IJCRT.ORG ISSN : 2320-2882
INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

ABNORMAL EVENT DETECTION IN HUMAN BEHAVIOUR

[1] V. Prema, [2] P. Sai Mounish, [3] S. Roach Amalan, [4] K. Ranjith

[1] Assistant Professor, [234] UG Scholar

[1234] Department of Computer Science and Engineering,

SRM Valliammai Engineering College, Kattankulathur, Chengalpattu District, India.

Abstract: Smart CCTV systems rely on abnormal event detection, human behavior detection, and object recognition to enhance surveillance capabilities. These systems can identify abnormal events, detect unusual human behavior, and monitor the state of the environment. Machine vision and machine learning techniques are commonly used to detect and identify anomalies in CCTV video feeds. Typically, these systems process video frames individually using supervised learning to train the algorithms. However, since anomalies can be diverse and difficult to predict, unsupervised and semi-supervised learning methods are being adopted to train the system.By using these techniques, the workload of manual detection and alert creation for abnormal events in CCTV feeds can be minimized or eliminated. Additionally, the storage efficiency of the system is improved by storing abnormal events in their original quality and storing normal scenarios in lower quality for archiving purposes.

***Keywords* – Human abnormal behavior, Suspicious behavior, Violence**

I. INTRODUCTION

Nowadays, CCTV cameras are ubiquitous and can be found in virtually every corner of our surroundings. While their primary purpose is post-event analysis, where the recordings are used to investigate incidents that have already occurred, there is a growing demand for more proactive systems. To address this need, technologies such as machine vision and advanced machine learning algorithms are being integrated to create new systems that can detect anomalies and alert authorities in real-time. One promising application of these technologies is crowd behavior analysis and object detection, which can be used to detect theft in crowded environments. However, analyzing crowd behavior is a challenging task because people may be located at different positions and move in different directions. As a result, identifying effective features of the crowd and performing higher-level analysis of their behavior can be a daunting task.

1.1 Human actions and behavior

Human behavior refers to the ability of individuals or groups to respond to various internal and external stimuli throughout their lifetime, encompassing mental, physical, and social capacities. It is influenced by both genetic and environmental factors, which shape an individual's response to different situations. Additionally, thoughts and feelings play a significant role in driving behavior, providing insight into a person's attitudes and values. Furthermore, personality traits also impact human behavior, as different individuals exhibit unique personality types that result in diverse actions and behaviors. Overall, human behavior is a complex and multifaceted phenomenon that reflects a combination of genetic, environmental, psychological, and social factors.

1.2 Human abnormal behavior

Abnormal actions can be classified into two categories: global and local. Local abnormal actions involve the actions of one or two individuals that differ from the behavior of others in the area (such as loitering, falling, fighting, or showing signs of illness). Global abnormal actions involve the behavior of groups of people, such as during violent incidents or when trying to escape during a natural disaster. Abnormal behavior recognition is comprised of two modules: action representation and abnormal action detection. Detecting abnormal behavior in human surveillance is a crucial challenge for ensuring public safety due to the sheer volume of video data. Occasionally, abnormal activities occur, making it difficult to identify specific frames of abnormal action. This often results in the need for additional manpower to verify video streams, which can be expensive, inefficient, and time-consuming. Fig. 1.2.1 and 1.2.2 shown the various normal and abnormal human behaviors in a different context.

