Suspicious Object Detection Using Images Using YoloV8 Model

Abstract

The rapid advancement of computer vision technologies has necessitated the development of efficient and accurate suspicious object detection models. In this paper, we propose a yolov8 model with integrated digital filter. The proposed model enhances the input image quality by distinguishing objects through effective noise suppression, adaptability, and interpretability. Following this, YOLOv8 is deployed for feature extraction and suspicious object identification. The model achieved a mean average precision (mAP) of 88% and an execution time of 5.3 seconds, thereby outperforming existing state-of-the-art methods.

Keywords— Object Detection; Suspicious; Images; Real-time Videos; IoT; Machine Learning.

Prati Dubey
Computer Science and Engineering
RNTU
Bhopal, India
pratidubey.245@gmail.com

Rakesh Kumar
Computer Science and Engineering
RNTU
Bhopal, India

I. Introduction

The field of detecting suspicious human behavior through video surveillance is rapidly advancing, particularly within the realms of image analysis and computer vision. The primary objective is to differentiate between routine and irregular activities carried out by people in public places. Activities like walking or waving hands are usually considered routine and non-threatening, while irregular behaviors, such as theft or potential attacks, pose risks to security [1]. The demand for video monitoring is on the rise, especially in high-risk locations like financial institutions, government properties, and transportation centers. Conventional monitoring methods often involve constant human oversight, which is not only expensive but also less efficient. Therefore, there's a growing need for smart, self-reliant monitoring systems capable of autonomously identifying, following, and categorizing unusual behaviors [2]-[4]. The end goal is to transition from passive to active monitoring systems that can autonomously issue warnings, either through alarms or notifications, upon detecting suspicious behaviors. These could include anything from identifying abandoned packages to monitoring health emergencies or detecting violent activities. With the increasing risk of attacks in public areas, there's an urgent focus on creating real-time smart systems that can quickly identify unattended baggage and notify the security staff [5]-[8]. To achieve this, these systems typically follow a series of operations such as identifying foreground objects, detecting specific objects, extracting features, classifying objects, and analyzing them. Numerous machine learning techniques, ranging from Support Vector Machines and Haar classifiers to Bayesian methods and K-Nearest Neighbors, are utilized for classifying objects and recognizing activities [9]. However, there are still obstacles to overcome. For instance, varying lighting conditions, object overlaps, background noise, low-quality resolution, and real-time processing are some of the challenges that persist. Additionally, existing machine learning models have limitations in accurately identifying multiple activities concurrently, which affects their overall effectiveness. Hence, while smart monitoring systems present a more proficient way to keep public and sensitive spaces secure, there are still areas that require further innovation.

II. Related Work

Object detection and suspicious activity recognition from images have been subjects of great interest in the research community [10]-[21]. As surveillance systems evolve to include more intelligence, the application of object detection methods to identify suspicious objects and activities becomes critically important. This paper aims to present an overview of key contributions and methodologies proposed in recent studies.

