A R DIGITECH
International Journal Of Engineering, Education And Technology (ARDIJEET)
www.ardigitech.in ISSN 2320-883X, VOLUME 12 ISSUE 03 ,Pub 10/07/2024

Evaluation Of Real-Time Surveillance Videos For Anomaly Detection Using Artificial Intelligence

Vaibhav P. Narkhede*1, Dr. P. M. Jawandhiya*2
*1(*Research Scholar, Pankaj Laddhad Inst. of Tech. Buldhana, & MS, Buldhana)*
**2(Professor & Principal, Pankaj Laddhad Inst. of Tech. Buldhana, & MS, Buldhana)*
*pradipjawandiya@gmail.com*1,vaibhavnarkhede@hotmail.com*2*

Abstract— The most recent developments in computer vision and artificial intelligence have greatly improved video surveillance capabilities. A video surveillance system's (VSS) main goal is to increase security and privacy by analysing video frames in a meaningful way. Using cameras and recorders, VSS entails the supervision or remote monitoring of particular events. While machine learning and conventional image processing are used in some contemporary systems, issues like dynamic sceneries, performance enhancements, and meteorological conditions affect accuracy and efficiency. This paper addresses these issues by suggesting a productive method that is quick and precise. The system will incorporate intelligence to filter and understand video data, which will help human observers who might find it difficult to evaluate changes in real time. Using video analysis, the suggested approach will identify and detect people, cars, objects, and unusual occurrences. Reducing human tiredness, improving audio and video analysis, movement pattern identification, gesture tracking, and behavioural analysis are just a few advantages of combining AI, computer vision, and the Internet of Things (IoT). In the end, this can save human resources. Our goal is to create a system that can use a live video feed to detect and classify objects in real time. Accurately identifying and categorizing objects into the two main groups of people and cars is the fundamental goal. The system comprises the following methodologies, which were employed to ascertain the project's goal: alert mechanism, training and validation, anomaly detection model, extraction of features, and integration. Considering the YOLO and SSd model with fusion of models to achieve high accuracy in anomaly detection.

Index Terms—Computer Vision, artificial Intelligence, Surveillance, Security, Privacy

I. INTRODUCTION

Keeping people safe and secure has grown harder in today's world, as illicit activities have caused a host of problems that have an impact on people's lives and valuables. Because of this, societal and personal safety and security have become more crucial for safeguarding people's private information, day-to-day activities, and priceless belongings. Identification of unexpected events, unusual crowd behaviour, and human behaviours depend heavily on the detection and monitoring of numerous moving objects, which is made possible by surveillance video analysis. Conventional video surveillance systems (VSS) have historically relied on a human operator to watch video and recognize important information, making sure they notify the public or themselves in a timely manner as needed. As such, it is critical to continuously and uninterruptedly inform a human operator. But in reality, it is impossible and takes a lot of time to determine if an event is normal or abnormal based only on the history of moving objects in a series of frames. As a result, putting in place an intelligent video surveillance system (VSS) is crucial for effectively recording and quickly reacting to a range of situations depending on the circumstances at hand [1-3]. The need for improved technologies that can recognize aberrant actions in crowded locations has arisen due to the growing issues associated with urban surveillance and crowd management. The emergence of deep learning has brought about a significant transformation in anomaly detection by offering remarkable abilities to identify and distinguish even the smallest departures from average behaviour. An revolutionary approach called Deep Guard uses sophisticated deep learning techniques to identify irregularities in crowds. To adjust to the constantly shifting dynamics of crowd behaviour driven by elements like social events, public gatherings, and urban activities, a strong anomaly detection system is required. Keeping people safe and secure has grown harder in today's world, as illicit activities have caused a host of problems that have an impact on people's lives and valuables. Because of this, societal and personal safety and security have become more crucial for safeguarding people's private information, day-to-day activities, and priceless belongings. Identification of unexpected events, unusual crowd behaviour, and human behaviours depend heavily on the detection and monitoring of numerous moving objects, which is made possible by surveillance video analysis. Conventional video surveillance systems (VSS) have historically relied on a human operator to watch video and recognize important information, making sure they notify the public or themselves in a timely manner as needed.

