

Suspicious Activity Detection in Adverse Weather Conditions using YOLOv7

Premanand Ghadekar¹, Sahil Jagtap², Bhushan Sadmake³, Nishka Mane⁴, Ketan Singh⁵ and Bhavika Chavan⁶

¹⁻⁶Department of Information Technology, Vishwakarma Institute of Technology, Pune, India'

Email: premanand.ghadekar@vit.edu, {sahil.jagtap21, bhushan.sadmake21, nishka.mane211, ketan.singh21, bhavika.chavan21}@vit.edu

Abstract— This research focuses on developing a robust and efficient system for suspicious activity detection in adverse weather conditions using the YOLOv7 model. The proposed system intends to annotate potentially hazardous actions within difficult environmental variables such as rain, fog, and crowded settings, including instances of gun use, fighting, and holding weapons like swords. We have designed the model to perform well in challenging weather circumstances, boosting its applicability in real-world scenarios, by utilizing the cutting-edge YOLOv7 architecture. We demonstrate the system's effectiveness in reliably detecting and localizing suspicious actions through thorough testing and evaluation, even in challenging weather and busy locations. The research advances public safety measures by developing a tool that can improve security monitoring and reaction under adverse weather situations.

Index Terms— Gun detection, knife detection, fight detection, sword detection, Computer Vision, Surveillance, YOLOv7 network.

I. INTRODUCTION

In recent years, advances in computer vision and deep learning have greatly impacted many things, especially public safety. Identifying and mitigating suspicious activity during inclement weather conditions poses a public safety challenge. Bad weather conditions such as rain and fog affect the effectiveness of traditional surveillance, often preventing potential threats from being detected. This research addresses the urgent need for efficient and robust anti-virus systems that can perform well in extreme weather conditions using the YOLOv7 model. The purpose of this research is to use the capabilities of the advanced technology YOLOv7. The model detects objects identified as suspicious activity in bad weather conditions. Suspicious activity detected included a variety of incidents, including incidents involving weapons such as firearms, physical confrontations such as fights, and incidents involving items of a capital dangerous nature such as swords. This research aims to improve public safety by examining these activities, especially in harsh weather conditions such as rain and fog and crowded environments where traditional observations are impossible. The use of YOLOv7 provides an effective means for immediate, efficient, and accurate detection, meeting the need for urgent public safety measures. In this work, we extend the work of YOLOv7 to take on the difficult task of detecting suspicious activity in severe weather conditions. We collect and interpret information containing numerous visuals depicting extreme weather conditions such as rain, fog, and crowded environments. Using this data, we trained the YOLOv7 model to detect and identify suspicious situations, including the presence of guns, fights, and swords in these negative situations.

By doing this, we aim to contribute to improving public safety by timely detection and intervention in severe weather conditions and threatening situations with large numbers of people. The program represents a significant step towards improving public safety and security through deep learning in extreme weather conditions.

