# Dangerous Objects Detection Using Deep Learning and First Responder Drone

Zeyad AlJundi
https://orcid.org/0009-0009-5158-4249
*Naif Arab University for Security Sciences, Saudi Arabia*

Saad Alsubaie
*Naif Arab University for Security Sciences, Saudi Arabia*

Muhammad H. Faheem
https://orcid.org/0000-0002-1643-6728
*Naif Arab University for Security Sciences, Saudi Arabia*

Raha Mosleh Almarashi
*Naif Arab University for Security Sciences, Saudi Arabia*

Emad-ul-Haq Qazi
https://orcid.org/0000-0003-1448-3632
*Naif Arab University for Security Sciences, Saudi Arabia*

Jong Hyuk Kim
https://orcid.org/0000-0003-4505-1103
*Naif Arab University for Security Science, Saudi Arabia*

## ABSTRACT

Detecting dangerous objects, such as firearms or knives, is crucial for public safety or accurate situational assessment in crime scenes in law enforcement applications. Drones as first responders have been actively utilized for this purpose, showing significant benefits in law enforcement with fast and early detection of such objects. However, automated detection is still challenging, particularly with low-quality drone cameras that operate in low illumination conditions. We evaluate the performance of four popular AI deep learning models to automate the detection of dangerous objects recorded from low-quality drone cameras. The results show that the YOLOv5s model achieves the best detection performance, yielding mAP50 results of 0.964 for color and 0.949 for infrared videos, which are excellent performances considering the low-quality and low-resolution dataset. The trained network model is further implemented as an online web application where law enforcement officers can upload videos taken from drones or CCTV.

## KEYWORDS

Drone as a First Responder (DFR), Dangerous Object Detection, Deep Learning, Security and Law Enforcement

## INTRODUCTION

Drone as a first responder (DFR) has emerged as a new technological tool in security and law enforcement, in which drones are deployed to respond to an emergency call in police departments, arriving at the scenes before the ground officers while streaming live footage to the officers en route. This system enables the ground responders to prepare for what they might encounter at the scene, such as the presence of firearms. For example, the Chula Vista Police Department in the United States has adopted this technology since 2018, deploying DFR from the rooftop for over 20 thousand emergency calls, arriving at the scene within 93 seconds on average while assisting in arrests of over 28 thousand suspects (Chula Vista Police Department, 2024). Since deploying drones is typically less expensive than ground patrols or helicopters, police can allocate resources more effectively for their budgets. Although DFRs are still in their infancy, they can be utilized for crime scene investigation, disaster response, critical infrastructure protection, crowd safety, and so on.

DOI: 10.4018/IJDCF.367034

Processing videos captured from the drone is also crucial for accurate situational assessment and decision-making in law enforcement. Most security systems have relied on camera operators' bare-eye inspection of the videos. However, this kind of task is typically very tedious and dull, as it requires processing a large volume of live or recorded footage, thus reducing the work performance. In recent years, artificial intelligence (AI) technologies have made significant progress in understanding visual data, such as object detection and scene understanding (Soori et al., 2023; Zhu et al., 2021a). Therefore, integrating this AI capability into drones can significantly enhance the performance of security systems.

Fast and robust detection of dangerous objects, like firearms and knives, is critical in the DFR systems to support the decision-makers in taking the necessary security measures. Although many datasets and AI models are available for general object detection, firearm and dangerous object datasets collected from drones are very limited. For example, gun detection datasets (Delong et al., 2021) collected gun images from internet movie databases aiming for embedded applications. Closed-circuit television (CCTV) footage is also utilized for weapon detection (Hnoohom et al., 2021), in which various deep-learning modules are evaluated. VisDrone-DET2021 releases a dataset aiming at drone object detection (Delong et al., 2021), which contains videos collected from drones over various urban areas with vehicles and pedestrians and applies deep learning models.

These public datasets are helpful but need to reflect the realistic DFR scenarios in the context of law enforcement. DFRs are typically equipped with low-cost cameras and operate from a distance for public safety. In addition, other factors, such as low-light or night-time conditions, affect the detection performance. This study collects new datasets from a realistic DFR scenario under low-light conditions to fill these gaps and evaluate AI models for detecting dangerous objects. The key contributions of this work are:

- Collection of new dangerous objects dataset (firearms, knives) from a realistic DFR scenario.
- Evaluation of four different AI models (“You Only Look Once” version 8 [YOLOv8], small “You Only Look Once” version 5 [YOLOv5s], faster region-based convolutional neural network [Faster R-CNN], and a visual geometry group 19 layers deep [VGG19]).
- Development of an online application that can process uploaded videos to detect firearms and knives, aiming to support law enforcement agencies in the future in automating the evaluation of videos in security.

This is the first study on dangerous object detection from a DFR in the security field.

## RELATED WORK

Lightweight convolutional neural network architecture for efficient aerial image classification on unmanned aerial vehicles (UAVs) is proposed by Kyrkou and Theocharides (2020), in which EmergencyNet allows the network to process multiresolution features without increasing the number of parameters. A “You Only Look Once” version 5 (YOLOv5)-based algorithm is also proposed by Zhang et al. (2023), which is tailored for small object detection in UAV images, with potential applications in tasks like traffic monitoring, search and rescue, and environmental surveillance. This algorithm enhances the original YOLOv5 architecture, incorporating a space-to-depth convolution module, various attention mechanisms, and an improved multiscale detection module. Sun et al. (2022) introduce a novel network UA-CMDet (University at Albany Detection and tracking), for detecting vehicles in aerial images by harnessing the capabilities of both RGB (red, green, and blue) and IR (infrared) cameras, leveraging uncertainty-aware learning to extract complementary insights from these two imaging modalities. TPH-YOLOv5 (transformer prediction head) is an enhanced version of the YOLOv5 object detection model introduced by Zhu et al. (2021b), which explicitly addresses two critical challenges: the variations in object scale and the presence of motion blur in images. This improvement is achieved through the integration of a transformer-based prediction head. By employing this transformer prediction head, the model gains the ability to capture long-range dependencies within the feature map, thus enhancing its object detection capabilities for objects of different scales and those affected by motion blur.

