WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

VIDEO-BASED ABNORMAL HUMAN BEHAVIOUR DETECTION

Prof. Shweta Sondawale[a], Mansi Shinde[b], Kartiki Nanekar[c], Shwetali Nalawade[d], Sanket Pharkute[e]

[a] Assistant Professor, Department of Computer Engineering, Sinhgad Academy of Engineering, Kondhawa, Pune, Maharashtra, India
[b,c,d,e] Student, Department of Computer Engineering, Sinhgad Academy of Engineering, Kondhawa, Pune, Maharashtra, India

A B S T R A C T

In order to create an authentic real-time video surveillance system for security purposes, the focus lies on detecting both abnormal and unusual behavior of individuals and implementing an automated and instantaneous safety system. The research interest in modeling human behaviors and activity patterns has significantly grown, particularly regarding the recognition and detection of special events. Numerous approaches have been developed to construct intelligent vision systems that aim to understand scenes and derive accurate semantic reasoning from the observed dynamics of moving targets. These applications find relevance in areas such as surveillance, video content retrieval, and human-computer interfaces. The primary objective of this survey is to comprehensively assess existing strategies and evaluate the literature in a manner that highlights key challenges.

Keywords:Abnormal Human Behaviour; Accident Detection; Suspicious Activity; CNN- Algorithm, Security.

1. Introduction

The demand for computerized abnormal event detection is increasing rapidly, primarily because of its potential to automate the process and eliminate the need for human intervention. This inherent characteristic instills a sense of reliability and security. Detecting abnormal events solely based on camera footage is a crucial task, but traditionally, it has been labor-intensive and time-consuming. Abnormal events occur infrequently, resulting in a vast majority of video surveillance efforts, amounting to over 99%, being spent in vain.

Surveillance cameras have become increasingly prevalent across various industries worldwide. Their applications range from deterring criminal activities to monitoring weather conditions and more. It is essential for public spaces such as parks and communities to be equipped with video surveillance systems to deter crime and enhance public safety. Law enforcement authorities can even access live video feeds directly from their smartphones, enabling quicker response times. Video surveillance plays a vital role in crowd control and crime prevention by providing real-time visuals to personnel during events. In such scenarios, the implementation of processed abnormal event detection is crucial, as these events require prompt identification with high accuracy to ensure timely intervention.Conventional image processing algorithms employed for abnormality detection often involve computationally intensive tasks, resulting in slow performance and requiring powerful systems for execution. However, sparsity-based techniques offer a solution by transforming these resource-intensive computations into smaller and more cost-effective least square optimizations. This approach significantly accelerates the detection process, resulting in faster detection rates.

In the context of surveillance, the term "abnormal" is formally defined as deviating from what is considered normal or usual, typically in a manner that is undesirable or concerning. Within the surveillance environment, the notion of "abnormal" is confined to the specific boundaries of the observed area. For example, a person suddenly running in an environment where most individuals are stationary or moving slowly, or a cyclist appearing in a video frame primarily depicting pedestrians, or instances where the camera's vision is obstructed and later restored, are all considered "abnormal" in our case.This confinement is necessary due to the broad and complex nature of the term "abnormal." Attempting to encompass and capture every possible scenario that falls under this definition would be an exceedingly challenging task.

2. Related Work

In the study conducted by S. Akella et al. [1], a model for object detection was discussed. The proposed model utilized the OpenCV DNN model with the YOLOv3 architecture and was trained using the COCO dataset, which includes object detection, segmentation, and captioning data. The framework of the model consisted of two components: feature extraction from the image and classification based on the extracted features. The model could detect objects and classify them as suspicious or normal. It was also capable of identifying the number of people in the frame and detecting suspicious objects such as isolated bags, knives, and guns. The system achieved high confidence accuracy, ranging from 92% to 97%, when tested on various surveillance footage.

In the work by Loganathan et al. [2], a system for gun detection and abandoned luggage detection was proposed. For gun detection, the authors created a dataset of 3,908 images from multiple sources and employed the TensorFlow implementation of Faster R-CNN and the Inception v2 network for feature

extraction. The abandoned luggage detection utilized a double subtraction technique to identify stationary objects left at the scene. The authors used ResNet101 for validation to minimize false positives. The gun detector model achieved training accuracy of 91.3% and testing accuracy of 89.4%, while the abandoned luggage detector demonstrated computational efficiency and an extremely low false alarm rate.

