VFAST Transactions on Software Engineering
https://vfast.org/journals/index.php/VTSE@ 2026, ISSN(e): 2309-3978, ISSN(p): 2411-6246
Volume 14, Number 1, January-March 2026 pp:264-277

Object Detection: Surveillancing Cities for Citizens' Safety and Protection Using Advanced Deep Learning Tools

Munam Ali Shah [1]*, **Tehreem Tahir** [1,2]

[1]Department of Computer Networks and Communication, College of Computer Science and Information Technology, King Faisal University, Al-Ahsa, Saudi Arabia; [2]Department of Computer Science, COMSATS University Islamabad, Pakistan

Keywords: *Object detection; Weapon detection; deep learning; Convolutional Neural Network; Multi Layer CNN, YOLO;*

Journal Info:
Submitted:
February 27, 2026
Accepted:
March 17, 2026
Published:
March 28, 2026

Abstract Weapon incidents result in casualties and loss of precious human lives every year. According to the Center for Disease Control (CDC) report, over 18500 people died in year 2025 in the USA due to gun violence. It is a well-known fact that the death rate from weaponization is increasing significantly. Firearm-related violence affects every aspect of human security, from personal safety issues like domestic violence and road rage to broader social and financial consequences. It can also escalate into large-scale armed conflicts that cause massive violence and account for many deaths. It is evident to surveillance our cities and societies for weapon objects using artificial intelligence (AI) and deep machine learning tools. The existing machine learning tools are not efficient at detecting different kinds of weapons. In this research, the contribution is threefold. Our first contribution is that we thoroughly investigated weapon detectio systems that uses deep learning and can based automated weapon detection system that allows the system to detect multiple weapons at the same time such as knife, handgun, and rifle automatically. We have used You Only Look Once (YOLO) v4 and Convolutional Neural network (CNN) for our experimentation. We have obtained several weapon images which are publicly available and have developed the datasets. Secondly, we have compared the performance of our deep learning models on multiple combinations of datasets (fewer images to several thousand images. The experimental evaluation has shown that the YOLOv4 outperformed the CNN. Lastly, we have proposed a method to improve the accuracy of the CNN and this has been accomplished by adding N number of CNN layers N times where N = 1,3,5,...19. This has resulted in reducing the complexity of the model and improving the accuracy and efficiency. The proposed model can be applied on surveillance systems, closed-circuit televisions and other object detection systems and many human lives can be saved by timely detection of weapons.

***Correspondence author email address:** mashah@kfu.edu.sa
DOI: 10.21015/vtse.v14i1.2358

1 Introduction

It has been observed that crimes involving arms and ammunition are increasing day by day. These incidents pose serious threats to human life and represent a major danger to public safety. Weapons should never be permitted in public places, as their presence negatively affects society, especially children. They contribute to human rights violations and moral degradation, including killing, kidnapping, burglary, child maiming, and

gender-based violence such as racism and sexual abuse. There is no region in the world completely free from the challenges and consequences of firearms. These violent incidents highlight the urgent need for an efficient surveillance system. Firearm-related violence affects every aspect of human security, from personal safety and crime prevention to the development of improved assistive technologies for visually impaired individuals [2], highlighting the versatile necessity of robust object detection systems. However, most existing surveillance systems are not fully automated and require continuous human monitoring for weapon detection. Closed-circuit television (CCTV) administrators often have to observe 25 to 30 screens at once. They must carefully monitor suspicious and hazardous activities affecting people or property. As the number of screens increases, human concentration begins to decline, and attention is divided across multiple feeds. It becomes difficult for any operator to maintain the same level of focus across all screens at all times; therefore, automated weapon detection is essential to reduce reliance on constant human supervision.

Research has been carried out using machine learning and deep learning algorithms. Artificial Intelligence (AI) allows computers and robots to act like humans, and they are controlled by other computers or algorithms. To detect weapons in a public place, machine learning is used which enables the machines to learn and act how humans detect arms and ammunition. The existing research lags in using advanced deep learning models which can be learned automatically and has the ability to learn and improve its functions by examining the algorithms. In most research, object detection is carried out using deep learning models—especially convolutional neural networks (CNNs). These networks are widely used for image recognition and can identify objects within images or video frames. Another type of object detection model is You Only Look once (YOLO) which analyzes the complete image at once. Annother type of deep learning image detection technique is Region based Convolutional Neural Network (RCNN). In the RCNN, the technique locate objects within the image. RCCN has two variations: i) Mask RCNN and ii) Faster RCNN. The former technique creates mask of images whereas, latter is computationally faster than other techniques as indicated in the literature [1, 3–5]. The summarized version of this research has been uploaded online and is accessible through [25].

In this paper, we have conducted our research on automated weapon detection system using deep learning object detection models such as CNN and YOLO. We have enhanced this research by adding new experiments and findings. We have identified the research gap i.e., the existing research lags in investigating the performance of weapon detection using N layers of CNN N number of times. We have experimented with different values of N i.e., 1, 2, 3 ..19. Using 6 different small, medium and large size datasets, we have identified some interesting facts about CNN and YOLOv4. Several new findings are results from part of this work.

We have investigated the multiple layers of CNN and found the threshold between accuracy and time consumption among multiple layers of CNN. We have applied N times N numbers of layers to improving the accuracy of CNN without any trade off. In the CNN layers, the dataset goes through several layers and gets trained.

After training datasets on multiple layers of CNN, we observed a drop in the accuracy of the model on the 19th layer and a rise in time consumption. However, on the 16th layer, CNN had the highest accuracy and stable time consumption. In other words the value of N is 16 number of layers of CNN, which is perfect balance between time and accuracy and provides accuracy without a threshold.

The rest of the paper is organized as follow. Section 2 consists of background studies related to this research. In Section 3, the proposed methodology has been presented. In Section 4, the performance analysis of different deep learning technique is provided. Section 5 presents the discussion and research findings. The conclusions and future direction of the research are presented in Section 6.

2 Related Studies

Convolutional neural networks (CNNs) identify where objects are located in an image by drawing bounding boxes around them and determine what each object is. The convolutional layers help reduce the complexity of an image data and preserves important information.

Regions with convolutional neural networks (R-CNN),

combines rectangular region proposals with convolutional neural network features. R-CNN is a two-stage detection algorithm.

The YOLO (You Only Look Once) technique uses one forward pass to detect and recognize various objects in an image. Object detection in YOLO is done as a regression problem which means that it predicts numerical values directly, such as where the object is located (coordinates of the bounding box); how big the object is and how confident it is about the detection and provides the class probabilities of the detected images. The class probabilities estimate how likely the detected object belongs to each possible category (knife, pistol, gun etc.)

Different machine learning and deep learning models have been used to detect the weapons such as knife, pistol and riffle etc. Experimentation has been carried out using YOLO, CNN, RCNN, FCNN, Mask RCNN, and Visual Geometry Group Net (VGGNET). Table 1 provides a brief comparison of different deep learning classifiers. It can be observed that in most cases, the YOLO has outperformed other models [3, 4, 21, 24, 36].

