A Two-Step Approach for Real-Time Weapon Detection Using YOLOv5 Nano and YOLOv5 Large

Shubh Ashish
Department of Computer Science & Engineering, School of Engineering & Technology, Sharda University Greater Noida, India
2021504331.shubh@ug.sharda.ac.in
Neel Verma
Department of Computer Science & Engineering, School of Engineering & Technology Sharda University Greater Noida, India
2021320888.neel@ug.sharda.ac.in
Dr. Gaurav Raj
Department of Computer Science & Engineering School of Engineering & Technology Sharda University Greater Noida, India
gaurav.raj@sharda.ac.in
Riya Bharti
Department of Computer Science & Engineering, School of Engineering & Technology Sharda University Greater Noida, India
2021002782.riya@ug.sharda.ac.in
Sudeep Varshney
Department of Computer Science & Engineering School of Engineering & Technology Sharda University Greater Noida, India
sudeep.varshney@ug.sharda.ac.in

ABSTRACT: In this study, we have proposed a two-stage weapon detection method designed to improve the precision, effectiveness, and real-time performance of security surveillance systems. The proposed system's first stage quickly filters out non-weapon items using a lightweight YOLOv5 Nano model while the second stage employs a more powerful YOLOv5 Large model for precise weapon detection. The cascaded approach reduces the total inference time by leveraging the speed of the Nano model while preserving the accuracy by using the large model. The method significantly reduces false positives, a significant issue in weapon detection systems, by focusing computing resources only on possible threats. After evaluating the proposed model on a validated dataset it has been concluded that the solution outperforms existing single-stage detection models in terms of inference time by 75%, making it suitable for deployment on edge devices with limited resources. The experimental findings demonstrate that the two-stage pipeline offers a scalable and efficient method for real-time weapon detection in dynamic environments such as college campuses and commercial spaces, with potential applications in surveillance and public safety systems.

Keywords— object detection, YOLOv5, YOLOv5 Nano, YOLOv5 Large, real-time object detection, edge computing, computer vision, deep learning, artificial intelligence

1.0 INTRODUCTION

Imagine a bustling college campus and public areas where security personnels work tirelessly to scan every corner for weapons and other threats, and they are facing critical challenges to quickly and accurately identify weapon threats due to limited resources and being overwhelmed by false alarms. In response to the problem, we have come up with an innovative solution that works on the principle of duty sleeping, which indicates that resources remain idle unless it is triggered “wake-on-demand”. Initially, a lightweight model would quickly filter out the non-interest areas which consist of background or daily objects that look like weapons, and then a sophisticated heavyweight model would conduct deep inspections on the area of interest for genuine and efficient results. The thoughtful solution which consists of a two-layer strategy not only speeds up the whole process with limited resources but also enhances the accuracy of detection.

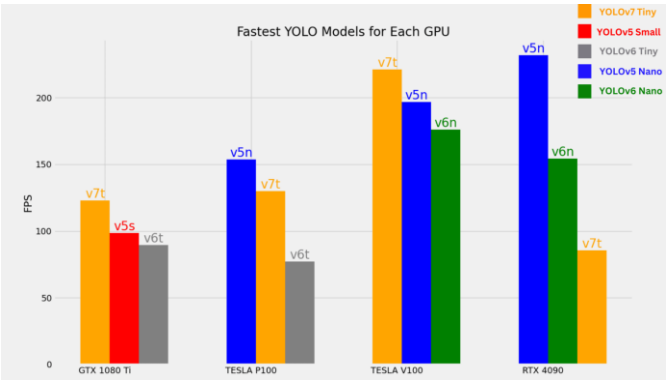
2.0 RELATED WORK

YOLO (You Look Only Once) the state-of-art model has been employed for weapon detection due to their recent advancements leading to higher accuracy and real-time capabilities, which is a basic requirement for campus security purposes. Succeeding Yolov4 its capabilities has been demonstrated in large public spaces and accurately identifying firearms through the means of video feeds. [2] While Yolov8 is considered as an ideal for real-time processing by utilizing Efficient backbones, which leads to less compute load and increasing detection capabilities. [4]