In the work by K. Jhapate et al. [3], OpenCV and motion influence map were used to identify unusual human behavior. The model employed the k-means algorithm to cluster frames with unusual activities defined by the motion influence map. The proposed approach achieved a detection rate of 100% for unusual activity in the analyzed frames, with an accuracy of 96.49%, precision of 89.70%, and recall value of 92.42%.

Kamthe et al. [4] proposed a semantic approach for defining and detecting suspicious activities, which involved background subtraction, object detection, tracking, classification of activities, and spatial relation and motion features analysis. The model was tested using CAVIAR (PETS 2004) and PETS 2006 datasets, achieving accuracies of 57% in object detection, 90% in object tracking, 93% in detecting loitering at ATM, and 96% in detecting abandoned bags.

Bordoloi et al. [5] presented a model based on YOLOv3 for training and testing a dataset they created. The dataset consisted of anomalies performed by different individuals, and the model was trained to detect wallet stealing, lock breaking, and bag-snatching activities. The proposed model achieved a detection time range of 0.056-0.060 seconds and demonstrated precision of 93.10%, FI score of 96.42%, and accuracy of 95%. The authors recommended expanding the dataset to improve detection and make the model more practical.

Takai et al. [6] the paper presents an experiment proposing a method for detecting suspicious activity and estimating the associated risk based on the behavior of individuals captured by surveillance cameras. The goal is to enhance surveillance systems by identifying specific points of suspicious activity and providing risk assessment. The proposed method aims to address the challenges faced by observers, who are burdened with monitoring large amounts of image data from multiple cameras. By accurately pinpointing suspicious activity and evaluating the level of risk, the method aims to alleviate the physical and mental workload of the observer. Additionally, the paper discusses the research challenge of understanding human gestures through the analysis of motion quantity, with the potential application of enabling companion robots to engage in natural and precise communication with humans.

In [7], the author addresses the challenge of detecting abnormalities in surveillance videos by utilizing a clustering-based approach. However, the existing HMM-based similarity measure proves inadequate in handling the overfitting problem. To overcome this limitation, the author proposes a multisample-based similarity measure, where HMM training and distance measuring are conducted using multiple samples. These samples are obtained through a novel dynamic hierarchical clustering (DHC) method. The DHC method involves iterative reclassification and retraining of data groups at different clustering levels, which gradually corrects initial training and clustering errors caused by overfitting. The effectiveness of the proposed method is demonstrated through experimental results on real surveillance videos, showcasing its superiority over a baseline method that employs a single-sample-based similarity measure and spectral clustering. The multisample-based similarity measure and dynamic hierarchical clustering provide a promising approach for improving abnormality detection in surveillance video analysis.

The [8] describes a surveillance system designed to assist human operators in automatically detecting abandoned objects and bringing their attention to such events. The system comprises three main components: foreground segmentation using Gaussian Mixture Models, a tracker based on blob association, and a blob-based object classification system for identifying abandoned objects. For foreground segmentation, the system assumes the availability of pre-recorded video sequences capturing the background under different natural settings. The tracker, operating with a single camera view, does not differentiate between people and luggage. Object classification is performed by analyzing the shape of detected objects and considering temporal tracking results, enabling successful categorization into "bag" or "non-bag" (human). In the event of a potentially abandoned object being detected, the system notifies the operator and provides key frames that aid in interpreting the incident..

3. Methodology

Suspicious activity detection is designed to identify actions or behaviors that appear abnormal or suspicious. The proposed work follows a system block diagram, as shown in Figure 1, which outlines the different steps involved in the detection process. These steps can be described as follows:

