

Research Article

Real-Time Surveillance Using Deep Learning

Muhammad Javed Iqbal,¹ Muhammad Munwar Iqbal ,¹ Iftikhar Ahmad,² Madini O. Alassafi ,² Ahmed S. Alfakeeh ,² and Ahmed Alhomoud³

¹Department of Computer Science, University of Engineering and Technology, Taxila 47050, Pakistan

²Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia

³Faculty of Computing and Information Technology, Northern Border University, Rafha, Saudi Arabia

Correspondence should be addressed to Muhammad Munwar Iqbal; munwariq@gmail.com

Received 4 July 2021; Accepted 10 August 2021; Published 17 September 2021

Academic Editor: Farhan Ullah

Copyright © 2021 Muhammad Javed Iqbal et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

It is crucial to ensure proper surveillance for the safety and security of people and their assets. The development of an aerial surveillance system might be very effective in catering to the challenges in surveillance systems. Current systems are expensive and complex. A cost-effective and efficient solution is required, which is easily accessible to anyone with a moderate budget. In aerial surveillance, quadcopters are equipped with state-of-the-art image processing technology that captures detailed photographs of every object underneath. A quadcopter-based solution is proposed to monitor desired premises for any unusual activities, like the movement of persons with weapons and face detection to achieve the desired surveillance. After detection of any unusual activity, the proposed system generates an alert for security personals. The proposed solution is based on quadcopter surveillance and video streaming for anomaly detection in the received video streams through deep learning models. A well-known FasterRCNN algorithm is modified for fast learning with feature reduction in the initial feature extraction step. Five different kinds of CNNs were evaluated for their ability to identify objects of interest in surveillance images. ResNet-50-based FasterRCNN with the highest average precision performed as an excellent solution for threat detection. The average precision of the system is 79% across all categories achieved.

1. Introduction

These days security is the critical concern of everyone around the globe. Every citizen wants to cease the possibility of any upcoming threat to himself or his possessions. This tremendous escalation in the requirement of security needs for individuals and an organization needs a proper and up-to-date surveillance system. The existing systems are expensive and complex to operate. Hence, a system that can account for insecurities regarding security and of less complexity is the need of time. A cost-effective, simple, and efficient surveillance system is necessary for every organization, either in the public sector or private sector, for internal monitoring and the analysis of post events later onwards [1]. The drone is globally recognized as an aerial vehicle without any pilot on board. They are known for their use in numerous diverse fields, namely, Security, Aerial

Photography, Surveying, and Surveillance. This drone-based security surveillance system provides the user ultimate control, providing surveillance updates directly to the user's owned systems. Using a quadcopter can conveniently start the surveillance process in areas where we cannot place security cameras. Also, with the help of the quadcopter, we can deliver goods where human access is beyond the bounds of possibility, as in explosions, volcano eruptions, and numerous other disasters, thereby minimizing the risk to human life.

Doing remote surveillance is one of the major problems that Pakistan is confronting these days due to the unsettling political situation and banned outfits. The basic architecture of the working drone system is given in Figure 1. Adversaries had targeted our country since 1990. The target killing and bomb blasts seized the growth of many business sectors, leaving inhabitants to face devastation and inflation. A

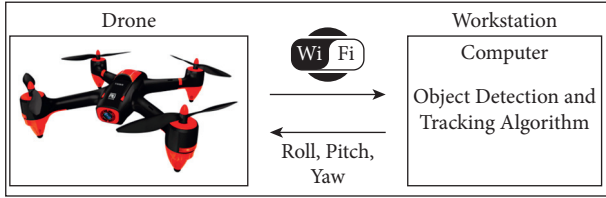


FIGURE 1: Basic architecture of drone surveillance system.

stranded economy is resultant due to losing thousands of precious lives and property worth millions of dollars. Unfortunately, remote surveillance has remained unaddressed so far. To facilitate the country in attaining proper security level requires deploying a number of security personnel, which is quite undesirable. This deployment will lay hold of the country's major resources, including human resources and finances.

Moreover, even by incorporating human resources, the surveillance issue still is alarming as staffing is unlikely to perform impregnable duties. We cannot incorporate individuals to fight against unenviable situations, including harsh, barren deserts, unbeatable oceanic waves, and blazing hot regions. Therefore, the idea here is to perform surveillance conveniently using drones because drones can undergo continuous and spot-on surveillance. In a nutshell, drones are the comprehensive and cheaper solution to all surveillance and security-related issues. One of the primary objectives behind developing this quadcopter security system was to assist security personnel [2]. However, the deployment of security personnel is not the ultimate solution to deal with security threats. An aerial surveillance system aids in advancing security procedures as real-time data are directly transmitted to user systems, including smartphones and computer servers. In addition to that, it continuously performs surveillance in open areas, especially in areas that are undesirable to human existence. This system helps the user to keep track of different angles of the spotted location. A dimensional view of the spotted area helps figure out minute details that lead to a more secure and hazard-free surrounding. We aim to provide the most appropriate and inexpensive solution while simultaneously deploying drones rather than personnel for surveillance, which is the most inexpensive solution to security-related issues.