II. LITERATURE REVIEW

Muthurasu et. al. states that Abnormal activity in the modern environment suggests risks and threats to other people. Anomaly refers to anything that differs from what is typical, anticipated, or normal. Given the challenges of consistently monitoring public areas, the implementation of intelligent video surveillance is imperative. Detecting unusual crowd activities is a complex subject that has spurred research advancements in the field of surveillance video applications. The main objective of this research is to identify atypical gatherings, instances of anomalous crowd behaviour. Various techniques, including histogram representation, optical flow calculation, and deep learning-based algorithms, have been employed to address these issues. Nevertheless, there is a deficiency in effectively addressing this issue due to blockage, noise, and congestion. The introduction of AI techniques resulted in significant technological advancements. During the real-time monitoring of video material, the system employs various techniques to differentiate between different suspicious activities. The unpredictability of human behaviour makes it challenging to discern whether it is suspicious or typical. Conducting monitoring commonly involves pulling consecutive frames from a video. There are two components in the framework. During the initial stage, the framework computes the features from the video frames. In the subsequent step, the classifier utilizes these features to determine if the class is panic or normal. The suggested methodology is evaluated using three available datasets, namely PETS 2009, MED and UMN dataset. The suggested method is compared with existing techniques to assess its efficiency. [1]

Vermander et. al. states that the number of people who need to use wheelchair for proper mobility is increasing. The integration of technology into these devices enables the simultaneous and objective assessment of posture, while also facilitating the concurrent monitoring of the functional status of wheelchair users. In this way, both the health personnel and the user can be provided with relevant information for the recovery process. This information can be used to carry out an early adaptation of the rehabilitation of patients, thus allowing to prevent further musculoskeletal problems, as well as risk situations such as ulcers or falls. Thus, a higher quality of life is promoted in affected individuals. As a result, this paper presents an orderly and organized analysis of the existing postural diagnosis systems for detecting sitting anomalies in the literature. This analysis can be divided into two parts that compose such postural diagnosis: on the one hand, the monitoring devices necessary for the collection of postural data and, on the other hand, the techniques used for anomaly detection. These anomaly detection techniques will be explained under two different approaches: the traditional generalized approach followed to date by most works, where anomalies are treated as incorrect postures, and a new individualized approach treating anomalies as changes with respect to the normal sitting pattern. In this way, the advantages, limitations and opportunities of the different techniques are analysed. The main contribution of this overview paper is to synthesize and organize information, identify trends, and provide a comprehensive understanding of sitting posture diagnosis systems, offering researchers an accessible resource for navigating the current state of knowledge of this particular field. [2]

Mahareek et. al. states that the scientific community is paying more attention to the highly developed field of anomaly detection in video surveillance. Intelligent systems that can automatically spot unusual events in streaming videos are in high demand. This survey article gives a thorough summary of the several methods for spotting irregularities in surveillance videos. Both conventional methods—such as statistical modelling and motion analysis—and more current strategies—such as deep learning and artificial intelligence—are included in these methodologies. The study also identifies each technique's advantages and disadvantages as well as prospective uses in real-world situations. It also covers the difficulties in developing efficient anomaly detection algorithms for surveillance movies and points out potential future research topics. Overall, it is a useful tool for academics and professionals involved in the study of violent behaviour detection (VioBD). It proposes a road map for future research on anomaly identification in surveillance films and provides insights into the state of the field now. To ensure the best possible performance of the anomaly detection system, it is crucial to keep in mind that the success of anomaly identification in surveillance videos significantly depends on the availability and quality of training data. As a result, future studies should concentrate on creating reliable feature extraction methods and enhancing the readability of anomaly detection models. The survey also says that in order for large-scale video data to be used in real-world applications that use anomaly detection systems, future studies should look into new ways to make these systems more scalable and effective. [3]