II. LITERATURE SURVEY

[1] In this study, three models were used for gun detection, knife detection, and fight detection: basic CNN, ResNet50v2, and VGG19. Among these models, the VGG19 model achieved the highest accuracy of 95.4827. However, despite its high validation accuracy, the VGG19 model sometimes incorrectly identifies activities in unseen image data. This problem is commonly referred to as overfitting, which is a known issue in machine learning models. One possible solution to address this issue is to increase the training data if additional computing power is available. Additionally, it should be noted that live detection of these activities is not currently possible due to the lack of an annotated dataset, and the models developed in this study are solely intended for classification purposes.[2]In this paper, the formulation of suspicious action detection for security systems is discussed, acknowledging the complexity and variety of human behaviors in natural environments. The achieved accuracy of the detection system is approximately 95%, with YOLOv3 outperforming R-CNN which is faster in terms of processing speed for one image detection. However, the current feature extraction method only provides accurate results in controlled environments, and incorporating better feature extraction methods can enhance the accuracy of the system. Despite the limited quantity of training data, the test results still showed some mismatch in comparison to the ground truth. To improve the detection and practicality of the model, future work includes extending the training dataset to include suspicious videos of different activities and resolutions. Additionally, designing more sophisticated algorithms for real-time applications can further enhance the system's performance.[3]The paper introduces a method for detecting anonymous behavior, focusing on automated recognition of unusual patterns through video surveillance and anomaly detection. The proposed YOLO-Robbery network enhances security in supermarkets by identifying threatening activities and automatically alerting shop owners. The process involves real-time footage collection, frame transformation, multi-scale retinex preprocessing, and augmentation for data enrichment. Utilizing the YOLO V7 network, features are extracted to effectively recognize and classify threats in supermarkets. Fine-tuning of hyperparameters using Greedy Snake Optimization and training on the DCSASS dataset results in an overall accuracy of 99.15%. Comparative analysis demonstrates superior performance, with the proposed method surpassing CLSTM-NN, J. DCNN, and ANFIS by 13.15%, 2.15%, and 6.24%, respectively.[4] The inefficiency of manually sorting *Caenorhabditis elegans* (*C. elegans*), a valuable genetic and disease research model organism, is addressed in this study. To automate the sorting process, microfluidic-assisted sorting chips are leveraged, taking advantage of their benefits. The article presents the YOLOv7 deep learning model for the automatic detection and identification of *C. elegans*. On a dataset consisting of 3931 annotated worms, the model outperforms YOLOv5 and Faster R-CNN. The model demonstrates exceptional performance, achieving an impressive mean average precision (mAP) of 99.56% with an IoU 0.5 as the minimum value for intersection over the union. Moreover, it exhibits strong generalization capabilities, as evidenced by a remarkable mAP@0.5 score of 94.21% when tested on an independent validation dataset. Beyond mere identification, the model excels in accurately evaluating various phenotypic attributes such as size, movement speed, and fluorescence. This multifaceted identification approach significantly enhances sorting efficiency and holds great potential for advancing automated, integrated microfluidic platforms in *C. elegans* research.[5] Surveillance has experienced significant growth, expanding beyond traditional observation to incorporate modern technologies and advanced prediction methods. The utilization of intelligent video surveillance systems (IVS) that employ image recognition has become crucial for the prevention of crime and the enhancement of public safety. However, the challenge of processing real-time data and analyzing image contents has impeded the progress of developing solutions that are both user-friendly and cost-effective. This literature study presents a synthesis of findings from 57 IEEE publications published between 1977 and 2015, with a specific focus on approaches for detecting suspicious activity, identifying anomalous human behavior, tracking abandoned items, and monitoring unattended baggage. The investigation reveals the prevalence of intelligent methodologies in augmenting the intelligence of surveillance systems, such as neural systems, fuzzy logic, support vector machines, genetic algorithms, and others. The review provides valuable insights into the utilized approaches, as well as their strengths and limitations. [6] Object recognition is crucial in computer vision, but it faces major hurdles when applied to real-world pictures with noise, blurring, and rotational jitter. To address these issues and allow real-time object recognition, the YOLO technique, which is based on convolutional neural networks (CNNs) has emerged as a viable option. This study of the literature dives into the field of object detection, concentrating on the numerous