Security is essential in every domain, be it public or private. For example, the increase in crime rate in a crowded event, be they concerts, airports, marriages, or other public gatherings or lonely areas where there are a lot of blind areas that cannot be monitored or are general blind spots in terms of security and public safety. Many solutions have been presented regarding the problems mentioned above. But none of them provides a clear and definitive solution to the issues regarding security and safety. Weapon detection or anomaly detection is one of the major problems most security institutions are trying to tackle every day. Weapon or anomaly detection system identifies abnormal events differently from existing patterns because any anomaly is a pattern that is different from a set of standard patterns. Weapon detection can be related to a mass shooting

incident, a bank robbery, or any other more public disaster, or it can be in the form of concealed weapon detection in airports, luggage, or other forms of searches. Millions of dollars have been put into the weapon detection field, and many solutions have been created or implemented. One aim of the study was to provide computationally less expensive algorithmic arrangements for detection purposes. A precise ablation can chalk out the weak areas, but in our case, our intuition was based on the fact that researchers like SSD, YoloV3 etc., have proposed many other fast detectors and that reduction in the number of input features can reduce the processing time. The search for an ideal method is still going on. The proposed system provides a security framework to monitor the defined premises to alert the security team after identifying any unusual activity, including any suspicious individual or item in a locality. Afterwards, subsequent operations undergo ensuring effective crisis management with proficiency without any dilemma or restriction. The system under discussion provides feasibility in case of emergencies with the efficiency of overcoming crises in the first place. Consequently, providing users with means of dealing with disasters effectively and proficiently.

The main contributions of our study can be described as follows:

- (i) Construction of data set that addresses four types of entities concerning drone surveillance
- (ii) A less computationally expensive FasterRCNN discussed
- (iii) A near Real-time performance for object detection
- (iv) Four existing deep neural networks are also used and compared for their capability to categorize

This article is arranged as follows. Section 2 briefly discusses the background information and related literature review. Section 3 provides the material and methods used to implement this research work. Results and discussion are given in Section 4. The final Section 5 presents the conclusion of this work.

2. Literature Review

Currently, many robotic applications are being developed for doing tasks autonomously without any commands from a human. A system enabling a robot to surveillance, like detecting and tracking an object in motion, will carry us to a more advanced task. For taking the role of UAV (Unmanned Aerial Vehicle) AR [3], the drone comes in place as a flying robotic platform. As far as the implementation part is concerned, they developed an algorithm to detect and track an object by analyzing its shape and color. Unmanned Aerial Aircraft Systems (UAS) are being divided into different transportation engineering areas. Traffic analysis and work on the grounds are highly compatible with the systems. Such studies are helpful for understanding and working UAS in transportation. In the study by Berrahal et al. [2], the authors proposed a cooperative border surveillance solution that relies on WSN (Wireless Sensor Network), implemented to detect and track transfer and UAVs in the form of

quadcopters. A heuristic-based scheduling algorithm is used by increasing the rate of trespassers spotted by the quadcopters. Also, techniques to localize sensors using RFID technology and optimal computing positions and relay data between isolated islands of nodes were used for evaluating the performance of the mentioned WSN-based surveillance system. In another work [3], the authors started their studies by defining the capabilities of aerial surveillance systems and suggested several technologies that can be used in it. They also discussed how drones could be used in aerial surveillance. The work also found that surveillance can be utilized in peace retaining activities and time monitoring of a place at any time of the day. The intention was to provide speedy and green surveillance at a low-cost charge so that it could be used extensively at the nonpublic, institutional, and governmental stage.

Gadda and Patil [4] stated that security issues of borders are increasing with rapid tensions across the countries. Terrorist attacks at the border are risking the lives of common people living near the borders. Monitoring of terrorism activities and security lapses becomes quite challenging and challenging in changing weather conditions. Designing an UAV for monitoring the border area from long distance conveyed to the observer. In this study [5], the IR remote is used for controlling the quadcopter. The Audio-Visual will be transferred to the computer through a wireless camera. The work presented in this article focused on trends of image processing for UAV (Unmanned Aerial Vehicles) and quadcopter data. Examples for totalling the total number of vehicles, edge detection, and evasion algorithms were discussed. SIFT (scale-invariant features transform) matching and the median filter are mainly used in UAVs image processing and visualization of new urban buildings and mechanisms that control the quadcopter. In this work, a simple camera-based navigation system for an autonomous quadcopter was proposed. Authors observed that additional infrastructure like artificial landmarks and GPS is not required in navigation methods and algorithms [6]. Factors affecting the computational complexity were independent of the environment size and factors like sensing only one landmark at a time. FPGA-based embedded realization of the methods was discussed with the most computationally demanding phase. In the study by Haque et al. [7], they proposed quadcopter as a low-weight and low-cost autonomous flight Unmanned Aerial Vehicle (UAV) for delivering parcels ordered online. Usage of Google maps to navigate to destination are the significant functions of the quadcopter. The authors also discussed the capability of delivering orders through quadcopter using an online process and coming back to the original point or starting place.

The work by Ding et al. [8] presented that those earlier systems were expensive and limited to outdoor environments with access to Global Positioning Systems (GPS). This work aimed to develop a design for a quadcopter to create a map of an unknown space using commercial off-the-shelf (COTS) to reduce cost and for mapping SLAM (Simultaneous Localization and Mapping) for creating an automatic map. Lastly, it was

suggested that the SLAM algorithm might run efficiently on the Android phone. This work by Ding et al. [8] stated that applications in communications, photography, agriculture, and mainly surveillance make drones as mini-UAVs a bit more attractive. State-of-the-art studies for that amateur drone with a vision named Dragnet, tailoring the recent Cognitive Internet of Things framework. At the same time, the detection and classification of authorized and unauthorized drones are done, which eventually allow only authorized ones to fly over.

