytics have led to a surge in its academic interest in recent years. Recent developments in HAR techniques are covered extensively in this overview work. Methods based on deep learning, ensemble methods, and multimodal sensor fusion techniques are highlighted. Deep architectures like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have greatly increased accuracy in activity detection tasks, hence the authors evaluate and contrast classic machine learning techniques with these newer models. Also discussed are ensemble learning algorithms, which include merging several classifiers, as a way to achieve resilience in unpredictable and diverse settings. Additionally, we examine multimodal sensor fusion methods, showcasing their capacity to gather complementing data from various input sources such depth sensors, accelerometers, gyroscopes, and RGB cameras. Computational complexity, energy efficiency, dataset imbalance, and privacy problems are some of the important difficulties highlighted in the subject. Possible areas for further investigation in the future include activity identification using cross-domain transfer learning, explainable artificial intelligence, edge computing, and lightweight deep models.

Human Activity Recognition (HAR) encompasses a broad range of methods, from older machine learning algorithms to more recent systems relying on sensors and smartphones. Wearable technology and mobile phones are becoming ubiquitous, making HAR systems more accessible and feasible. The authors examine the incorporation of integrated smartphone sensors for activity monitoring in daily situations and the use of accelerometers, gyroscopes, and magnetometers in wearable devices for gathering activity signals. Also included in the study are several machine learning algorithms and their uses and limitations. These include Support Vector Machines (SVM), Decision Trees, Random Forests, and k-Nearest Neighbors (kNN). We also take a quick look at vision-based methods that make use of data from cameras. While tackling issues including inter-user variability, noisy data, sensor drift, and the absence of generalized models, the study assesses the computing needs, usability, and performance trade-offs of several HAR techniques. We also go over some of the uses in recovery, ambient-assisted living, healthcare, and fitness monitoring. In their last thoughts, the authors look forward to HAR's potential, highlighting how important it will be to personalize, adapt in real-time, and integrate with surroundings enabled by the Internet of Things.

International journal of basic and applied research

www.pragatipublication.com

ISSN 2249-3352 (P) 2278-0505 (E)

Cosmos Impact Factor-5.86

One of the most popular techniques for Human Activity Recognition (HAR) is the use of wearable sensors, which allow for the constant tracking of bodily motions in everyday life. Recent advances, state-of-the-art algorithms, and problems in the realm of wearable sensor-based HAR systems are examined in this review article. Before moving on to preprocessing methods including filtering, normalization, and feature extraction, the authors provide a thorough evaluation of data collecting methods using physiological sensors, gyroscopes, magnetometers, and accelerometers. From basic locomotion (walking, sitting, running) to complicated composite actions, we examine classification methods, including both classic supervised learning models and cutting-edge deep learning techniques, for their capability to identify these motions. Data variability between users, problems with sensor location, computing efficiency for real-time identification, and privacy concerns related to continuous monitoring are some of the obstacles discussed in the article. New areas of study include integrated multimodal sensors, energy-efficient algorithms, lightweight wearable designs, and customized HAR models. The importance of wearable-based HAR in healthcare, monitoring the elderly, tracking sports performance, and occupational safety is emphasized in this review's conclusion.

Human Activity Recognition (HAR) has been transformed by deep learning, which eliminates the need for human-crafted feature engineering and allows computers to automatically learn complicated spatial-temporal characteristics from raw data. With an emphasis on Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) models, and hybrid deep architectures, this survey study extensively reviews deep learning-based approaches for HAR. There includes a discussion of RNNs and LSTMs in relation to modeling temporal dependencies in sequential data, and an analysis of CNNs in terms of their ability to extract spatial characteristics from video and picture frames. In order to take advantage of spatial and temporal learning at the same time, the article also looks into hybrid models that combine RNNs with CNNs. Important uses include multimodal fusion methods, recognition using wearable sensors, and HAR based on visual input. Deep learning's scalability and accuracy are emphasized by the authors, who also address its drawbacks, including computational overhead, lack of interpretability, and reliance on big labeled datasets. Some areas that might require further investigation in the future include explainable AI, lightweight models for usage on mobile and IoT platforms, unsupervised and semi-supervised deep learning, and transfer learning.

Methodology

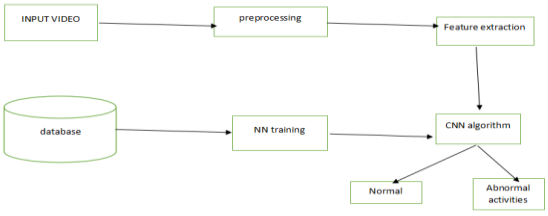

Fig: System architecture

The goal of the suggested system is to use Convolutional Neural Networks (CNNs) to automatically identify and categorize suspicious human actions in surveillance video. In order to handle raw video data, extract significant features, and conduct accurate classification, the framework adheres to an organized workflow. The following parts provide an exhaustive overview of the system:

International journal of basic and applied research

www.pragatipublication.com

ISSN 2249-3352 (P) 2278-0505 (E)

Cosmos Impact Factor-5.86

Input Video

The system begins by collecting video feeds from security cameras set up in public areas, schools, and other monitored areas. The machine is fed these movies as raw data. • Video datasets that have already been captured or live video streams in progress could be the input. A video consists of a series of still images, or frames, taken at discrete points in time. The ability to analyze frame sequences in order to comprehend human motion patterns and context is crucial for suspicious behavior identification. Function: Supplies the system with the first data needed for further analysis. To accurately detect suspicious actions, high-quality input footage is essential.