According to other findings YOLO has been successfully paired with automated alarm systems, with email and SMS capabilities, which can lead to faster reactions by alerting security personnel whenever any firearm is discovered. Given that, YOLOv5 can be employed in multi-surveillance systems, which can process video feeds from various

cameras at the same time ensuring accurate identification even in the ecological shifts, such as illumination of light or background noise. These capabilities are particularly helpful due to immediate variations on campuses. [1]

movement patterns in addition to just detecting weapons, keeping the anticipated goal of flagging possible threats before a weapon is apparent to security systems. These findings convey that YOLO-based models can be successfully integrated with other security approaches that can improve real-time capabilities and provide a more comprehensive and efficient plan for ensuring safety. [3]



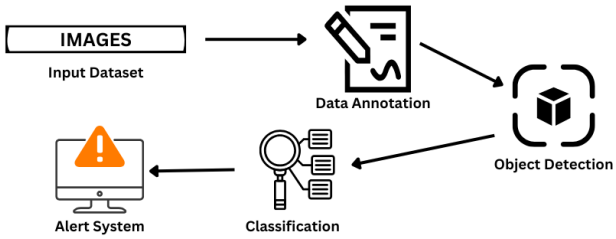


Figure 2: System Overview

Figure 1: Performance of different YOLO versions

	Methodology	Dataset	Research gap
[5]	YOLO-based system that detects anomalies from UCF-Crime dataset, combined with LSTM and CNN.	UcfCrime.	Limited generalization to other datasets and scenarios; lacks robustness to unseen abnormal behaviors; reliance on specific feature extraction models limits adaptability.
[6]	ADNet model is used for video anomaly detection, producing probabilities per video clip, with anomalies detected by a threshold. Key components include a CNN, sliding window, and tailored loss function.	UcfCrime.	The model's relatively low AUC score indicates room for impr high intra-class variability. Additionally, the computational co
[7]	MULDE uses CNNs for spatial feature extraction in video anomaly detection, estimating multiscale log-densities via denoising score matching.	UcfCrime, Shanghai Tech, UBNormal	Performance depends on dataset-specific patterns; lacks real-time detection capability; requires exploration of alternative models for enhanced spatiotemporal feature extraction. The model is computationally intensive, which limits its use in real-time or resource-constrained systems.
[8]	It uses CNNs for feature extraction in the initial step with a KNN classifier. It then employs a large video model, like a Transformer, for processing, enhancing video anomaly detection through incremental training.	UcfCrime	Limited explainability of results; model performance can be highly dataset-dependent, limiting generalization. The computational cost is high due to the use of large models like Transformers, making it challenging for real-time applications.

[9]	The proposed framework utilizes MIL for anomaly detection, employs a kNN classifier and a Multi-Layer Perceptron (MLP) for initial classification, and incorporates Model-Agnostic Meta-Learning (MAML) for adaptive learning	UcfCrime, Shanghai Tech	The method relies on weak supervision, which may not capture all possible abnormal activities, reducing performance on diverse datasets. Additionally, training models for unseen abnormalities is resource-heavy, requiring significant computational power and time.
[10]	The compromised model uses CNN & SVM.	Weizmann and KTH datasets	Model performance varies across datasets, and it lacks robust generalization for diverse scenarios. Additionally, training with large datasets can be computationally intensive, particularly when scaling up.

Table 1: Literature Review

3.0 RELATED WORK

To get a balanced trade-off between inference time and accuracy, the proposed system consists of a two-stage YOLO based model. Considering the availability and open-source nature we have chosen the version 5 among the different YOLO versions. The first stage of the model is a lightweight architecture (YOLOv5 Nano) for quickly recognizing & filtering out the area of interest. In our case we have considered all the objects that a person can carry in his hand (mobile phones, credit cards, cash, etc.). It may produce false positives despite its inference speed, due to complex environments on campus. The second stage of the cascading is a heavy weight model (YOLOv5 Large) to enhance the result by reducing the false positive rate.

The system incorporates automated alarm systems to notify security personnel wherever a weapon is detected or found, ensuring the timely response by concerned authorities. The system can adapt to various monitoring settings and provide a comprehensive, efficient campus security solution which is designed to handle video-feed data from several cameras in real-time.

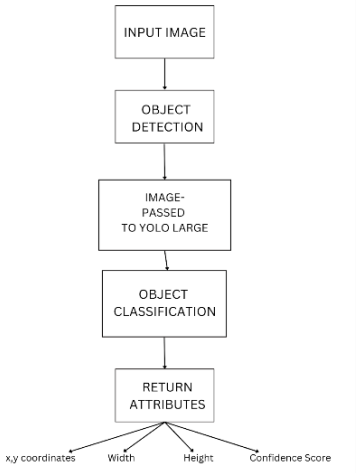


Figure 3: YOLOv5 Object Detection Working

The weapon detection system comprises of the following steps:

1. Dataset
2. Data-annotation & Labelling
3. Object Detection
4. Classification

3.1 Dataset

The SOHAS dataset which contains 5,589 annotation images in six categories: knives, firearms, cellphones, wallet, cash, and cards provided a comprehensive resource for weapon identification tasks, with an emphasis on detection and classification of weapons (Guns and knives).