Amato et al. [9] discussed that Public security and safety are ensured through wireless communication technologies and support mobility. In the study by Alford et al. [10], the authors stated that the primary cause of security and critical information could be transmitted among the entities involved, dealing with cyber issues and mainly risking privacy of UAV equipped with communication hardware and directed to specific positions for privacy and safety. For the past few decades, Iraq and Syria among the most affected countries due to terrorism by ISIS were liberated, and internationally displaced persons (IDP) were going back to their homes. Infrastructure loss in the form of destroyed buildings and operators are exposed to undefined situations. Situations like this make innovations in the search and clearance of operational technology for ensuring safety and peace. In the study by Bangare et al. [11] focussed on ways to save homes from fire and smoke. The primary concern was home monitoring, appliances controlling, and door latches control from faraway regions.

With the use of their device, people will be able to manage our home appropriately from remote places. In the study of Munagekar [12], researchers focused on detecting places where criminal talk happens and tracing the robber quickly and efficiently. The purpose of this work was to minimize the efforts and finances of the security forces. Canny edge detection algorithm was used to develop this system, which was further used to detect and catch the robbers and stop unusual happenings. Usage of canny edge filter reduces the postprocessing size of objects, which saves much memory of hard disk. The focus of Alshammari and Rawat work was on a multicamera video system for the betterment of security [13]. The multicamera video system can make a widespread impact within the security industry. The technical aspect of this may help many people to counter the protection challenges of their everyday lives.

The proposed technique may locate and understand human behaviour from videos taken from cameras mounted at the wall. The proposed method consists of the detection and monitoring of any targets. In this article [14], a live surveillance system was proposed for weapon and abnormal events detection. The system was composed of three modules. The detection was done using the convolutional neural networks in the first processing module, whereas the second module manages guns and tracking. The alarm operations were done in the last processing module. In their work, shape detection and object detection techniques were experimented to find detection accuracy. The authors claimed that the proposed method substantially reduces the crime rate and the time required to trap the criminal. In this

article, Mahalanobis et al. [15] presented autodetection and monitoring strategies based on three significant aspects. Three different aspects of this work were the automatic goal detection and reputation strategies with monitoring, the handover, and smooth tracking of objects within a community and the improvement of actual-time communication and messaging protocols using COTS networking components.

The comprehensive review of the various security and surveillance techniques shows that this area has the potential to work. The mechanism of surveillance can be enhanced using modern deep learning approaches. This work proposed a technique for the detection of arms and other objects using advanced deep learning models.

3. Materials and Methods

This work proposed a methodology to detect various entities in a video. Figure 2 shows the block diagram of the methodology and phases involved in the proposed security and surveillance system. The proposed system consists of the following main phases: Data acquisition/collection of the data set containing the instances of entities to be detected. Next, the preprocessing phase transforms the images into reduced features for efficient processing and the Training Phase for model building. The fourth phase consists of Query Image entity prediction, in which query image is processed and presented to a trained model for the prediction of potential. Following is a detailed description of each phase.

3.1. Data Acquisition. The data acquisition phase is responsible for collecting and acquiring videos from surveillance drones to construct the required data set. The videos in the data set are obtained from different environments through a surveillance drone camera. The data set consists of videos of different persons with different poses, with and without weapons, etc. The persons include guards and common people. Later the captured videos are transformed into image frames, which are then taken from the data set for further processing. A good data set plays an important core role in the development of the proposed surveillance system. A machine learning model for the detection will be trained by giving them image frames from the developed data set. In the next phase, necessary preprocessing steps are applied to the image frames. The detail of the preprocessing step is as follows.

3.2. Preprocessing. In this preprocessing phase, several steps are taken to facilitate the model training process. The key steps include resizing the data set image frames, extracting image features using pretrained VGG16 CNN, and later PCA is applied on features extracted from the desired input image. The detail of each step within preprocessing is discussed next.

3.2.1. Resizing of Image Frames and Data set Construction. The original size of the captured videos/image frames was $2980 * 2021$ pixels is shown in Figure 3 and is captured from

a hovering quadcopter. Image with such a large number of pixels is computationally expensive to be modelled to learn the presence of entities in images. So, the next natural step is to reduce the size of an image to make it computationally tractable. So, all the extracted images in the data set are resized to $600 * 800$ dimensions as shown in Figure 4. The presence of a data set is a key step for judging the quality of algorithms. For that purpose, we also built a data set consisting of 1805, 1700, 1600, and 1550 training images for objects like Face, Weapon, Guard, and Intruder, respectively. Similarly, 695, 550, 500, and 500 testing images are selected for Face, Weapon, Guard, and Intruder, respectively.

The images used for feature extraction should be in grayscale format. This step converts the images into grayscale. For example, Figure 5 shows the color image, whereas Figure 6 shows the same transformed image in the grayscale form.