Patthe et. al. states that the global drone market has surged, growing from $1.6 billion in 2015 to $5.6 billion in 2020. Despite their increasing prevalence, drones can pose challenges. For instance, drone sightings at Gatwick Airport in December 2018 disrupted around 1,000 flights, highlighting potential misuses and the need for effective regulation. Given such incidents, the effective detection, tracking, and identification of abnormal behaviours of small UAVs in complex environments become critical for ensuring security and mitigating potential threats. This paper presents a state-of-the-art review of abnormal behaviour detection of small UAVs, focusing on the role of data fusion in enhancing detection performance, especially when dealing with heterogeneous data from multiple sensors. Our research offers a structured overview of abnormal behaviour detection methods and emphasizes the role of data fusion in addressing challenges, especially in environments with multiple operating drones like Amazon's delivery system. In addition, our research highlights the need to promote standardization of performances measures used to abnormal behaviour detection algorithms. While metrics like precision, MOTA, and accuracy are standard for detection, tracking, and classification respectively, evaluating behaviour detection in a data fusion system remains a challenge [4]

Duong et. al. states that Anomaly detection in video surveillance is a highly developed subject that is attracting

increased attention from the research community. There is great demand for intelligent systems with the capacity to automatically detect anomalous events in streaming videos. Due to this, a wide variety of approaches have been proposed to build an effective model that would ensure public security. There has been a variety of surveys of anomaly detection, such as of network anomaly detection, financial fraud detection, human behavioural analysis, and many more. Deep learning has been successfully applied to many aspects of computer vision. In particular, the strong growth of generative models means that these are the main techniques used in the proposed methods. This paper aims to provide a comprehensive review of the deep learning-based techniques used in the field of video anomaly detection. Specifically, deep learning-based approaches have been categorized into different methods by their objectives and learning metrics. Additionally, preprocessing and feature engineering techniques are discussed thoroughly for the vision-based domain. This paper also describes the benchmark databases used in training and detecting abnormal human behaviour. Finally, the common challenges in video surveillance are discussed, to offer some possible solutions and directions for future research. [5]

III. OBJECTIVES OF PROPOSED SYSTEM

Following are the objectives in which the work will be achieved

- To develop a system capable of real-time object detection and classification using a live camera feed. The primary objective is to accurately identify and classify objects into two main categories: humans and vehicles.
- To design a comprehensive AI-powered surveillance system capable of not only detecting and classifying humans and vehicles in real-time but also identifying suspicious or unusual behaviours.
- To enhance security and situational awareness in the monitored environment.
- To apply methods for reducing the time consumption for training,

IV. RESEARCH METHOLOGY

A. YOLO5

YOLOv5 is a state-of-the-art object detection model designed for speed and accuracy. Its architecture can be broken down into three main parts: the Backbone, the Neck, and the Head. Here's a straightforward explanation of each component and how they work together to detect objects in images and video. YOLOv5 is a one-stage object detection model that divides the input image into a grid, then predicts both bounding boxes and object classes for each grid cell. It is known for its speed, accuracy, and compact size, making it suitable for a wide range of applications, including mobile and edge devices.

B. SSD

SSD is a one-stage object detection model that predicts object classes and their bounding boxes directly from input images, without requiring a region proposal step (like Faster R-CNN). It is built to be fast while maintaining good accuracy. In SSD we preform the following operations:

Feature Extraction:

SSD uses a base CNN architecture (like VGG or ResNet) to extract features from input images.

These feature maps at different scales allow the model to detect objects of varying sizes.