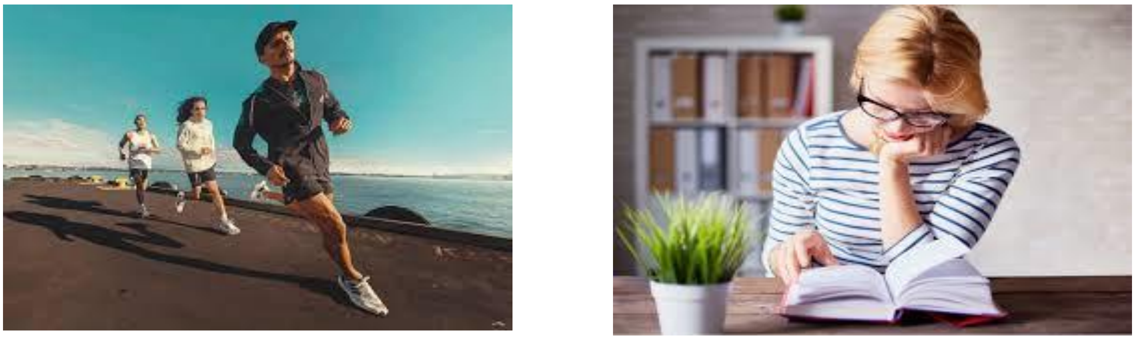

Fig. 1.2.1 Normal Behavior

Fig. 1.2.2 Abnormal Behavior

II. STUDY ON RELATED WORKS

Detecting abnormal events often involves first learning normal patterns and then identifying deviations from those patterns as anomalies due to the limited availability of abnormal training data. Local features of videos are typically extracted and used to train a normality model. Trajectory analysis is a common approach, but it may not be practical in crowded environments. Other methods involve using spatiotemporal video volumes with dense sampling or interest point selection to identify spatial anomalies in low-level visual features like histograms of oriented gradients, oriented flows, and optical flow. The effectiveness of these methods depends on the quality and completeness of the training data. Deep learning techniques, such as using a 3D ConvNet for classification or an end-to-end convolutional autoencoder for anomaly detection, have also been used. However, while multiple frames can be used as input, 2D convolutions only operate spatially, compressing temporal information, making LSTM models better suited for learning temporal patterns and forecasting time series data. Convolutional LSTMs have been proposed for learning the regular temporal patterns found in videos. The proposed approach in this work involves a two-stream architecture that uses crowd density heat-maps and optical flow to identify anomalous events. A network with convolutional LSTM layers processes inputs from each modality and describes their spatiotemporal patterns. The network has been trained to recognize panic and fight situations.

III. METHODOLOGY

3.1 Semi-supervised Learning

Semi-supervised learning is a type of machine learning that falls between supervised and unsupervised learning. This approach involves training a model using both labeled and unlabeled data. The goal of semi-supervised learning is similar to that of supervised learning - to train a function that can accurately predict the output variable based on the input variables. However, semi-supervised learning differs from supervised learning in that it utilizes a dataset that includes both labeled and unlabeled data. This approach is particularly useful when a large amount of unlabeled data is available, but labeling all of it would be expensive or difficult. Semi-supervised learning can be used to make use of this unlabeled data to improve the accuracy of the model while minimizing the need for labeled data.

3.2 Unsupervised Learning

Unsupervised learning is a machine learning technique used to analyze and group unlabeled datasets. This method utilizes algorithms to identify hidden patterns or data clusters without human assistance. Unsupervised learning is an ideal option for tasks such as exploratory data analysis, client segmentation, cross-selling strategies, and image identification, as it can identify similarities and differences within the data. Unsupervised learning models are used for three main tasks: clustering, association, and dimensionality reduction. This approach offers an exploratory means of viewing data, enabling businesses to identify patterns in large datasets more quickly than with manual observation.Overall, unsupervised learning provides a powerful tool for identifying patterns and structures in data, without requiring the input of labeled data or human supervision.

3.3 Convolutional Neural Networks

Convolutional neural networks (CNNs) are deep learning algorithms that are highly effective at processing and recognizing images. These networks consist of several layers, including convolutional layers, pooling layers, and fully connected layers. The central component of a CNN is its convolutional layers, which utilize filters to extract features such as edges, textures, and shapes from the input image. The output of the convolutional layers is then fed into pooling layers, which down-sample the feature maps and preserve the most important data while reducing the spatial dimensions. Finally, one or more fully connected layers are applied to the output of the pooling layers to predict or classify the image. By using a large dataset of labeled images, CNNs can be trained to identify patterns and features associated with specific objects or classes. These networks can be used to classify new images and can also be employed for other tasks such as object detection or image segmentation by extracting features. Overall, CNNs are a powerful tool for processing and analyzing images, and their ability to learn and extract relevant features from data makes them a valuable tool in a wide range of applications.