Tian et al [11] proposed a dual examination approach, designed to identify missing targets in suspect regions. This study aimed to improve single-stage identification by sharing dual choices that optimize feature-level multi-instance detection modules. Rani et al [12] introduced a novel item identification method using wireframe-based properties. They utilized cellular logical array processing for identifying pictures' aesthetic and geometrical properties. This research laid particular emphasis on deep neural network architectures and employed Fast R-CNN for object identification. Almahasneh et al [13] put forth a multi-task deep learning system which took advantage of picture band interdependence. The study modified an instructional method based on weak labels to overcome issues in obtaining dense AR annotations for controlled machine learning. Ge et al [14] presented a new architecture for aircraft identification in SAR images, termed the spatial orientation focus augmentation network. Based on YOLOX, the architecture aims to enhance performances by integrating various new features. Posilović et al [15] critically evaluated several deep-learning anomaly identification algorithms and discussed their pros and cons in depth. They reported an average ROC AUC efficiency of around 82%, providing an insight into the efficacy of existing methods. Hirooka et al [16] utilized transfer learning-based multi-channel attentiveness networks in convolutional neural frameworks. This ensembling approach aimed at retrieving more contextualized data for more accurate object identification. Yuan et al. [17] proposed an approach based on the Multi-Path Extraction Network (MPEN), aiming at efficient anomalous multi-object identification. This approach utilized YOLO v3 as its base network, emphasizing its versatility in object identification. Song et al [19] introduced a hierarchical design that employs geographic priors and multilevel key point characteristics for quickly locating similar regions and efficiently detecting targets. Javed et al [22] offered a novel real-time solution for object recognition in digital forensics. By utilizing deep learning algorithms, the method aims to provide high-level illustrations of photos containing suspicious objects. Yang et al [23] combined convolutional neural network (CNN) techniques with spatiotemporal data for achieving autonomous object identification. Their approach consists of two main parts: crude identification and thorough identification. Chen et al [24] proposed an identification algorithm that utilized two different CNNs. Their algorithm aims at the precise location of smaller size objects and also handles objects placed at random alignments. While many advancements have been made, challenges such as real-time processing, poor resolution, and handling of multiple activities simultaneously still exist. Moreover, ambiguity in recognition results remains a hurdle in achieving higher recognition accuracy.

III. Methodology Used

The proposed model for suspicious object detection aims to blend the strengths of two advanced techniques: Image Digital Filter (IDF) and YoloV8's Feature Pyramid Network as presented in fig 1. The first part of the model employs IDF, a specialized digital filter focuses on pre-processing the input image to enhance its quality. This is coupled with the digital filter's capabilities for noise suppression and adaptability to varying conditions, which means it can work well even in poorly lit or cluttered environments. Once IDF has improved the image, the model employs YoloV8 for feature extraction and object identification. YoloV8 belongs to the well-known YOLO (You Only Look Once) family, acclaimed for its quick and real-time capabilities in identifying objects. It employs a Feature Pyramid Network to recognize objects with varying sizes and orientations, making it particularly adept at detecting objects that might be partially hidden, at different distances, or clustered together. Our proposed model seeks to augment the capabilities of YoloV8 by incorporating techniques based on IDF for image refinement and noise reduction. This blended approach is designed to tackle a diverse array of real-world difficulties, including inconsistent lighting and a wide variety of object types and behaviors that may be deemed suspicious.