modifications attempted to improve the performance of the YOLO network. These changes are intended to overcome the challenges given by real-world circumstances and improve the algorithm's performance. By critically evaluating these changes, this analysis offers insight into the changing environment of object detection using YOLO and highlights the ongoing attempts to improve its capabilities. [7] In reaction to the worrisome exponential growth in crime rates, surveillance systems have become increasingly common in public places such as malls, transit hubs, schools, and airports. Manual monitoring for detecting suspicious activity has become impossible and inefficient due to the constant 24x7 video feed supplied by these cameras. As a result, demand for Intelligent surveillance systems has increased. This study describes a thorough method for automatically detecting and flagging various unusual actions in video data. Object tracking, moving object identification, and analysis of behavior for activity recognition are the three key components of the proposed framework. A feature extraction procedure determines essential parameters such as speed, direction, centroid, and dimensions, allowing for object tracking across video frames. The system successfully separates distinct activities and assesses whether a certain action belongs within the sphere of normality or deviates towards abnormal behavior by using domain-specific knowledge rules. Empirical investigation validates the framework's efficacy, with a detection accuracy of up to 90% achieved. The incorporation of intelligent surveillance systems into our security infrastructure answers the critical requirement for efficient, automated identification of suspicious activity, therefore improving public safety and security. This study makes an important addition by establishing a solid framework and demonstrating its effectiveness via empirical data. The three-pronged method of object identification, tracking, and behavior analysis provides a comprehensive answer to the difficult issue of detecting and reporting abnormal actions in video feeds. As society grapples with security issues, this paradigm represents a potential step towards the creation of smarter, more proactive monitoring systems that will provide a safer environment for everybody. [8] The new YOLO-GNS technique used in the work by Qiu et al. (2023) addresses the difficult challenge of special unmanned detection from UAV views. To improve feature extraction and lower computing costs, their technique proposes the Single Stage Headless (SSH) context framework. With better detection accuracy and frame rates than previous derivative models, YOLO-GNS performs better. The use of YOLO-based models for special vehicle recognition from UAVs by Xu (2018) and Changchun (2018) has also contributed to this field. These authors use cutting-edge approaches and optimizations that are particularly effective at recognizing small, swiftly moving objects. Together, these developments highlight the expanding importance of deep learning networks, especially YOLO-based models, in the field of special vehicle recognition from the perspective of UAVs, with widespread applications in surveillance, traffic monitoring, and disaster response. [9] CNN-based approaches are increasingly used in object identification and picture categorization. However, due to domain shifts in input pictures, these models frequently fail to recognize distorted images. To solve this, the authors present an integrated denoising object detector that performs admirably on a multi-noise challenge dataset. They use numerous WBF-trained denoisers to produce an integrated denoising model. On the test set, the model achieves an am(0.5:0.95) of more than 56.2%. To increase detection accuracy, the scientists additionally use ensemble models, data augmentation, and motion deblurring. They restore low-resolution pictures using the Real-organ model and test the performance of their denoising approach. Using the ensemble technique, the detector predicts more accurate anchors with more confidence. The authors summarize the key findings of their detection models and stress the significance of the denoiser in the detection process. Overall, their approach outperforms the challenge dataset and effectively eliminates numerous noises. [10] The paper describes an advanced multi-modal deep learning fusion approach for detecting illegal retail networks operating on social media platforms. This innovative system is specifically designed to identify drug-related content and individuals involved in drug sales within the social media landscape. It can recognize various types of narcotics mentioned in social media posts and gather relevant data on the individuals engaged in drug distribution. Illicit activities such as drug trafficking, human trafficking, and money laundering pose serious threats to both individuals and society as a whole. Law enforcement agencies are actively combating these criminal activities and are remaining vigilant in their efforts. Furthermore, the illicit drug trade often intersects with other criminal enterprises, further complicating the issue. The use of social media platforms for illegal trafficking activities has emerged as a pressing concern for law enforcement, necessitating the development and implementation of advanced techniques and tools to detect and intervene in these activities. [11] The realm of pothole detection research has made significant strides in response to the increasing demand for automated solutions, thereby promoting safer driving practices and fostering the evolution of more intelligent global transportation systems. Within this context, machine learning has emerged as a pivotal tool, with a particular emphasis on harnessing deep learning-based object detection algorithms. This scholarly paper conducts an extensive examination of an array of pothole detection systems that have been devised, casting light on the broad spectrum of machine learning and object detection methodologies deployed to tackle this pressing issue. These methodologies encompass diverse