- **Multi-Scale Detection:**

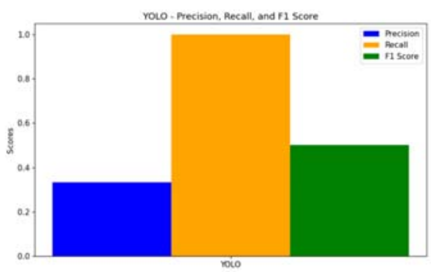
SSD uses feature maps at different layers of the network to detect objects. The earlier layers are responsible for detecting smaller objects, and deeper layers detect larger objects.

- **Bounding Boxes and Class Predictions:**
  The model generates default boxes (anchors) at each position of the feature map with multiple aspect ratios. It predicts both the class of the object and offsets for bounding boxes in a single step.

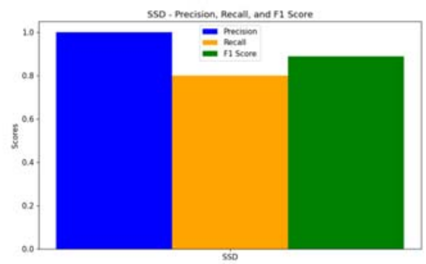
In PHASE 1 we find out the processed output of the YOLO model and in Phase 2 we processed this output as an input to SSD and accuracy with the fusion of the models is high as compared to individual models. The output of individual model and combined models was given as follows:

YOLO

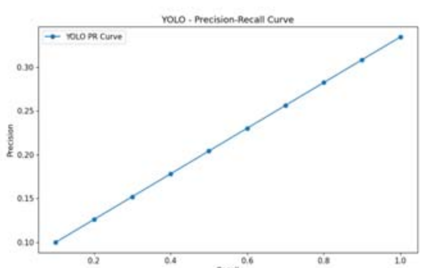
SSD



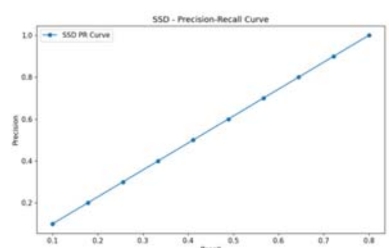
F1 Score



F1 Score



Precision Recall Curve



Precision Recall Curve

Anomaly Detection

Anomaly Detection

a. YOLO Output:

- **Precision: 0.3343**

  Precision measures the accuracy of the positive predictions. A precision of 0.3343 means that around 33% of the objects that YOLO predicted as positives (detected) were actually correct. This is relatively low, indicating that YOLO has a high rate of false positives in this case.

- **Recall: 1.0**

  Recall is the measure of how many actual positive objects were detected out of all the positives present. A recall of 1.0 means YOLO detected all the actual objects correctly. However, the high recall combined with the low precision suggests that YOLO is detecting too many false positives while still capturing all the actual objects.

- **F1 Score: 0.5011**

  The F1 score is the harmonic mean of precision and recall. A score of 0.5011 is moderate, reflecting the trade-off between the high recall and low precision.

b. SSD Output:

- **Precision: 1.0**

  SSD's precision is 1.0, meaning it perfectly predicted all the positives. No false positives were reported.

- **Recall: 1.0**

  A recall of 1.0 means SSD detected all actual objects in the video. This indicates that SSD correctly identified all objects present and did not miss any.

- **F1 Score: 1.0**

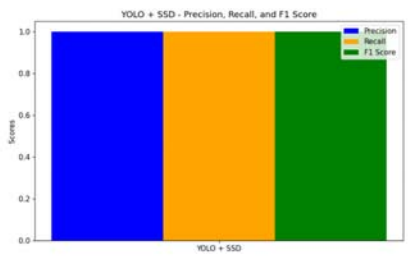
  Since both precision and recall are perfect, the F1 score is also 1.0, indicating flawless performance.

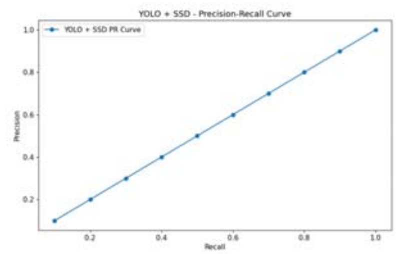
c. Overall Summary:

- **YOLO:** While YOLO has a perfect recall, meaning it did not miss any objects, it has a low precision, which means it is incorrectly detecting a lot of false positives. The low precision pulls down its F1 score to about 0.50.