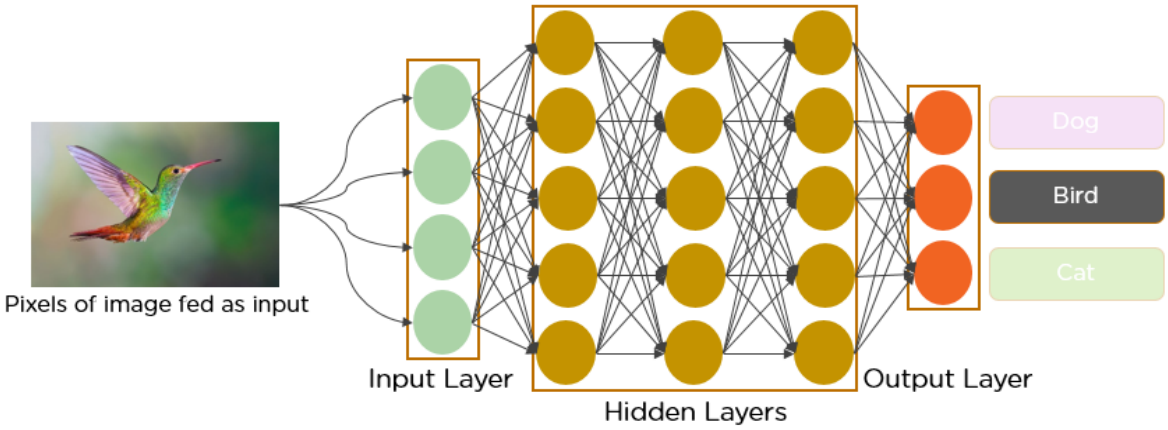

Fig 3.3.1 Convolutional Neural Networks

3.4 LSTM Architecture

In a traditional neural network, the final outputs are typically not used as inputs for the following step. However, in many real-world scenarios, the output is not only influenced by external inputs but also by the previous output. For instance, when people read a book, comprehending each sentence not only depends on the current list of words but also on the understanding of the previous sentence or the context established by prior sentences. Humans do not start thinking from scratch every second. However, traditional neural networks lack the concept of "context" or "persistence," which limits their ability to use context-based reasoning. To address this limitation, recurrent neural networks (RNNs) have been introduced, which feature feedback loops to enable the persistence of information.

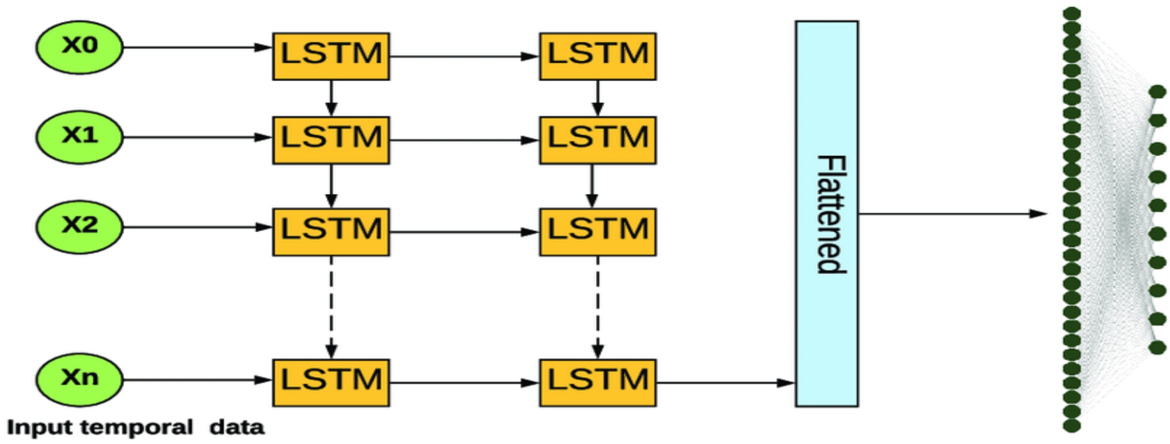

Fig 3.4.1 Architecture of the LSTM model for human activity recognition

IV. PROPOSED SYSTEM

Our system automates the process of identifying abnormal events in CCTV camera feeds through the use of CNN and LSTM technologies, which can detect anomalies in both supervised and unsupervised manners. Alert messages are sent to relevant authorities once an event is detected, and original quality video snippets of the abnormal event are stored in high and low qualities, respectively, while events considered normal are stored separately. By implementing this system, we reduce the time taken and human workload for detecting anomalies, as well as increase storage efficiency. The system comprises various modules, including the Video Compressor, Anomaly Detector, Storage Management, and Alert Management modules, which need to be integrated to create a complete system. The Video Compressor is the first module and is responsible for compressing the original video to a low quality and resolution for archiving. The Anomaly Detector module detects any kind of anomalies in the live feed from the CCTV system and is trained by unsupervised learning to detect all types of anomalies, both predetermined and undetermined. The video snippet where the anomaly is detected is stored in the original video quality and resolution. The Storage Management module manages the storage of both the low-quality archived video and original quality anomalous video snippets, which are stored separately for future reference. The Alert Management module is used to manage and send alerts to relevant personnel once an anomaly is identified.