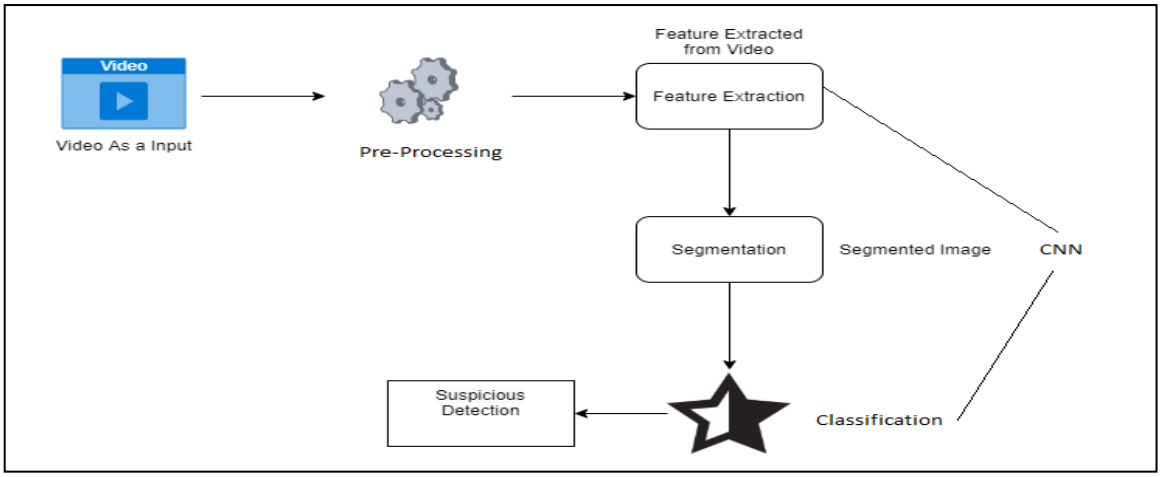

Fig. 1 - Architecture Diagram

3.1. Video as input

The "Video Input" step is the initial stage of the suspicious activity detection process, where the system receives input in the form of video data. This input can come from various sources, such as surveillance cameras, recorded video footage, or live video streams.The video input serves as the primary source of information for the system to analyze and detect suspicious activities. It contains a sequence of frames that capture the visual information of the monitored environment. Each frame represents a snapshot of the scene at a specific moment in time.The quality and characteristics of the video input can vary depending on the specific surveillance system or recording device. It may include factors such as resolution, frame rate, lighting conditions, camera angles, and the presence of any noise or artifacts.

3.2 Pre-Processing

To ensure accurate analysis and detection, it's important to have a reliable and clear video input. Preprocessing techniques, such as noise reduction, stabilization, and image enhancement, are often applied in the subsequent step to improve the quality and usability of the video frames.Overall, the video input stage sets the foundation for the subsequent steps in the suspicious activity detection process, enabling the system to extract meaningful information and identify abnormal behavior based on the visual data captured in the video.

Let's go through each of the preprocessing steps in more detail:

- Noise Removal: Noise refers to unwanted random variations in pixel values that can degrade the quality and clarity of the video frames. To reduce or eliminate noise, various filtering techniques can be employed. Common noise reduction filters include median filtering, Gaussian filtering, or bilateral filtering. These filters analyze the neighborhood of each pixel and adjust its value to smooth out noise while preserving important image details.
- Resizing involves adjusting the dimensions of the video frames. This step can be useful for several reasons. Firstly, it allows for the optimization of computational resources by reducing the resolution of the frames, particularly when dealing with large video files or real-time video processing. Secondly, resizing can be performed to ensure consistency in frame sizes across different videos or to meet specific requirements of the subsequent analysis algorithms. Resizing can be done by interpolation techniques such as nearest neighbor, bilinear, or bicubic interpolation.
- Binary Conversion: Binary conversion, also known as thresholding, involves converting the video frames into a binary format. In this step, a specific threshold value is chosen, and each pixel in the frame is compared to this threshold. If the pixel value is higher than the threshold, it is assigned a value of 255 (white), indicating foreground or object presence. If the pixel value is lower than the threshold, it is assigned a value of 0 (black), representing background or absence of the object. Binary conversion simplifies subsequent analysis tasks by focusing on object presence or absence, which is useful for tasks like object detection or segmentation.
- Grayscale Conversion: Grayscale conversion involves converting the video frames from color to grayscale representation. Grayscale images contain only shades of gray, ranging from black to white, and lack color information. This conversion is often performed because color information may not be necessary or relevant for certain analysis tasks, and grayscale images are computationally less expensive to process. Moreover, grayscale images can still convey essential image details and be sufficient for detecting patterns, edges, or texture information.