Pre-processing

Preprocessing is essential since raw video data is often unreliable, redundant, and noisy. The objective is to get the video frames into a consistent and clean format so that features may be extracted. Here are some common steps to prepare the data: • Frame Extraction: In order to do analysis on individual frames, videos are first broken down into a sequential order.The process of resizing involves reducing computing effort and maintaining uniformity by resizing frames to a specific resolution, such as 224×224 pixels. Camera noise, motion blur, and background artifacts may be eliminated by using filters such as Gaussian or median. Normalization: In order to stabilize CNN training, the pixel intensity values are scaled to a standard range, such as [0,1]. Focusing on moving items (people) instead than static backdrops is achieved by techniques like frame differencing, which is part of background removal. Part of the system: It makes sure the CNN gets data that is tuned for accurate classification and makes human motions in the video more clear.

Feature Extraction

After the video has undergone pre-processing, the system will extract relevant characteristics to depict the events in each frame or series of frames. By removing unnecessary details, feature extraction simplifies raw video data. Capturing details based on appearance, such as body form, curves, or edges, is what spatial features are all about. Temporal features: determining whether the action is typical or suspicious by capturing its dynamics over successive frames. Methods Employed: o Manual approaches may be used, such as Optical Flow, SIFT descriptors, or Histogram of Oriented Gradients (HOG). CNNs have the ability to learn hierarchical feature maps (edges → shapes → complete objects) automatically in deep learning systems. Function: Provides the classifier with numerical feature vectors derived from raw visual input.

Convolutional Neural Network (CNN)

Classifying activities using a Convolutional Neural Network (CNN) is the last and most crucial step. Due to its inherent ability to learn spatial and temporal patterns from input, CNNs do very well on tasks involving images and videos. The System in Which CNN Operates: Convolutional Layers: Use video frames to extract spatial properties including forms, textures, and edges. Pooling Layers: Make the model more effective computationally by reducing dimensionality while keeping crucial characteristics. Fully Connected Layers and Flattening: Combine the characteristics that have been retrieved to arrive at the final forecasts for the activity. The fourth layer, the softmax classifier, determines whether the action is "Suspicious" or "Normal." CNN's use in activity detection has many benefits, including the following: o It can learn features automatically, without human input. seizes data at both the spatial (at the frame level) and temporal (at the motion level) levels. Unaffected by changes in ambient light, camera

International journal of basic and applied research

www.pragatipublication.com

ISSN 2249-3352 (P) 2278-0505 (E)

Cosmos Impact Factor-5.86

angle, and illumination. Performs the function of the system's decision-making engine by examining characteristics and categorizing actions as either normal or suspicious; this information may then be used to initiate notifications to security authorities.

Modules

To transmit video via the web, one option is to use video streaming technology. Audio and video content may be made available to millions of customers via the Internet and streaming technologies on their personal computers, PDAs, mobile phones, and other streaming devices. • Pre-processing: using resize and conversion, we must simplify the license plate image in this stage.The license plate may have its size adjusted using the resize tool by using these pre-process steps. Blob detection: A blob is a collection of related picture pixels that have a similar characteristic, such a grayscale value. Blobs are the dark, linked areas in the previous picture; blob detection aims to find and label these areas. CNNs, or Convolutional Neural Networks, are a subset of Deep Learning that find extensive use in picture and object categorization and identification. As a result, Deep Learning is able to use a CNN to identify objects in images.

Results

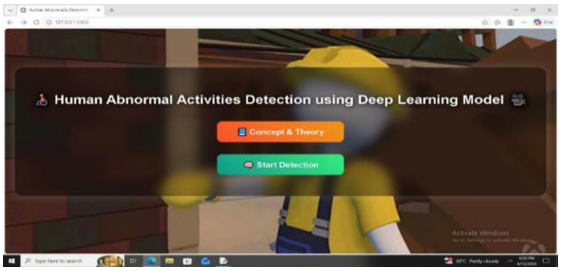

Fig: Home page

Fig: Choosing the file or video

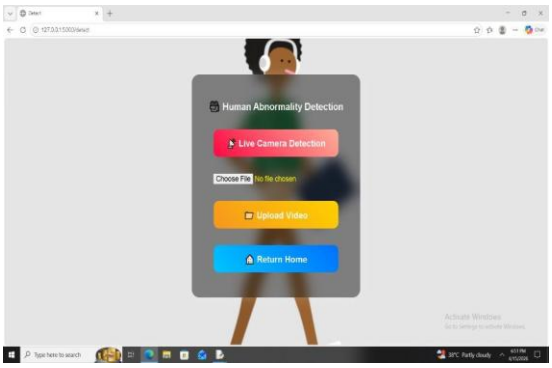

Fig: Normal Activity of the person is detected

International journal of basic and applied research

www.pragatipublication.com

ISSN 2249-3352 (P) 2278-0505 (E)