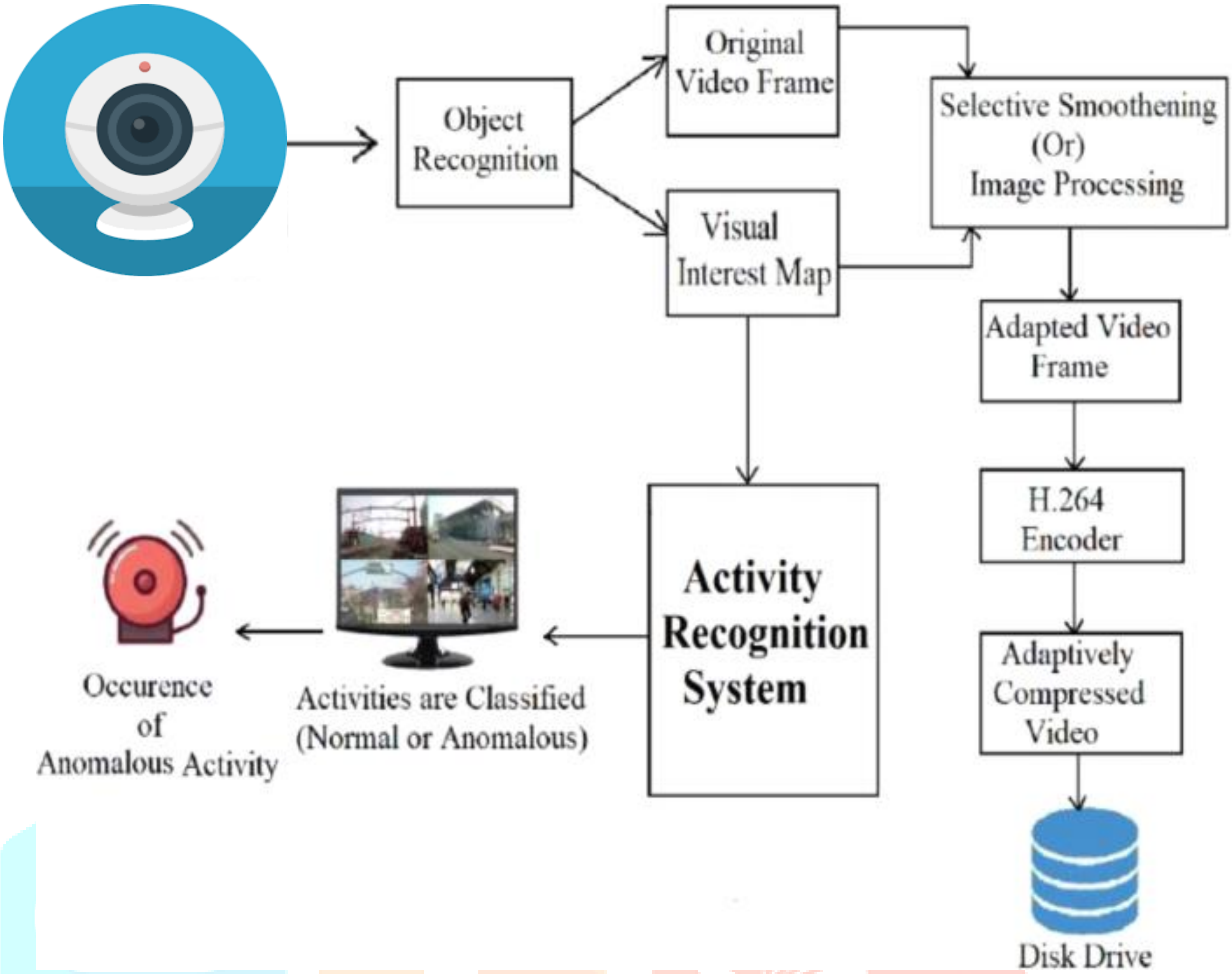

Fig 4.1 Block diagram for proposed system

V. IMPLEMENTATION

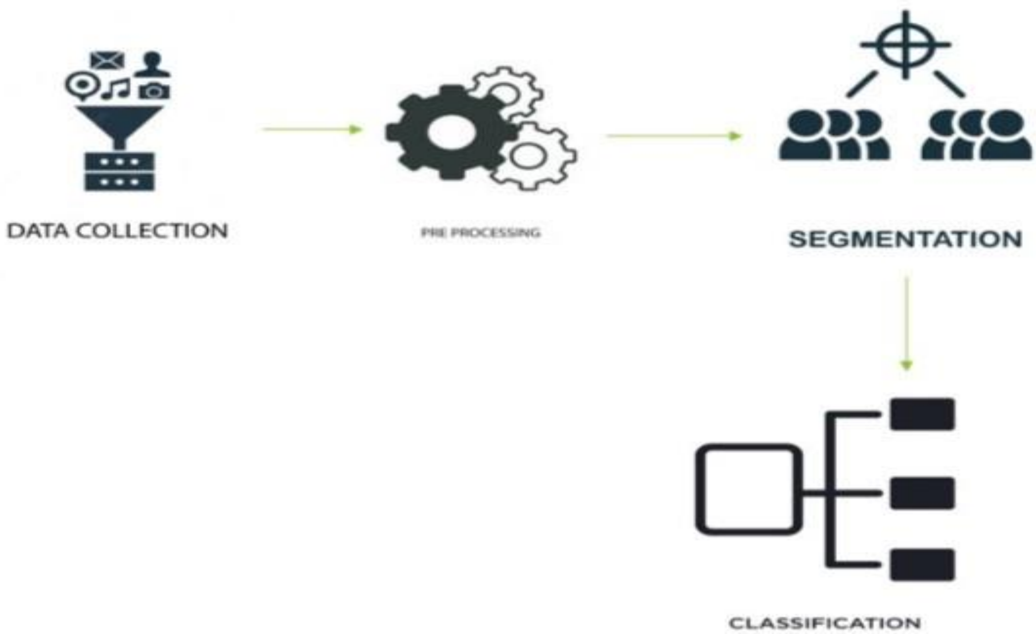

Fig 5.1 Modules used in this system

5.1 Data Collection

For our anomaly detection system, we have chosen to use two datasets - the UCSD anomaly detection dataset and the Avenue dataset. The UCSD dataset is composed of videos that were captured by a camera placed at a height overlooking a pedestrian walkway. The videos primarily feature pedestrians, with abnormal events being non-pedestrian entities such as bikers, skaters, and small carts, as well as unusual pedestrian motion patterns like walking across the walkway or on the grass surrounding it. The UCSD dataset is divided into two parts, Ped1 and Ped2. We will use both Ped1, Ped2, and the Avenue dataset to train and test our system.

5.2 Pre-Processing

To train our model, we are using a set of normal video frames that are arranged in sequences. The model is designed to learn how to reconstruct these sequences. However, we are only using every 5th alternate frame from the video sequence initially. This is being done to reduce both the processing time and the memory usage of the system.

5.3 Segmentation

To optimize the training process, we utilize Adam Optimizer with a learning rate of 0.0001. If the training loss stops decreasing, the learning rate is decreased using a decay of 0.00001. Additionally, we set the epsilon value to 0.000001 to prevent division by zero. For initialization, we use the Xavier algorithm, which helps ensure that the signal neither becomes too small nor too large as it passes through each layer.

5.4 Classification