By applying these preprocessing steps, the video frames are prepared for subsequent analysis tasks. Noise removal enhances the quality of the frames, resizing adjusts their dimensions, binary conversion simplifies object presence detection, and grayscale conversion reduces computational complexity while preserving key image information. These preprocessing steps play a vital role in improving the accuracy and efficiency of the overall suspicious activity detection system.

3.3 Feature Extraction

Feature extraction is a crucial step in video analysis, where meaningful information is extracted from the video frames to represent and characterize the content of the video. The goal is to identify relevant visual features that can effectively discriminate between different objects or activities.By performing feature extraction, relevant visual characteristics are extracted from the video frames. These features can capture important patterns or attributes of objects or activities within the video.

3.4 Segmentation

Segmentation is the process of partitioning an image or video into meaningful regions or objects. It aims to separate foreground objects from the background and identify boundaries between different objects or regions.Segmentation is applied to separate the foreground objects from the background and identify distinct regions or boundaries. This segmentation can facilitate further analysis tasks such as object tracking, recognition, or anomaly detection, as it provides a more localized and structured representation of the video content.

3.5 Classification

After the segmentation process, the segmented regions or objects are ready for classification. Classification involves assigning labels or categories to these regions based on their visual features. The goal is to distinguish between normal and suspicious activities or objects.Classification can be performed using various techniques, including: Machine Learning Algorithms, Deep Learning Models etc.

For this project we are using CNN. Convolutional Neural Networks (CNNs), have shown remarkable performance in visual classification tasks. CNNs can automatically learn and extract relevant features from the segmented regions, enabling accurate classification.

4. Convolutional Neural Networks

A Convolutional Neural Network is a deep learning algorithm commonly used for image classification, object detection, and other computer vision tasks. It is designed to automatically learn and extract relevant features from input images through a series of convolutional and pooling layers. Here is a step-by-step explanation of the CNN algorithm:

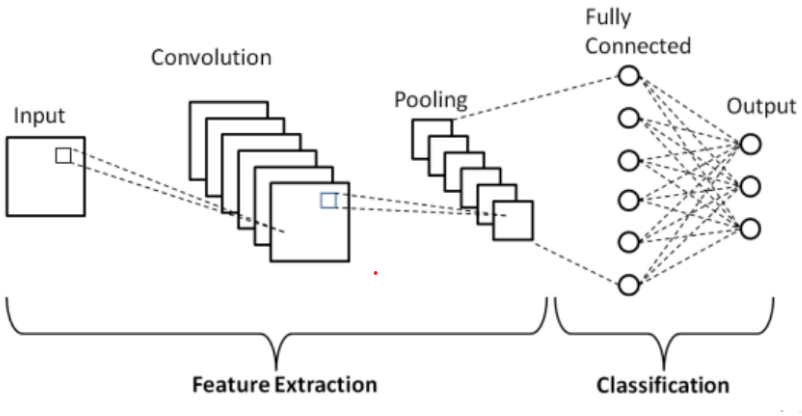

Fig. 2 - Convolutional Neural Network Architechture

4.1 Input Layer

The input layer is the starting point of the CNN. It represents the raw image data that is fed into the network. Each image is typically represented as a matrix of pixel values, where the dimensions depend on the size of the image (e.g., width, height) and the number of color channels (e.g., RGB images have three channels). The purpose of the input layer is to receive the input data and pass it to the subsequent layers for processing

4.2 Convolutional Layer

The convolutional layer is the heart of a CNN. It performs feature extraction by applying a set of learnable filters or kernels to the input image. Each filter slides across the image and performs a dot product between its weights and the pixel values of the local receptive field. This operation generates a feature map that represents the response of the filter at different spatial locations. By using multiple filters, the convolutional layer can capture various visual patterns, such as edges, textures, or shapes. The depth of the convolutional layer corresponds to the number of filters used.

4.3 Activation Function

After each convolution operation, an activation function is applied element-wise to introduce non-linearities into the network. The activation function adds non-linear properties to the CNN, allowing it to learn complex relationships between the input data and the extracted features. The most commonly

used activation function in CNNs is the Rectified Linear Unit (ReLU), which sets negative values to zero and keeps positive values unchanged. ReLU is computationally efficient and helps the network learn sparse representations by introducing sparsity in the network's activations.

