International Journal of Advanced Engineering Application
ISSN: 3048-6807, Volume No.1, Issue No1, May-2024

Suspicious activity detection for prevention of violence using YOLO and CNN algorithms

Atharva Pathak[1] , Shreyas Patil[2], Arnav Sinha[3], Parag Pandharpote[4], Lata Sankpal[5]
[1,2,3,4,5]Dept of Computer Engineering PVG's COET and GKPWIOM, Pune. India.
atharvapathak27@gmail.com [1] , patilsd2002@gmail.com [2], arnavsinha24@gmail.com [3],
9764504951s@gmail.com [4], latasankpal5@gmail.com [5].

Abstract—In response to the surge in violent incidents, this paper introduces an advanced surveillance system addressing the need for proactive threat detection. Leveraging You Only Look Once (YOLO) and Convolutional Neural Network (CNN) algorithms, our system achieves real-time identification of weapons, recognizes riots, detects suspicious bags, identifies their owners, and uncovers camera tampering. The outcome of this research contributes to the development of an intelligent surveillance framework that not only detects potential threats but also generates prompt alert notifications, enabling swift response measures to prevent or mitigate violent situations.

Index Terms—Object detection, knife/gun detection, suspicious luggage detection and its owner identification, riots detection, camera tampering detection, alert, email notification, Yolo object detection, Convolutional Neural Network.

I. INTRODUCTION

In recent years, there has been an alarming surge in violent activities, encompassing a spectrum from killings and riots to bomb blasts, necessitating an urgent need for more effective surveillance and crime prevention measures. Despite the widespread deployment of Closed-Circuit Television (CCTV) cameras, they rely on manual inspections. Additionally, a prevailing reluctance among the public to promptly report crime incidents to law enforcement is observed. Nations worldwide are struggling with the complexity of controlling violent riots, while the presence of unattended bags in public spaces, such as airports and malls, raises concerns about potential threats. It is increasingly documented that criminals strategically break or tamper with surveillance cameras to evade detection, emphasizing the imperative to develop a system capable of not only detecting and preventing violent incidents but also identifying instances of camera tampering. This research proposes the integration of advanced algorithms such as You Only Look Once (YOLO) for weapon and suspicious bag detection and Convolutional Neural Network (CNN) for riot identification and camera tampering. Moreover, the proposed system utilizes image-based detection, capturing snapshots at specific intervals in targeted locations to enhance the overall efficacy of surveillance, emphasizing the advantages of image-based processing for fast execution and low system requirements.

II.LITERATURE REVIEW

In recent years, the application of deep learning algorithms has gained prominence in the field of violence detection, with a particular emphasis on video processing techniques for identifying weapons and motion detection to detect abandoned bags [1] [2] [3]. While previous works have contributed significantly to enhancing security measures, the literature reveals a predominant focus on identifying camera tampering through techniques such as blockage, defocusing, and spray paint on lenses. Notably, existing methods lack a specific focus on

detecting the breaking of camera lenses using sharp objects, a crucial aspect in countering deliberate acts of sabotage against surveillance systems [4]. Moreover, for abandoned luggage detection, our proposed method analyzes bounding boxes generated after detecting both bags and individuals. This not only offers a robust and computationally efficient solution but also leverages previously captured images to identify the owner of an abandoned bag, enhancing security and potentially preventing false alarms.

Furthermore, while existing riot detection systems primarily identify general unrest, our research addresses the critical need to distinguish severe riots characterized by human harm, killings, and the presence of blood, distinguishing it from peaceful protests [5] [6] [7]. Additionally, our innovative approach emphasizes the capture of images over traditional video processing, optimizing system performance for limited hardware resources. This shift not only allows for efficient storage of relevant data but also streamlines the computational load. In terms of response mechanisms, our system goes beyond traditional alerts by storing images with timestamps, providing a comprehensive record of detected incidents. Furthermore, our system is equipped to generate immediate email notify- cations and activate a siren in response to the identification of potential threats, including the presence of weapons, abandoned bags, riots, and instances of camera tampering. This comprehensive approach positions our system as a valuable contribution to the evolving landscape of deep learning-based surveillance systems, addressing critical gaps in the current state of the art.

