Applied Computer Science, vol. 20, no. 3, pp. 1–16
doi: 10.35784/acs-2024-25

Submitted: 2024-03-04 | Revised: 2024-06-12 | Accepted: 2024-06-18

Keywords: violence prediction system, YOLO v8, ontology, surveillance cameras, anomaly prediction

Esraa A. MAHAREEK [0000-0002-9042-248X]*, *Doaa R. FATHY* [0000-0002-5625-0282]*,
Eman K. ElSAYED [0000-0001-7870-927X]**,
Nahed M. ElDESOUKY [0009-0008-4547-3051]*,
Kamal A. ELDAHSHAN [0000-0002-9953-5480]*

VIOLENCE PREDICTION IN SURVEILLANCE VIDEOS

Abstract

Forecasting violence has become a critical obstacle in the field of video monitoring to guarantee public safety. Lately, YOLO (You Only Look Once) has become a popular and effective method for detecting weapons. However, identifying and forecasting violence remains a challenging endeavor. Additionally, the classification results had to be enhanced with semantic information. This study suggests a method for forecasting violent incidents by utilizing Yolov9 and ontology. The authors employed Yolov9 to identify and categorize weapons and individuals carrying them. Ontology is utilized for semantic prediction to assist in predicting violence. Semantic prediction happens through the application of a SPARQL query to the identified frame label. The authors developed a Threat Events Ontology (TEO) to gain semantic significance. The system was tested with a fresh dataset obtained from a variety of security cameras and websites. The VP Dataset comprises 8739 images categorized into 9 classes. The authors examined the outcomes of using Yolov9 in conjunction with ontology in comparison to using Yolov9 alone. The findings show that by combining Yolov9 with ontology, the violence prediction system's semantics and dependability are enhanced. The suggested system achieved a mean Average Precision (mAP) of 83.7 %, 88% for precision, and 76.4% for recall. However, the mAP of Yolov9 without TEO ontology achieved a score of 80.4%. It suggests that this method has a lot of potential for enhancing public safety. The authors finished all training and testing processes on Google Colab's GPU. That reduced the average duration by approximately 90.9%. The result of this work is a next level of object detectors that utilize ontology to improve the semantic significance for real-time end-to-end object detection.

[*] Al-Azhar University, Faculty of science, Mathematics Department, Egypt, esraa.mahareek@azhar.edu.eg, doaaelzalbany@azhar.edu.eg, nahedeldesouky5922@azhar.edu.eg, dahshan@gmail.com
[**] Canadian International College, Dean of School of Computer Science, Egypt, eman_k_elsayed@ciccairo.com

1. INTRODUCTION

In recent years, the need for accurate violence prediction system in video surveillance has grown significantly. A multitude of uses, including crime prevention, security monitoring, and public safety, depend on the ability to identify and stop violent situations in real time. The employment of sophisticated object identification and tracking systems has become essential to achieving this. In the meantime, Ultralytics presented a faster and more accurate version of the YOLOv5 algorithm. YOLOv5 based learning studies were performed out by (Han et al., 2023) to enhance the object detection accuracy of 32x32 pixels or less. Next, they examined the inference performance of the YOLOv5 model in simulating several violent situations with a specific dataset that included subjective enhancing, utilising traditional image processing techniques. It is clear from looking at Ultralytics' work on YOLOv5 that their algorithm has demonstrated encouraging gains in accuracy and speed over its predecessors.

This work uses the features of YOLOv8 (Solawetz, 2023) which has improved architecture, a real-time object detection system based on convolutional neural networks, to address the problems associated with violence prediction in surveillance videos.

The goal of integrating ontology with YOLOv8 in violence prediction is to enhance the reliability and semantics of identifying and categorising items connected to violent incidents in surveillance videos. The violence prediction system performs better when ontology is added because it provides a structured format to show how it understands and the relationships between various concepts related to violence in surveillance videos. This makes it easier to distinguish between events that represent threats from those that don't. Therefore, combining YOLOv8 with ontology, holds great potential in accurately predicting and preventing violent incidents in surveillance videos.

The primary objectives of this research paper are as follows:

- Develop a violence prediction system using YOLOv8 to enhance violence and armed people detection.
- Integrate ontology to provide contextual information and improve the accuracy of violence prediction.
- Evaluate the proposed system and compare its performance with existing approaches.
- All the training and testing were done on GPU with the support of the Google Colab environment.

The proposed system can detect 9 different classes of objects (armed man, body, face, hand, pistol, grenade, knife, uniform, and rifle). Then apply semantic prediction process to determine whether they represent a threat or not.

The suggested system was evaluated using a test dataset of surveillance films and found that it predicts violence with 83.7% mAP.

In the real-world, the system ability was also tested to identify violence events in surveillance footage. According to the study's findings, the suggested approach is useful for the public’s safety.

The rest of the paper is organized as follows: Section 2 discusses Literature review. Section 3 describes the proposed system in detail. Section 4 presents the results and discussion. Finally, Section 5 represents conclusion and future works.

2. LITERATURE REVIEW

Real-time object detection: Real-time object detection strives to quickly identify and pinpoint objects, a critical need for practical purposes. In recent years, there have been significant attempts to create effective detectors (Li et al., 2020; Lin et al., 2017; Redmon et al., 2016; Redmon & Farhadi, 2017; Tian et al., 2022; Zhang et al., 2019; Zheng et al., 2020). In particular, the YOLO series (Bochkovskiy et al., 2020; Li et al., 2023; Redmon et al., 2016; Redmon & Farhadi, 2017, 2018) (Wang et al., 2024) (Glenn, 2022) are notably mainstream. YOLOv1, YOLOv2, and YOLOv3 recognize the standard detection structure that includes three components: backbone, neck, and head. YOLOv4 (Bochkovskiy et al., 2020) and YOLOv5 (Glenn, 2022) incorporate the CSPNet (Wang et al., 2020) architecture in place of DarkNet (Redmon, 2016), along with data augmentation techniques, improved PAN, and a wider range of model scales. YOLOv6 introduces BiC and SimCSPSPPF for the neck and backbone, along with anchor-assisted training and self-distillation methods. YOLOv7 (Wang et al., 2022) presents E-ELAN for enhanced gradient flow path and investigates multiple trainable bag-of-freebies approaches. YOLOv8 introduces the C2f module for efficient feature extraction and integration. Gold-YOLO (Wang et al., 2023) offers an enhanced GD mechanism to increase the ability for multi-scale feature fusion. YOLOv9 (Wang et al., 2024) introduces GELAN to enhance the structure and PGI to enhance the training procedure.