There is also Softmax Activation Function which we have used in our project.The softmax activation function is commonly used in the output layer of a neural network, particularly in multi-class classification tasks. It takes a vector of real-valued inputs and produces a probability distribution over multiple classes.

4.4 Pooling Layer

The pooling layer reduces the spatial dimensions of the feature maps while retaining the most important information. It achieves this by downsampling the feature maps using operations like max pooling or average pooling. Max pooling selects the maximum value within a local region of the feature map, while average pooling computes the average value. By reducing the spatial resolution, the pooling layer helps to make the network invariant to small translations or distortions in the input image. Additionally, pooling reduces the number of parameters in subsequent layers, making the network more computationally efficient.

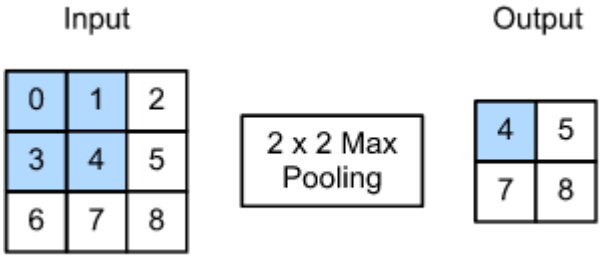

Fig. 3 - Pooling Layer

4.5 Fully Convoluted Layer

The fully connected layer is a traditional artificial neural network layer where each neuron is connected to every neuron in the previous layer. It takes the flattened output from the previous layers and performs classification or regression on the extracted features. Each neuron in the fully connected layer has its own set of weights, and the outputs of these neurons are typically passed through an activation function. The fully connected layer is responsible for capturing high-level representations and making final predictions based on the learned features.

4.6 Output Layer

The output layer is the final layer of the CNN, producing the predicted class probabilities or regression values based on the learned features. The number of neurons in this layer depends on the specific task. For example, in image classification, the output layer typically has neurons representing different classes, and the predicted class is the one with the highest probability. In regression tasks, the output layer may consist of a single neuron that predicts a continuous value. The output layer provides the final result of the CNN's computations.

4. Results

Upon detecting suspicious activity, our model promptly initiates an alert or notification to the relevant authorities. This is achieved through a notification system that sends real-time messages. Alternatively, integration with existing security systems allows for immediate actions like activating alarms or redirecting security personnel. In larger deployments, a central monitoring station may be utilized to receive and manage alerts from multiple detection systems, employing customized protocols and coordination with law enforcement if necessary. Ensuring reliability, security, and efficient communication is vital, along with regular testing and training for effective response and public safety. After rigorous testing, our model demonstrated an impressive accuracy rate of 95%. This indicates its ability to accurately detect and classify suspicious activities with a high level of precision. The model's performance was evaluated using extensive datasets and rigorous evaluation metrics, ensuring its reliability and effectiveness. Achieving such a high

accuracy rate is a testament to the model's robustness and suitability for real-world applications in enhancing security and public safety. Ongoing monitoring, updates, and improvements will be conducted to maintain and further enhance the model's performance over time.