3.2.2. Features Extraction from the Image Frames and PCA. After resizing step, we need to extract dense VGG16 features from the images [16]. The VGG-16 is a well-known pre-trained CNN. VGG-16 model is trained on a subset of the ImageNet database of over one million images to classify images into 1k categories. The data set images and their respective annotation files are given for features extraction. PCA is a feature or dimensionality reduction technique that transforms original features and produces a new set of variables, namely, principal components (PC) [17]. Each PCs is a linear combination of the original variables, which are orthogonal to each other. So finally, extracted features are reduced in size using the PCA method to decrease the computational burden on the training model [18].

Reduced features from the image frames and their corresponding entity's locations and category identification are needed to be learned by some ML or CNN-based algorithm. Traditionally, Viola-jones [19] algorithm is used by researchers for object detection. However, for our purpose, we have chosen the FasterRCNN object detector [20], which is better and well known for its applications in different fields like cars/pedestrians detection [21], cancerous cell type identification [22], and even Urdu-text detection [22]. The basic working of FasterRCNN is shown in Figure 7. They are a little different from standard types of shallow networks because they have two outputs. The two kinds of outputs are classification for a given entity, and the other is the bounding box (location coordinates) prediction. Figures 8 and 9 show the example of bounding boxes created in the MATLAB application for our surveillance data set.

Therefore, for training, extracted reduced features and corresponding are given to the Faster-RCNN model. The learning rate was tuned to 0.001, with "Adam" as learning optimizer, and the value of minibatch size was 1. The model is trained for a different number of epochs to see its learning capability for different entities. The model is trained and tested from 50 to 200 epochs. The key categories are identified and located in our proposed methodology based on Four (4) classes/categories. These four categories are

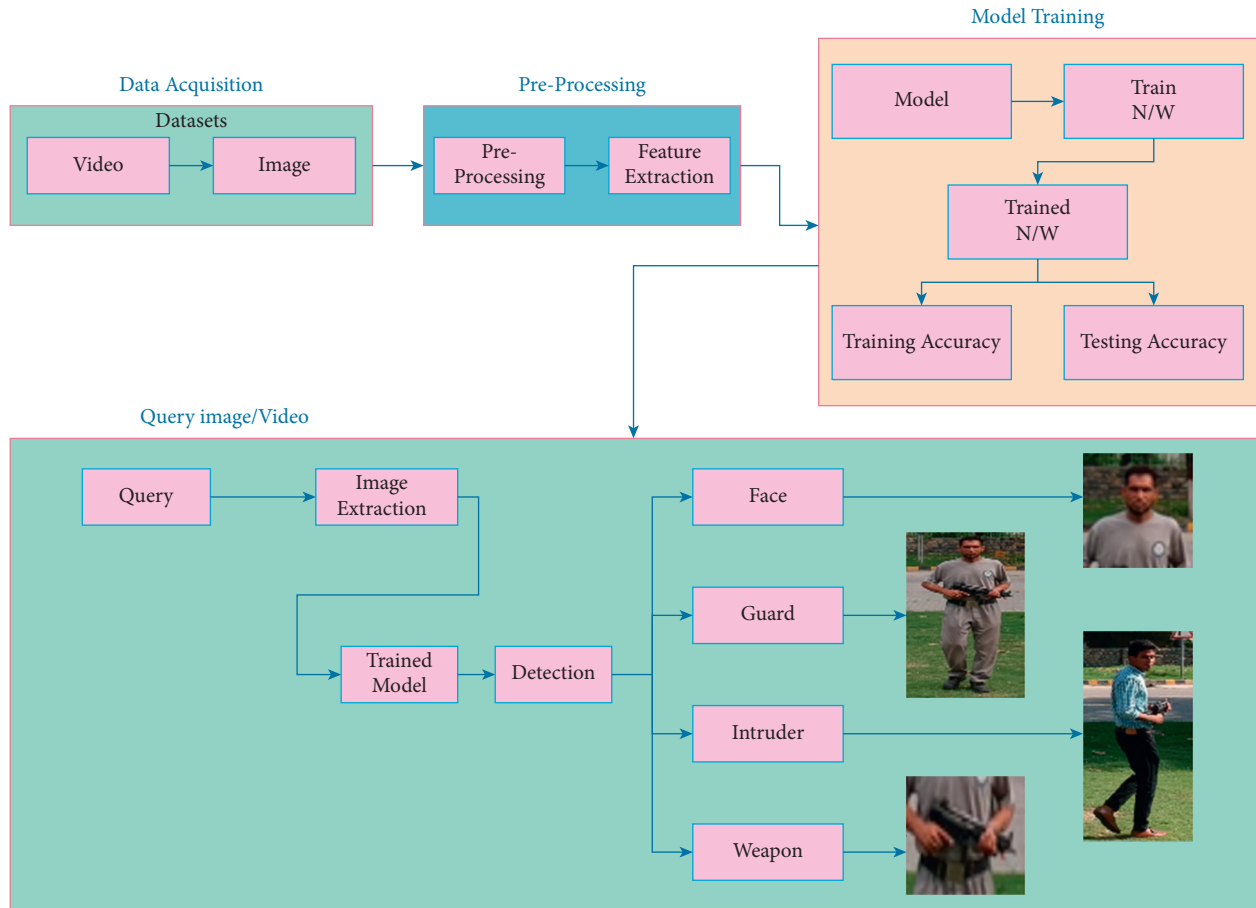


FIGURE 2: Block diagram and various phases of proposed methodology.