Faster R-CNN and RetinaNet are also studied to identify different waterbird species from aerial images by Kabra et al. (2022). Yang et al. (2022) introduced object detection in aerial drone images by fusing the RGB and IR imagery, in which RGB images offer excellent visibility in the daytime. In contrast, IR images prove effective in low-light or obstructed conditions. A deep learning approach for object detection and scene understanding in adverse weather conditions is investigated by Sharma et al. (2022) by customizing the YOLOv5 and synthetic dataset containing images of vehicles and pedestrians in challenging weather scenarios like rain, snow, and fog. Jung et al. (2022) enhanced the performance of the original YOLOv5 model by different performance metrics, including precision, recall, F-1 score, and mean average precision at 50% (mAP50), demonstrating improved performance in the loss function and mAP50. Weapon detection using CCTV or surveillance cameras has been investigated (Delong et al., 2021; González et al., 2020; Hnoohom et al., 2022; Narejo et al., 2021). Narejo et al. (2021) introduced and assessed weapon detection systems integrated into an intelligent surveillance environment. They leveraged the “You Only Look Once” version 3 (YOLOv3) model, a deep learning architecture known for real-time object detection, as the central algorithm, emphasizing its suitability for real-time monitoring in intelligent surveillance systems due to its speed and efficiency. Real-time gun detection from CCTV using deep learning models has been studied, showing promising performance (Hnoohom et al., 2022). This comprehensive literature review in the dangerous object detection field suggests a heavy reliance on general object benchmark datasets. It needs dedicated datasets for weapons captured by drones in a law enforcement context. The review also revealed several limitations in utilizing low-quality videos under low illumination conditions, warranting further investigation to enhance the effectiveness of these systems.

Table 1 summarizes the related work, focusing on the algorithmic model, the dataset, evaluation metrics, the accuracy achieved, the processing speed measured in frames per second, and the limitations.

Table 1. Comparison of recent related work on weapon detection

| Paper/ Year | Model | Dataset | Evaluation method | Accuracy | Speed (Frames-Per-Second) | Limitations |
|---|---|---|---|---|---|---|
| (Kyrkou & Theocharides, 2020) | Emergency net convolutional neural network Algorithm | The aerial image database for emergency response | N/A | 95.7% | 20 times faster and suitable for affordable, low-energy devices. | Need to improve performance |
| (Zhang et al., 2023) | YOLOv5s-based, with several improvements to the original architecture | VisDrone-DET2019 dataset | Precision, recall, and mAP | mAP of 41.8% on the visdrone-DET2019 dataset, which is 7.8% better than the baseline YOLOv5 model | 54 FPS | UAV images can be oriented in various directions, and excessive background information may be present within the horizontal object frame, potentially causing objects to be missed during detection |
| (Sun et al., 2022) | UA-CMDet | Drone vehicle dataset and data collected by authors | (mAP) | mAP of 64.01% | N/A | Improve detection accuracy, sensitive to occlusion and camouflage |
| (Zhu et al., 2021b) | TPH-YOLOv5 | Vis drone 2021 test-challenge dataset | mAP | 39.18% on the vis drone dataset | N/A | Need to improve performance |
| (Kabra et al., 2022) | Faster R-CNN RetinaNet | Over 10,000 aerial images of water birds | mAP | 67.9% for Faster R-CNN and 63.1% for RetinaNet | N/A | Need to improve performance |
| (Yang et al, 2022) | RGB/IR fusion and deep learning object detection, YOLOv4 | Simulated and real RGB/IR image pairs, Vis drone dataset | mAP | Different accuracy between different aata sets | 28ms per image pair | Superior performance during daytime operations compared to nighttime scenarios. |

*continued on following page*

**Table 1. Continued**

| Paper/ Year | Model | Dataset | Evaluation method | Accuracy | Speed (Frames-Per-Second) | Limitations |
|---|---|---|---|---|---|---|
| (González et al., 2020) | Faster R-CNN | UGR- handgun dataset designed for region proposals (Split1), US - mock attack camera 1 and 7, edgecase - synthetic gun detection data, US - unity synthetic dataset, UGR - handgun test set, and UGR - handgun dataset for the region proposals approach (Split2) are included. | FPN AP | Different accuracy between different datasets | N/A | The performance gets better when actual images are combined with digital images, and the outcomes are enhanced when the object is positioned closer to the camera. |
| (Jiang et al., 2022) | MIDD, CSRNet, ECFFNet CGFNet SwinNet and CAVER | VT821 VT1000 and VT500, RGB-T SOD dataset VT723 | Precision recall | Different accuracy between different datasets | N/A | Weather fluctuations affect the outcome of the object's discovery |
| (Zhu et al., 2021a) | Faster R-CNN, cascade R-CNN, YOLOv3, RefineDet and CenterNet | ImageNet video detection, UA-DETRAC detection, MOT17Det, Okutama-action, UAVDT-DET, and DroneSURF datasets are all included | AP scores | 14.44% | N/A | Blur movement in videos, discontinue recording |
| (Sharma et al., 2022) | Modified YOLOv5 | Rob flow self-driving car dataset | mAP | 72.3% | N/A | Need to improve performance |
| (Jung et al., 2022) | YOLOv5_Ours modified | VisDrone dataset | mAP | 95.5% | Real-time | Need to improve performance |
| (Narejo et al., 2021) | YOLOv3 | Custom dataset of weapon images | mAP | 98.89% | Real-time | Limited dataset |