Weapons detection: Khalid et al. (2023) suggested a methodology for quickly identifying handguns that may be applied to surveillance applications utilizing alarm-based systems. They made use of YOLOv5 that had incredibly fast and effective outcomes. Ashraf et al. (2022) used web scraping techniques to create a novel dataset called "DIAT-Weapon" Dataset for weapon object recognition. The 2712 photos in the DIAT-Weapon Dataset are primarily categorized into six categories: Camera, Handgun, Dagger, Sword, Rifles, and Sticks. The YOLOv4 models have been developed and fine-tuned to accurately recognize and position these six sorts of hazardous objects. Verma and Jayant (2022) used Convolutional Neural Networks (CNNs) to improve real-time object detection in surveillance system, and with updated hyperparameters, the best result was obtained on Overfeat-3, with 93% training accuracy and 89% test accuracy.

To improve the ability to predict violence, certain researchers have investigated combining YOLO-based object detection with ontological reasoning. A weapon identification model was proposed by Lai and Maples (2017) utilizing feature extraction. The CSBO algorithm is used to identify the important features, which are then provided to the PELSF-DCNN classifier together with the output of Yolov9 by Raman Dugyala et al. (2023). To determine the number of guns in the frame, the confidence score is finally computed. While Abdullah N. Arslan et al. (2012) provides a weapon ontology-based method for weapon identification from a single image. Expert-selected ontological nodes retain convex hull (CH) sequences for their offspring, while ontological leafs are designated with object boundary sequences. The CH sequences are formed from the CHs of the objects, whilst the latter are generated from the vertices of object boundaries. An active contour model extracts the object's boundary and CH. Cyclic sequence alignment, which offers a scaling and rotational invariant matching, is used in top-down ontology search.

For the goal of threat assessment, Christian F. Hempelmann et al. (Lou et al., 2023) investigated the use of conceptual ontology and visual hierarchy in relation to weapons. The threat assessment algorithm was refined, mostly to account for features connected to ammunition.

Restrictions of current methods: The studies that were evaluated emphasize the capability of YOLO models in detecting weapons, yet they mainly concentrate on detection rather than addressing the wider issue of predicting violence in surveillance videos. The combination of YOLO and ontological reasoning is promising, yet current methods are limited in focus and fall short of offering a complete predictive solution for real-life violence. Furthermore, there is limited assessment of the functionality of these systems in practical settings, which hinders their real-world utility.

The system proposed in this research seeks to overcome these challenges by creating a violence prediction system that merges YOLOv9's object detection capabilities with ontological reasoning to improve the meaning and precision of violence prediction in surveillance footage. The thorough assessment of the new system on a fresh dataset and in real-life situations is anticipated to offer useful perspectives on enhancing public safety by accurately predicting violence.

3. YOLO (YOU ONLY LOOK ONCE)

The field of object recognition and computer vision has been transformed by the YOLO technique. YOLO is a quick and effective method for identifying things in photos or real-time video shows. Yolo takes a single-pass technique to object detection, which makes it exceptionally fast, in contrast to typical methods that require numerous passes over an image. An image is divided into a grid by YOLO, which then forecasts bounding boxes and class probabilities for every grid cell. YOLO may detect several things in a single frame at the same time thanks to this technique. YOLO has found use in many different fields, such as robots, augmented reality, autonomous cars, and surveillance systems. It is an essential tool in computer vision applications that need prompt and accurate object recognition because of its speed, accuracy, and capacity for real-time object detection.

YOLO combines algorithmic and architectural improvements to deliver fast and accurate real-time object detection. First, YOLO uses a one-shot detection method in which it uses the full image to predict bounding boxes and class probabilities in a single pass. As a result, the laborious region proposal methods required by conventional object identification systems are no longer necessary. YOLO achieves real-time speed and greatly decreases processing cost by examining the entire image at once.

Second, YOLO creates a grid from the input image, and each grid cell's job is to identify things that are inside its bounds. With this grid-based method, YOLO can effectively detect objects of various sizes and scales. Furthermore, YOLO forecasts bounding boxes together with corresponding confidence ratings and class probabilities, facilitating precise object location and categorization.

Moreover, the core architecture of YOLO is a deep convolutional neural network (CNN). Robust object detection is made possible by this network's ability to learn rich hierarchical information from the input image. Large-scale datasets are used to train the network, which helps it generalize well to different item categories and environmental settings.

YOLO uses techniques like feature pyramid networks, which allow multi-scale object identification, and anchor boxes, which capture objects of various shapes and aspect ratios, to further improve accuracy. YOLO can more precisely handle objects with different sizes and orientations because of these strategies. YOLO's deep CNN architecture, effective single-shot detection method, grid-based object localization, and other optimization approaches all contribute to its speed and accuracy. Because of these features, YOLO is an effective tool for real-time object recognition in a variety of settings.

This project suggested Yolov9 (Wang et al., 2024) for two specific motives. Initially, YOLOv9 functions as a live object recognizer, allowing for fast identification of weapons in real-time situations by quickly analyzing video or image feeds. Additionally, YOLOv9 incorporates programmable gradient information (PGI) to ensure reliable gradient updates during training, leading to improved accuracy and robustness of the model. Furthermore, YOLOv9 makes use of a compact network design called Generalized Efficient Layer Aggregation Network (GELAN) to enhance parameter usage and computational effectiveness while maintaining detection accuracy. The YOLOv9's lightweight design enables its use on resource-constrained devices like surveillance cameras or embedded systems. In general, YOLOv9 presents the benefits of immediate operation, dependable training, and effective deployment, positioning it as a favorable option for weapon detection tasks.