To test the system, each video is evaluated individually. The UCSD Ped1 dataset provides 36 testing videos with 200 frames each, but since we are only selecting the 5th alternate frame, we end up with a total of 40 frames from each video. Similarly, in UCSD Ped2, there are 12 testing videos with varying numbers of frames, while in Avenue Dataset, we have 21 testing videos of varying duration. This selection of alternate frames helps reduce processing time and memory usage. Although selecting all frames would likely produce better results, it is not practical due to the significant amount of time required for computation. To obtain all sequences of 4 consecutive frames (after selecting 5th alternate frames), we use the sliding window technique. For each t between 0 and 36 in UCSD Ped1 dataset, we calculate the regularity score, Sr(t), for the sequence that begins at frame(t) and ends at frame(t+3). If the reconstruction error value is greater than the threshold, the system sends an alert signal. Additionally, the system stores the frames of original quality and resolution from a predefined number of frames before the occurrence of the anomaly to a predetermined number of frames after it in a video format for future reference. We consider it acceptable to store 10 to 20 frames before the occurrence of the anomaly and 100 to 120 frames after it to obtain a clear understanding of the anomaly and how it occurred.

VI. RESULT

The initial step of the system requires a manual input of a video. This input video is then processed to extract multiple images that contain feature points indicating human actions. By analyzing these features, the system can identify the type of abnormal event that is occurring.