*continued on following page*

**Table 1. Continued**

| Paper/ Year | Model | Dataset | Evaluation method | Accuracy | Speed (Frames-Per-Second) | Limitations |
|---|---|---|---|---|---|---|
| (Kaya et al., 2021) | VGGNet architecture | dataset is created, encompassing seven distinct weapon categories. | mAP | 98.40% | N/A | Not real time |
| (Krišto et al., 2020) | Faster R-CNN, SSD, Cascade R-CNN, and YOLOv3 | Coco dataset in different conditions | AP | YOLOv3 97.93%<br>Faster R-CNN 98.86% | YOLOv3- 0.036<br>Faster R-CNN 0.141 | Small dataset |
| (Bhatti et al., 2021) | VGG16, Inception-V3, Inception-ResnetV2, SSDMobileNetV1, Faster-RCNN Inception-ResnetV2 (FRIRv2), YOLOv3, and YOLOv4 models are included | Collected and constructed in different phases and imfdb database | mAP precision | YOLOv4 -91.73%<br>YOLOv4 -93% | N/A | |
| (Song et al., 2020) | SOD-Model | RGBN-SOD | Precision | Supervised 0.8502<br>unsupervised 0.8266 | Supervised 0.8940<br>unsupervised 0.8281 | Did not define the algorithms used |
| (Ruprah & Shrivastav, 2023) | Faster R-CNN with media pipe | Manually annotated dataset of images and videos containing people and weapons | Accuracy, recall, precision, and F1-score | The Faster R-CNN model achieves accuracy rates of 97% for guns, 93% for knives, and 90% for person detection | N/A | Lack of annotated datasets |

*continued on following page*

**Table 1. Continued**

| Paper/ Year | Model | Dataset | Evaluation method | Accuracy | Speed (Frames-Per-Second) | Limitations |
|---|---|---|---|---|---|---|
| (Hashi et al., 2023) | VGG19, ResNet50, GoogleNet, YOLOv6 | Open images dataset V6. | mAP | ResNet50 achieved the highest mean accuracy of 0.92%, with VGG19 following closely at 0.91%, and GoogleNet at 0.89%. | N/A | The model requires a large amount of training data |
| (Mohiuddin et al., 2024) | YOLOv5 YOLOv6, YOLOv8, Quicker RCNN | The images were collected from a variety of sources, including online databases and security cameras | mAP, F1-Score | mAP of 91.73 percent and F1-score of 91% | Real-time camera | Needs to improvement performance |

Note. mAP50 = mean average precision at 50%; YOLOvX = You only look once version X; UA-CMDet = Uncertainty-aware cross-modality vehicle detection; TPH-YOLOv5 = Transformer prediction head YOLOv5; Faster R-CNN = faster region-based convolutional neural network; RGB/IR = Red, green and brown/infrared; UGR = University of Granada; VGG19 = Visual geometry group 19; SOD = Salient object detection; SSD = Single shot detector; MIDD = Multi-interactive dual-decoder; CSRNet = Context-guided stacked refinement network; ECFFNet = Effective and consistent feature fusion network; CGFNet = Cross-guided fusion network; SwinNet = Swin transformer network; CAVER = Cross-modal view-mixed transformer; RGB-T = Red, green, blue and thermal; CNN = Convolutional neural network; UA-DETRAC = University at Albany Detection and tracking; MOT17Det = Multiple object tracking 2017 dataset; UAVDT-DET = Unmanned aerial vehicle detection and tracking dataset

## EVALUATION OF DEEP LEARNING MODELS FOR DANGEROUS OBJECT DETECTION

This section outlines the procedures for implementing a dangerous object detection system utilizing advanced deep learning models. It consists of collecting drone data, processing and annotating it, training deep learning models, and evaluating model performance. The trained network model is implemented as an online application for further real-time detection.

Four videos from a drone are collected in a mockup DFR scenario with six different model weapons (guns, pistols, a knife), as shown in Figure 1 and Figure 2.

**Figure 1. Samples of color (left) and infrared (IR) video (right) with different weapons (guns, pistols, and a knife)**

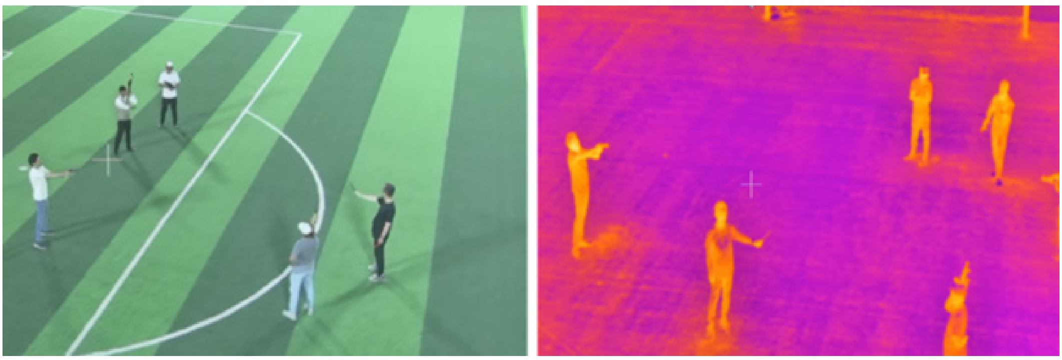


**Figure 2. A quadrotor drone equipped with a color/infrared (IR) camera for data collection**

Two videos are from a color camera, and two are captured using an IR camera, which is used for training and validation, as summarized in Table 2.

**Table 2. Description of data collected**

| # | Type | Altitude (m) | Zoom Level | Camera Resolution | Size | Duration (MM:SS) | Frames |
|---|---|---|---|---|---|---|---|
| 1 | Color | 20m | No zoom | 1920x1080 | 286 MB | 7:34 | 2900 |
| 2 | IR | 20m | No zoom | 640x480 | 237 MB | 4:29 | 1190 |

*Note*. IR = infrared.

This dataset was collected at nighttime with illumination, using a drone equipped with a color/IR camera, recording 20 meters above ground level and without zoom. A DFR drone was flown around the scene, following a circular trajectory to detect the objects from varying perspective angles.