Face, Weapon, Intruder, and guard. Figure 10 shows how the 4 classes or categories are identified from a given image. The detection of weapons was at top priority, along with intruders. Based on our experiments, as shown in Table 1, we were successful regarding weapons but not best for an intruder. This shows that we further need to enhance our training and perhaps need to expand the data set. ConvNets are extensively used in deep learning [23] to solve various types of similar problems.

The biological structure inspires convolutional neural networks (ConvNet). A ConvNet comprises several layers, such as convolutional, max-pooling or average-pooling layers, and fully-connected layers. Ultimately, the trained model is used to enquire about the training and testing accuracy. Figure 11 presents the architecture of RCNN-based image region classification model. A trained detector was able to detect the Civilian, Guard, Person Face, and intruder.

For detection of the image, query videos and trained models are given for processing. It detects the images and shows the information of that image.

Figure 8 shows the label or ground truth of a civilian. Similarly, Figures 9 and 12 show the labels of guard and threat, respectively. These videos are labelled with the help of a ground truth labeller. FasterRCNN [24] model extracts potential regions for learning to detect whether a person is a

civilian, guard, or a threat at later stages with this labelled data.

Generally, the Object training algorithm can be subdivided into two subparts. One is a feature extractor, whereas the other is a classification and regressor (i.e., bounding box predictor), as shown in Figure 7. Although objects are initially detected through a trained object detector, the final decision regarding the presence of objects is finalized using Algorithm 1.

3.3. Pretrained Neural Networks. The other 4 pretrained neural networks used as a feature extracted in our experiments are discussed, and explanations are given to understand their role in object categorization with key information about them.

3.3.1. SqueezeNet. SqueezeNet is a deep neural network that uses the concept of squeeze block and expands blocks [25]. It has 80.3% accuracy for ImageNet. SqueezeNet objective was to build a smaller neural network with a reduced number of parameters compared with other architectures. The SqueezeNet layer involves layers of fire modules and several max-pooling layers as shown in Figure 13. The network takes an input of 227×227 pixels and can classify 1k categories.



FIGURE 3: Image with original size (2980 * 2021).



FIGURE 4: Resized image (600 * 800).

3.3.2. *GoogleNet*. GoogleNet is 144-layer pretrained CNN and takes in an image with the dimension of 224×224 pixels [26]. It introduced the concept of inception blocks in its architecture and is shown in Figure 14. It is the winner of the ILSVRC 2014 competition with a top-5 error rate of 6.67%. Nowadays, we can have 2 types of pretrained GoogleNet, one is trained on the ImageNet data set, and the other is trained on the Places365 data set.

3.3.3. *ResNet-18*. ResNet is residual network, and it works based on the concept of skip module [27]. ResNet-18 is a deep neural network (DNN) trained on more than 1 million images [25]. ResNet-18 originally has 18 layers, and its

detailed architecture is described in Figure 15. It can categorize images into 1 k classes, enabling this network to learn rich feature representations of various objects. ResNet-18 expects input images of size 224×224 pixels.

3.3.4. *ResNet-50*. ResNet-50 is an extension of ResNet-18 and is trained on the ImageNet data set of more than 1 million images [28]. This network has 50 layers and can classify images into 1 k classes. ResNet-50 processes an input image with a dimension of 224-by-224 pixels.

In its architecture, as shown in Figure 16, builders of this network have further connected the layers by bypassing 2 layers, and this step continues till fully-connected layer (F).



FIGURE 5: Color image representation.

Thus, bilayer jumping and connection between layers can be considered a Network-In-Network model and a key intuition that improves accuracy.

4. Results and Discussion

We trained and tested our built model using 4-different type object detectors trained on 50, 100, 150, and 200 epochs to know how well our proposed model object detector methodology works. For each model corresponding to epochs, the average precision for each kind of entity is determined. Table 1 succinctly describes the experimental outcome. Although Figure 17 graphically describes the results, increasing epochs improve the detection precision of each entity. As a result, our model best detects weapons and faces.

On the other hand, a guard is the third most effectively identified entity, whereas the lowest results were far intruder. As intruders can be a person, but on the other hand, a guard is also the person. The only difference is the presence or absence of a weapon, which in our case seems tricky for the learning algorithm to pick. There is a subtle difference between guard and intruder, which can hopefully be overcome with more training examples and extending methodology.

The results of the study by Angelov et al. [29] were compared based on speed, accuracy, and memory usage. Figure 17 shows the results for Detection Average Precision for four categories. We have compared the proposed method with previous work [30] and Table 2 and Figure 18 show that the method mentioned earlier [28, 29] run faster to some extent with fewer memory requirements, but these produce false positives. Both methods proposed in earlier studies [28, 29] approximate the background as a plane, so they are too oversimplified, and the AP for various object categories is shown in Figure 19. The proposed method uses less resources, and its performance is comparable with others.

The number of objects per frame vs the percentage of frames in the training, validation, and testing sets for object detection in videos is described in Figure 20.

4.1. Impact of Different Convolutional Neural Networks Features. Finally, detailed experiments were carried out for each kind of intended object shown in Table 3. The experiments were carried out using reduced vgg16 features for training the FasterRCNN. Later combinations of different CNN features are used for identifying the impact of various models on the detection of various intended entities and the time taken to train each model.

The models are compared using hold out ratios and again 4-different well-established models like SqueezeNet, GoogleNet, ResNet-18, and ResNet-50. The comparison shows the consistency of the data set for different target categories.