S Gawade et al. [1] proposed a CNN-based model for weapon detection. They used a custom dataset for training which consisted of three classes if weapon images: i) long guns - 2497 images; ii) small guns - 3876 images; and iii) knives - 3641 images [1]. The model achieved an accuracy of 85% which is quite good. In [3], the authors have proposed a weapon detection model based on YOLOv4 using a dataset that contained 8000. This work has presented an interesting finding i.e., by decreasing the resolution of images, the accuracy will be improved. In their experiments, they have improved the accuracy from an initial 90.4% to 92.1% with an average loss of 4.75. S. Narejo et al. proposed a deep learning model in which they compared the performance of YOLOv2, YOLOv3, and CNN. YOLOv3 outperformed the other models with 98.89% mAP. They used a custom dataset for training and testing purposes [4]. In another research, the authors worked on CNN using VGGNET as the base model. They created a custom dataset [5]. They proposed a Deep CNN model, they compared the proposed model with VGG-16, ResNet- 101, and ResNet-50 to compare the accuracy level of their proposed model with other similar techniques. The proposed VGG based CNN model's accuracy was highest at 98.41%. R. Grag et. al [6] carried out a comparative research by evaluating the performance of YOLOv3 and YOLOv4. They also used custom dataset for weapon detection. YOLOv3 gained 90% accuracy and YOLOv4 gained 98% accuracy.

In [8] , authors presented a CNN-based model for RCNN for weapon detection. This model used a custom pretrained dataset for model training then fed it into a single shot detector CNN after that it detects the weapon and sends a mail alert to the concerned authority. They promised to achieve high accuracy. N. Hnoohm and other carried out research on and proposed a model in which they compared different algorithms such as CNN, Faster RCNN, and VGGNet and they used multiple datasets of public image dataset of weapons for training purposes. Faster RCNN outperformed other algorithms with the highest of 79% Mean Average Precision (mAP) [9]. U. V. Navalgund et al proposed CNN based YOLO model [10]. They used acustom dataset for training purposes. The model gained 81.41 accuracy. The limitation was limited computational resources and real-time requirements. F. Gelana carried out research on a model based on CNN for weapon detection. They used sliding window and gaussian blur to soften edges and they used a custom dataset [11]. The model gained 93.84% accuracy. This research had high computational, power, and memory consumption issue. In a research [12], authors proposed a model in which they compared Faster RCNN and CNN. They used a custom dataset. Faster RCNN outperformed CNN with 94.25%. The limitation was CNN followed by SVM could be a disadvantage. In 2022, N. U. Haq et al. proposed a CNN-based OWAD model. They used a cus tom dataset for training purposes. After training, they compared this model with RCNN [13]. The model gained 82% accuracy and RCNN gained 72% accuracy. In [14], Sidharth G proposed a YOLO-based CNN model. They used a custom dataset to train a model. They created a user interface for the detection system and gained 92% accuracy. Roberto Olmos et al. presented a novel multi-confirmation-level alarm system based on CNN and Long Short Term Memory networks (LSTM). They used a custom dataset for training [15]. Their proposed system MULTICAST reduced by 80% the number of false alarms with respect to the Faster RCNN based-single image detector. Abhinav Juneja carried out research

Authors	Description of Research	Evaluation Measures	Deep Learning Model	Dataset Used	Time Efficiency	Limitations
[1] S. Gawade et al., 2022	Designed a system which activates an alarm after detection of a weapon	mAP 85%	CNN	Custom dataset	Low	Computational cost is high
[3] S. Ahmed et al., 2022	Implemented on different datasets with different image resolutions	92.1% mAP	YOLOv4	Custom dataset	Moderate	Limited computational resources are used
[4] S. Narejo et al., 2022	Compared YOLOv2, YOLOv3 and CNN	98.89% mAP	YOLOv3, YOLOv2 and CNN	Custom dataset	Low	Computational time is too high
[5] N. Dwivedi et al., 2021	Implemented CNN using VGGNet	98.41% Accuracy	VGGNet, CNN	Custom dataset	Low	Images were mistakenly classified
[6] R. Garg et al., 2021	Compared YOLOv3 and YOLOv4	90%, 98% Accuracy	YOLOv3, YOLOv4	Custom dataset	Low	Processing time is high

Table 1. Comparison of Deep Learning-Based Weapon Detection Methods

Table 2. Comparative table addressing validation of the proposed model

Ref.	YOLOv4	CNN	Multiple CNN layers	Multiple datasets
[7]	Yes	No	No	Yes
[20]	No	Yes	No	No
[37]	Yes	Yes	No	No
[19]	Yes	No	No	Yes
[30]	No	Yes	No	No
This research	Yes	Yes	Yes	Yes

on CNN-based SSD model for weapon detec tion. They used a custom dataset to train the model. Their model 92% accuracy [16]. Similar work on object detection has been carried out in [17].

Figure 1 presents the taxonomy of object detection. Furthermore, it helps to understand the sequence how a weapon detection model is selected. We briefly review some of these models in the following section.

For weapon detection in image datasets, different classifiers such as CNN, RCNN, YOLOv4 have been used. In literature [5–7], as shown in Table 2, CNN has been used with other base models such as YOLO to achieve better accuracy, however, it increases computational cost. There is a tradeoff between the accuracy and computational efficiency.

3 Proposed Methodology

We believe that there is a need to investigate the accuracy and efficiency levels of multiple layers of CNN and YOLOv4, on multiple datasets with different number of layers of the CNN. This is the main reason, we have prepared different datasets with images of different resolution and have investigated different models. Importantly, we have investigated the multiple layers of CNN and identified a threshold value for accuracy and time consumption amongst multiple layers of CNN. We have investigated the point where the CNN gives best values for accuracy and if more layers are added in the CNN, the computational cost of the model will exceed the accuracy. In short, we have applied N times N numbers of layers for improving the accuracy of CNN without any trade-off.

We looked at CNN's various layers and discovered the cutoff point between time consumption and accuracy. In order to increase CNN's accuracy without sacrificing anything, we implemented N times N layers. The dataset is trained over multiple levels in the CNN layers.

We saw a decrease in the accuracy of the model on the 19th layer and an increase in time consumption after training datasets on several CNN layers. CNN, on the other hand, had the best accuracy and steady time consumption on the sixteenth layer. In other words, the value of N is 16 CNN layers, which offers accuracy without a threshold and strikes the ideal balance between time and accuracy.