## Data Collection and Annotation

The following process is to extract image frames from the videos. The authors defined a sampling frequency of 15 frames per second to obtain clear, discernible frames where the objects (weapons) are distinctly visible. Out of 2,900 daytime video frames, only 1,190 were approved for training. As for the night video, 2,346 out of 2,690 frames were approved. The reason is due to the presence of unclear objects in frames from the extracted videos, as some frames suffered from the motion blur of the camera or a lack of clarity in the image. The datasets are split into two sets, one for training (80% of frames) and the other for validation (20%).

For the annotation, three object classes are defined and labeled; the knife is described as class 0, the pistol as class 1, and the gun as class 2. Accurate annotation is crucial, as it directly affects the performance of the AI models. Figure 3 and Figure 4 show some annotated objects from the color and IR images, showing that some of the weapons are vaguely visible and indistinguishable in images due to the distance to the objects coupled with the low image resolution and low illumination, showing that reliable object detection is quite challenging.

Figure 3. Sample labeled gun, pistol, and knife from the color images showing many objects in low-resolution

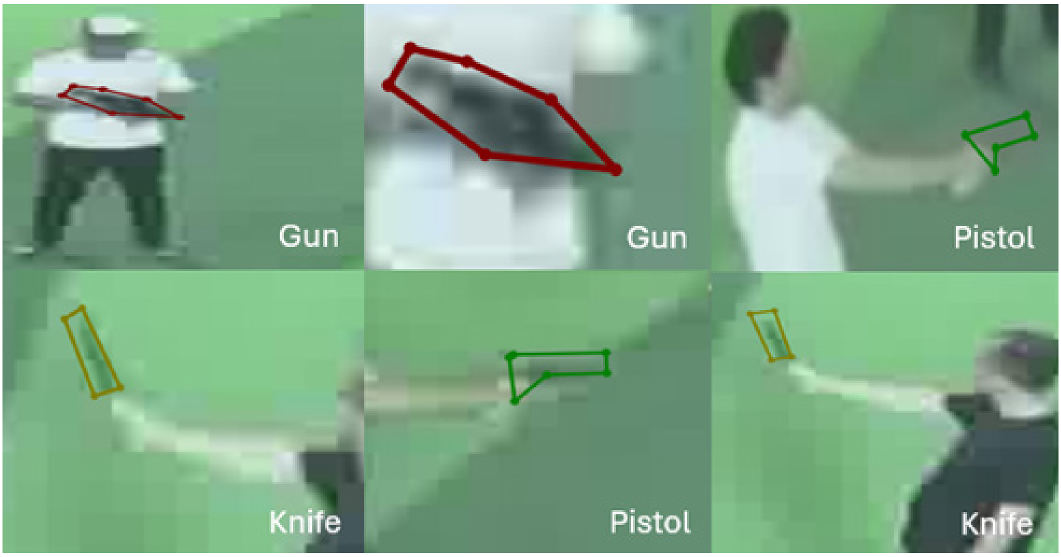
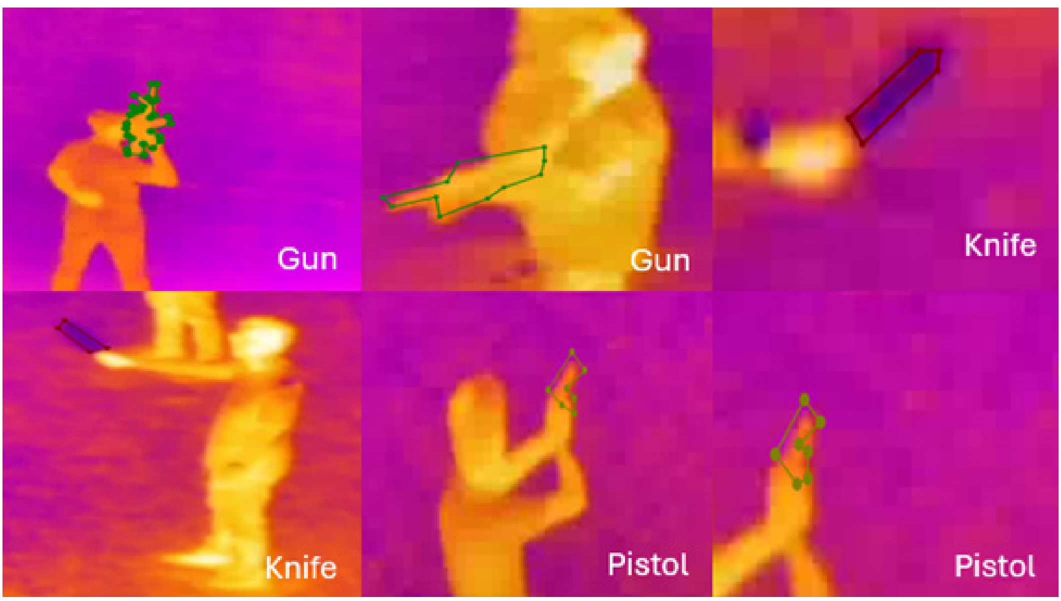



Figure 4. The sample labeled gun and knife from infrared (ir) images showed many low-resolution objects




## Selection of Deep Learning Models and Performance Metrics

Four popular deep-learning models, YOLOv8, YOLOv5s, Faster R-CNN, and VGG19 are selected to evaluate the performance of the drone dataset. The annotated datasets train the models while tuning the hyperparameters to obtain the best accuracy. The rationales for selecting these four models are discussed below.

YOLOv8 is a convolutional neural network model known for its ability to balance speed and accuracy. YOLOv8 has many features that make it a powerful and efficient model, incorporating a path aggregation network, which enhances feature aggregation, and a feature pyramid network layer for multi-scale object detection, thus making it an excellent candidate for the weapon detection problem. Although YOLOv5s is an older model than YOLOv8, it shows a unique balance of accuracy and speed, which is vital for real-time weapon detection applications. YOLOv5s is a miniature version of YOLOv5 regarding the number of network layers and parameters. The YOLO frameworks analyze the input image at once to predict the locations and classes of objects. This speeds up the detection process and makes it ideal for situations that require quick reactions. During the model training phase, YOLOv5s model parameters are periodically adjusted to fit the data used, enabling it to learn accurate predictions of object locations and classify them efficiently.