4.2. Detailed Results of Training and Testing for Various Object Detectors. Object entities were divided into two subcategories of train and test sets. Train and test set images contained a different number of objects per category. Four main types of object detectors were trained and tested with different parameters as given in Table 3. Furthermore, the results have been graphically visualized in Figure 21. Different models show consistent APs across different categories except for SqueezeNet. The SqueezeNet model has a small parameter size and is the most efficient, as checked in the existing literature. In fact, there is a tradeoff between model size and efficiency.

The image issues of the blur, jaggedness, and other problems, as we are using deep learning models. It implicitly accepted that deep learning models efficiently address occlusion, blur, jaggedness, and other problems. That is why, we have not discussed those tasks that come directly under



FIGURE 6: Grayscale image representation.

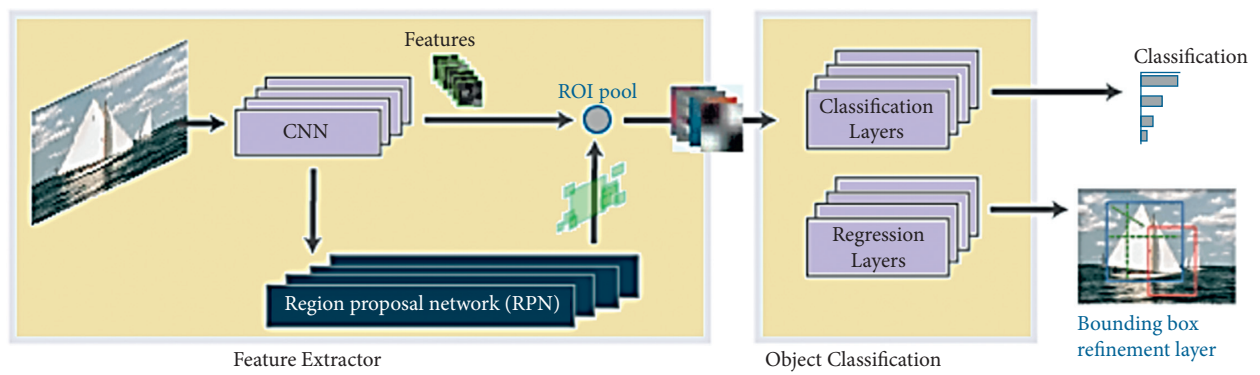


FIGURE 7: Fast R-CNN structure.

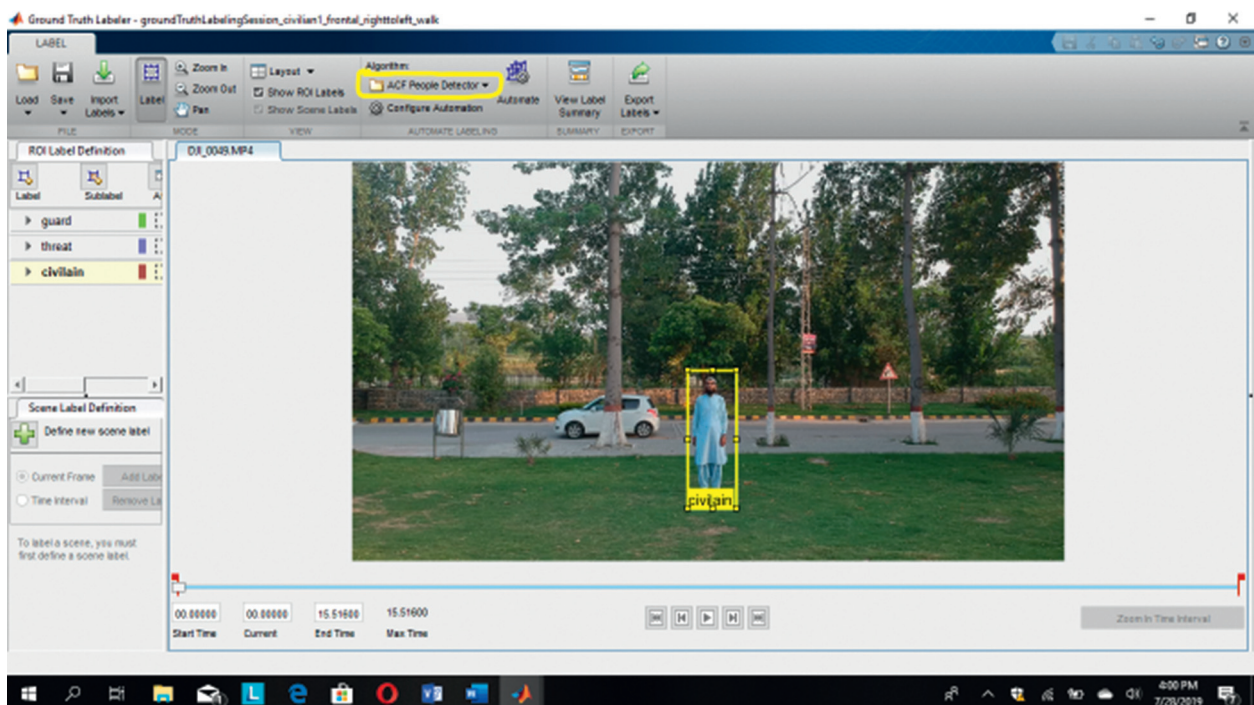


FIGURE 8: Ground truth labeller annotation of civilian.