The authors chose YOLOv9 as their baseline model because of its impressive balance between latency and accuracy, as well as its availability in different model sizes. The authors utilize semantic consumption within TEO for training yolov9 and develop a model design focused on efficiency and accuracy, resulting in their violence prediction model. They test the suggested detector on their VP and UCF-crime datasets using the identical training-from-scratch configuration (Songire et al., 2023; Wang et al., 2024). Additionally, the models' latencies were measured using a T4 GPU running TensorRT.

4. VIOLENCE PREDICTION SYSTEM

Figure 1 represents the steps for building the violence prediction system. The proposed system involved the following steps:

1. Dataset preparation: the authors collected 8739 images of various weapons and armed people from various websites. Then, labelling the dataset images: This process involves tagging each instance of a weapon in the images or videos with relevant metadata such as the type of weapon, its position, and its orientation. Finally, Pre-processing the dataset images: by using filtering and contrast enhancement.
2. Model training: the authors trained the YOLOv9 model using the collected dataset.
3. Creating the ontology: as a case study the authors created a part of TEO to represent the 9 classes, its individuals, and the relationships between them.
4. Integrating the model with ontology: The trained YOLOv9 model was integrated with ontology to enhance the semantic violence prediction of the classification results.

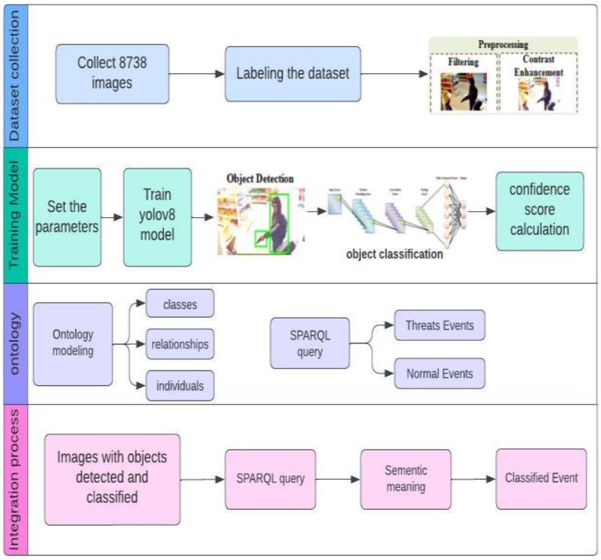


Fig. 1. Proposed system steps

4.1. Dataset preparation

Keywords such as "shooting," "handgun," and "knife" were used to search for images on Google, Pinterest, and the Open photos datasetv7 website. The authors also use text searches in various languages such as French, Russian, Chinese, and others to retrieve a maximum number of photos. There are a total of 8739 images in the VP Dataset. The dataset includes different situations, lighting conditions, and object angles to guarantee the versatility of the proposed system.

The dataset is partitioned into 9 categories (armed man, uniform, body, face, hand, pistol, grenade, knife, and rifle). The dataset is marked to generate accurate bounding box annotations for weapons. Every labeled frame includes the related class label and additional details like position, timestamp, and nearby objects demonstrated in figure 2. Afterward, the dataset is uploaded to the Roboflow website in the URL project (Mahareek, 2024) for filtering and improving contrast.

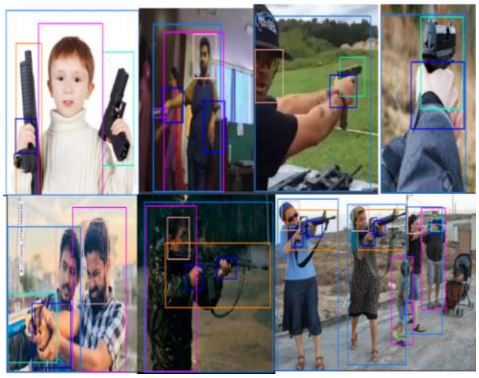


Fig. 2. Shows selection of photos from VP Dataset (Mahareek, 2024)

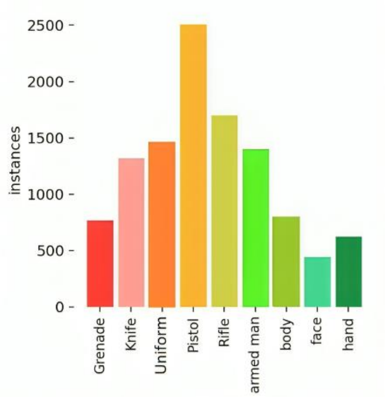


Fig. 3. Number of instances for each class in VP Dataset

The dataset is divided into training, validation, and testing sets. The size of training set is about 73% of the VP Dataset with 6351 images. The validation set is about 20% of the VD Dataset with 1776 images used for hyperparameter tuning and performance evaluation. While testing set is 7% of the dataset with 612 images. Figure 3 represents the number of instances for each class in VP Dataset.

4.2. Training process

For weapons and armed people detection and classification, the authors used the newest Yolov9 algorithm. Yolov9 is an efficient single shot detector that can be used for detection, segmentation, and classification tasks (Lou et al., 2023). It requires fewer parameters for training (Zhang et al., 2024). Yolov9 has greater mAPs and slower inference speed on the COCO dataset, it is thought to represent the new state of the art (Reis et al., 2023). It provides excellent real-time accuracy in object recognition and classification.

The code from the Ultralytics repository is implemented. Preprocessing the movies to extract individual frames is the initial step in applying Yolov9 to the VP Dataset. These frames are fed to the Yolov9 model, which detects and classifies objects. The resolution of the input image is 640x640, and other parameter settings were consistent with the default settings of Yolov9.

Training and testing were done on Google Colab's GPU to reduce the training time. The system was trained with a 64-batch size, a 0.00001 learning rate, and 100 epochs.