Faster R-CNN is also evaluated as the model can detect and locate objects accurately. It adopts the region proposal network, which scans the entire image to find regions that may contain objects using anchors, then takes regions of interest, classifying each region into a specific class and adjusting the bounding box size to fit the detected object precisely. The key benefit of Faster R-CNN is detecting multiple types of objects with extreme accuracy. The model improves detection speed over previous models by reducing the need to recalculate features for each region proposal. VGG19 is also one of the popular models used in image analysis. The model contains 19 weight layers, which include convolutional layers and fully connected neural network layers.

Performance metrics are required to measure the effectiveness of each model in detecting objects, such as precision, recall, mAP, and intersection over union scores. Precision measures the true positives out of all detections, which consists of the true positives and false positives (also known as false alarms) by the detector. It is helpful if the false alarms need to be reduced from the detector. Recall (or sensitivity) is another measure of true positives out of all actual cases, which consists of the true positives and false negatives (also known as missed detections). If the missed detections are crucial to the systems, for example, missed weapons whose consequences can be catastrophic, the recall metric is more relevant. To balance the precision and recall, mAP is typically used, the average precision value of different possible points for each class based on changing the recall value threshold.

The mAP metric provides a comprehensive measure of model performance across all classes and is the key criterion in evaluating object detection performance nowadays. Intersection over union also measures the object detection performance by comparing a predicted bounding box and a true bounding box. That is, if the detector has a larger intersection area, the accuracy is better. mAP50 calculates the mAP value with an intersection over the union threshold of 50%, and this metric is used in this work.

The selected four AI models are evaluated using the prepared dataset. A graphics processing unit laptop computer is used, which has an Intel(R) Core (TM) i9-13980HX processor with 2.20 GHz. The device has 32.0 GB of random-access memory and an NVIDIA GeForce RTX 4090 graphics processing unit card. The datasets are split into two sets, one for training (80%) and the other for validation (20%). During this process, the hyperparameters of each model are systematically tuned to explore the maximum potential of each model and determine the best parameter configuration.

## EVALUATION OF DANGEROUS OBJECT DETECTION

Table 3 summarizes the mAP50 performance of AI models using the color and IR datasets, which excludes the results from the VGG19 model as it performed poorly.

**Table 3. Comparison table for mean average precision at 50% (mAP50) based on "you only look once" version 8 (YOLOv8) experiments for nighttime video frames**

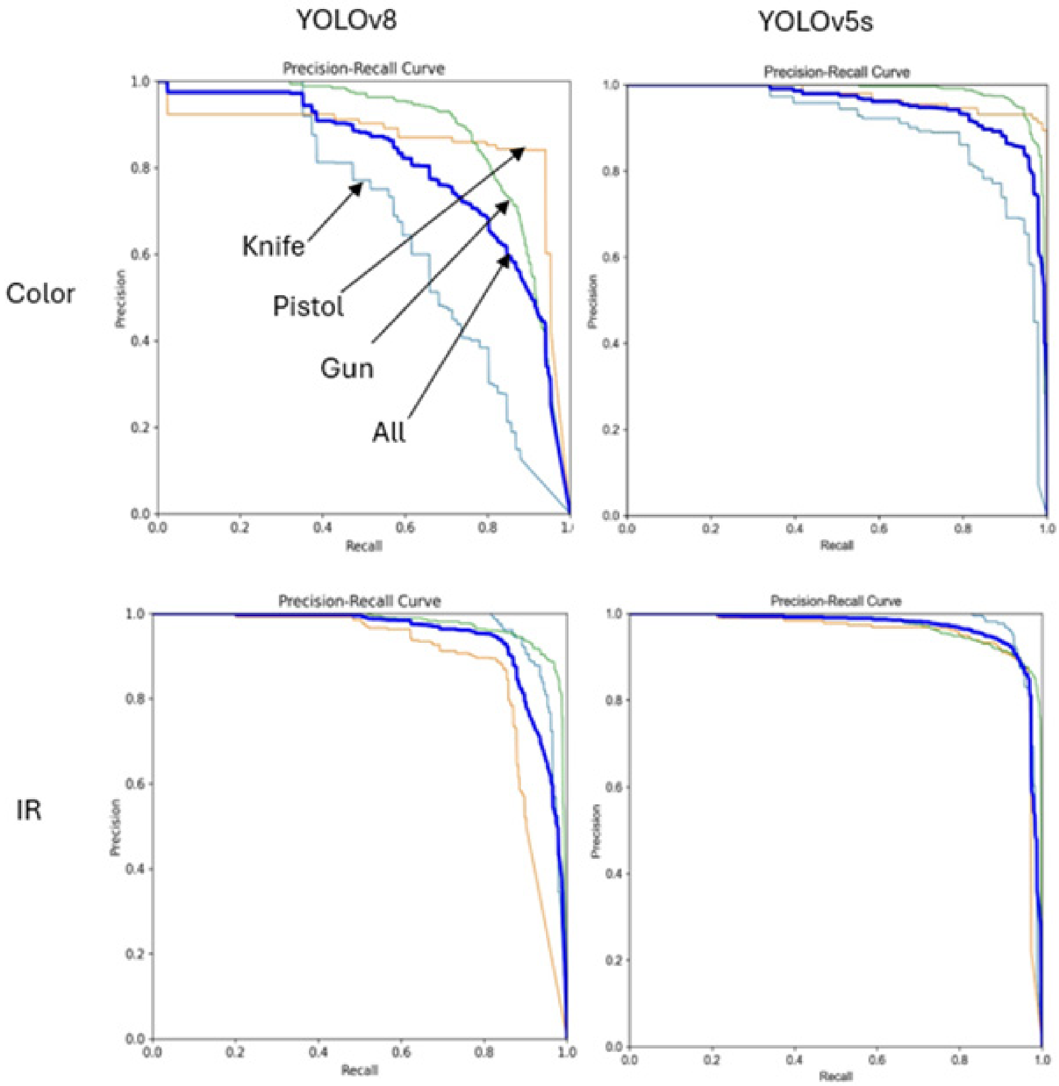
| Model | Epochs | Epochs/ Batch Size | Time (hours) | mAP50 | | | |
|---|---|---|---|---|---|---|---|
| | | | | Knife | Pistol | Gun | All Class |
| **YOLOv8** | Color | 300/32 | 0.459 | 0.669 | 0.884 | 0.894 | 0.816 |
| | IR | 300/32 | 0.842 | 0.963 | 0.884 | 0.973 | 0.940 |
| **YOLOv5s** | Color | 300/32 | 0.520 | 0.897 | 0.969 | 0.982 | 0.949 |
| | IR | 300/32 | 1.978 | 0.974 | 0.949 | 0.969 | 0.964 |
| **Faster R-CNN** | Color | 100/4 | 6.250 | 0.87 | 0.45 | 0.73 | 0.68 |
| | IR | 100/4 | 0.77 | 0.87 | 0.73 | 0.73 | 0.75 |