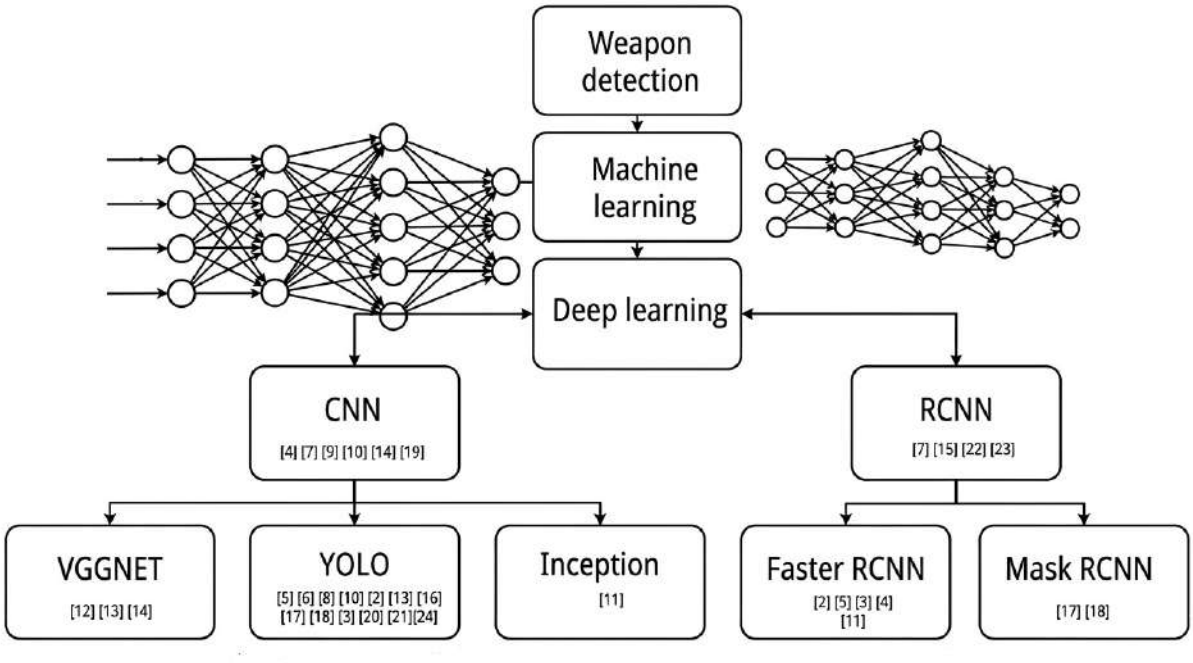

Figure 1. Taxonomy of Machine Learning Classifiers

Furthermore, we also analyzed the performance of YOLOv4 and compared it with multiple layers of CNN. We carry out this research on CNN based object detection which detects the weapons automatically with better mAP of CNN at an acceptable computational cost. The objective is not to achieve highest accuracy with highest computational cost. The main idea is to develop a model with reasonably high accuracy in weapon detection with acceptable computational cost. Table 2 shows the related work using different combinations of YOLO and CNN.

3.1 Data Preprocessing

The proposed research follows a classical machine learning approach. Firstly, we collected different images of different weapons such as knife, gun and rifle. The blurred, duplicate and mislabeled images were removed. Then we ensured the uniformity of all images by using a common size for all images. This helped us in reducing the computation cost of running different algorithms.

3.2 Model Architecture

After the preprocessing, we trained our cleaned data on CNN and YOLOv4 that is accessible through these URLs mentioned in the Data Availability section. We have used six datasets in our this research which we have created after gathering images from different online sources, and after that, we have selected two models for the training purpose of images, CNN model using multiple layers and YOLOv4. Figure 2 presents the architecture of our proposed model used in this research. First, we feed input data into CNN model and we train our CNN model upto N times N number of layers, the dataset goes through several layers and is being trained. In the prediction, the phase dataset is tested on any single random image to check the accuracy of the trained dataset. To get the appropriate accuracy and the performance, we used N times N number of CNN layers. Because it becomes more difficult for a model to detect and train as the complexity increases, adding more layers will lessen the complexity of the model by allowing it to train with a wider range of data and with the support of numerous (N) pooling layers, time complexity will decrease, and image size will shrink.

3.3 Hyperparameter Configurations

In this stage, we set the external controls that govern how and what our model learns. Initially, we were not sure, how many layers of CNN will be sufficient

so we started with single layer of CNN and kept in increasing CNN layers at each round until desirable results were achieved. When the image is fed into CNN first it goes into convolutional layers which divide an image into features then output of the convolutional layers is forwarded to *ReLU* layer as input *ReLU* layer is responsible for removing negative values, *ReLU* layer is applied on each divided feature one by one then fed into Max pooling layer which is used to shrink image's size after that we will repeat all process again from the convolutional layer as we proposed to apply multiple layers. We will continue to apply layers on CNN until unless we will find a threshold where we will have a balance between accuracy and time consumption. After gaining acquired results from CNN we will train the same datasets on YOLOv4. In YOLO after feeding the dataset, it is forwarded the to backbone layer which is responsible for feature formation then there is the neck layer in which features are aggregated, and then the head layer comes which is also called the detection layer after going through these layers data is being trained and then in the last step it is being predicted in the end, will compare the accuracy and performance of CNN with YOLOv4 after being trained on the same datasets as shown in Table 5.

3.4 A Comparative Analysis of Deep Learning Techniques

This section presents the compariosn if different object detection techniques. We have used six different datasets. The datasets are publically available and can ve accessed through the URLs provided at the end of the paper. These datasets have been compiled from different online sources to train them on various object detection. The details of the datasets used in this study is provided in Table 3. The datasets used in this study have been classified into three categories:

- Knife
- Handgun
- Rifle

Each dataset contains same percentage of images from all three weapon classes. In our experiments, there is 20% testing images and 80% training images used. Note that in Table 3, the first three datasets are created with low resolution images. This is to investigate the performance of the deep learning algorithm with respect to different resolutions and whether or not the accuracy is improved by lowering the image resolutions. The other datasets i.e., D4, D5 and D6 are prepared to test and check the efficiency of algorithms as shown in Table 4. Moreover, we carried out experiments on YOLOv2 and YOLOv3 on a single dataset D4 to see if the YOLOV4 really works better than the YOLOv2 and YOLOv3. After the comparison of YOLO classifiers, our main aim has been to carry out experiments on YOLOv4 and CNN for their comparison and performance evaluation.

Table 3. Detailed information about datasets

No	Number	No of images	Resolution
1	D1	10	512x512
2	D2	10	640x640
3	D3	10	1900x1200
4	D4	200	Min 512 Max 1900
5	D5	1000	Min 512 Max 1900
6	D6	4000	Min 512 Max 1900

Table 4 shows the performance comparison of different object detection algorithms. It also shows the computational time each algorithm has consumed on a specific dataset. Observe that the YOLOv4 has outperformed other algorithms with less time consumption when compared with other algorithms. YOLOV2 has gained 23% after being trained for 3 hours on D1, YOLOv3 gained 27% accuracy after being trained for 1 hour, YOLO gained 34% accuracy after being trained for 30 minutes and CNN gained 23% accuracy after being trained on 6 layers and consumed 10 minutes. Also note that the YOLOv2 gained 25% accuracy on D2 with the same time consumption as D1. YOLOv3 gained 27% accuracy using the same time as D1. Whereas the YOLOv4 achieved 37% accuracy within 30 minutes. The CNN gained 27% accuracy after being trained on six layers and the results were generated only in 10 minutes. On D3, the YOLOv2 gained 21% accuracy when it was trained for 3.5 hours, YOLOv3 gained 28% of accuracy with 1.5 hours of time consumption, YOLOv4 gained 31% accuracy with 45 minutes of time consumption and CNN gained 21% accuracy after being trained on 6 layers for 10 minutes.