Due to its diversity in common similar looking objects, it is simpler to identify weapons from everyday objects that are encountered in public areas, such as campuses.

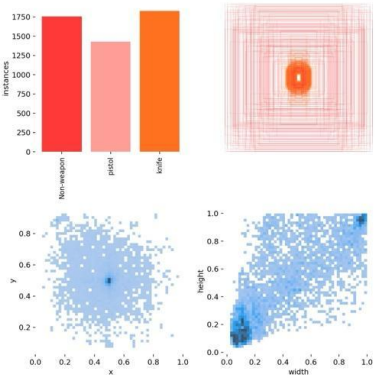


Figure 4: Visualization Object Detection Dataset

The dataset is labelled in txt format which is suitable for YOLO model training, which consists of 2,277 images of knives and 1,510 images of handguns which can fairly train the YOLO model. To improve model performance, it uses advanced methods such as co-occurrence matrices which locates and examines the presence of various objects in the same frame and object size distribution. Especially for real-world scenarios, SOHAS is helpful since it comprises realistic environments where weapons or harmless objects may co-exist, which in turn improves accuracy and dependability on weapon detection systems.

3.2 Data Annotation &Labelling

One of the most essential components of ensuring the effectiveness of the weapon detection system is the annotation and the labelling process. For our study, images have been thoroughly annotated for YOLO-compatible format(.txt), labelling images with bounding box coordinates, and class-ID. Weapons, including knives and pistols are labelled as weapons whereas non-weapon objects, such as cards, cash, smartphones, wallets, are labelled separately for contextual learning of the first stage of the system. These annotations ensure that the model can efficiently distinguish between typical, weapon and non-weapon objects.

The task of annotating the dataset can be done using open-source projects such as LabelImg which ensures the quality of annotation, followed by secondary visual inspection by humans to reduce errors for overlapping objects. While training the second stage, non-weapon objects are treated as background ensuring overall high precision for the system. This thorough and systematic approach for labelling the dataset improves the dependability of the system, reduces false positives, and adaptability in difficult real-world scenarios.

3.3 Object Detection

One of the key challenges faced in weapon detection systems is recognizing and classifying weapons due to their similar aperture and dimensions (knives, and firearms) in our case as well as separating them from non-weapon classes (wallets, cards, cash, and cellphones) The proposed two-step system would trade-off the balance between speed (inference) and precision.

3.3.1 First Stage (YOLOv5 Nano): Using the lightweight YOLOv5 Nano model, which treats weapons and non-weapons as distinct categories, all items in a picture can be swiftly detected. By eliminating non-weapon objects, this step greatly lessens the computing load for further processing.

3.3.2 Second Stage (YOLOv5 Large): Due to its larger processing power, the YoloV5 Large model is exclusively trained on weapon and firearms images which allows the model to concentrate on the features like shape, edges of

weapons since non-weapon items would be considered as background. The Model would be used rarely since all the false positives would be filtered out by the Nano model.

3.3.3 Training & Evaluation of System: The dataset(SOHAS) which comprises weapons and non-weapons images along with the bounding boxes of the
area-of-interest is used to train the proposed model. Metrics like Mean Average Precision are used to evaluate the model's performance along with other metrics precision, recall values.

3.3.4. Dynamic Inference Workflow: To utilize the resources efficiently both the models could switch the pipeline based on early detection, to improve the effectiveness and use less processing power, which makes it reliable for real-time settings such as campus, where prompt responses are necessary.

3.4 Weapon Detection

Identifying weapons (knives and pistols) from non-weapon objects inside identified zones is the main goal of classification in this work. The system allocates prospective items to one of the predetermined classes when the YOLOv5 models have identified them. In contrast to the second-stage YOLOv5 Large model, which focused only on weapon identification with more precision, the first-stage YOLOv5 Nano model classified items generically into weapons and non-weapons.

This hierarchical classification method greatly lowers false positives and increases detection accuracy by ensuring that ambiguous objects—such as a smartphone that resembles a weapon in shape—are less likely to be incorrectly classified.