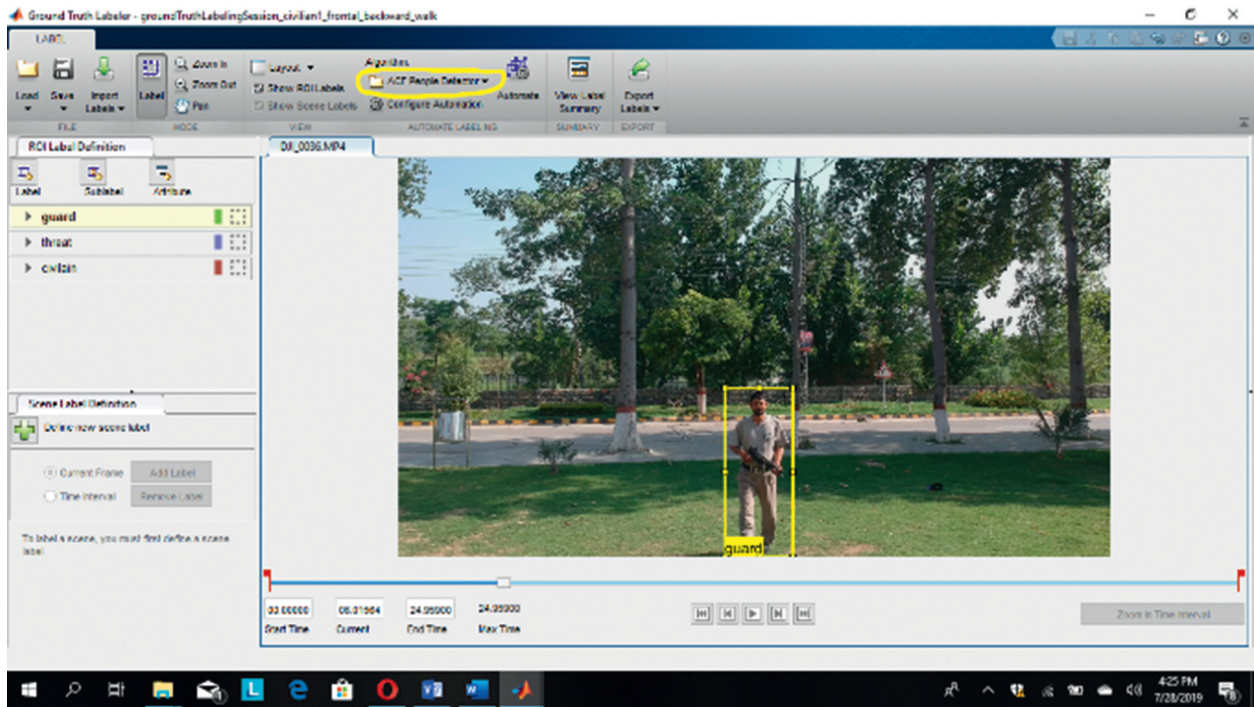


FIGURE 9: Ground truth labeller annotation of guard.

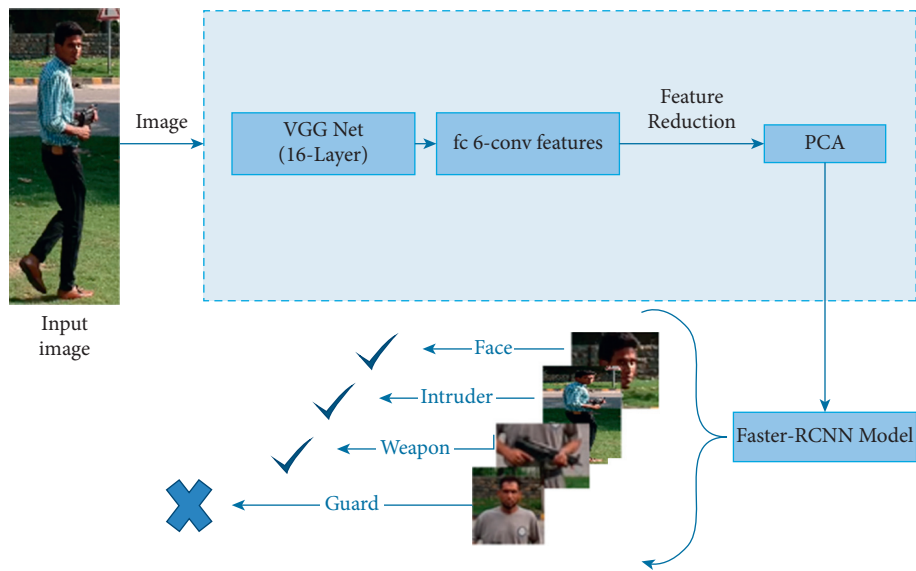


FIGURE 10: Reduced VGG16 features and detector model.

the umbrella of image processing. The experiments and results depict that the detection of the security surveillance using the drone is effective. The security surveillance is performed through deep learning architecture. CNN's Deep

Learning architecture SqueezeNet, GoogleNet, ResNet-18, and ResNet-50 are used. ResNet-50, based on FasterRCNN, is performed excellent at the data set and surpassed the results generated.