When we used D4 and trained YOLOV2, it gained

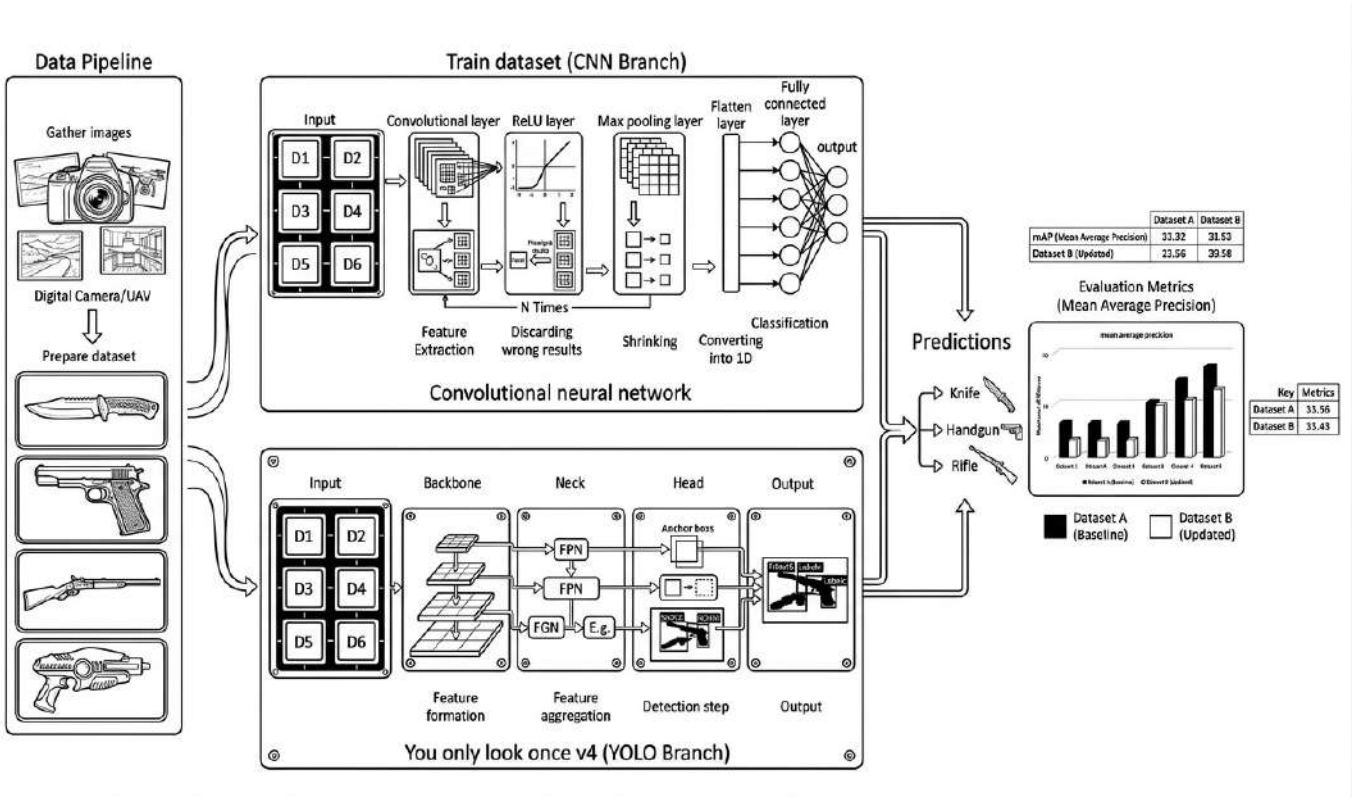

Figure 2. Architecture of the proposed model

47% accuracy, however, the computation time was a little over 6 hours.Whereas, a better accuray and time efficiency was achieved by YOLOV3 when trained on D4. The YOLOV3 gained 49% of accuracy and 5.5 hours of processing time. Lastly, YOLOv4 accomplished an accuracy of 52% and 5 hours computational time. The results of CNN were notably different than the YOLO. CNN achieved 35% accuracy with first three layers. A better accuracy level i.e., 38% was achieved when three more layers were added and only an hour was consumed to process the datasets. The lesser computational time is a surpringing finding.

The results clearly indicate that the CNN is taking less computational time. This is the reason, we carried out more experiments only on the CNN to improve the accuracy. The results of different layers of the CNN are provided in Table 5. We did not continued the experiments on YOLOv2 and YOLOv3 as they computational cost was getting higher with slight improvement in accuracy.

In our next experiments, we computed and compared the accuracy of CNN and YOLOv4. We added multiple layers of CNN to improve its accuracy. Initially, we had no knowledge about adding the number of CNN layers. We started by adding 1 layer and calculated its accuracy. We kept adding CNN layers one by one and it was the 16th layer which gave us the highest accuracy with reasonable computational cost.

After the 16th layer, there was no improvement in the accuracy and the computational time was too high. Therefore, we stopped adding more layers of the CNN. Table 5 is showing some interesting results of the accuracy and the computational cost. It can be seen that on Dataset 1, the YOLOv4 was 34% accurate and consumed a thirty minutes. On other hand, CNN gained 37% accuracy after the implementation of sixteen layers with 45 minutes. This means that adding more and more layers on small datasets in not effective in case of CNN. On Dataset 2, the YOLOv4 gained 37% accuracy with 30 minutes of time consumption, and CNN gained 33% accuracy after being trained on sixteen layers with 45 minutes of time consumption. Similarly, on Dataset 3, the YOLOv4 gained 31% accuracy with 45 minutes of time consump-

Table 4. Preliminary results of weapon detection on model using D1 D2, D3 and D4

YOLO Models				CNN		
Dataset	Model	Accuracy	Time	Layers	Accuracy	Time
D1	YOLOv2	23%	3 hrs	3	17%	10 min
	YOLOv3	27%	1 hr	6	23%	10 min
	YOLOv4	34%	30 min			
D2	YOLOv2	25%	3 hrs	3	13%	10 min
	YOLOv3	22%	1 hr	6	27%	10 min
	YOLOv4	37%	30 min			
D3	YOLOv2	21%	3.5 hrs	3	15%	10 min
	YOLOv3	28%	1.5 hrs	6	21%	10 min
	YOLOv4	31%	45 min			
D4	YOLOv2	47%	6 hrs	3	35%	1 hr
	YOLOv3	49%	5.5 hrs	6	38%	1 hr
	YOLOv4	52%	5 hrs			

Table 5. Evaluated results of weapon detection of the proposed solution using datasets D1–D6.

Dataset	**YOLOv4**		**Convolutional Neural Network**		
	Accuracy	Time	3–9 layers (Acc / Time)	12–16 layers (Acc / Time)	19 layers (Acc / Time)
D1	34%	30 min	17–25% / 15 min	29–37% / 45 min	34% / 1 h
D2	37%	30 min	13–21% / 15 min	31–33% / 45 min	31% / 1 h
D3	31%	45 min	15–27% / 15 min	27–35% / 45 min	32% / 1.5 h
D4	52%	5 h	35–44% / 1.5 h	48–55% / 3.5 h	55% / 4 h
D5	73%	6 h	59–65% / 3 h	69–71% / 5 h	71% / 6 h
D6	87%	8 h	67–73% / 6 h	78–85% / 8.5 h	83% / 9 h

tion and CNN gained 35% accuracy after being trained on sixteen layers for 45 minutes.

On Dataset 4, the YOLOv4 gave an accuracy of 52% with 5 hours of time consumption. Whereas, the CNN gained 55% accuracy on the same dataset with 3.5 hours of time consumption having 12 layers. YOLOv4 gained 73% accuracy on Dataset 5 with 6 hours of time consumption. In contrast, the CNN had 71% accuracy with 5 hours of time consumption with 16 layers. YOLOv4 gained 87% accuracy with 8 hours of time consumption. On the other hand, the CNN gained 85% accuracy with 8.5 hours of time consumption. We carried out more experiments and applied more layers up to the 19th layer, however, there was no notable improvement in the accuracy level and the computational cost became too high. The bold values in Tables 5 represent the highest accuracy in each of the six datasets. The datasets are accessible through these URLs: https://github.com/tahreemkhann/Weapon-detection-dataset. https://drive.google.com/drive/folders/1av4-nFgi1W-UO_goz-TcN-FqxxmPtdFb.