Note. *IR = infrared; mAP50 = mean average precision at 50%; YOLOv8 = "You Only Look Once" Version 8; YOLOv5s = small "You Only Look Once" version 5; Faster R-CNN = faster region-based convolutional neural network.*

Various tuning numbers of epochs and batch sizes were tested to fine-tune the model: Epoches = (100, 150, 300), and batch size = (15, 32, 64). For YOLOv8 and YOLOv5s, epochs of 300 and a batch size of 32 showed the best performance in terms of accuracy and training speed, while Faster R-CNN used epochs of 100 and a batch size of four. It can seen from Table 3 that YOLOv5s showed the best detection performance, achieving 94.9% and 96.4% for all classes in color and for the IR dataset, respectively.

The Faster R-CNN model performed poorly, as can be seen from Table 3, which was tuned with epochs of 100 and a batch size of four, as well as the stochastic gradient descent algorithm with a learning rate of 0.01, momentum value of 0.9, and weight decay of 0.0005 to ensure stable convergence. The detection performance achieved, however, was 68% and 75% for all classes in color and IR images, respectively. In YOLOv5s, the gun class showed the highest detection score in the color dataset, while the knife class had the highest score in the IR dataset. This is due to the distinctive shape feature of the gun class in color and the metallic contrast of the knife class in IR images. Figure 5 shows the precision-recall curve results over the varying detection threshold values for the three AI models.

**Figure 5. Precision-recall curves for "you only look once" version 8 (YOLOv8) and small "you only look once" version 5 (YOLOv5s) and color (top) and infrared (IR) data (bottom)**




*Note. YOLOv8 = "You Only Look Once" version 8; YOLOv5s = small "You Only Look Once" version 5; IR = infrared. This figure shows that YOLOv5 performed best, accurately detecting the gun and pistol accurately in the infrared (IR) dataset. Faster region-based convolutional neural network (Faster R-CNN) and visual geometry group (VGG) tesults are not shown, as they performed poorly.*

If the curve is closer to the (recall, precision) = (1, 1) point, it detects better, minimizing the false alarms and missed detections. It can be observed that YOLOv5s performs better than YOLOv8, detecting the gun/pistol well in color and the knife in the IR dataset.

From this comparative performance analysis of four different AI models, the YOLOv5s model shows the best weapon detection, with a mAP50 of 0.964 compared to 0.940 from YOLOv8. This result indicates that the YOLOv5s more efficiently handles the low-light conditions in weapons detection. YOLOv8 contains much larger network parameters than YOLOv5s, which requires extensive data for training. If the training size is moderate or small, as in this scenario, YOLOv5s can perform better. Faster R-CNN delivers consistent results for the weapon category with a receiver operating charasteristic (ROC) close to 1.00, demonstrating its potential capability to distinguish between true and false positives. Still, the mAP is low, with 0.77 for all classes. YOLOv5s achieved the highest mAP of 0.974 for the knife class. This reflects the efficiency of models in the detection knife class at night, and the reason is that the thermal camera distinguished the knife because it was made of metal, which makes it easier for the models to detect with high accuracy. Also, YOLOv5s proved effective in detecting pistols with a mAP of 0.972 and a call rate of up to 0.988. This is because, during the labeling process, the pistol was visible in many frames, leading to accuracy in the annotation, which made the accuracy in detecting it high.

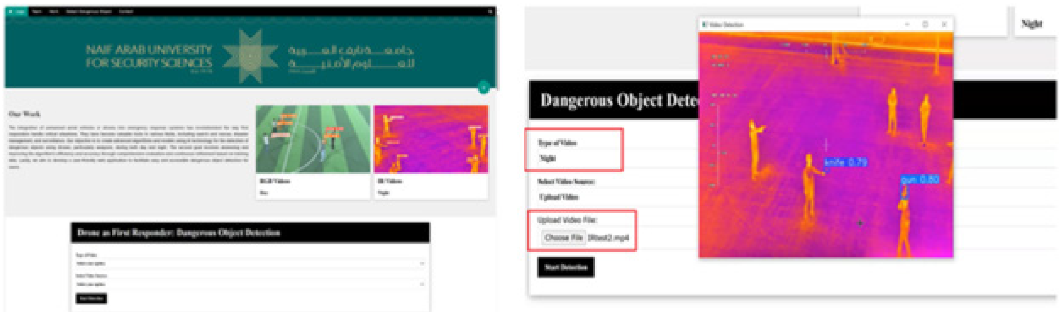
YOLOv8 needed more data for training than what was used for the experiments, as approximately 2,364 frames were used for night data and 1,190 for daytime data, and this is considered a small amount of training dataset for this model, which explains why YOLOv5s obtained higher accuracy results. Regarding the VGG19 model, the low results can be interpreted for two reasons. The first and main reason could be that the model is primarily designed for classification, not for detection, and the second reason could be that the data used in experiments is of low quality, affecting the model's performance. This reason also explains the low performance of Faster R-CNN in detection. The YOLOv5s model is then chosen to predict weapons on videos used for validation, and some of the detection results are shown in Figure 6, demonstrating its ability to detect guns, pistols, and knives from low-resolution images.