One of the most used evaluation metrics for object detection is mAP. At a predetermined Intersection over Union (IoU) threshold, mAP computes the average precision (AP) across all classes. Determining true positives and false positives for object detection is necessary to define precision. When the IoU between the predicted box and the ground reality is higher than the predetermined IoU threshold, it is deemed a true positive; when it is lower, it is deemed a false positive. By iterating over a range of thresholds and averaging them, the mean for each class can be calculated. The authors use increments of 0.05 for mAP50-95, beginning at an IoU threshold of 0.5 and ending at 0.95. The class AP is the average precision over this range. After completing this for each class and averaging them, the mAP50-95 can be calculated.

4.3. Ontology modeling

To facilitate semantic reasoning, ontology defines ideas, relationships, and attributes within a domain (Elsayed & Fathy, 2020a). This allows for an organized representation of knowledge. Ontology is used to build a semantic framework in the context of violence prediction, which improves comprehension of events that are identified. The proposed TEO is an Ontology that contains information on violence event domain. The authors proposed a semantic analysis of the labels of predicted objects to better understand the nature of events. Deep learning output needed to be enriched by semantics (Elsayed & Fathy, 2020b).

The TEO ontology defines terms associated with violent events, such as "Weapons' appearance" and "Armed people”. It defines a particular knowledge about weapons and their characteristics. Also, it defines a surrounding circumstance such as whether the armed person is wearing a uniform or not.

With the help of violence event ontology at reference (Arslan et al., 2015), the athors built a part of VEO as shown in Fig. 4 using protégé 5.6. Armed people are defined as a person or a part of a person and weapon. The proposed ontology classifies each event as threatening or non-threatening and this depends on whether or not a uniform is detected in the video. For example, if the object detector finds an armed person who carries any kind of weapons, and he wears uniform then the TEO will classify this event as a normal event. While if the object detector finds the armed person without uniform, then the TEO will classify the event as a threat event.

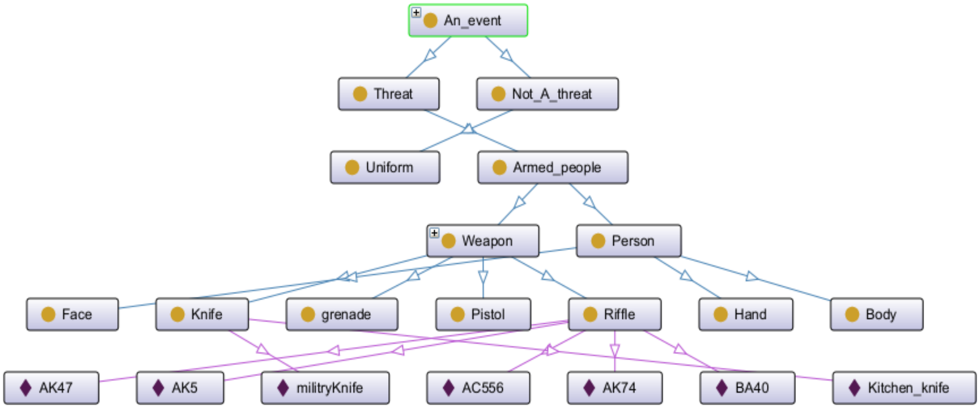


Fig. 4. Part of Threat Event Ontology (TEO)

4.4. Integrating with ontology

Yolov9 detection and classification are enhanced by ontology to incorporate semantic information. This process could be known as a semantic prediction process. Ontology is used to improve the semantic meaning of the items that are recognized, which increases the accuracy of recognizing and classifying violent behaviors. The systematic representation of concepts and their relationships offered by ontology improves the semantics and meaning of the objects that are detected.

Potentially violent situations can be more carefully interpreted when detected items and behaviors are examined considering the established ontology. By considering the connections and background data that the ontology provides, this integration improves the precision of violence prediction system.

The integration of the Yolov9 model with ontology is performed by mapping the output of Yolov9 model detection to the associated TEO ontology topics. The output of object detection process (object name) represents as individual in the TEO ontology.

In semantic prediction process the result of the object detection "weapons and persons objects" used in searching TEO Ontology to get all semantic meanings about that detection. So, each object label has its corresponding individual value to retrieve all semantic data of this object. The result of ASK SPARQL query determines if the event represents a threat or not. Thus, if the detected object was an armed person and wearing a military uniform, as he most likely, it does not represent a threat or violence event. But if the detected object was a person and not wearing a military uniform it represents a violent event. So, the violence prediction system is more accurate to identify potential threats situation by integrating ontology with the output of Yolov9.

5. RESULTS AND DISCUSSION

5.1. VP Dataset

The performance of the violence prediction system is evaluated using precision, recall, and mAP metric. That is a widely used evaluation metric in the field of computer vision for measuring the performance of object detection algorithms. Precision is the ratio of predicted positive examples to all detected objects, and recall rate is the ratio of the number of correctly detected objects to the number of all labeled objects. The mAP is the mean value of the average precision across all object categories present in the dataset.