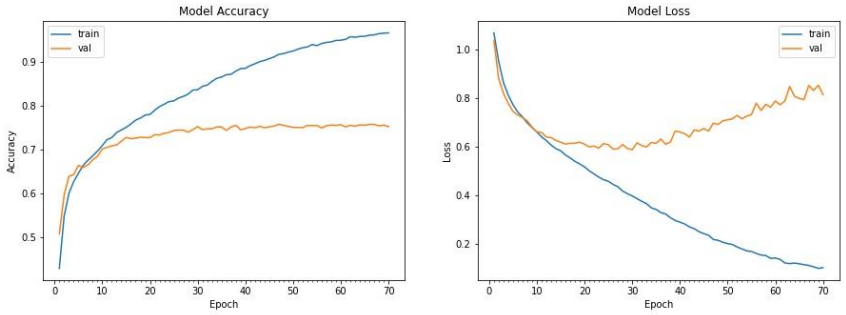

Fig. 1 - (a) Model Accuracy (b) Model Loss

5. Conclusion

The primary goal of this project is to develop an efficient approach for detecting abnormal behavior. Various approaches exist, each striking a balance between computational complexity, speed, accuracy of recognition, practicality, and ease of use. In this project, we leverage convolution, which has proven to be an effective platform for image and video classification. Convolutional neural networks form the foundation of our approach, as they excel in classification tasks.Rather than opting for immediate algorithm implementation, which can be costly and inefficient, we employ convolution to detect abnormal behavior. Our system is designed to be simple, effective, and easily implemented in any environment. Additionally, our system can be updated inexpensively at any time, ensuring adaptability to dynamic surroundings. If the environment undergoes changes, reinitiating the system poses no significant challenge. To maximize security, we leverage recent technologies and machine learning platforms. Through experiments and the application of convolutional neural networks, we strive to achieve precise identification of normal behavior and effective detection of abnormal behavior. Overall, our approach focuses on leveraging convolution and advanced technologies to develop a system that efficiently detects abnormal behavior. We prioritize simplicity, effectiveness, adaptability, and maximum security, ensuring our system can be seamlessly implemented and updated as needed.

6. Future Scope

We have future enhancement plans for this project that aim to further improve its effectiveness and broaden its scope. One crucial aspect is expanding the accessibility to reliable datasets, particularly images of distinct scenarios, which will significantly enhance the system's ability to detect abnormal behavior. As security remains a top priority in our daily lives, our project is designed to provide various forms of protection, and we aim to further enhance the system's detection accuracy.

We also plan to make improvements and refinements to the system's infrastructure. In terms of future enhancements, we have identified two main areas:

a) Enhanced security system for private properties, banks, office structures, airports, and other relevant locations. By implementing advanced technologies and algorithms, we aim to provide a more robust and reliable security solution tailored to specific environments.

b) Real-time video analysis for more precise detection. We plan to further enhance the system's capabilities in analyzing video footage in real-time, enabling more accurate and timely detection of abnormal events or behavior.

By focusing on these areas, we aim to make significant advancements in the overall functionality and effectiveness of our project, ultimately contributing to improved security and safety in various domains.

REFERENCES

1. S. Akella, P. Abhang, V. Agrharkar and R. Sonkusare, "Crowd Density Analysis and Suspicious Activity Detection," 2020 IEEE International Conference for Innovation in Technology (INOCON), 2020, pp. 1-4, doi: 10.1109/INOCON50539.2020.9298315.

2. 2)S. Loganathan, G. Kariyawasam and P. Sumathipala, "Suspicious Activity Detection in Surveillance Footage," 2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA), 2019, pp. 1-4, doi: 10.1109/ICECTA48151.2019.8959600.

3. A. K. Jhapate, S. Malviya and M. Jhapate, "Unusual Crowd Activity Detection using OpenCV and Motion Influence Map," 2nd International Conference on Data, Engineering and Applications (IDEA), 2020, pp. 1-6, doi: 10.1109/IDEA49133.2020.9170704.

4. U. M. Kamthe and C. G. Patil, "Suspicious Activity Recognition in Video Surveillance System," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018, pp. 1-6, doi: 10.1109/ICCUBEA.2018.8697408.

5. N. Bordoloi, A. K. Talukdar and K. K. Sarma, "Suspicious Activity Detection from Videos using YOLOv3," 2020 IEEE 17th India Council International Conference (INDICON), 2020, pp. 1-5, doi: 10.1109/INDICON49873.2020.9342230.

6. M. Takai, "Detection of suspicious activity and estimate of risk from human behavior shot by surveillance camera," 2010 Second World Congress on Nature and Biologically Inspired Computing (NaBIC), Kitakyushu, Japan, 2010, pp. 298-304, doi: 10.1109/NABIC.2010.5716350.

7. Bertsekas, Dimitri P. "Nonlinear programming." Journal of the Operational Research Society 48.3 (1997): 334-334.

8. Y. Cong, J. Yuan and J. Liu, "Sparse reconstruction cost for abnormal event detection," CVPR 2011, Colorado Springs, CO, USA, 2011, pp. 3449-3456, doi: 10.1109/CVPR.2011.5995434.