III. METHODOLOGY

At a higher-level project is divided into 4 modules: Weapon detection, abandoned luggage detection, riot detection, and camera tampering detection. Fig1 Demonstrates a high-level view of the project.

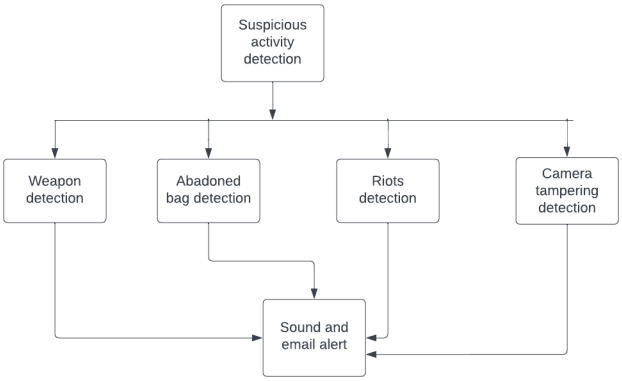

Fig. 1. High-level architecture.

A. Module 1: Weapon detection

In Module 1, our methodology centers on weapon detection, specifically focusing on identifying sharp weapons like knives and guns. Utilizing the YOLOv8 custom object detection algorithm, the model is trained on a dataset comprising 9918 images, each meticulously annotated to ensure precise identification of weapons. With an achieved accuracy of 81 percent, the model is adept at discerning potential threats within the surveillance area.

The operational sequence of this module involves the peri- odic capture of images from either a webcam or an attached camera at predefined intervals. These images are then fed as input to the trained YOLOv8 model, which conducts real- time analysis to detect the presence of weapons within the captured frames. Upon positive detection, the module triggers an immediate alert and generates an email notification to the designated authorities, ensuring a prompt and effective response.

B. *Module 2: Abandoned Luggage Detection*

Module 2 of our surveillance system is dedicated to aban- doned luggage detection, employing a nuanced

approach based on the analysis of bounding boxes. The YOLOv4 algorithm, trained on the COCO dataset, serves as the foundation for

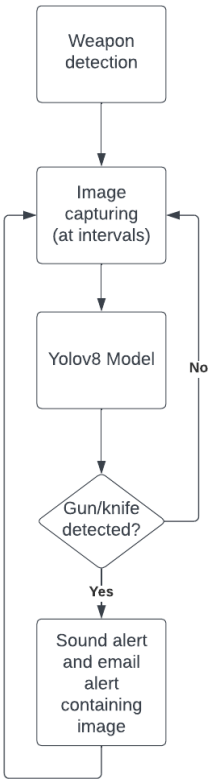

Fig. 2. Weapon detection module.

Fig. 3. Pistol Detected

Fig. 4. Knife Detected

this module, providing robust object detection capabilities. The model has been trained to distinguish between images containing only individuals, those featuring solely bags, and those presenting both a person and a bag.

Upon capturing an image, the module initiates a series of analyses. If the image exclusively contains a person and no bag, it is classified as safe. If the image exclusively contains a bag, a warning is issued, prompting the individual to collect the bag within a predetermined period. After the predetermined period is over, again another image is captured to verify if the bag is collected or not, failure to do so within the allocated time frame triggers an alarm for immediate attention.

In the case of an image containing both a person and a bag, the module examines whether the bounding boxes of the person and the bag overlap. Overlapping boxes indicate that the bag likely belongs to the person, and therefore, it is not considered abandoned. If the boxes do not overlap, a warning is issued, urging the individual to collect the bag within the predetermined time. If an abandoned bag is detected after this stage, our system traces back and finds the last/recent 5 images captured. These images, taken at intervals, are likely to contain the owner of the bag. The system then compiles these 5 images along with the current image and sends them as an email alert to the security personnel. Failure to retrieve the bag results in the generation of an email alert and a sound alarm, enhancing the system's responsiveness to potential security threats associated with abandoned luggage. This meticulous approach contributes to the overall effectiveness of our surveil- lance system in mitigating risks and ensuring public safety.

Fig. 5. Abandoned Luggage Fig.

6. Non Abandoned Luggage

A. *Module 3: Riots Detection*

Module 3 of our surveillance system is dedicated to the crucial task of riot detection, utilizing images from the UCLA dataset, which encompasses both riots and normal scenarios.