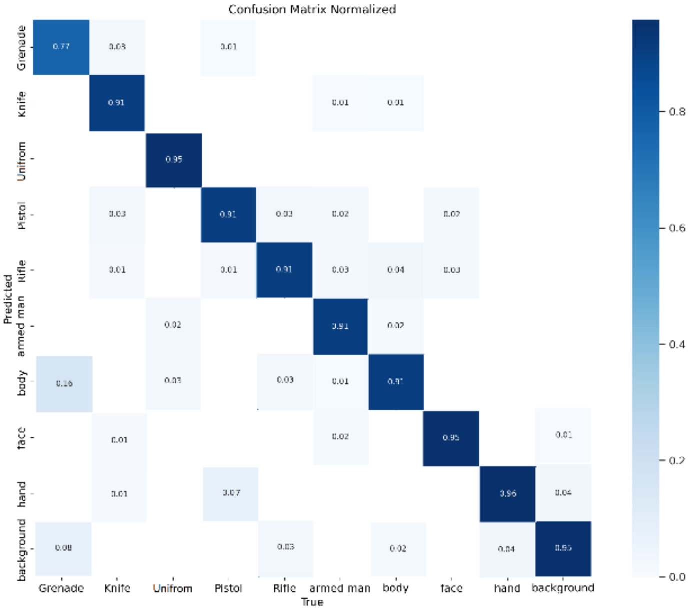
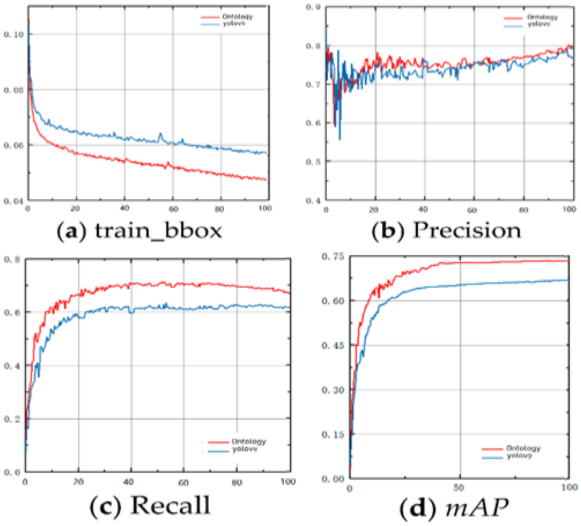


Fig. 5. Shows the confusion matrix in training process for VP Dataset

Figure 5 shows the confusion matrix during training process on Yolov9 model before the integration with ontology. The modifications to the Yolov9 training procedure and typical signs of the proposed method are displayed in Fig. 6. The training's confidence loss performance curve is displayed in Fig. 6a. It is evident that our method's confidence loss converges more quickly and progressively drops to 0.04 always less than Yolov9 without combining with ontology.



(a) train_bbox (b) Precision

(c) Recall (d) *mAP*

Fig. 6. Training curve of ontology and Yolov9 compared with Yolov9 only

The precision, recall, and mAP performance curves are represented in Fig. 6. These metrics allow for the evaluation of the model's effectiveness in classification issues. The model's detection accuracy is increased. Figure 6 b–d demonstrates that after fewer training cycles, the precision, recall, and mAP reached a stable state, surpassing those of Yolov9 alone. The suggested method's training speed and efficiency were likewise quicker than Yolov9 alone.

To investigate the effects of each module on the training process, after executing several experiments using the VP Dataset. We compared the precision , recall, mAP, and Epoch time in minutes of the suggested technique with YOLOv5m (Benjumea et al., 2021), YOLOv5l (Benjumea et al., 2021), YOLOv5x (Benjumea et al., 2021), YOLOv6s (Li et al., 2022), YOLOv6l (Li et al., 2022), YOLOv7-tiny (Wang et al., 2023), YOLOv7, and YOLOv8 (Reis et al., 2023)to highlight its positive aspects in training process for 100 epoch using GPU with help of google colab (Bisong, 2019) . The image size in all techniques was 640×640.

The method of combining Yolov9 with ontology has achieved better results. Table 1 illustrates that the proposed method outperformed in terms of precision, recall, and mAP. That is because combining Yolov9 output with TEO correct and riche understanding of violent events with semantic.

But, due to the semantic prediction process that was caused by the SPARQL query that was implemented to increase accuracy; the suggested work is slower than Yolov9.

Tab 1. Using training and validation dataset, the suggested approach compared to various YOLO versions

Method	Precision	Recall	mAP	Epoch Time
Yolov5m	67.9%	57.0%	51.9%	11.5min
Yolov5l	69.7%	58.1%	65.1%	12.5 min
Yolov5x	72.4%	60.1%	69.6%	13 min
Yolov6s	67.2%	62.6%	67.1%	10 min
Yolov6l	70.9%	63.1%	68.3%	11 min
Yolov7-tiny	76.3%	60.9%	69.9%	12min
Yolov7	77.2%	63.7%	70.8%	12.5 min
Yolov8	79.8%	64.4%	72.1%	8 min
Yolov9	84.6%	73.2%	80.4%	7 min
Proposed method	88.0%	76.4%	83.7%	7.1 min

To guarantee the impartiality and precision of the tests, we employed identical testing data and parameter configurations on 7% of the dataset with 612 images. The results in table 2 show that the recommended technique outperformed previous versions of Yolo, showing its effectiveness in identifying between threats and non-threats.

Tab 2. Using a testing dataset, the suggested approach compared to various YOLO versions

Method	Precision	Recall	mAP
Yolov5m	50.6%	49.5%	45.1%
Yolov5l	52.7%	58.1%	51.1%
Yolov5x	61.9%	55.7%	59.0%
Yolov6s	58.2%	58.6%	55.6%
Yolov6l	59.9%	60.1%	57.3%
Yolov7-tiny	62.2%	56.7%	60.0%
Yolov7	64.5%	57.3%	61.4%
Yolov8	68.6%	59.01%	65.1%
Yolov9	72.5%	62.1%	69.9%
Proposed method	79.5%	71.6%	73.9%

The method presented in tables 1 and 2 demonstrates the system's effectiveness in weapon detection and providing relevant context. It reached 88% precision, 76.4% recall, and 83.7% mAP in the training process on VP dataset. While in the testing phase, it obtained precision of 79.5%, recall of 71.6%, and mAP of 73.9% on the same dataset. Figure 7 represent the result in table 2 in chart to show the performance of the proposed method.

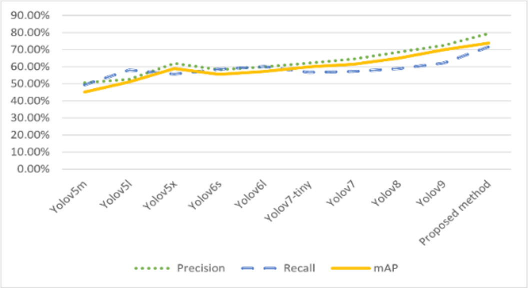


Fig. 7. Obtained results from test dataset

5.2. UCF-Crime dataset

The proposed system applied also to the UCF-Crime dataset. UFC is a popular benchmark dataset for evaluating techniques for violence detection and prediction is UCF-Crime. The UCF-Crime dataset consists of a wide range of videos that show different types of criminal activities, including shooting, fighting, and arresting. There is a label on every video stating whether or not it contains violence. Because of changes in lighting, camera angles, and occlusions, violence prediction in this dataset is difficult.