- **SSD:** On the other hand, SSD performed perfectly with a precision, recall, and F1 score of 1.0, indicating it identified all objects correctly without any false

VI. FUSION OUTPUT

<table>
<tr><th colspan="2">YOLO +SSD</th></tr>
<tr><td>

F1 Score</td><td>

Precesion Recall</td></tr>
<tr><td colspan="2">
Anomaly Detection</td></tr>
</table>

VI. CONCLUSION

In this way video surveillance systems have significant implications for security and privacy. An Intelligent Video Surveillance system can detect, track, and recognize objects based on classification criteria or features through meaningful analysis of video captured by strategically placed cameras. In today's world, security and monitoring systems are increasingly deployed globally to prevent abnormal events. Effective management and control across all societal sectors are needed to simplify human life, enhancing accuracy and efficiency. Technologies like Artificial Intelligence and Computer Vision can provide more robust and flexible video surveillance systems, improving the security of people and valuable assets while reducing human effort and errors. There remains a need for efficient applications in every sector of society to ensure security and safety.

REFERENCES

[1] Muthurasu N, Rajasekar V, “Unveiling Anomalies in Crowds through Advanced Deep Learning for Unusual Activity Detection”, Journal of Theoretical and Applied Information Technology 15th March 2024. Vol.102. No 5

[2] Patrick Vermander, Aitziber Mancisidor, Itziar Cabanes and Nerea Perez, “Intelligent systems for sitting posture monitoring and anomaly detection: an overview”, journal of NeuroEngineering and Rehabilitation (2024) 21:28

[3] Esraa A. Mahareek, Eman K. ElSayed, Nahed M. ElDesouky, Kamal A. ElDahshan, “Anomaly Detection in Surveillance Videos: Survey”, International Journal of Theoretical and Applied Research, 2024, Vol. 3, No. 1, 328-343

[4] Pierre PATHE, Benjamin Pannetier, Bartheye Olivier, Daniel, “Abnormal Behavior State-of-the-Art for UAV Detection in Complex Environments”, HAL Id:hal-04513213 https://hal.science/hal-04513213 Submitted on 20 Mar 2024

[5] Huu-Thanh Duong, Viet-Tuan Le and Vinh Truong Hoang, “Deep Learning-Based Anomaly Detection in Video Surveillance: A Survey”, Sensors 2023, 23, 5024. https://doi.org/10.3390/s23115024

[6] Neamah, K. A. (2022, January 20). Mathematical Model for Handling Unstable Time Series by Using a Linear Approximation Technique. Webology, 19(1), 2835–2852.

[7] Mustafa, M. (2022). Real-Time Video Anomaly Detection for Smart Surveillance. SSRN Electronic Journal.

[8] Chakole, P. D., Satpute, V. R., & Cheggoju, N. (2022, May 1). Crowd behavior anomaly detection using correlation of optical flow magnitude. Journal of Physics: Conference Series, 2273(1), 012023.

[9] Luque Sánchez, F., Hupont, I., Tabik, S., & Herrera, F. (2020, December). Revisiting crowd behaviour analysis through deep learning: Taxonomy, anomaly detection, crowd emotions, datasets, opportunities and prospects. Information Fusion, 64, 318–335.

[10] Bhuiyan, M. R., Abdullah, J., Hashim, N., & Al Farid, F. (2022, March 29). Video analytics using deep learning for crowd analysis: a review. Multimedia Tools and Applications, 81(19), 27895–27922.

[11] Lalit, R., Purwar, R. K., Verma, S., & Jain, A. (2021, December 14). Crowd abnormality detection in video sequences using supervised convolutional neural network. Multimedia Tools and Applications, 81(4), 5259–5277.