4 Performance Analysis of Different Deep Learning Techniques

In this section, we discuss the performance of different techniques used in this study. We have used grahs to visualize the experimental findings. Two types of graphs are used to present the findings: i- mean average preci-

sion and ii- time consumption.

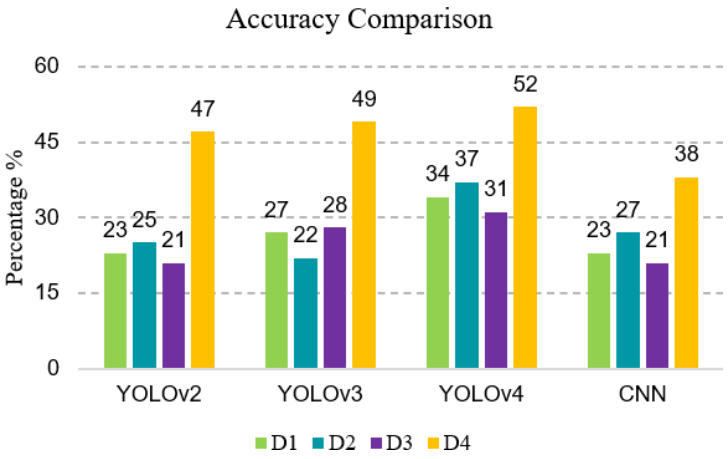

Figure 3. Accuracy percentage of different deep learning models on different datasets.

In Figure 3, we compare the accuracy of different models. The x-axis is indicating the models being compared and the y-axis is showing the mean rank difference. This comparison is plotted for four datasets, i.e., D1, D2, D3, and D4. On D1, the YOLOv2 gave an accuracy of 23% while YOLOv3 27 percent accuracy, YOLOv4 34% accuracy, and CNN 23% accuracy. On Dataset 2 YOLOv2 gained 25 percent accuracy, YOLOv3 22 percent accuracy, YOLOv4 37 percent accuracy, and CNN 27 percent accuracy. On Dataset 3 YOLOv2 gained 21% accuracy, YOLOv3 28% accuracy, YOLOv4 31% accuracy, and CNN 21% accuracy. On Dataset 4, YOLOv2 gained 47% accuracy, YOLOv3 49% accuracy, YOLOv4 52% accuracy, and CNN 38 percent accuracy. YOLOv4 got the highest accuracy when compared with other models on available datasets.

Figure 4 is showing the computational cost of each of the deep learning model for all 4 datasets. The time consumption is represented in hours. It can be observed that the computational cost of CNN remains less even after the implementation of 6 layers. In the same figure, we compared the time consumption of YOLOv2, YOLOv3, YOLOv2, and CNN. On D1, YOLOv2 consumed 3 hours, YOLOv3 consumed 1 hour, YOLOv4 consumed 0.5 hour and CNN took 0.167 hour to train. On D2, YOLOv2 consumed 3 hours, YOLOv3 consumed 1 hour, YOLOv4 consumed 0.5 hour and CNN took 0.167 hour for training the model. To train the models on D3, YOLOv2 consumed 3.5 hours, YOLOv3 consumed 1.5 hours, YOLOv4 consumed

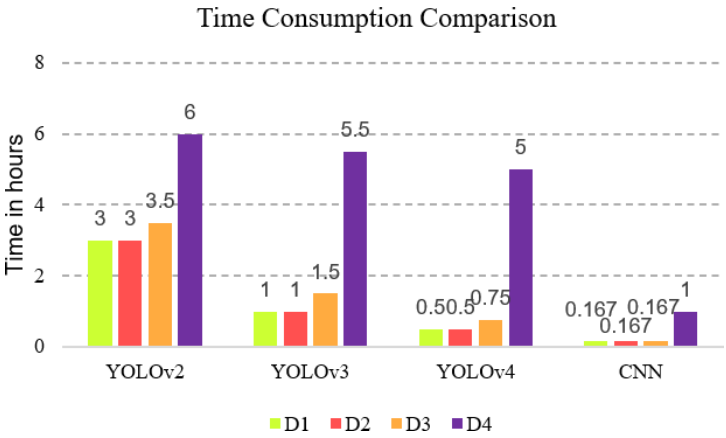

Figure 4. Time consumption graph of object detection models

0.75 hour and CNN took 0.167 hour to train on Dataset 3. Lastly, on D4, YOLOv2 consumed 6 hours, YOLOv3 consumed 5.5 hours, YOLOv4 consumed 5 hours and CNN took 1 hour to train on Dataset 4.

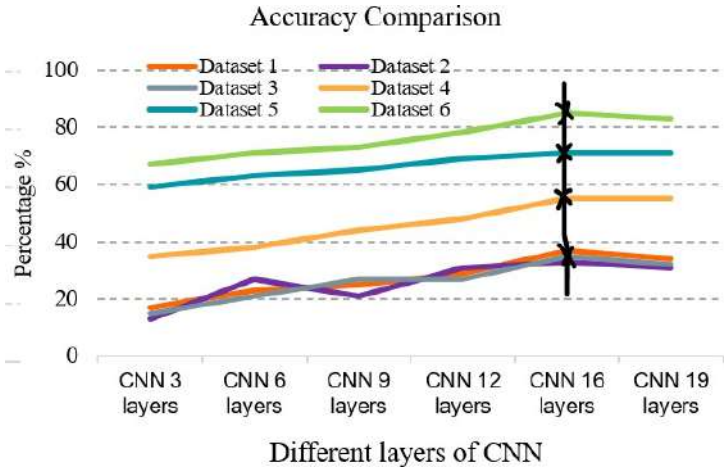

Figure 5. Accuracy % of different CNN layers on Different Datasets

A comparison of accuracy level achieved by different layers of the CNN on different datasets has been plotted in Figure 5. Different colors of lines are indicating different layers of the CNN. Note that the datasets with a larger number of images have higher accuracy. The obvious reason is that the model is better trained when it is provided larger number of images. On D1, the accuracy on with different number of layers range from 20% to a maximum of 40%. Note that using D3 and onwards datasets, after the third dataset, there is a continuous raise in accuracy. The maximum accuracy is achieved is

with D6 which is comprised of 4000 images. The important thing to observe here is that adding up to 16 layers of the CNN – there was an improvement in the accuracy for almost all datasets. After the 16 layers, when we applied more layers, i.e., 19 CNN layers, we faced two issues: 1) there was not any improvement in the accuracy; 2) the computational cost of the model went too high.