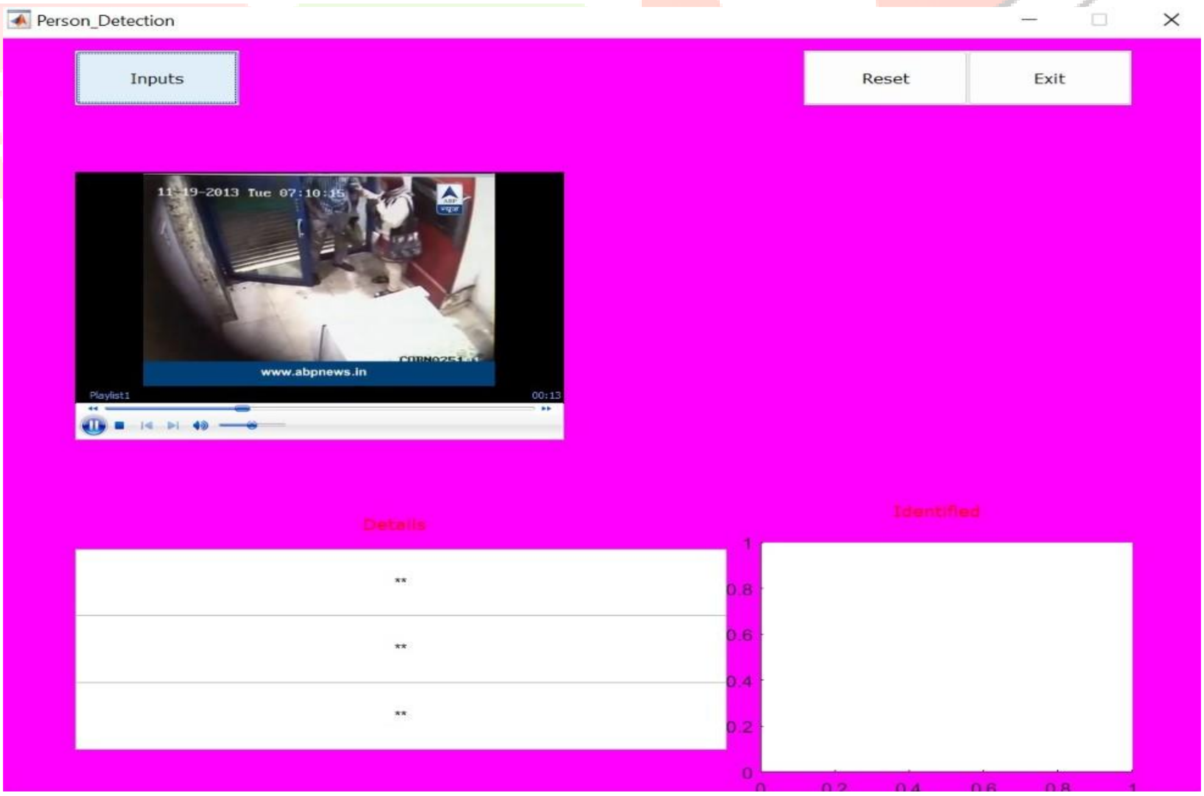

Fig 6.1 UI for providing input

To extract the identified feature points, the input video is processed and 20 frames are selected. These frames contain the specific features that were identified in the video, which can then be used to determine the type of abnormal event occurring.

Fig 6.2 Covertion of video into frames

Once the type of anomaly is identified, the system retrieves corresponding information from a MongoDB database that has been preloaded. This information is then processed and presented to the user as an output to discover details of the people involved in the event.