We choose to use transfer learning and initialize our models using pre-trained weights. These weights come from a model that used the VP Dataset for training.

We test the proposed method on subset of UCF-crime classes like normal, arrest, fighting, and assault videos. It was about 20% of the full size of the dataset which is 217 hours of videos (Mahareek et al., 2024) to test the proposed method.

During the test phase we use the trained model parameters obtained from the VP Dataset during the training phase. That is to decrease prediction errors and improve its performance. The testing is applied to every frame of the videos, and classifying the detected objects into different classes, including armed man, body, face, hand, pistol, grenade, knife, uniform, and rifle.

After performing testing on the UCF-Crime dataset, we evaluate the ability of YOLOv8 model. YOLOv8 is evaluated using the UCF-Crime dataset, and performance parameters such mAP, recall, and precision are measured. These metrics shed light on how well the model can identify and categorize violent behavior. Table 3 construct a comparison between testing results of the proposed yolov8 combining with TEO and previous versions of Yolo for UCF-Crime dataset. The table shows the superiority of the proposed method.

Tab 3. Comparation for the test results on UCF-Crime dataset

Method	Precision	Recall	mAP
Yolov5m	61.9%	55.9%	52.8%
Yolov5l	62%	58.7%	55.4%
Yolov5x	67.4%	60.1%	59.2%
Yolov6s	67.2%	60.6%	59.9%
Yolov6l	70.9%	61.9%	61.4%
Yolov7-tiny	71.3%	60.9%	62.9%
Yolov7	72.5%	61.7%	66.6%
Yolov8	78.9%	69.4%	68.0%
Yolov9	85.1%	79.1%	78.0%
Proposed method	88.9%	84.3%	85.2%

The proposed system could be applied in real time. Violence prediction system architecture in real time is shown in fig. 8. Firstly, the input video is divided to frames like (Mahareek et al., 2024). After object detection and classification using yolov9, the semantic prediction method is applied on detected frames to classify the event in the video and take suitable action.

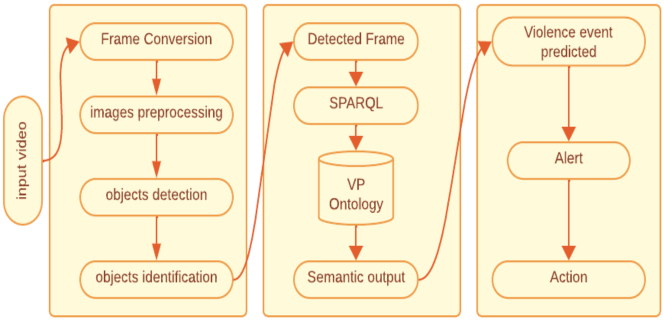
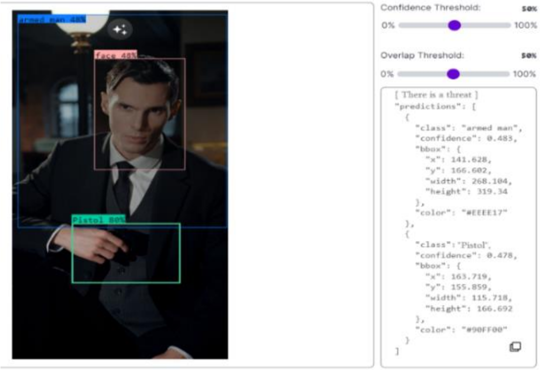


Fig. 8. Violence prediction system architecture in real time

Figure 9 and 11 present threat examples of violence prediction system using 50% for confidence threshold and overlap threshold. The figures show that the system detects objects in each frame and analyzes them to distinguish between the threat events and non-threat events. Figure 10 shows a non-threat example of violence system because the system detected people who were armed but wearing uniforms, so the system did not classify them as a threat.

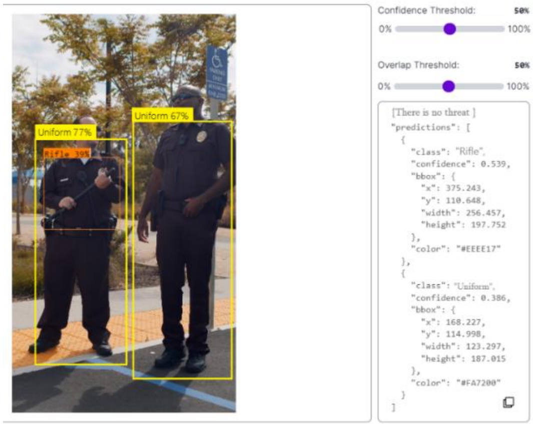


Confidence Threshold: 50%

Overlap Threshold: 50%

```
[ There is a threat ]
"predictions": [
  {
    "class": "armed man",
    "confidence": 0.483,
    "bbox": {
      "x": 141.628,
      "y": 166.602,
      "width": 268.104,
      "height": 319.34
    },
    "color": "#EEEE17"
  },
  {
    "class": "Pistol",
    "confidence": 0.478,
    "bbox": {
      "x": 163.719,
      "y": 155.859,
      "width": 115.718,
      "height": 166.692
    },
    "color": "#90FF00"
  }
]
```