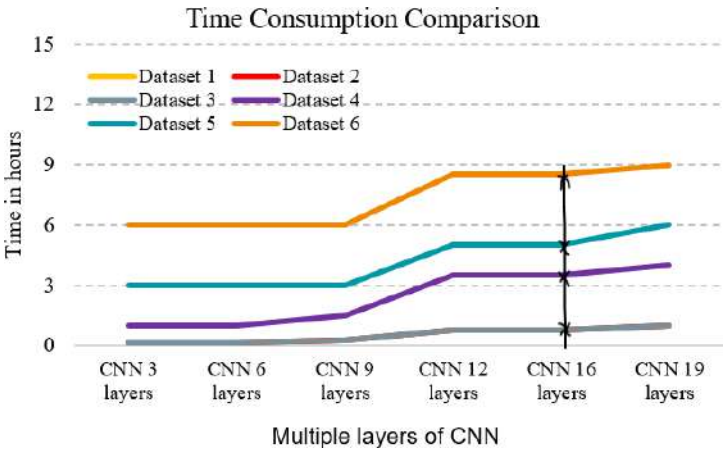

Figure 6. CNN Layers Comparison on Different Datasets

In Figure 6, the comparison of time consumption on different layers of CNN on different datasets has been presented. Different colors of lines are indicating different layers of CNN. The x-axis is showing different layers of CNN and the y-axis is showing the time in hours. On Datasets 1, 2, and 3 we have time consumption between few seconds to less 2 hours. On D4, the six layers and the nine layers of the CNN had a computational cost of up to a maximum of two hours. Whereas, on D4, six layers and nine layers have a time consumption between 2 to 4 hours. Note that the 16 layers and 19 layers have a higher time consumption which is between 4 hours to 6 hours on D5. Lastly, on D6, 16 layers and 19 layers of the CNN consumed a maximum amount of time which is more than 8 hours. On the 16th layer, CNN has the highest accuracy but as the number of layers are increased from 16 to 19, there is no significant improvement in accuracy and a notable raise in the time consumption.

We summarize our experiments in Figure 7 and 8. It can be observed in Figure 7 that in all the datasets, when we compare YOLOv4 and CNN, the YOLOv4 provided a better accuracy. We believe that YOLOv4 works well with smaller datasets. This version of YOLO has a better speed and accuracy in object detection when compared

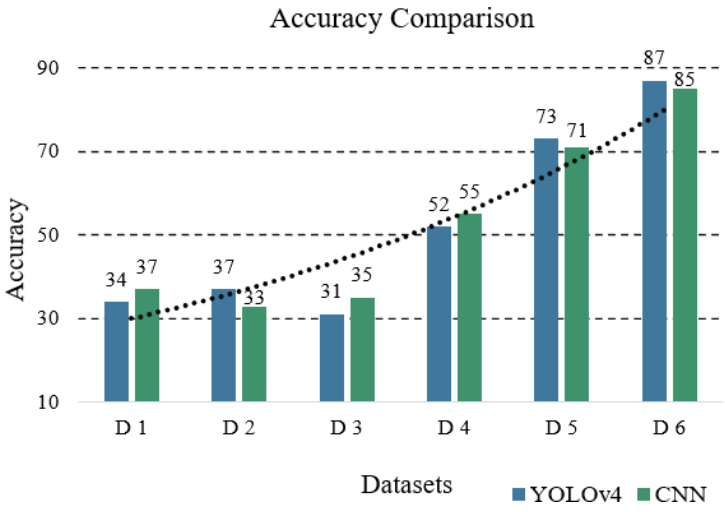

Figure 7. Accuracy Comparison Graph of YOLOv4 and CNN.

with CNN and its multiple layers.

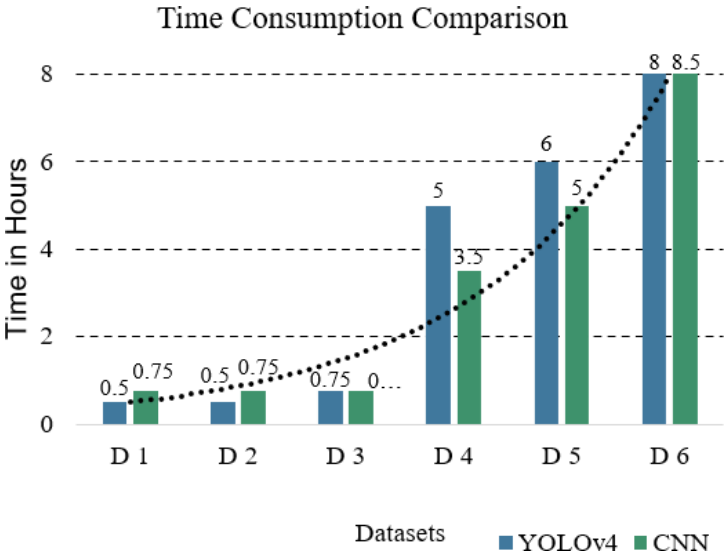

Figure 8. Time consumption graph of YOLOv4 and CNN.

Lastly, Figure 8 shows the computational time comparison of YOLOV4 and CNN. We can see that on D5, the YOLOv4 consumed more time than the CNN but has a better accuracy than the CNN. It is perceived that YOLO is more accurate but slow. There is a tradeoff between computational time and accuracy. Similarly, onD6, the YOLOv4 has better accuracy and lesser computational time. This result also indicate that the CNN took less time up to six layers. Afterwards, as the number of the layers are increased in CNN, the accuracy improves but so is the computational complexity. Note that after

applying up to sixteen layers, with the aim to improve the accuracy of the CNN, it is still not better than YOLOv4. On smaller datasets, CNN performed way better than YOLOv4 but not this is not the case with larger datasets. Hence, YOLOv4 is a wise choice for larger datasets, and CNN with fewer layers is a better choice for smaller datasets.

5 Discussion and Research Findings

This research revealed some interesting findings. We observed a tradeoff between the accuracy and computation cost. It can be observed that adding more layers of CNN improved the accuracy but this is not always the case. After applying maximum number of layers, there was a stage when no accuracy improvement could be achieved and the computation cost was getting higher. The proposed research yielded better results in object detection when compared with [44] and [45]. Following are four discussion points and research findings:

- The experiment demonstrated that YOLOv4 consistently outperformed CNN-based models, achieving 32% mAP on the smallest dataset and 87% mAP on the largest dataset, whereas the baseline CNN achieved only 15% mAP and 65% mAP, respectively.
- Increasing CNN depth from 1 to 19 layers improved performance by reducing model complexity through progressive pooling and feature reduction; however, even a 16-layer CNN failed to surpass YOLOv4's accuracy, especially on larger datasets.
- Comparative analysis of YOLO variants showed that YOLOv4 surpassed YOLOv2 and YOLOv3 in both accuracy and robustness, confirming it as the most effective architecture among the evaluated detectors.
- While CNNs trained significantly faster, their accuracy remained inferior to YOLOv4 on medium and large datasets; however, CNNs with fewer layers performed better than YOLOv4 on smaller datasets, indicating dataset-size sensitivity.

6 Conclusion

This study used six datasets of various sizes and image resolutions to compare the performance of CNN-based and YOLO-based methods for weapon detection. The tests assessed the processing cost and detection accuracy of various YOLOv4 and CNN architecture configurations. With a maximum the mean average precision, or m of 87% on the largest dataset, the experimental results show that YOLOv4 consistently beat CNN models with respect of detection accuracy, thus surpassing the performance recorded in previously discussed related works. YOLOv4 achieved 32% mAP on smaller datasets, while CNN's three-layer design yielded 15% mAP. CNN accuracy was still lower than YOLOv4 on large-scale datasets even after expanding the architecture to nineteen layers, however increasing CNN depth improved performance.

CNN models had a lower detection performance than YOLOv4, although requiring less training time. It's interesting to note that while YOLOv4 showed higher scalability and resilience as dataset size increased, CNN with fewer layers performed relatively better on smaller datasets. Additionally, YOLOv4 fared better overall in detection than YOLOv2 and YOLOv3, according to a comparative analysis of YOLO variations.

Overall, the results show that YOLOv4 is the best model for large scale weapon detection tasks, outperforming CNN-based models and previously published methods in terms of accuracy, while lightweight CNN architectures are still appropriate for smaller datasets requiring less processing power.