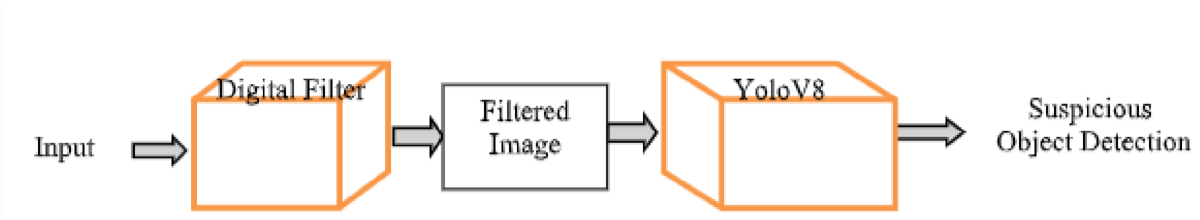

Figure 1: Proposed Model

In 2016, a team of researchers unveiled the YOLO (You Only Look Once) algorithm for identifying objects in visual media such as photos and videos. The approach utilizes a convolutional neural network (CNN) to simultaneously estimate the location and category of each object present in an image. It uses a grid system to segment the input image, and each grid cell is tasked with forecasting these attributes for the objects that fall within its area. A distinctive aspect of the YOLO framework is its incorporation of "anchor boxes," which enhances the precision of object detection. Trained on an extensive set of annotated images, the YOLO algorithm stands out for its speed and high accuracy. The YOLO (You Only Look Once) model for object detection consists of 24 convolutional layers and 2 fully connected layers, as presented in fig 2. To manage computational complexity, some of the convolutional layers use 1x1 reduction layers. The output of the last convolutional layer is a tensor of shape (7, 7, 1024), which is then flattened. Two fully connected layers produce linear regression parameters that are reshaped to (7,7,30), allowing for two bounding box predictions per grid cell. To compute the loss for a true positive, the model selects the bounding box with the highest Intersection over Union (IoU) value compared to the ground truth. This approach specializes the bounding boxes in their predictions, improving size and aspect ratio estimations over time. YOLO uses sum-squared error to measure the difference between its predictions and the actual values.

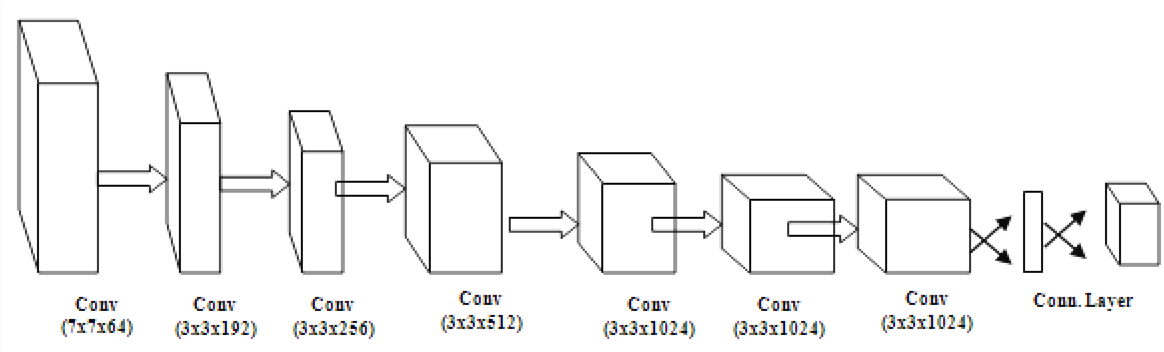

Figure 2: Architectural Diagram of YOLO

Here, classification loss l_c, localization loss l_l and confidence loss l_{co} combined called as loss function represented as:

$$Loss = l_c + l_l + l_{co} \tag{1}$$

$$l_c = \sum_{c \in classes} (P_c - A_c)^2$$

Where, P_c = predicted class and A_c is the actual class.

$$l_l = \lambda_{cord} \sum_{i=0}^{I} \sum_{j=0}^{J} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] + \lambda_{cord} \sum_{i=0}^{I} \sum_{j=0}^{J} [(w_i - \widehat{w}_i)^2 + (h_i - \hat{h}_i)^2] \tag{2}$$

Where, λ_{cord} is the loss of bounding box coordinates

$$l_{co} = \sum_{i=0}^{I} \sum_{j=0}^{J} [(C_i - \hat{C}_i)^2] \tag{3}$$

Where, C_i is the confidence score of box j in cell i

IV. Results and Discussion

The designed framework is implemented in Python using Google Colab. The backend for this implementation is TensorFlow. The total data set is split into two parts, with 70% dedicated to training and 30% dedicated to testing. Adam optimizer with a learning rate of 0.0001 is utilized for training. Training for all networks takes place on a Tesla P100-PCI-E GPU for a total of 100 iterations. The paper presented the result using following parameters:

Mean average Precision (mAP): To evaluate mAP, first precision need to be evaluated as:

$$Precision = \frac{(TP)}{(TP + FP)} \tag{4}$$

Then, mAP is mathematically represented as:

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{5}$$

Fig 3 presents the training graph of the proposed model. Table 1 provides a performance evaluation of a model on testing samples based on three key metrics: Loss, mAP (Mean Average Precision), and Execution Time. The table shows a loss value of 0.68839. A lower loss value is generally better, indicating that the model makes more accurate predictions. Mean Average Precision (mAP) serves as a common standard for assessing the effectiveness of object detection algorithms. With a score of 0.87641, the model demonstrates high proficiency in detecting objects. Scores can range from 0 to 1, with values approaching 1 signifying superior performance. The table also notes an execution time of 5.3 seconds, a vital statistic for applications that require swift decision-making. Reduced execution time enables the model to generate predictions more rapidly, making it highly applicable for real-time or near-instantaneous tasks. In summary, the model exhibits impressive accuracy (as highlighted by its mAP score) and minimal latency, making it well-suited for time-sensitive tasks.

Table 2 offers a side-by-side analysis of cutting-edge object detection algorithms, specifically juxtaposing YOLO-V3 [25] with the novel YOLO-V8 approach. Two pivotal metrics—Execution Time and Mean Average Precision (mAP)—are used for this comparison. YOLO-V3, a frequently employed object detection algorithm, takes 135.2 seconds for execution, whereas the innovative YOLO-V8 dramatically slashes this time to just 5.3 seconds. This implies that YOLO-V8 excels in speed and is likely more apt for real-time or nearly instant object detection assignments. Regarding accuracy, YOLO-V3 scores an mAP of 65.7%, a commendable but not optimal figure. On the other hand, the YOLO-V8 model boasts an mAP of 88%, denoting a substantial boost in detection precision. To sum up, the table strongly suggests that YOLO-V8 outperforms the established YOLO-V3 in both speed and accuracy, positioning it as a formidable contender in the field of state-of-the-art object detection.