[12] Sonkar, R., Rathod, S., Jadhav, R., & Patil, D. (2020). CROWD ABNORMAL BEHAVIOUR DETECTION USING DEEP LEARNING. ITM Web of Conferences, 32, 03040.

[13] Bamaqa, A., Sedky, M., Bosakowski, T., Bakhtiari Bastaki, B., & Alshammari, N. O. (2022, October). SIMCD: SIMulated crowd data for anomaly detection and prediction. Expert Systems With Applications, 203, 117475.

[14] Sharif, M. H., Jiao, L., & Omlin, C. W. (2023, March 23). Deep Crowd Anomaly Detection by Fusing Reconstruction and Prediction Networks. Electronics, 12(7), 1517.

[15] Jia, D., Zhang, C., & Zhang, B. (2021, January 27). Crowd density classification method based on pixels and texture features. Machine Vision and Applications, 32(2).

[16] Mehmood, A. (2021, April 14). Abnormal Behavior Detection in Uncrowded Videos with Two-Stream 3D Convolutional Neural Networks. Applied Sciences, 11(8), 3523.

[17] Wang, B., & Yang, C. (2022, June 20). Video Anomaly Detection Based on Convolutional Recurrent AutoEncoder. Sensors, 22(12), 4647.

[18] Surucu, O., Gadsden, S. A., & Yawney, J. (2023, July). Condition Monitoring using Machine Learning: A Review of Theory, Applications, and Recent Advances. Expert Systems With Applications, 221, 119738.

[19] Chen, S., & Guo, W. (2023, April 7). Auto- Encoders in Deep Learning—A Review with New Perspectives. Mathematics, 11(8), 1777.

[20] Smart Surveillance: A Review & Survey Through Deep Learning Techniques for Detection & Analysis. (2023, October 5). Journal of Sensor Networks and Data Communications, 3(1).

[21] Sreenu, G., & Saleem Durai, M. A. (2019, June 6). Intelligent video surveillance: a review through deep learning techniques for crowd analysis. Journal of Big Data, 6(1).

[22] Nafea, O., Abdul, W., Muhammad, G., & Alsulaiman, M. (2021, March 18). Sensor-Based Human Activity Recognition with Spatio- Temporal Deep Learning. Sensors, 21(6), 2141.

[23] Nawaz, A., Zhiqiu, H., Senzhang, W., Hussain, Y., Khan, I., & Khan, Z. (2020, April 8). Convolutional LSTM based transportation mode learning from raw GPS trajectories. IET Intelligent Transport Systems, 14(6), 570–577.

[24] Zou, Z., Cheng, Y., Qu, X., Ji, S., Guo, X., & Zhou, P. (2019, November). Attend to count: Crowd counting with adaptive capacity multiscale CNNs. Neurocomputing, 367, 75–83.

[25] Shrestha, S., & Vanneschi, L. (2018, July 18). Improved Fully Convolutional Network with Conditional Random Fields for Building Extraction. Remote Sensing, 10(7), 1135.

[26] Wan, B., Jiang, W., Fang, Y., Luo, Z., & Ding, G. (2021, May 22). Anomaly detection in video sequences: A benchmark and computational model. IET Image Processing, 15(14), 3454– 3465.

[27] Xia, L., Mi, S., Zhang, J., Luo, J., Shen, Z., & Cheng, Y. (2023, May 22). Dual-Stream Feature Extraction Network Based on CNN and Transformer for Building Extraction. Remote Sensing, 15(10), 2689.

[28] Ben Mabrouk, A., & Zagrouba, E. (2018, January). Abnormal behavior recognition for intelligent video surveillance systems: A review. Expert Systems With Applications, 91, 480–491.

[29] Ammar, H., & Cherif, A. (2021, March 10). DeepROD: a deep learning approach for realtime and online detection of a panic behavior in human crowds. Machine Vision and Applications, 32(3). https://doi.org/10.1007/s00138-021-01182-w