Fig. 9. Threat examples of violence prediction system architecture in real time

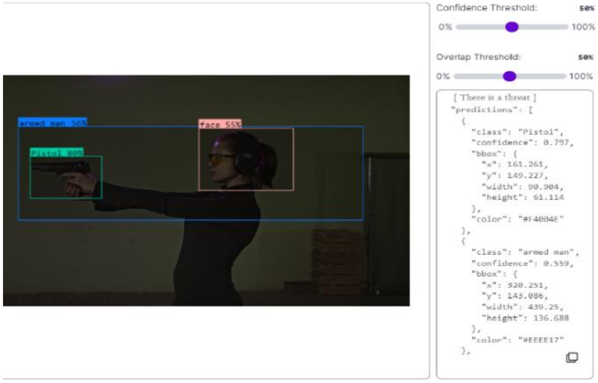


Confidence Threshold: 50%

Overlap Threshold: 50%

```
[There is no threat ]
"predictions": [
  {
    "class": "Rifle",
    "confidence": 0.539,
    "bbox": {
      "x": 375.243,
      "y": 110.648,
      "width": 256.457,
      "height": 197.752
    },
    "color": "#EEEE17"
  },
  {
    "class": "Uniform",
    "confidence": 0.386,
    "bbox": {
      "x": 168.227,
      "y": 114.998,
      "width": 123.297,
      "height": 187.015
    },
    "color": "#FA7200"
  }
]
```

Fig. 10. Non-threat examples of violence prediction system architecture in real time



```
[ There is a threat ]
"predictions": [
  {
    "class": "Pistol",
    "confidence": 0.797,
    "bbox": {
      "x": 161.261,
      "y": 149.227,
      "width": 90.904,
      "height": 61.114
    },
    "color": "#F4004E"
  },
  {
    "class": "armed man",
    "confidence": 0.559,
    "bbox": {
      "x": 320.251,
      "y": 143.086,
      "width": 439.25,
      "height": 136.688
    },
    "color": "#EEEE17"
  },
```

Fig. 11. Threat examples of violence prediction system architecture in real time

6. CONCLUSION

Yolov9 and TEO ontology are combined to create a unique framework for violence prediction in surveillance films presented in this paper. Real-time object detection coupled with semantic knowledge representation is a valuable addition to the field of video analytics for public safety and security, as it improves the interpretability and accuracy of violence predictions.

The results revealed that the suggested method-combining Yolov9 with ontology-performed better at predicting violence and grasping semantics than did employing Yolov9 alone. The mAP of Yolov9 without ontology was 72.1%, while the mAP of the suggested system was 74.7%. This improvement demonstrates how the integrated technique between YOLOv9 and ontology may be used to identify weapons in security cameras and identify violence events posed by armed individuals. The research report detailed the development process, which included steps like collecting datasets, labelling data, pre-processing, training models, and evaluating them. With the help of GPU in Google Colab, the training process was accelerated, and the training time was significantly reduced.

The system is able to distinguish between events that indicate possible risks and those that do not. That is by including contextual information from the ontology. Using ontology increases the precision of violence prediction process.

The authors recommend improving the semantic violence prediction system in surveillance videos even more. First, there is a potential to grow the dataset by adding more pictures and videos that represent a wider range of environments, lighting, and possible dangers. To obtain even greater accuracy and quicker processing times, the Yolov9 model can also be improved. It is also possible to investigate other deep learning architectures and algorithms for object detection. Moreover, adding real-time video streaming capabilities to the system would allow for ongoing surveillance and prompt threat identification. Also, a greater understanding of violent events is possible by the ontology's ability to be enlarged to include greater complexity links and relationships between objects and events.

REFERENCES

Arslan, A. N., Hempelmann, C. F., Attardo, S., Blount, G. P., & Sirakov, N. M. (2015). Threat assessment using visual hierarchy and conceptual firearms ontology. *Optical Engineering*, *54*(5), 053109. https://doi.org/10.1117/1.oe.54.5.053109

Arslan, A. N., Sirakov, N. M., & Attardo, S. (2012). Weapon ontology annotation using boundary describing sequences. *2012 IEEE Southwest Symposium on Image Analysis and Interpretation* (pp. 101-104). https://doi.org/10.1109/SSIAI.2012.6202463

Ashraf, A. H., Imran, M., Qahtani, A. M., Alsufyani, A., Almutiry, O., Mahmood, A., Attique, M., & Habib, M. (2022). Weapons detection for security and video surveillance using CNN and YOLO-V5s. *Computers, Materials and Continua*, *70*(2), 2761–2775. https://doi.org/10.32604/cmc.2022.018785

Benjumea, A., Teeti, I., Cuzzolin, F., & Bradley, A. (2021). YOLO-Z: Improving small object detection in YOLOv5 for autonomous vehicles. *ArXiv, abs/2112.11798*. https://doi.org/10.48550/arXiv.2112.11798

Bisong, E. (2019). *Building Machine Learning and Deep Learning models on Google Cloud platform: A Comprehensive Guide for Beginners*. Apress Berkeley.

Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. *ArXiv, abs/2004.10934*. https://doi.org/10.48550/arXiv.2004.10934

Dugyala, R., Vishnu Vardhan Reddy, M., Tharun Reddy, C., & Vijendar, G. (2023). Weapon detection in surveillance videos using YOLOV8 and PELSF-DCNN. *4th International Conference on Design and Manufacturing Aspects for Sustainable Energy (ICMED-ICMPC 2023)* (pp. 01071). E3S Web of Conferences. https://doi.org/10.1051/e3sconf/202339101071

Elsayed, E. K., & Fathy, D. R. (2020a). Semantic Deep Learning to translate dynamic sign language. *International Journal of Intelligent Engineering and Systems*, *14*(1), 316-325. https://doi.org/10.22266/IJIES2021.0228.30

Elsayed, E. K., & Fathy, D. R. (2020b). Sign language semantic translation system using ontology and Deep Learning. *International Journal of Advanced Computer Science and Applications*, *11*(1), 141-147. https://doi.org/10.14569/IJACSA.2020.0110118

Glenn, J. (2022, November 22). *Yolov5 release v7.0*. https://github.com/ultralytics/yolov5/tree/v7.0

Han, J., Liu, Y., Li, Z., Liu, Y., & Zhan, B. (2023). Safety helmet detection based on YOLOv5 driven by super-resolution reconstruction. *Sensors*, *23*(4), 1822. https://doi.org/10.3390/s23041822

Khalid, S., Waqar, A., Ain Tahir, H. U., Edo, O. C., & Tenebe, I. T. (2023). Weapon detection system for surveillance and security. *2023 International Conference on IT Innovation and Knowledge Discovery (ITIKD 2023)* (pp. 1-7). IEEE. https://doi.org/10.1109/ITIKD56332.2023.10099733

Lai, J., & Maples, S. (2017). Developing a real-time gun detection classifier. *Stanford University*.