**Figure 6. Small "you only look once" version 5 (YOLOv5s) prediction results for color video (top) and infrared video (bottom)**


## Web Application Development

An easy-to-use web application was developed in which law enforcement agency users can upload videos extracted from the drone and perform the necessary checks to determine whether they contain any dangerous weapons. This web application was linked to the model that obtained the highest accuracy, the YOLOv5s. Through this web page, the user can choose the type of video they want to upload, color or IR, as shown in Figure 7. Once a type is chosen, it is linked to the required model, and then the user uploads the video for weapon detection.

**Figure 7. Detection videos using a web application**




## DISCUSSION

The evaluation results show that YOLOv5s performs best in detecting each class (gun, pistol, knife) using the mAP50 metric. It is somewhat surprising that the YOLOv8 has more advanced network features than the YOLOv5s model, and Faster R-CNN also showed reliable performance in the handgun dataset. The key reason for the underperformance of YOLOv8 and Faster R-CNN is that they have a more significant number of network parameters, requiring more datasets for training. Also, the firearm objects seen from the drone camera are very small and typically occluded by hands with low illumination, leading to a low signal-to-noise ratio. This suggests that more drone-based datasets are required to make the model more robust, and a customized network, utilizing small-scale networks, can be more suitable for the firearm detection problem. Collecting more drone-based datasets is being investigated from different operating conditions as an ongoing work. This includes the data augmentation techniques, utilizing the generative adversarial network to synthesize the gun images. Detecting real firearms from fake or model guns is also a necessary capability for DFRs, as it can significantly de-escalate the situation if model guns are present rather than real ones. Another critical aspect of drone-based real-time surveillance is privacy and data security. The International Criminal Police Organization (2023) has recently proposed the Toolkit for Responsible AI Innovation in Law Enforcement (AI Toolkit), which embeds ethical principles in designing, developing, and deploying AI applications. Drone-based surveillance systems are developed following this framework due to their potentially intrusive nature. Data encryption and incorporating hashing algorithms, such as the message-digest algorithm, should be incorporated in application programming interface (API) development to address data security.

## CONCLUSIONS AND FUTURE WORK

The authors evaluated the performance of four popular AI deep learning models (YOLOv8, YOLOv5s, Faster R-CNN, VGG19) to automate the detection of dangerous objects recorded from a first responder drone equipped with a low-quality color and IR camera. The results showed that the YOLOv5s model achieves the best detection performance, yielding mAP50 results of 0.964 for color and 0.949 for IR videos, which are excellent performances considering the low-quality and low-resolution datasets. This result suggests that YOLOv5s is quite feasible for detecting dangerous objects from DFRs, even using low-quality videos and under realistic outdoor conditions. An online web application was developed using the trained YOLOv5s model to assist law enforcement agencies. The color and thermal images can be further fused to enhance the AI detection performance, which is the direction of future research.

## COMPETING INTERESTS

No conflicts of interest.

## FUNDING

This work was supported by the Security Research Center at Naif Arab University for Security Sciences (Project No. NAUSS-23-R14). The authors would like to express thanks to the Vice Presidency for Scientific Research at Naif Arab University for Security Sciences for their support on the work.

## PROCESS DATES

December 27, 2024
Received: September 4, 2024, Revision: December 2, 2024, Accepted: December 15, 2024

## CORRESPONDING AUTHOR

Correspondence should be addressed to Jong Hyuk Kim (Saudi Arabia, jkim@nauss.edu.sa)

## REFERENCES

Bhatti, M. T., Khan, M. G., Aslam, M., & Fiaz, M. J. (2021). Weapon detection in real-time CCTV videos using deep learning. *IEEE Access: Practical Innovations, Open Solutions*, *9*, 34366–34382. DOI: 10.1109/ACCESS.2021.3059170

Chula Vista Police Department. (2024). Drone program. https://www.chulavistaca.gov/departments/police-department/programs/uas-drone-program

González, J. L. S., Zaccaro, C., Álvarez-García, J. A., Morillo, L. M. S., & Caparrini, F. S. (2020). Real-time gun detection in CCTV: An open problem. *Neural Networks*, *132*, 297–308. DOI: 10.1016/j.neunet.2020.09.013 PMID: 32977275

Hashi, A. O., Abdirahman, A. A., Elmi, M. A., & Rodriguez, O. E. R. (2023). Deep learning models for crime intention detection using object detection. *International Journal of Advanced Computer Science and Applications*, *14*(4), 300–306. DOI: 10.14569/IJACSA.2023.0140434

Hnoohom, N., Chotivatunyu, P., & Jitpattanakul, A. (2022). ACF: An armed CCTV footage dataset for enhancing weapon detection. *Sensors (Basel)*, *22*(19), 7158. DOI: 10.3390/s22197158 PMID: 36236253

International Criminal Police Organization. (2023). Toolkit for responsible AI innovation in law enforcement (AI toolkit). https://www.interpol.int

Jiang, X., Zhu, L., Hou, Y., & Tian, H. (2022). Mirror complementary transformer network for RGB-thermal salient object detection. *arXiv preprint,* arXiv:2207.03558.