Fig. 7. Abandoned bag detection.

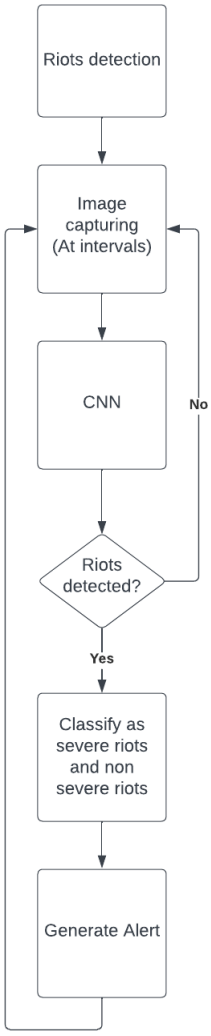

Fig. 8. Riots detection

The dataset is enriched by incorporating real-life riot images gathered from news resources.

For continuous monitoring, images are captured at specific intervals, forming the basis for the CNN (Convolutional Neural Network) algorithm employed in this module. Unlike con- venational riot detection systems, our approach is unique in its ability to differentiate between severe riots, characterized by incidents such as killings and the presence of blood, and peaceful protests. With a validated accuracy of 67 Percent, the model exhibits a reliable performance.

Fig. 9. Serve riots (contains blood)

Fig. 10. Riots (Do not contain blood)

A. Module 4: Camera Tampering Detection

Module 4 is designed for the detection of camera tam- pering attempts, addressing the unavailability of a specific dataset for breaking cameras using sharp objects. To overcome this limitation, a dataset was self-created by simulating real- time instances of breaking cameras with sharp objects and obstructing their views. In total, 1335 images were self- collected to train and evaluate the CNN (Convolutional Neural Network) algorithm. The CNN model, trained on this syn- thesized dataset, exhibits a robust accuracy of 94 Percent. This accuracy attests to the model's effectiveness in discerning instances of camera tampering, showcasing its reliability in real-world scenarios. The model is finely tuned to recognize patterns associated with obstructed views and damaged camera lenses, providing a high level of sensitivity to potential security threats.

Upon capturing images at predefined intervals, the module utilizes the trained CNN algorithm to analyze the frames for signs of camera tampering. In the event of a tampering attempt, the module triggers an immediate alert, notifying relevant author

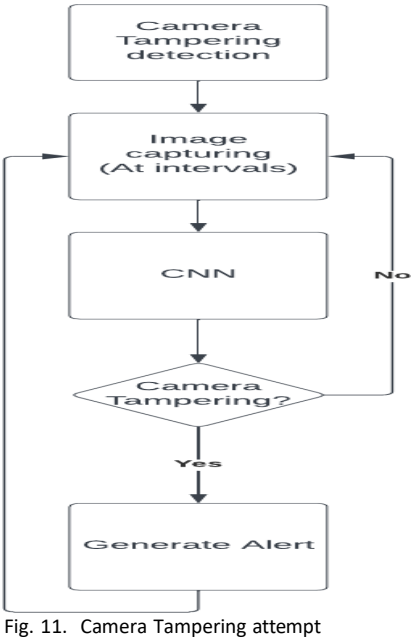

Fig. 11. Camera Tampering attempt

Fig. 12. Camera Tampering attempt

Fig. 13. Camera Blocking attempt

I. CNN TRAINING

CNNs are a class of deep neural networks adept at handling grid-like data such as images. Comprising convolutional lay- ers, activation functions, pooling layers, and fully connected layers, CNNs inherently capture hierarchical spatial features. In the context of riot and camera tampering detection, these layers play a pivotal role in recognizing patterns indicative of tumultuous events.

A. Convolutional Layers

CONVOLUTIONAL NEURAL NETWORK LAYERS

1) **Convolutional Layers:** The initial layers consist of convolutional operations, where filters are applied to the input image to extract spatial features. These filters detect patterns such as edges, textures, and shapes.
2) **Activation Functions:** Non-linear activation functions like Rectified Linear Unit (ReLU) are often applied after convolutional layers to introduce non-linearity into the model, enabling it to learn complex relationships within the data.
3) **Pooling Layers:** Pooling layers, typically max pooling, follow convolutional layers to down-sample the spatial dimensions of the feature maps, reducing computational complexity while retaining essential features.
4) **Flattening Layer:** After several convolutional and pool- ing layers, the architecture typically includes a flattening layer to convert the 2D feature maps into a 1D vector, preparing the data for fully connected layers.
5) **Fully Connected Layers:** Fully connected layers act as a decision-making component, where learned features from the previous layers are used to classify the input into different classes. The output layer usually employs a softmax activation function for multi-class classification, providing class probabilities.