data sources, spanning from imagery to sensor-generated data, and have been adapted to diverse geographical regions, addressing the unique challenges presented by each locale..[12] A completely automated computer-based system was created in this study to recognize common weapons, especially pistols and rifles. Significant progress has been made in object identification and recognition thanks to recent developments in deep learning and transfer learning. Using a customized dataset, the YOLO V3 "You Only Look Once" item recognition model was created. YOLO V3 beats YOLO V2 and traditional convolutional neural network (CNN) models, according to the training results. Notably, while transfer learning is employed to train the models, our technique doesn't demand powerful GPUs or a lot of computing power. By adding this model to our monitoring system, we aim to prevent manslaughter or mass killings and save lives. The probability of an assault or injury to human life can be decreased by using our suggested method in advanced surveillance and security robots to identify weapons or harmful objects. [13] The studies conducted by Svanström in 2022 focus on the research inquiry related to the detection and tracking of drones. To address this challenge, the authors employ a wide range of sensory modalities. They methodically design and thoroughly evaluate a multi-sensor system for detecting drones. This system includes a wide array of sensory inputs, such as video cameras, microphone sensors, thermal infrared cameras, and a fish-eye camera. Additionally, they explore the integration of supplementary technologies, including GPS receivers, radar modules, and ADS-B receivers. Another noteworthy aspect of their research is the creation of a publicly available dataset replete with comprehensive annotations for multi-sensor data. [14]The emergence of video conferencing technology, though available for several decades, has witnessed a substantial surge in popularity, especially in recent times. However, a conspicuous gap in this arena has been the absence of labeled datasets specifically tailored for detecting abnormal events within video conferencing contexts. Addressing this pressing need, a groundbreaking dataset has been meticulously curated, offering meticulous annotations related to anomalies and uncommon occurrences, and it is made accessible to the broader public. Importantly, the existing landscape currently lacks open-source methods designed explicitly for the surveillance and identification of abnormal incidents occurring within video conferencing sessions. In response, the study harnessed two pivotal tools: a pre-trained YOLO 3 model and a RetinaFaceNet integrated with a ResNet-50 network. These tools facilitated the precise extraction of bounding boxes encompassing individuals' faces within each image collage. Impressively, the study yielded notable outcomes, achieving a recall rate of 72.5% and a precision rate of 92.3%, predominantly relying on the utilization of facial expressions' embedding vectors and statistical profiling. [15]The papers discussed here delve into the realm of car detection models featuring attention mechanisms. In the work by Xu (2023), a novel cross-domain car detection model is proposed, incorporating an integrated convolutional block attention mechanism. This innovative approach demonstrates a remarkable 40% enhancement in performance when compared to models lacking this attention framework. Zhang (2021) contributes a multiscale parallel vehicle detection model that leverages attention mechanisms, achieving superior detection results in comparison to mainstream models. Chen (2020) presents a novel approach to enhance the accuracy of vehicle local region detection.

III. METHODOLOGY

Suspicious Activity Recognition using YOLO (You Only Look Once) object detection is a system that aims to detect and identify potentially suspicious activities in real time using computer vision techniques.

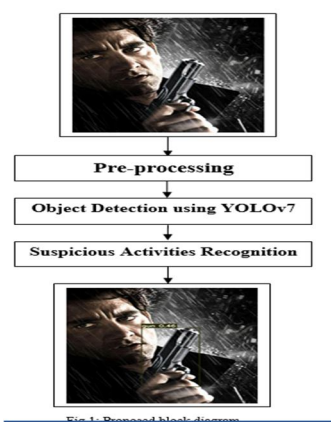


Fig.1: Proposed block diagram

A. System block diagram

The suggested system block diagram is displayed in Fig 1. The system typically consists of the following components:

1. *Video or image Input:* The system takes video streams as input from surveillance cameras or other sources. These video streams can be captured in real-time or pre-recorded.
2. *Pre-processing:* The video frames are pre-processed to enhance their quality and prepare them for object detection. Common preprocessing steps may include resizing, normalization, and noise reduction.
3. *Object Detection with YOLO:* The YOLO algorithm is utilized to identify and pinpoint objects within video frames. This algorithm partitions the image into a grid and makes predictions regarding bounding boxes and the likelihood of object classes within each grid cell. It possesses the capability to detect multiple objects simultaneously, exhibiting a notable level of accuracy and efficiency.
4. *Suspicious Activity Recognition:* Once objects are detected, the system applies specific rules or algorithms to determine whether the observed activities are suspicious. These rules can be based on predefined criteria, such as abnormal behavior, specific poses, or interactions between objects and individuals.