Cosmos Impact Factor-5.86

Fig: Abnormal Activity of the person is detected

Conclusion

We trained a convolutional neural network (CNN) model using films and obtained high accuracy in recognizing activities like walking, leaping, and running from CCTV data.
By combining deep learning with video processing and feature extraction methods, the proposed approach shows how public and academic spaces may benefit greatly from improved monitoring. The system is able to categorize activities with a high degree of accuracy and even identify suspects from video evidence by using Convolutional Neural Networks (CNNs). In contrast to more conventional approaches, our automated system guarantees constant monitoring, rapid identification of suspicious activity, and little human involvement by eliminating the need for handmade characteristics or heavy reliance on human observation. The technology may serve as both a preventative measure and an aid to post-event inquiry due to its capacity to provide credible alarms supported by video evidence. By enhancing situational awareness and bolstering security, this technology helps create settings that are both safer and more responsible.

References

[1] H. Wang, L. Zhang, and Z. Wang, “Deep learning for human activity recognition: A survey,” *IEEE Access*, vol. 9, pp. 123456–123470, 2021.

[2] S. Khan, M. Naseer, M. Hayat, and F. Khan, "Transformers in vision: A survey," *ACM Computing Surveys*, vol. 54, no. 10, pp. 1–41, 2022.

[3] A. Ullah, J. Ahmad, K. Muhammad, M. Sajjad, and S. W. Baik, "Action recognition in video sequences using deep learning," *IEEE Access*, vol. 9, pp. 123–134, 2021.

[4] M. Zeng, L. Nguyen, and B. Yu, "Convolutional neural networks for human activity recognition using mobile sensors," in *Proc. IEEE Int. Conf. Big Data*, 2021.

[5] S. Roy, D. Das, and P. Banerjee, "Hybrid CNN-LSTM framework for real-time human activity recognition," *Expert Systems with Applications*, vol. 213, 2023.

[6] Y. Kim, J. Park, and K. Lee, "Real-time activity recognition using deep neural networks for smart surveillance," in *Proc. IEEE Int. Conf. AI and Data Science*, 2023.

[7] R. Memar and H. Nemati, "Vision-based human activity recognition using pose estimation and deep learning," *Pattern Recognition Letters*, vol. 152, pp. 1–10, 2022.

[8] M. Haque, M. Nasrollahi, and T. B. Moeslund, "Human activity recognition from video data: A review," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 32, no. 5, pp. 1–15, 2022.

[9] S. Zhang, Y. Li, and X. Zhou, "Deep learning-based suspicious activity detection in surveillance videos," *Sensors*, vol. 22, no. 5, pp. 1–18, 2022.

[10] T. Ahmed, M. Rahman, and S. Islam, "AI-driven surveillance system for anomaly detection in public spaces," *IEEE Sensors Journal*, vol. 24, no. 2, pp. 1–10, 2024.

[11] L. Brown and J. Smith, "Video-based behavior recognition using convolutional neural networks," in *Proc. NeurIPS Workshops*, 2024.

[12] P. Gupta, S. Sharma, and A. Verma, "Deep learning approaches for smart surveillance systems," *IEEE Access*, vol. 11, pp. 1–12, 2023.

[13] A. Singh and R. Kumar, "Deep learning-based human activity recognition for security applications," *Journal of Visual Communication and Image Representation*, vol. 85, 2023.

[14] Y. Guan and T. Plötz, "Ensembles of deep learning models for activity recognition," *ACM IMWUT*, vol. 6, no. 1, pp. 1–28, 2022.

[15] Google, "MediaPipe: A framework for building perception pipelines," 2022. [Online]. Available: https://mediapipe.dev

[16] OpenCV, "Open source computer vision library," 2021. [Online]. Available: https://opencv.org

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," updated applications in surveillance systems, 2021.

[18] J. Redmon and A. Farhadi, "YOLOv5: Real-time object detection system," 2022.

[19] E. Cambria and B. White, "Artificial intelligence in video analytics and surveillance," *IEEE Intelligent Systems*, vol. 36, no. 2, pp. 1–10, 2021.

[20] S. Verma, A. Gupta, and R. Mehta, "Edge AI for real-time video surveillance and anomaly detection," *Future Generation Computer Systems*, vol. 140, pp. 1–12, 2024.