Future research will look into improved YOLO variations including YOLOv7, YOLOv8, YOLOv11, and YOLO26, as well as sophisticated CNN-based detectors like RCNN [46], Faster RCNN, and Mask RCNN [47]. In next-generation YOLO systems, hyperparameter optimization and latency reduction via enhanced prediction algorithms will receive particular attention.

Author Contributions

Munam Ali Shah: Conceptualization, supervision, methodology, and review and editing of the manuscript. **Tehreem Tahir:** Data collection, implementation, experimental analysis, and writing the original draft. Both authors have read and approved the final version of the manuscript.

Conflict of interest

The authors declare no conflict of interest.

Data availability

The datasets used in this study are available and publicly accessible through the following URLs: https://github.com/tahreemkhann/Weapon-detection-dataset https://drive.google.com/drive/folders/1av4-nFgi1W-UO_goz-TcN-FqxxmPtdFb

Acknowlegdement

This research has been supported by the Deanship of Scientific Research (DSR), King Faisal University, Al-Ahsa, Saudi Arabia.

References

[1] A. Lamas et al., "Human pose estimation for mitigating false negatives in weapon detection in video-surveillance," Neurocomputing, vol. 489, pp. 488– 503, 2022, doi: 10.1016/j.neucom.2021.12.059.

[2] S. Naz and F. Jabeen, "Towards improved assistive technologies: classification and evaluation of object detection techniques for users with visual impairments," *VAWKUM Transactions on Computer Sciences*, vol. 12, no. 2, pp. 165–177, 2024.

[3] A. Warsi, M. Abdullah, M. N. Husen, M. Yahya, S. Khan, and N. Jawaid, "Gun Detection System Using Yolov3," 2019 IEEE 6th Int. Conf. Smart Instrumentation, Meas. Appl. ICSIMA 2019, no. August, pp. 27–29, 2019, doi: 10.1109/ICSIMA47653.2019.9057329.

[4] Y. Huang, X. Fu, and Y. Zeng, "Anchor-Free Weapon Detection for X- Ray Baggage Security Images," IEEE Access, vol. 10, no. August, pp. 97843– 97855, 2022, doi: 10.1109/ACCESS.2022.3205593.

[5] R. Reddy, K. Gyan Vallabh, and S. Sharan, "Multiclass weapon detection using multi contrast convolutional neural networks and faster region-based convolutional neural networks," 2021 2nd Int. Conf. Emerg. Technol. INCET 2021, pp. 1–8, 2021, doi: 10.1109/INCET51464.2021.9456407.

[6] M. Attique and M. Habib, "Weapons Detection for Security and Video Surveillance Using CNN and YOLO-V5s," Comput. Mater. Contin., vol. 70, no. 2, pp. 2761–2775, 2022, doi: 10.32604/cmc.2022.018785.

[7] A. Goenka, "Weapon Detection from Surveillance Images using Deep Learning," 2022 3rd Int. Conf. Emerg. Technol., vol. 1, no. 1, pp. 1–6, 2022, doi: 10.1109/INCET54531.2022.98242.

[8] H. Lokhande, "A review on weapon detection and alert system using deep neural networks," Irjmets, vol. 4, no. 6, pp. 1–4, Jun. 2022.

[9] N. Hnoohom, P. Chotivatunyu, N. Maitrichit, V. Sornlertlamvanich, S. Mekruksa vanich and A. Jitpattanakul, "Weapon Detection Using Faster R-CNN Inception V2 for a CCTVSurveillance System," 2021 25th International Computer Science and Engineering Conference (ICSEC), 2021, pp. 400-405, doi: 10.1109/ICSEC53205.2021.9684649.

[10] U. V. Navalgund and P. K., "Crime Intention Detection System Using Deep Learning," 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET), 2018, pp. 1-6, doi: 10.1109/ICCSDET.2018.8821168.

[11] Gelana and A. Yadav, "Firearm Detection from Surveillance Cameras Using Image Processing and Machine Learning Techniques," Smart Innovationsin Communication and Computational Sciences, pp. 25–34, Nov. 2018, doi:10.1007/978-981-13-2414-73.

[12] R. Reddy, K. Gyan Vallabh and S. Sharan, "Multiclass Weapon Detection using Multi Contrast Convolutional Neural Networks and Faster Region-Based Convolutional Neural Networks," 2021 2nd International Conference for Emerging Technology (INCET), 2021, pp. 1-8, doi: 10.1109/INCET51464.2021.9456407.

[13] N. U. Haq, M. M. Fraz, T. S. Hashmi, and M. Shahzad, "Orientation aware weapons detection in visual data: a benchmark dataset," Computing, Jun. 2022, doi: 10.1007/s00607-022-01095-0.

[14] G.Sidharth,"Weapon detection and suspicious activity using cnn and yolo 4," Mar. 2022, (accessed Dec. 31, 2022),(http://dspace.srmist.edu.in/jspui/ bitstream/123456789/45864/1/P12469.pdf).

[15] R. Olmos, S. Tabik, F. Perez-Hernandez, A. Lamas, and F. Herrera, "MULTICAST: MULTI Confirmation-level Alarm SysTem based on CNN and LSTM to mitigate false alarms for handgun detection in video-surveillance," arXiv :2104.11653 [cs], May 2021.

[16] A. Juneja, S. Juneja, and S. Jain, "Real Time Object Detection using CNN based Single Shot Detector Model," Journal of Information Technology Management, vol. 13, no. 1, p. 63, 2021, doi: 10.22059/jitm.2021.80025.

[17] S. Naz and F. Jabeen, "Towards Improved Assistive Technologies: Classification and Evaluation of Object

Detection Techniques for Users with Visual Impairments," VAWKUM Transactions on Computer Sciences, vol. 12, no. 2, pp. 165–177, Dec. 2024, doi: https://doi.org/10.21015/vtcs.v12i2.1911.

[18] H. V Lokhande, "A review on weapon detection and alert system using deep neural networks," irjmets, vol. 4, no. 06, pp. 1–4, 2022.

[19] S. Ahmed, M. T. Bhatti, M. G. Khan, B. Lövström, and M. Shahid, "Development and Optimization of Deep Learning Models for Weapon Detection in Surveillance Videos," Appl. Sci., vol. 12, no. 12, p. 5772, 2022, doi: 10.3390/app12125772.

[20] S. Gawade, R. Vidhya, and R. Radhika, "Automatic Weapon Detection for Surveillance Applications," SSRN Electron. J., 2022, doi: 10.2139/ssrn.4143822.

[21] S. Narejo, B. Pandey, D. Esenarro Vargas, C. Rodriguez, and M. R. Anjum, "Weapon Detection Using YOLO V3 for Smart Surveillance System," Math. Probl. Eng., vol. 2021, 2021, doi: 10.1155/2021/9975700.

[22] N. Hnoohom, P. Chotivatunyu, N. Maitrichit, V. Sornlertlamvanich, S. Mekruksavanich, and A. Jitpattanakul, "Weapon Detection Using Faster R- CNN Inception-V2 for a CCTV Surveillance System," ICSEC 2021 - 25th Int. Comput. Sci. Eng. Conf., pp. 400–405, 2021, doi: 10.1109/ICSEC53205.2021.9684649.

[23] U. V. Navalgund and P. K. Priyadharshini, "Crime Intention Detection System Using Deep Learning," 2018 Int. Conf. Circuits Syst. Digit. Enterp. Technol. ICCSDET 2018, 2018, doi: 10.1109/ICCSDET.2018.8821168.