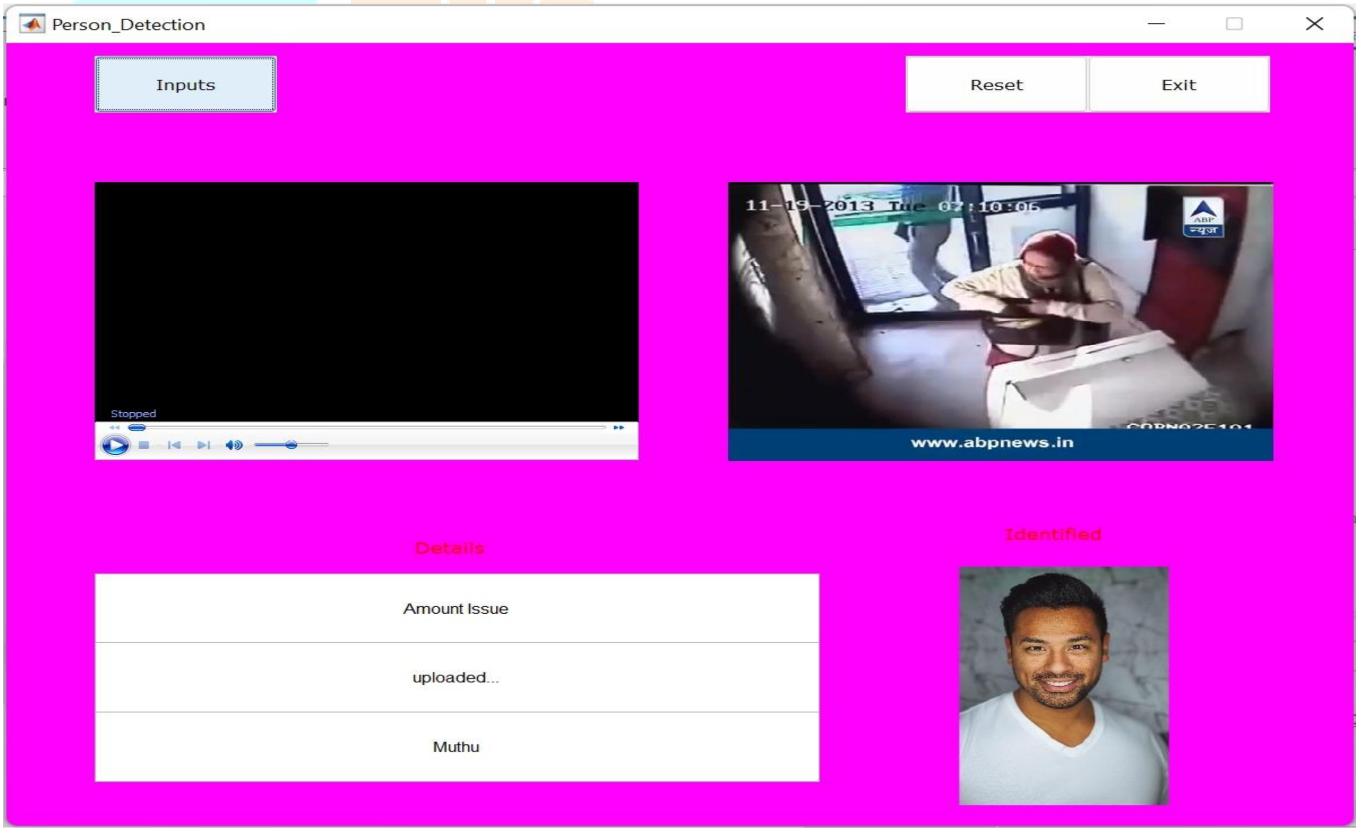

Fig 6.3 Final Output

VII. CONCLUSION

Detecting abnormal events is a vital part of a smart CCTV system as it enables automatic detection of anomalies and triggers relevant notifications. To detect various types of anomalies, current systems typically use supervised learning models and require significant computational resources. However, since anomalies can fall into different categories, it is not feasible to train the system to recognize them all. In such cases, unsupervised learning is used to effectively train the system. By implementing this system, storage efficiency is also achieved, as only abnormal frames are saved in high quality while recordings are saved in lower quality. In the future, combining supervised and unsupervised learning techniques could improve the system's performance. Additionally, integrating methods for identifying different types of anomalies and object detection could enhance the system's functionality.