II. YOLO OBJECT DETECTION

The YOLOv8 object detection model leverages a deep convolutional neural network for real-time object identification and localization.

1. *Backbone: Feature Extraction and Refinement*

1.1 Focus Layer: This initial layer efficiently scales the input image while reducing channels, optimizing further processing.

1.2 CSPDarknet53: This backbone, a modified Darknet ar- chitecture, utilizes 53 convolutional layers, employing" Cross Stage Partial Connections" for enhanced information flow.

1.3 SPPF (Spatial Pyramid Pooling): This layer extracts fea- tures at multiple scales through diverse kernel sizes, enabling accurate detection of both large and small objects.

2. *Neck: Bridging the Gap Between Backbone and Head*

2.1 C3 and C5 Upsampling: These layers increase the resolution of feature maps from the backbone, facilitating robust object detection at lower scales.

2.2 C2f Module: This unique module combines high-level C3 features with rich context from C5 using channel attention, significantly improving detection accuracy.

3. *Head: Generating Predictions and Refining Outputs*

3.1 P5, P4, and P3 Prediction Layers: These layers receive upsampled features from different levels and predict bounding boxes, confidence scores, and class probabilities for each grid cell within those features.

3.2 Detection Output: The final output combines predictions from all P layers, filtering out low-confidence detections and employing non-max suppression to ensure only one accurate bounding box per object.

4. System Accuracy

Model	Accuracy
Weapon detection	81%
Abandoned bag detection	65.7%
Riots Detection	67%
Camera Tampering detection	94%

Table I Table Of System Accuracy

References

[1] A. Datta, M. Shah and N. Da Vitoria Lobo, "Person-on-person violence detection in video data," 2002 International Conference on Pattern Recognition, Quebec City, QC, Canada, 2002, pp. 433-438 vol.1, doi: 10.1109/ICPR.2002.1044748.

[2] T. Santad, P. Silapasupphakornwong, W. Choensawat and K. Sookhanaphibarn, "Application of YOLO Deep Learning Model for Real Time Abandoned Baggage Detection," 2018 IEEE 7th Global Conference on Consumer Electronics (GCCE), Nara, Japan, 2018, pp. 157-158, doi: 10.1109/GCCE.2018.8574819.

[3] A. Warsi, M. Abdullah, M. N. Husen, M. Yahya, S. Khan and N. Jawaid, "Gun Detection System Using Yolov3," 2019 IEEE International Conference on Smart Instrumentation, Measurement and Application (ICSIMA), Kuala Lumpur, Malaysia, 2019, pp. 1-4, doi: 10.1109/ICSIMA47653.2019.9057329.

[4] E. Ribnick, S. Atev, O. Masoud, N. Papanikolopoulos and R. Voyles, "Real-Time Detection of Camera Tampering," 2006 IEEE International Conference on Video and Signal Based Surveillance, Sydney, NSW, Australia, 2006, pp. 10-10, doi: 10.1109/AVSS.2006.94.

[5] Sharath Kumar, Y.H., Naveena, C. (2023). A Deep Learning Based System to Estimate Crowd and Detect Violence in Videos. In: Biswas, A., Semwal, V.B., Singh, D. (eds) Artificial Intelligence for Societal Issues. Intelligent Systems Reference Library, vol 231. Springer, Cham.

[6] Won, Donghyeon, Zachary C. Steinert-Threlkeld, and Jungseock Joo. "Protest activity detection and perceived violence estimation from social media images." Proceedings of the 25th ACM international conference on Multimedia. 2017.

[7] M. A. Hanif, A. A. Butt and M. M. Khan, "Detecting riots using action localization," 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 2016, pp. 3031-3035, doi: 10.1109/ICIP.2016.7532916.