B. Implementation of YOLOv7 Algorithm

To develop a system for detecting suspicious activities using computer vision and YOLOv7 object detection, we considered these broad steps:

1. *Data collection and preprocessing:* To address the lack of available datasets on handguns, swords, knives, and fight detection in adverse climatic conditions such as fog and rain, we were compelled to gather our dataset. Given that we are utilizing the YOLOv7 object detection model for training purposes, a substantial amount of data is necessary to accurately define the specific task at hand. Acquiring a dataset tailored to our particular task proved to be a challenge due to the diverse range of images in various environments that needed to be incorporated into the training process. Relying solely on a publicly accessible search engine to find training images would have yielded only a limited set of images. Consequently, we relied on gathering data from publicly available datasets that were pertinent to our specific scenario. To construct our suspicious activity detector, we extracted frames from multiple sources to compile a comprehensive dataset. This involved obtaining images from the Roboflow databases, YouTube videos, and Google Images. In total, we amassed a collection of 10,000 images to train the model.
2. *Annotation Process:* As a preliminary measure, the data we acquired required annotation. A cumulative duration of 20 hours was dedicated to annotating all the image examples we obtained. This annotation process entailed the creation of bounding boxes around firearms, blades, and swords, as well as the identification of altercations between individuals depicted in the images. In our particular scenario, as we were specifically discerning four distinct activities, we established four categories to serve as labels for the detected objects. A selection of example images with questionable activities from this data set are displayed in Figure 2.



Fig. 2 A selection of example images with questionable activities from this data set are displayed

C. YOLO Model Training and Testing

To train and test a YOLO object detection model using the annotated dataset, it is necessary to follow the steps outlined below. YOLO is renowned for its ability to detect objects in real-time.

- Establish a training environment that includes the necessary dependencies, such as TensorFlow, Keras, or Darknet (the original framework for YOLO).
- Prepare the dataset in the required format for YOLO training, which typically involves using YOLO's labeled bounding box format (known as YOLO darknet format).
- Configure the YOLO model architecture based on specific requirements, including the number of classes and the desired backbone network (e.g., YOLOv3 or YOLOv4). In our case, we used YOLOv7.
- The training process can be initiated by utilizing the annotated dataset and making adjustments to hyperparameters, including the learning rate, batch size, and the number of training epochs.
- Continuously monitor the training process, evaluate the model's performance on a validation set, and save the best-performing model.

D. YOLO Algorithm

1. Grid System

- The input image is divided into an $S \times S$ grid.
- The task of estimating a predetermined number of boundaries falls to each grid cell.

2. Bounding Box Prediction

For a given grid cell at position (i, j) and bounding box index b, the predicted bounding box is represented as follows:

- Center coordinates (bx, by):

$$\begin{aligned} bx &= \sigma(t_x) + c_x \\ by &= \sigma(t_y) + c_y \end{aligned} \quad (1)$$

- Width and height (bw, bh):

$$\begin{aligned} bw &= p_w e^{t_w} \\ bh &= p_h e^{t_h} \end{aligned} \quad (2)$$

Here, σ is the sigmoid activation function.

3. Class Prediction

The class prediction for a bounding box is given by a softmax activation over all classes:

$$P(Class_i | Object) = \frac{e^{t_i}}{\sum_j e^{t_j}} \quad (3)$$

4. Objectness Score

Each grid cell predicts an objectness score:

$$P(Object) \times IOU_{pred}^{truth} = \sigma(t_o) \quad (4)$$

Where The anticipated bounding box and the ground truth box intersect at an intersection, or IOU.

5. Loss Function

The overall loss function is a combination of localization loss, confidence loss, and class loss:

$$Loss = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B L_{ij}^{coord} + \lambda_{conf} \sum_{i=0}^{S^2} \sum_{j=0}^B L_{ij}^{conf} + \sum_{i=0}^{S^2} L_i^{class}$$

Where:

- L_{ij}^{coord} is the localization loss for the bounding box (l, j, b).
- L_{ij}^{conf} is the confidence loss for the objectness score.
- L_i^{class} is the classification loss.