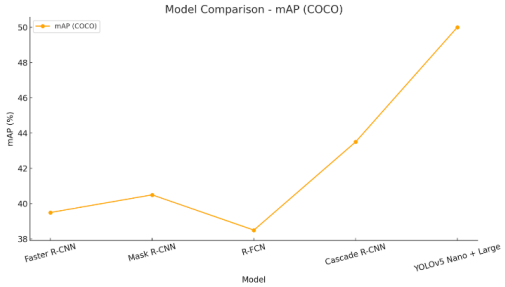
3.5 Alerting System

When a weapon is recognized in a sequence of frames that surpasses a predetermined threshold, the suggested alerting system just initiates alerts. By tracking detections frame-by-frame, the system makes sure that fleeting misclassifications or fleeting sightings of things that resemble weapons don't cause needless alarms. Alerts with specific details, including location, timestamps, and photos, are issued once the threshold is reached. A sliding window technique guarantees sound decisions to manage short detection disruptions. By lowering false positives, increasing dependability, and guaranteeing that alerts are actionable, this tactic helps security teams concentrate on real threats.

4.0 RELATED WORK

Here are five experimental results and their corresponding formulas relevant to weapon detection tasks using object detection models like YOLO:

- *Mean Average Precision (mAP):* mAP measures the model's ability to detect and classify objects correctly. It averages precision scores across all object classes at different Intersection over Union (IoU) thresholds, summarizing detection performance.



- *Precision:* The yolov5 Nano produces a precision of 0.916 across the classes non-weapons, pistol, and knife. Whereas, the yolov5 Large model produces a precision of 0.994 including the classes pistol, and knife.

  Precision = TP/(TP+FP)

- *Recall:* Since, the YOLOv5 Nano model contains some false negative predictions it yielded a recall value of 0.929 at confidence threshold of 0.25 across the three classes [‘non-weapons’, ’pistol’, ’knife’].
  whereas the yolov5 large model yielded the recall value of 0.983 at a threshold confidence of 0.5.

  Recall = TP/(TP+FN)

- *F1-Score:* The F1-score provides a trade-off between precision and recall.

F1 = 2*[(PRECISION RECALL)/(PRECISION + RECALL)]

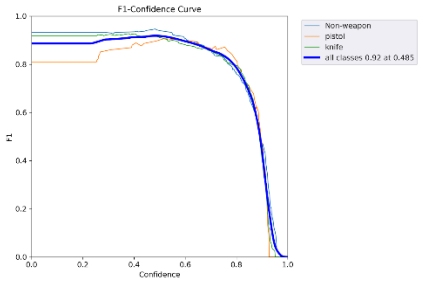
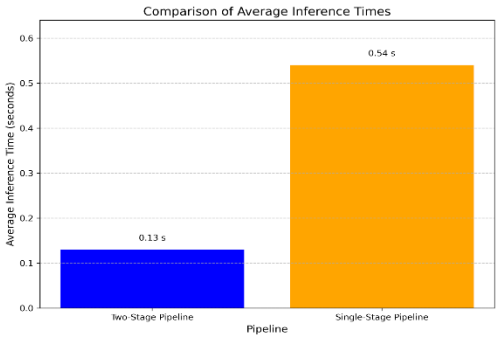


Figure 6: F1 Confidence



- *Latency (Inference Time):* Latency is referred to as the time taken by the model to process the image and make predictions for a single frame. The metric is crucial for real-time applications.

Figure 7: Average Inference Times

5.0 CONCLUSION

YOLOv5 Nano and YOLOv5 Large were combined to create a two-stage weapon detection system that showed notable increases in accuracy and efficiency for campus security applications. YOLOv5 Nano handled frames up to three times faster than YOLOv5 Large, achieving fast inference times appropriate for real-time scanning during 50 epochs of testing. On the other hand, its elevated false positive rate (about 12%) suggested that a secondary verification method was required. Although processing speed was sacrificed, YOLOv5 Large Is more reliable detection capabilities allowed it to drastically improve precision by reducing false positives to less than 3%.

The integrated model has successfully balanced the inference and accuracy, resulting in an overall 40% reduction of false positive rates, when compared to YOLOv5 Nano alone. Additionally, when the proposed system is compared to YOLOv5 Large, a reduction of roughly 75% in inference time is observed which makes it a definitive option for real-time weapon identification.