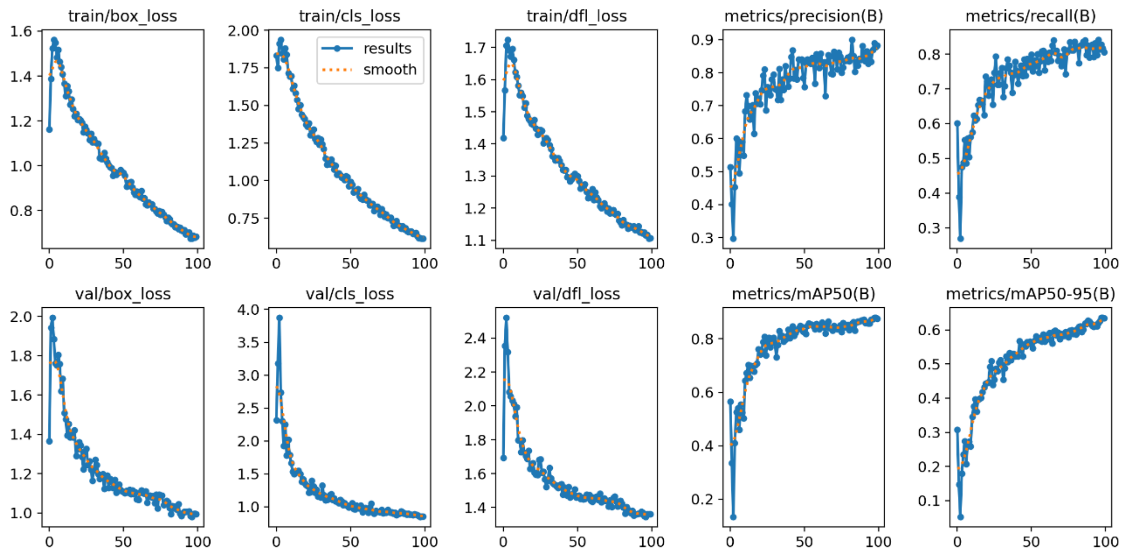

Figure 3: Training Performance

Table 1: Performance Evaluation on Testing Samples

Parameter	Value
Loss	0.688
mAP	0.88
Execution Time (in sec)	5.3

Table 2: Comparative State-of-Art

Ref	Methodology	Time	mAP
[25]	YOLO-V3	135.2	65.7%
Ours (Yolov8)		5.3	88%

V. Conclusion

In modern society, safety has become an increasingly pressing issue, especially in busy public venues like train terminals, airports, shopping centers, and densely populated zones. The ability to detect unattended objects is vital for enhancing the effectiveness of video monitoring systems. This article introduces an advanced model based on YOLOV8, representing a substantial advancement in the domain of identifying suspicious objects. By integrating the digital filter with YOLOv8, the model achieves a harmonious balance between image quality enhancement and high-speed, accurate object detection. The results are promising, with a substantial reduction in execution time to 5.3 seconds and an improved mean average precision (mAP) of 88%, far exceeding the performance metrics of the previous state-of-the-art model, YOLO-V3. This makes proposed model a strong candidate for real-time object detection tasks and opens avenues for future research in optimized, high-performance suspicious object detection algorithms.

References

[1] Z. Meng, M. Zhang, and H. Wang, “CNN with Pose Segmentation for Suspicious Object Detection in MMW Security Images,” Sensors 2020, Vol. 20, Page 4974, vol. 20, no. 17, p. 4974, Sep. 2020, doi: 10.3390/S20174974.

[2] Yang, Xi, Tan Wu, Lei Zhang, Dong Yang, Nannan Wang, Bin Song, and Xinbo Gao. "CNN with spatio-temporal information for fast suspicious object detection and recognition in THz security images." Signal Processing 160 (2019): 202-214.

[3] X. Yang, Z. Wei, N. Wang, B. Song, and X. Gao, “A novel deformable body partition model for MMW suspicious object detection and dynamic tracking,” Signal Processing, vol. 174, p. 107627, Sep. 2020, doi: 10.1016/J.SIGPRO.2020.107627.

[4] W. Cai, J. Li, Z. Xie, T. Zhao, and K. Lu, “Street object detection based on faster R-CNN,” Chinese Control Conf. CCC, vol. 2018-July, pp. 9500–9503, Oct. 2018, doi: 10.23919/CHICC.2018.8482613.

[5] K. V Shivthare, P. D. Bhujbal, A. P. Darekar, and Y. N. N, “Suspicious Activity Detection Network for Video Surveillance Using Machine Learning,” vol. 6, pp. 2456–0774, 2021, doi: 10.51319/2456-0774.2021.4.0017.

[6] H. Jain, A. Vikram, Mohana, A. Kashyap, and A. Jain, “Weapon Detection using Artificial Intelligence and Deep Learning for Security Applications,” Proc. Int. Conf. Electron. Sustain. Commun. Syst. ICESC 2020, pp. 193–198, Jul. 2020, doi: 10.1109/ICESC48915.2020.9155832.

[7] Cristyan Rufino Gil Morales, “Video Analysis to Detect Suspicious Activity Based on Deep Learning - DZone AI”, https://dzone.com/articles/video-analysis-to-detect-suspicious-activity-based Accessed: 2022-04-26.

[8] V. Singh, S. Singh, and P. Gupta, “Real-Time Anomaly Recognition Through CCTV Using Neural Networks,” Procedia Comput. Sci., vol. 173, pp. 254–263, Jan. 2020, doi: 10.1016/J.PROCS.2020.06.030.

[9] T. Saba, A. Rehman, R. Latif, S. M. Fati, M. Raza and M. Sharif, "Suspicious Activity Recognition Using Proposed Deep L4-Branched-Actionnet With Entropy Coded Ant Colony System Optimization," in IEEE Access, vol. 9, pp. 89181-89197, 2021, doi: 10.1109/ACCESS.2021.3091081.