(5)

6. Non-Maximum Suppression

To filter out redundant bounding boxes, non-maximum suppression is applied based on the confidence scores. While the fundamental idea of real-time object detection with a grid system is present in all versions, the architecture, optimizations, and additional features have evolved. A cutting-edge object identification model noted for its quickness and precision in real-time applications is the YOLOv7 method. Architecture Design: YOLOv7 follows a design that combines deep learning techniques for object detection. It includes a backbone network, often a variant of Darknet, for feature extraction, and utilizes multiple detection heads for predicting bounding boxes and class probabilities at different scales. YOLOv7 incorporates a soft label generation method, which aids in training the model effectively. The soft label method involves providing the model with more informative labels during training to enhance its performance. YOLOv7 employs gradient flow propagation paths to optimize the combination of the network with re-parameterized convolutions. This contributes to its speed and accuracy in real-time object detection. YOLOv7 is referred to as a "trainable bag-of-tricks," emphasizing its adaptability and ease of training for various computer vision tasks.

IV. RESULTS AND DISCUSSION

The results obtained with machine learning classifiers are taken into consideration in this part for the identification of four distinct suspicious human behaviours. Metrics like accuracy, precision, recall, and F1-score may be computed from the confusion matrix and provide important information about how well the classification model is doing. Therefore, it is essential to initially examine the confusion matrices of the model shown in figure 3.

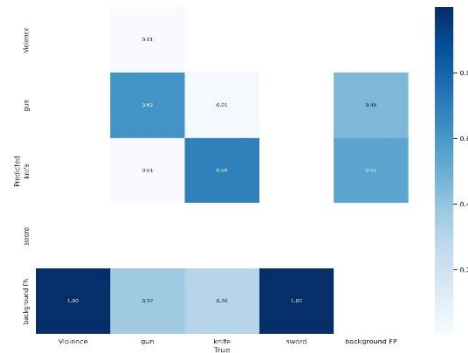


Fig3: Confusion matrix for the YOLOv7 model

From the analysis of this matrix, it is evident that the identification of fight (violence) and sword is inaccurate, resulting in a decrease in the overall accuracy of the model. The model only demonstrates satisfactory accuracy in detecting two activities, namely gun detection and knife detection.

To ensure accurate object detection and activity recognition, performance measurement is essential. In this case, the following parameters are employed for measurement:

- Precision is a metric used to evaluate the accuracy of predictions, specifically the percentage of correct predictions.
- Conversely, recall evaluates how well an item can be found inside a database. The formula for recall is given by:
- Precision and recall are weighted averages that are used to calculate the F1 score, a test's accuracy metric. It provides an overall assessment of the system's performance.
- Accuracy, The precision of a YOLO (You Only Look Once) model in detecting objects is commonly evaluated using Intersection over Union (IoU) criteria for Mean Average Precision (mAP) measures.. As YOLO is intended for object detection and localization in images, where multiple objects may exist, traditional classification accuracy is not directly applicable in measuring its accuracy.

$$mAP = \frac{AP@IoU=0.5 + AP@IoU=0.75 + \dots}{\text{Number of IoU thresholds}} \quad (6)$$

$$IoU = \frac{\text{Area of Intersection}}{\text{Area of Union}} \quad (7)$$

Mean Average Precision (mAP) is a metric that combines precision and recall to provide a comprehensive measure of the effectiveness of a model for object detection across multiple categories or classes. Table I shows the overall performance measured.

THE PERFORMANCE MEASUREMENTS ARE PRESENTED IN TABLE I.

Precision (P)	Recall (R)	F1 score	mAP@.5
90.5%	40%	55.4%	44.5%

The model was trained for only 30 epochs, resulting in relatively average accuracy. Improving accuracy can be achieved by extending the training duration through additional epochs. Training the model for more epochs allows it to further learn patterns and features in the data, potentially enhancing its performance. However, it's crucial to monitor for signs of overfitting and consider adjusting other training parameters, such as learning rates and regularization techniques, to ensure optimal model generalization. The predicted output images are shown in fig.4 which displays the gun detection , fig.5 which displays Knife detection , fig.6 displays the violence(fight) detection.