Li, C., Li, L., Geng, Y., Jiang, H., Cheng, M., Zhang, B., Ke, Z., Xu, X., & Chu, X. (2023). YOLOv6 v3.0: A full-scale reloading. *ArXiv, abs/2301.05586*. https://doi.org/10.48550/arXiv.2301.05586

Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., Ke, Z., Li, Q., Cheng, M., Nie, W., Li, Y., Zhang, B., Liang, Y., Zhou, L., Xu, X., Chu, X., Wei, X., & Wei, X. (2022). YOLOv6: A single-stage object detection framework for industrial applications. *ArXiv, abs/2209.02976*. https://doi.org/10.48550/arXiv.2209.02976

Li, X., Wang, W., Wu, L., Chen, S., Hu, X., Li, J., Tang, J., & Yang, J. (2020). Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. *ArXiv, abs/2006.04388*. https://doi.org/10.48550/arXiv.2006.04388

Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollar, P. (2017). Focal loss for dense object detection. *2017 IEEE International Conference on Computer Vision (ICCV)* (pp. 2999–3007). IEEE. https://doi.org/10.1109/ICCV.2017.324

Lou, H., Duan, X., Guo, J., Liu, H., Gu, J., Bi, L., & Chen, H. (2023). DC-YOLOv8: Small-size object detection algorithm based on camera sensor. *Electronics*, *12*(10), 2323. https://doi.org/10.3390/electronics12102323

Mahareek, E. A. (2024). *VP Dataset*. https://Universe.Roboflow.Com/al-Azhar-Unversity/Violence-Prediction-in-Surveillance-Videos.

Mahareek, E. A., Elsayed, E. K., Eldesouky, N. M., & Eldahshan, K. A. (2024). Detecting anomalies in security cameras with 3D-convolutional neural network and convolutional long short-term memory. *International Journal of Electrical and Computer Engineering*, *14*(1), 993–1004. https://doi.org/10.11591/ijece.v14i1.pp993-1004

Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, faster, stronger. *Proceedings. 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*, (pp. 6517-6525). IEEE. https://doi.org/10.1109/CVPR.2017.690

Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. *ArXiv, abs/1804.02767*. https://doi.org/10.48550/arXiv.1804.02767

Redmon, J. (2016). *Darknet: Open source neural networks in c*. http://pjreddie.com/darknet/

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *2016 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 779-788). IEEE. https://doi.org/10.1109/CVPR.2016.91

Reis, D., Kupec, J., Hong, J., & Daoudi, A. (2023). Real-Time flying object detection with YOLOv8. *ArXiv, abs/2305.09972*. https://doi.org/10.48550/arXiv.2305.09972

Solawetz, J. F. (2023, January 11). What is YOLOv8? The Ultimate Guide. https://blog.roboflow.com/whats-new-in-yolov8/

Songire, S. B., Chandrakant Patkar, U., Chate, P. J., Patil, M. A., Wani, L. K., Pathak, A. S., Bhardwaj Shrivas, S., & Patil, U. (2023). Using Yolo V7 development of complete vids solution based on latest requirements to provide highway traffic and incident real time info to the atms control room using Artificial Intelligence. *Journal of Survey in Fisheries Sciences, 10*(4S), 3444-3456.

Tian, Z., Shen, C., Chen, H., & He, T. (2022). FCOS: A simple and strong anchor-free object detector. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(4), 1922–1933. https://doi.org/10.1109/TPAMI.2020.3032166

Verma, R., & Jayant, S. (2022). Cyber crime prediction using Machine Learning. In M. Singh, V. Tyagi, P. K. Gupta, J. Flusser, & T. Ören (Eds.), *Advances in Computing and Data Sciences* (Vol. 1614, pp. 160–172). Springer International Publishing. https://doi.org/10.1007/978-3-031-12641-3_14

Wang, C. Y., Mark Liao, H. Y., Wu, Y. H., Chen, P. Y., Hsieh, J. W., & Yeh, I. H. (2020). CSPNet: A new backbone that can enhance learning capability of CNN. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* (pp. 1571-1580). IEEE. https://doi.org/10.1109/CVPRW50498.2020.00203

Wang, C., He, W., Nie, Y., Guo, J., Liu, C., Han, K., & Wang, Y. (2023). Gold-YOLO: Efficient object detector via Gather-and-Distribute mechanism. *ArXiv, abs/2309.11331*. https://doi.org/10.48550/arXiv.2309.11331

Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *ArXiv, abs/2207.02696*. https://doi.org/10.48550/arXiv.2207.02696

Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 7464–7475). IEEE. https://doi.org/10.1109/cvpr52729.2023.00721

Wang, C.-Y., Yeh, I.-H., & Liao, H.-Y. M. (2024). YOLOv9: Learning what you want to learn using programmable gradient information. *ArXiv, abs/2402.13616*. https://doi.org/10.48550/arXiv.2402.13616

Zhang, X., Fang, S., Shen, Y., Yuan, X., & Lu, Z. (2024). Hierarchical velocity optimization for connected automated vehicles with cellular vehicle-to-everything communication at continuous signalized intersections. *IEEE Transactions on Intelligent Transportation Systems*, *25*(3), 2944–2955. https://doi.org/10.1109/TITS.2023.3274580

Zhang, S., Chi, C., Yao, Y., Lei, Z., & Li, S. Z. (2019). Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. *ArXiv, abs/1912.02424*. https://doi.org/10.48550/arXiv.1912.02424

Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., & Ren, D. (2020). Distance-IoU loss: Faster and better learning for bounding box regression. *34th AAAI Conference on Artificial Intelligence (AAAI 2020)* (pp. 12993-13000). https://doi.org/10.1609/aaai.v34i07.6999