[10] Ramzan, Muhammad, Adnan Abid, Hikmat Ullah Khan, Shahid Mahmood Awan, Amina Ismail, Muzamil Ahmed, Mahwish Ilyas, and Ahsan Mahmood. "A review on state-of-the-art violence detection techniques." IEEE Access 7 (2019): 107560-107575.

[11] Tian, Gangyi, Jianran Liu, Hong Zhao, and Wenyuan Yang. "Small object detection via dual inspection mechanism for UAV visual images." Applied Intelligence (2022): 1-14.

[12] Rani, Shilpa, Deepika Ghai, and Sandeep Kumar. "Object detection and recognition using contour based edge detection and fast R-CNN." Multimedia Tools and Applications (2022): 1-25.

[13] Almahasneh, Majedaldein, Adeline Paiement, Xianghua Xie, and Jean Aboudarham. "MLMT-CNN for object detection and segmentation in multi-layer and multi-spectral images." Machine Vision and Applications 33 (2022): 1-15.

[14] Ge, Ji, Chao Wang, Bo Zhang, Changgui Xu, and Xiaoyang Wen. "Azimuth-sensitive object detection of high-resolution SAR images in complex scenes by using a spatial orientation attention enhancement network." Remote Sensing 14, no. 9 (2022): 2198.

[15] Posilović, Luka, Duje Medak, Fran Milković, Marko Subašić, Marko Budimir, and Sven Lončarić. "Deep learning-based anomaly detection from ultrasonic images." Ultrasonics 124 (2022): 106737.

[16] Hirooka, Koki, Md Al Mehedi Hasan, Jungpil Shin, and Azmain Yakin Srizon. "Ensembled transfer learning based multichannel attention networks for human activity recognition in still images." IEEE Access 10 (2022): 47051-47062.

[17] Yuan, Minghui, Quansheng Zhang, Yinwei Li, Yunhao Yan, and Yiming Zhu. "A suspicious multi-object detection and recognition method for millimeter wave sar security inspection images based on multi-path extraction network." Remote Sensing 13, no. 24 (2021): 4978.

[18] Tian, Gangyi, Jianran Liu, and Wenyuan Yang. "A dual neural network for object detection in UAV images." Neurocomputing 443 (2021): 292-301.

[19] Song, Zhina, Haigang Sui, and Li Hua. "A hierarchical object detection method in large-scale optical remote sensing satellite imagery using saliency detection and CNN." International Journal of Remote Sensing 42.8 (2021): 2827-2847.

[20] Bravo, Daniel Trevisan, Gustavo Araujo Lima, Wonder Alexandre Luz Alves, Vitor Pessoa Colombo, Luc Djogbenou, Sergio Vicente Denser Pamboukian, Cristiano Capellani Quaresma, and Sidnei Alves de Araujo. "Automatic detection of potential mosquito breeding sites from aerial images acquired by unmanned aerial vehicles." Computers, Environment and Urban Systems 90 (2021): 101692.

[21] Meng, Zhichao, Man Zhang, and Hongxian Wang. "CNN with pose segmentation for suspicious object detection in MMW security images." Sensors 20.17 (2020): 4974.

[22] Javed, Abdul Rehman, and Zunera Jalil. "Byte-level object identification for forensic investigation of digital images." 2020 International Conference on Cyber Warfare and Security (ICCWS). IEEE, 2020.

[23] Yang, Xi, Tan Wu, Lei Zhang, Dong Yang, Nannan Wang, Bin Song, and Xinbo Gao. "CNN with spatio-temporal information for fast suspicious object detection and recognition in THz security images." Signal Processing 160 (2019): 202-214.

[24] Chen, Chao, Jiandan Zhong, and Yi Tan. "Multiple-oriented and small object detection with convolutional neural networks for aerial image." Remote Sensing 11.18 (2019): 2176.

[25] Fang, Wei, Lin Wang, and Peiming Ren. "Tinier-YOLO: A real-time object detection method for constrained environments." IEEE Access 8 (2019): 1935-1944.