1) Gun Detection



Fig. 4: Detection of guns with their confidence scores.

2) Knife Detection



Fig. 5: Detection of knives with their confidence scores.

3) Violence detection



Fig. 6: violence detection based on their confidence ratings

Fig.7. A fundamental measure of robustness and efficacy is the Precision-Recall Curve, which is seen below. Fig.8. represents The F1-score-confidence curve illustrates the trade-off between precision and recall at varying confidence levels in a classification model, which is seen below.

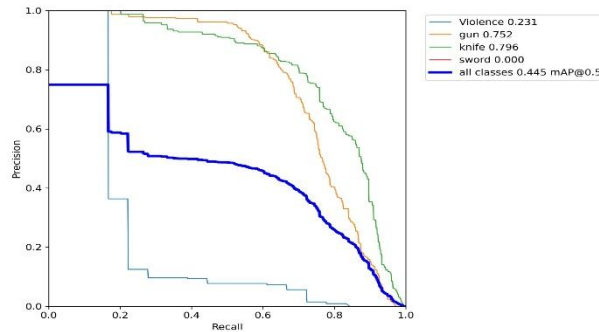


Fig. 7: Precision vs Recall curve.

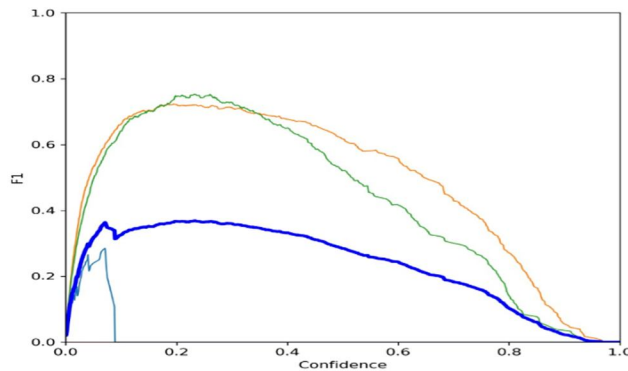


Fig. 8: F1-score - Confidence.

The proposed work has been compared to similar work conducted by Mounika[12]. Mounika's existing model was specifically developed for gun detection. In contrast, the proposed model can detect multiple weapons, including guns, swords, and knives. Additionally, Dr. Vaqar Ansari[1] has also conducted similar work in the field. Dr. Ansari's existing model was designed to classify suspicious activities related to guns, fights, and knives based on images. On the other hand, the proposed model not only classifies but also detects objects in both images and videos. A comparison with the previous study is shown in Table II.

TABLE II: COMPARISON WITH EXISTING MODEL

MODEL	Precision
OS+BE+AR+FE [15]	85%
The proposed model	90.5%

V. CONCLUSION

In conclusion, our work focuses on the use of YOLOv7 to detect suspicious activities, including the use of guns, fights, and swords, in adverse weather conditions such as rain, fog, and crowded scenes. We achieved good results using YOLOv7, detecting targets in real-time, and demonstrating its robustness in bad weather conditions. The system demonstrates the ability to detect potential hazards even in severe weather conditions, which is important for improving public safety. A safer society will result from the successful use of this model, which provides the way for future developments in surveillance systems that will work well in a variety of environmental scenarios.

FUTURE SCOPE

The project's future objective includes improving the YOLOv7 model's capacity to detect suspicious activities in challenging weather circumstances like rain, fog, or congested areas. Robust performance under a variety of

scenarios might be achieved by implementing real-time adaptation algorithms that dynamically modify model parameters based on weather data and crowd density. Additionally, adding multi-modal sensing capabilities, such as radar or infrared imaging, could improve the model's capability to detect weapons precisely in bad weather even more. The model's performance can be greatly enhanced, and its application in real-life situations can be expanded, through working with climatic help for real-time weather updates and investigating the prospect of utilizing cloud computing for heavier computations and data processing.