VIII. REFERENCES

[1] C. Lu, J. Shi, and J. Jia, “Abnormal event detection at 150 fps in Matlab,” in Computer Vision (ICCV), 2013 IEEE International Conference on. IEEE, 2013, pp. 2720–2727.
[2] Wang, T., Chen, J., & Snoussi, H. Online detection of abnormal events in video streams. Journal of Electrical and Computer Engineering, 2013.
[3] Lee, S. C., & Nevatia, R, Hierarchical abnormal event detection by real time and semi-real time multi-tasking video surveillance system. Machine Vision and Applications, 25(1), 133-143, 2014.
[4] Gu, X., Cui, J., & Zhu, Q, Abnormal crowd behaviour detection by using the particle entropy. Optik-International Journal for Light and Electron Optics, 125(14), 3428-3433, 2014.
[5] Varol, G., & Salah, A. A, Efficient large-scale action recognition in videos using extreme learning machines. Expert Systems with Applications, 42(21), 8274-8282, 2015.
[6] Yan, M., Meng, J., Zhou, C., Tu, Z., Tan, Y., Yuan, J. Detecting Spatiotemporal Irregularities in Videos via a 3d Convolutional Autoencoder. Journal of Visual Communication and Image Representation, 2020, 67, 47-59. https://doi.org/10.1016/j.jvcir.2019.102747
[7] Zhu, X., Liu, J., Wang, J., Li, C., Lu, H. Sparse Representation for Robust Abnormality Detection in Crowded Scenes. Pattern Recognition, 2014, 47(5), 1791-1799. https://doi.org/10.1016/j.patcog.2013.11.018
[8] Wang, T., Qiao, M., Zhu, A., Shan, G., Snoussi, H. Abnormalevent Detection via the Analysis of MultiFrame Optical Flow Information. Frontiers of Computer Science in China, 2020, 14(2), 304-313. https:// doi.org/10.1007/s11704-018-7407-3
[9] Liu, W., Luo, W., Lian, D., Gao, S. Future Frame Prediction for Anomaly Detection - A New Baseline. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018. https://doi.org/10.1109/ CVPR.2018.00684
[10] Hu, Z., Zhang, L., Li, S., Sun, D. Parallel Spatial-Temporal Convolutional Neural Networks for Anomaly Detection and Location in Crowded Scenes. Journal of Visual Communication and Image Representation, 2020, 67, 765-771. https://doi.org/10.1016/j.jvcir.2020.102765
[11] Adam, A., Rivlin, E., Shimshoni, I., Reinitz, D. Robust Real-Time Unusual Event Detection Using Multiple Fixed-Location Monitors. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(3), 555-560. https://doi.org/10.1109/TPAMI.2007.70825
[12] Afiq, A. A., Zakariya, M. A., Saad, M. N. M., Nurfarzana, A. A., Khir, M. H. M., Fadzil, A. F., Jale, A., Gunawan, W., Izuddin, Z. A. A., Faizari, M. A Review on Classifying Abnormal Behavior in Crowd Scene. Journal of Visual Communication and Image Representation, 2019, 58, 285-303. https://doi.org/10.1016/j.jvcir.2018.11.035
[13] Ben, M., A., Zagrouba, E. Abnormal Behavior Recognition for Intelligent Video Surveillance Systems: A Review. Expert Systems with Applications, 2018, 91(1), 480-491. https://doi.org/10.1016/j.eswa.2017.09.029
[14] Almazroey, A. A., Jarraya, S. K. Abnormal Events and Behavior Detection in Crowd Scenes Based on Deep Learning and Neighborhood Component Analysis Feature Selection. In: Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020), 2020, 258-267. https://doi. org/10.1007/978-3-030-44289-7_25
[15] Calderara, S., Heinemann, U., Prati, A., Cucchiara, R., Tishby, N. Detecting Anomalies in People’s Trajectories Using Spectral Graph Analysis. Computer Vision and Image Understanding, 2011, 115(8), 1099-1111. https://doi.org/10.1016/j.cviu.2011.03.003
[16] Chang, H., Wang, T., Li, A., Fang, H. Local Hyperspectral Anomaly Detection Method Based on Low-Rank and Sparse Matrix Decomposition. Journal of Applied Remote Sensing, 2019, 13(02), 1-8. https://doi. org/10.1117/1.JRS.13.026513
[17] Elbayoudi, A., Lotfi, A., Langensiepen, C. The Human Behavior Indicator: A Measure of Behavioural Evolution. Expert Systems WITH Applications, 2019, 118, 493-505. https://doi.org/10.1016/j.eswa.2018.10.022
[18] Ergen, T., Kozat, S. S. Unsupervised Anomaly Detection with LSTM Neural Networks. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(8), 3127-3141. https://doi.org/10.1109/TNNLS.2019.2935975
[19] Febin, I. P., Jayasree, K., Joy, P. T. Violence Detection in Videos for an Intelligent Surveillance System Using Mobsift and Movement Filtering Algorithm. Pattern Analysis and Applications, 2020, 23, 611-623. https:// doi.org/10.1007/s10044-019-00821
[20] Dwivedi, N., Singh, D. K., Kushwaha, D. S. Orientation Invariant Skeleton Feature (oisf): A New Feature for Human Activity Recognition. Multimedia Tools and Applications, 2020, 79, 1-36. https://doi.org/10.1007/ s11042-020-08902-w