Jung, H. K., & Choi, G. S. (2022). Improved YOLOv5: Efficient object detection using drone images under various conditions. *Applied Sciences (Basel, Switzerland)*, *12*(14), 7255. DOI: 10.3390/app12147255

Kabra, K., Xiong, A., Li, W., Luo, M., Lu, W., Yu, T., & Barman, A. (2022). Deep object detection for waterbird monitoring using aerial imagery. In *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE. DOI: 10.1109/ICMLA55696.2022.00073

Kaya, V., Tuncer, S., & Baran, A. (2021). Detection and classification of different weapon types using deep learning. *Applied Sciences (Basel, Switzerland)*, *11*(16), 7535. DOI: 10.3390/app11167535

Krišto, M., Ivasic-Kos, M., & Pobar, M. (2020). Thermal object detection in difficult weather conditions using YOLO. *IEEE Access: Practical Innovations, Open Solutions*, *8*, 125459–125476. DOI: 10.1109/ACCESS.2020.3007481

Kyrkou, C., & Theocharides, T. (2020). Emergencynet: Efficient aerial image classification for drone-based emergency monitoring using atrous convolutional feature fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 1687–1699. DOI: 10.1109/JSTARS.2020.2969809

Mohiuddin, M. A., Laxmikanth, M., Vardhani, G. H., Hari, B. D., Rani, K. S., Mouli, K. C., & Dhyani, S. (2024, July). Weapon detection in real-time CCTV videos using deep learning. In *AIP Conference Proceedings,* 3101(1). AIP Publishing. DOI: 10.1063/5.0222303

Narejo, S., Pandey, B., Esenarro Vargas, D., Rodriguez, C., & Anjum, M. R. (2021). Weapon detection using YOLO V3 for smart surveillance system. *Mathematical Problems in Engineering*, *2021*, 1–9. DOI: 10.1155/2021/9975700

Qi, D., Tan, W., Liu, Z., Yao, Q., & Liu, J. (2021). A gun detection dataset and searching for embedded device solutions. *arXiv preprint,* arXiv:2105.01058.

Ruprah, T., & Shrivastav, H. (2023). Crime prediction based on person-weapons relation using deep learning techniques. Preprint. DOI: 10.21203/rs.3.rs-2743470/v1

Sharma, T., Debaque, B., Duclos, N., Chehri, A., Kinder, B., & Fortier, P. (2022). Deep learning-based object detection and scene perception under bad weather conditions. *Electronics (Basel)*, *11*(4), 563. DOI: 10.3390/electronics11040563

Song, S., Miao, Z., Yu, H., Fang, J., Zheng, K., Ma, C., & Wang, S. (2020). Deep domain adaptation-based multi-spectral salient object detection. *IEEE Transactions on Multimedia*, *24*, 128–140. DOI: 10.1109/TMM.2020.3046868

Soori, M., Arezoo, B., & Dastres, R. (2023). Artificial intelligence, machine learning and deep learning in advanced robotics: A review. *Cognitive Robotics*, *3*, 54–70. DOI: 10.1016/j.cogr.2023.04.001

Sun, Y., Cao, B., Zhu, P., & Hu, Q. (2022). Drone-based RGB-infrared cross-modality vehicle detection via uncertainty-aware learning. *IEEE Transactions on Circuits and Systems for Video Technology*, *32*(10), 6700–6713. DOI: 10.1109/TCSVT.2022.3168279

Yang, L., Ma, R., & Zakhor, A. (2022). Drone object detection using RGB/IR fusion. *arXiv preprint,* arXiv:2201.03786.

Zhang, J., Wan, G., Jiang, M., Lu, G., Tao, X., & Huang, Z. (2023). Small object detection in UAV image based on improved YOLOv5. *Systems Science & Control Engineering*, *11*(1), 2247082. DOI: 10.1080/21642583.2023.2247082

Zhu, P., Wen, L., Du, D., Bian, X., Fan, H., Hu, Q., & Ling, H. (2021a). Detection and tracking meet drone challenges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(11), 7380–7399. DOI: 10.1109/TPAMI.2021.3119563 PMID: 34648430

Zhu, X., Lyu, S., Wang, X., & Zhao, Q. (2021b). TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. IEEE. DOI: 10.1109/ICCVW54120.2021.00312

*Raha Almarashi is an artificial intelligence researcher specializing in deep learning model development and deepfake detection. Her work focuses on advancing detection techniques to counter artificial intelligence-driven security threats. With a strong background in developing state-of-the-art models, Raha contributes to cutting-edge research in the field, publishing in renowned scientific journals.*

*Dr. Qazi Emad-Ul-Haq received his PhD in the field of artificial intelligence and machine learning. He did his postdoc in the field of cybersecurity and machine learning. He is currently working as an assistant professor at the Centre of Cybercrimes and Digital Forensics (CoECDF), Naif Arab University for Security Sciences (NAUSS), Riyadh, Saudi Arabia. He has more than 15 years of teaching and research experience. He has completed a number of international collaborative research projects with reputable universities. He holds one U.S. patent and has published many research articles in high-ranked peer-reviewed journals and well-recognized international conferences. His research interests include AI, machine learning, deep learning, security analytics, and cybercrime prevention.*

*Jonghyuk Kim is a Professor of Robotics and Autonomous Systems at the Naif Arab University of Security Science (NAUSS) in Saudi Arabia and has been an adjunct associate professor at the University of Technology Sydney. Dr. Kim has more than 20 years of experience in the navigation and control of autonomous vehicles, publishing over 120 research papers. Dr. Kim obtained his PhD in Field Robotics at the University of Sydney in 2004, Australia. He was a pioneer in the field of airborne simultaneous localization and mapping) for unmanned aerial vehicles (UAVs) operating in global navigation satellite system (GNSS) -denied environments. He received the prestigious Charles Sharpe Beecher Award and Prize from Institution of Mechanical Engineers United Kingdom for his contributions to decentralized data fusion for multiple UAVs. The Australian Academy of Technological Science and Engineering recognized him as an Emerging Researcher. He co-chaired the Australasian Conference in Robotics and Automation in 2009 and 2015 and has been the associate editor of IEEE/ICRA and IEEE/IROS numerous times.*