REFERENCES

- [1] Dr. Vaqar Ansari, Aditya Ghadge, Prasham Shah, Sumon Ghosh, Vaibhav Sanghavi. Suspicious Activity Detection Using Different Models. <https://www.ijraset.com/research-paper/suspicious-activity-detection-using-different-models>.
- [2] N. Bordoloi, A. K. Talukdar and K. K. Sarma, "Suspicious Activity Detection from Videos using YOLOv3," 2020 IEEE 17th India Council International Conference (INDICON), New Delhi, India, 2020, pp. 1-5, doi: 10.1109/INDICON49873.2020.9342230.
- [3] Vivek alias Meenatchisundaram Chidambaram, Karthik Painganadu Chandrasekaran. YOLO ROBBERY: Video Surveillance based Robbery Prediction and Alert System Using YOLO V7, 06 September 2023.
- [4] Zhang, Jie, Shuhe Liu, Hang Yuan, Ruiqi Yong, Sixuan Duan, Yifan Li, Joseph Spencer, Eng Gee Lim, Limin Yu, and Pengfei Song. 2023. "Deep Learning for Microfluidic-Assisted Caenorhabditis elegans Multi-Parameter Identification Using YOLOv7" *Micromachines* 14, no. 7: 1339.
- [5] G. Mathur and M. Bundele, "Research on Intelligent Video Surveillance techniques for suspicious activity detection critical review," 2016 International Conference on Recent Advances and Innovations in Engineering (ICRAIE), Jaipur, India, 2016, pp. 1-8, doi: 10.1109/ICRAIE.2016.7939467.
- [6] Rekha BS, Mariam A, Srinivasan GN, Shetty SA. Literature survey on object detection using YOLO. *International Research Journal of Engineering and Technology (IRJET)*. 2020 Jun;7(06).
- [7] Chaudhary S, Khan MA, Bhatnagar C. Multiple anomalous activity detection in videos. *Procedia Computer Science*. 2018 Jan 1;125:336-45.
- [8] Qiu, Zifeng, Huihui Bai, and Taoyi Chen. 2023. "Special Vehicle Detection from UAV Perspective via YOLO-GNS Based Deep Learning Network" *Drones* 7, no. 2: 117.
- [9] Ji B, Huang J, Huang W, He S. An Ensemble Model for Distorted Images in Real Scenarios. *arXiv preprint arXiv:2309.14998*. 2023 Sep 26.
- [10] Rupa, Anamika Paul. "Multi-Modal Deep Fusion for Activity Detection from Both the Virtual and Real-World." PhD diss., University of Maryland, Baltimore County, 2023.
- [11] Audu-war, Samuel Ochai, Sylvanus Okwudili Anigbogu, Kenechukwu Sylvanus Anigbogu, Gloria Nkiru Anigbogu, and Doris Chinedu Asogwa. "Pothole detection using image surveillance system: A review." *World Journal of Advanced Engineering Technology and Sciences* 9, no. 2 (2023): 214-222.
- [12] Sanam Narejo, Bishwajeet Pandey, Doris Esenarro vargas, Ciro Rodriguez, M. Rizwan Anjum, "Weapon Detection Using YOLO V3 for Smart Surveillance System", *Mathematical Problems in Engineering*, vol. 2021, Article ID 9975700, 9 pages, 2021. <https://doi.org/10.1155/2021/9975700>
- [13] Svanström, Fredrik, Fernando Alonso-Fernandez, and Cristofer Englund. 2022. "Drone Detection and Tracking in Real-Time by Fusion of Different Sensing Modalities" *Drones* 6, no. 11: 317.
- [14] Horowitz, Shmuel, Dima Kagan, Galit Fuhrmann Alpert, and Michael Fire. "Interruptions detection in video conferences." *arXiv preprint arXiv:2303.02052* (2023).
- [15] Revathi, A.R., Kumar, D. "An efficient system for anomaly detection using deep learning classifier". *SIViP* 11, pp. 291–299 (2017)