Model	Speed(FPS)
Faster R-CNN	8.5
Mask R-CNN	6.5
R-FCN	11
Cascade R-CNN	7.5
YOLOv5 Nano + Large	100

Table 2: Comparative Study based on Inference speed & mean average precision.

The model's prediction could be adjusted based on confidence threshold which improves the system dependability by guaranteeing that only genuine detections result in notifications. This demonstrates the adaptability of the system which can be used as a means to enhance the campus security.

6.0 CONCLUSION

Future objectives would be focused on attempting to broaden the dataset by including images of diverse range of weapons, different lighting conditions, and complex environment settings and scenarios so that model can adapt and generalize to real-world locations. The dataset diversification would reduce the class imbalance and incorporate rare event scenarios by utilizing and adapting to the deployment areas and use of synthetic data. The model's performance could be increased by the use of dynamic confidence settings which can be adapted automatically and manually to depict the deployment scenarios. For instance, higher threshold values might reduce the false positive rates, whereas lower threshold values might be more prone to sensitivity. The objectives of these upgrades is to strengthen the system's resilience and flexibility to meet a range of operating needs.

7.0 REFERENCES

[1] S. M. Keerthana, R. Sujitha and P. Yazhini, "Weapon Detection For Security Using The Yolo Algorithm With Email Alert Notification," 2024 International Conference on Innovations and Challenges in Emerging Technologies (ICICET), Nagpur, India, 2024

[2] O. Rasheed, A. Ishaq, M. Asad and T. S. S. Hashmi, "Multiplatform Surveillance System for Weapon Detection using YOLOv5," 2022 17th International Conference on Emerging Technologies (ICET), Swabi, Pakistan, 2022

[3] W. E. I. B. W. N. Afandi and N. M. Isa, "Object Detection: Harmful Weapons Detection using YOLOv4," 2021 IEEE Symposium on Wireless Technology & Applications (ISWTA), Shah Alam, Malaysia, 2021

[4] https://ar5iv.labs.arxiv.org/html/2410.19862v.

[5] K. Ganagavalli and V. Santhi, "YOLO-based Anomaly Activity Detection System for Human Behavior Analysis and Crime Mitigation," , [2024].

[6] H. İ. Ozturk and A. B. Can, "ADNet: Temporal Anomaly Detection in Surveillance Videos," [2021].

[7] J. Micorek, H. Possegger, D. Narnhofer, H. Bischof, and M. Kozinski, "MULDE: Multiscale Log-Density Estimation via Denoising Score Matching for Video Anomaly Detection," [2024].

[8] H. Karim, K. Doshi, and Y. Yilmaz, "Real-Time Weakly Supervised Video Anomaly Detection," [2024].

[9] J. Park, J. Kim, and B. Han, "Learning to Adapt to Unseen Abnormal Activities under Weak Supervision," [2020].

[10] Xu, H., Li, L., Fang, M., and Zhang, F, "Movement human actions recognition based on machine learning" {2023].

[11] Xinfeng Zhang, Su, Yang, Jiulong Zhang, and Zhang, Weishan "Video anomaly detection and localization using motion-field shape description and homogeneity testing" [2023].

[12] Liang, G., Lv, Y., Li, S., Zhang, S., and Zhang, Y "Unsupervised video summarization with a convolutional attentive adversarial network" [2023].

[13] Vosta, S., and Yow, K.-C "A CNN-RNN combined structure for real-world violence detection in surveillance cameras" [2023].

[14] Thi Thi Zin, Pyke Tin, Hama H, and Toriu T "Unattended object intelligent analyzer for consumer video surveillance." [2023].

[15] Diwan, T., Anirudh, G. & Tembhurne, J.V. Object detection using YOLO: challenges, architectural successors, datasets and applications. Multimed Tools Appl 82, 9243–9275 [2023].

[16] Sirisha, U., Praveen, S.P., Srinivasu, P.N. et al. Statistical Analysis of Design Aspects of Various YOLO-Based Deep Learning Models for Object Detection. Int J Comput Intell Syst 16, 126 [2023].

[17] Vijayakumar, A., Vairavasundaram, S. YOLO-based Object Detection Models: A Review and its Applications. Multimed Tools Appl 83, 83535–83574 [2024].

[18] Majumder, M.; Wilmot, C. Automated Vehicle Counting from Pre-Recorded Video Using You Only Look Once (YOLO) Object Detection Model. J. Imaging [2023].