[24] M. T. Bhatti, M. G. Khan, M. Aslam, and M. J. Fiaz, "Weapon Detection in Real-Time CCTV Videos Using Deep Learning," IEEE Access, vol. 9, pp. 34366–34382, 2021, doi: 10.1109/ACCESS.2021.3059170.

[25] T. Tahir, "Performance Evaluation and Comparison of YOLOv4 and Multiple Layers of CNN for Weapon Detection," Feb. 2023, doi: https://doi.org/10.36227/techrxiv.22060520.v1.

[26] J. Elsner, T. Fritz, L. Henke, O. Jarrousse, S. Taing, and M. Uhlenbrock, "Automatic Weapon Detection in Social Media Image Data using a Two-Pass Convolutional Neural Network," Eur. Law Enforc. Res. Bull., no. 4, p. tbc-tbc, 2018.

[27] J. Ruiz-Santaquiteria, A. Velasco-Mata, N. Vallez, G. Bueno, J. A. Alvarez-Garcia, and O. Deniz, "Handgun Detection Using Combined Human Pose and Weapon Appearance," IEEE Access, vol. 9, pp. 123815–123826, 2021, doi: 10.1109/ACCESS.2021.3110335.

[28] J. Salido, V. Lomas, J. Ruiz-Santaquiteria, and O. Deniz, "Automatic handgun detection with deep learning in video surveillance images," Appl. Sci., vol. 11, no. 13, 2021, doi: 10.3390/app11136085.

[29] S. Nikkath Bushra, G. Shobana, K. Uma Maheswari, and N. Subramanian, "Smart Video Survillance Based Weapon Identification Using Yolov5," 2022 Int. Conf. Electron. Syst. Intell. Comput., vol. 1, no. 1, pp. 1–7, 2022, doi: 10.1109/icesic53714.2022.9783499.

[30] F. Gelana and A. Yadav, Firearm Detection from Surveillance Cameras Using Image Processing and Machine Learning Techniques, vol. 851, no. January. Springer Singapore, 2020.

[31] T. S. S. Hashmi, N. U. Haq, M. M. Fraz, and M. Shahzad, "Application of Deep Learning for Weapons Detection in Surveillance Videos," 2021 Int. Conf. Digit. Futur. Transform. Technol. ICoDT2 2021, 2021, doi: 10.1109/ICoDT252288.2021.9441523.

[32] A. Singh, T. Anand, S. Sharma, and P. Singh, "IoT Based Weapons Detection System for Surveillance and Security Using YOLOV4," Proc. 6th Int. Conf. Commun. Electron. Syst. ICCES 2021, pp. 488–493, 2021, doi: 10.1109/ICCES51350.2021.9489224.

[33] N. U. Haq, M. M. Fraz, T. S. S. Hashmi, and M. Shahzad, "Orientation Aware Weapons Detection In Visual Data: A Benchmark Dataset," Comput., 2022, doi: 10.1007/s00607-022-01095-0.

[34] M. Brahmaiah, S. R. Madala, and C. M. Chowdary, "Artificial Intelligence and Deep Learning for Weapon Identification in Security Systems," J. Phys. Conf. Ser., vol. 2089, no. 1, 2021, doi: 10.1088/1742- 6596/2089/1/012079

[35] Y. Ma, H. Chen, and J. Huo, "Assault Rifle Detection and Identification Based on Convolutional Neural Network YOLOv3," 2021 3rd World Symp. Artif. Intell. WSAI 2021, pp. 1–4, 2021, doi: 10.1109/WSAI51899.2021.9486333.

[36] A. O. Ramon and L. Barba Guaman, "Detection of weapons using Efficient Net and Yolo v3," 2021 IEEE Lat. Am. Conf. Comput. Intell. LA-CCI 2021, pp. 0–5, 2021, doi: 10.1109/LA-CCI48322.2021.9769779.

[37] R. Garg and S. Singh, "Intelligent Video Surveillance Based on YOLO: A Comparative Study," 2021 7th IEEE Int. Conf. Adv. Comput. Commun. Control. ICAC3 2021, pp. 0–5, 2021, doi: 10.1109/ICAC353642.2021.9697321.

[38] A. H. Ashraf et al., "Weapons detection for security and video surveillance using CNN and YOLO-V5s," Comput. Mater. Contin., vol. 70, no. 2, pp. 2761–2775, 2022, doi: 10.32604/cmc.2022.018785.

[39] A. Belurkar, A. Waghmare, S. Mallick, N. Waghamode, and P. R. Totare, "Weapon Detection using Yolov4, CNN," Int. J. Res. Appl. Sci. Eng. Technol., vol. 10, no. 4, pp. 2058–2062, 2022, doi: 10.22214/ijraset.2022.41702.

[40] S. A. Ali Shah, M. Ahmad Al-Khasawneh, and M. I. Uddin, "Review of weapon detection techniques within the scope of street-crimes," 2021 2nd Int. Conf. Smart Comput. Electron. Enterp. Ubiquitous, Adapt. Sustain. Comput. Solut. New Norm. ICSCEE 2021, pp. 26–37, 2021, doi: 10.1109/ICSCEE50312.2021.9498007.

[41] M. V Girish, H. K. G. B. H, B. Prakash, and P. K. N, "Crime Detection In Videos And Alerting System Using Artificial Intelligence," Turkish J. Comput. Math. Educ. Vol . 12 No . 12 (2021), 738 -743 Res. Artic. Crime Detect. Videos Alerting Syst. Using Artif. Intell. Turkish J. Comput. Math. Educ. Vol . 12 No . 12 (, vol. 12, no. 12, pp. 738–743, 2021).

[42] S. Mahmud, S. Hossain, Y. Arafat, J. Rawnak Jahan, and R. Zannat, "Weapons Detection of Criminal Activities Based on Computer Vision," Http://Www.Sciencepublishinggroup.Com, vol. 2, no. 4, p. 64, 2021, doi: 10.11648/j.advances.20210204.11.

[43] V. Kaya, S. Tuncer, and A. Baran, "Detection and classification of different weapon types using deep learning," Appl. Sci., vol. 11, no. 16, 2021, doi: 10.3390/app111675355

[44] V. Arulalan, T. Javali, S. Jha and P. Garg, "Object Detection Using YOLOv4 and PSO CNN, Innovations and Advances in Cognitive Systems, ICIACS Conference, NatureSpringer, 2024,"

[45] K. Lu, F. Zhao, X. Xu and Y. Zhang, "An object detection algorithm combining self attention and YOLOv4 in traffic scenes," in PLOS One., vol. 18, no. 05, 2023, doi.org/10.1371/journal.pone.0285654.

[46] M. Ali Shah and T. Tahir, "A Novel Model for Pneumonia Detection Using Hybrid CNN Architecture," IPSI Transactions on Internet Research, Jan. 2025, doi: https://doi.org/10.58245/ipsi.tir.2501.04.

[47] M. I.-U. Haque, M. Fatima, F. Waqas, S. Fatima, M. Bukhari, and M. Ahmed, "A Next Generation Real Time Frame work for Drone Video Decoding Leveraging IoT-Enabled Communication Network," VAWKUM Transactions on Computer Sciences, vol. 13, no. 2, pp. 205–219, Dec. 2025, doi: https://doi.org/10.21015/vtcs.v13i2.2266.