TABLE 1: Detection average precision (A) for given entities.

Object type	AP @ 50 epochs	AP @ 100 epochs	AP @ 150 epochs	AP @ 200 epochs
Face	0.60	0.73	0.78	0.83
Weapon	0.40	0.50	0.58	0.90
Guard	0.66	0.69	0.71	0.76
Intruder	0.3	0.44	0.50	0.67

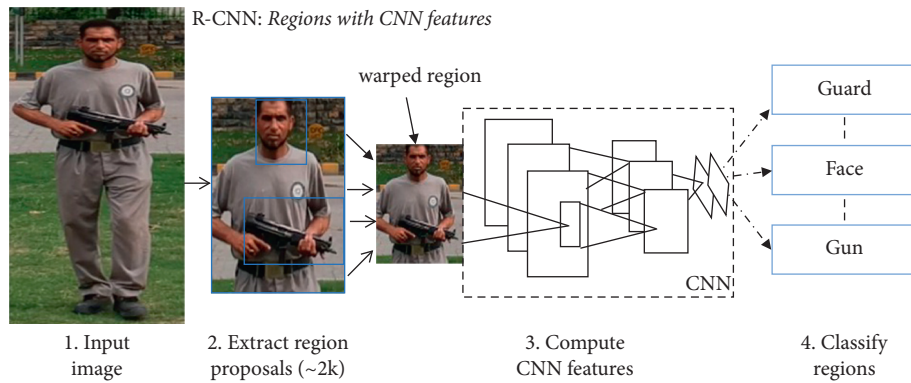


FIGURE 11: FasterRCNN-based image recognition model.

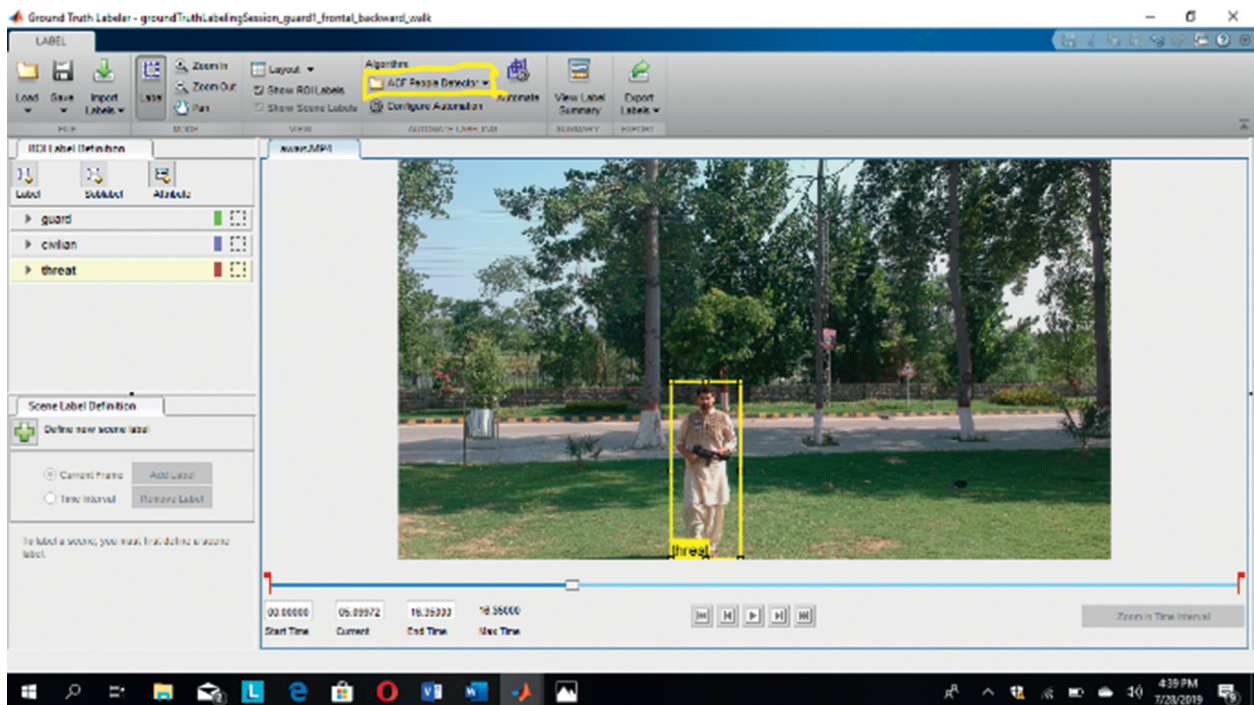


FIGURE 12: Ground truth labeller annotation of threat.

```

INPUT Image
OUTPUT Intruder Detected (Yes/No)
(1) procedure INTRUDER
(2)   for  $y = 1$  to  $images_n$  do
(3)     Predict candidate objects in  $y_n \rightarrow [F, iND, W, G]$ 
(4)     if  $F == true \ \&\& \ W == true \ \&\& \ G! = true$  then
(5)        $iND_{x,y} = true$ , highlight area(s) in  $y_n$ 
(6)       interrupt for  $y_n$  to main system
(7)     end if
(8)     if  $W == true$  then
(9)       Generate alert and keep-track of location in  $y_n$ 
(10)    end if
(11)  end for
(12) end procedure
    
```

ALGORITHM 1: Intruder detection algorithm.

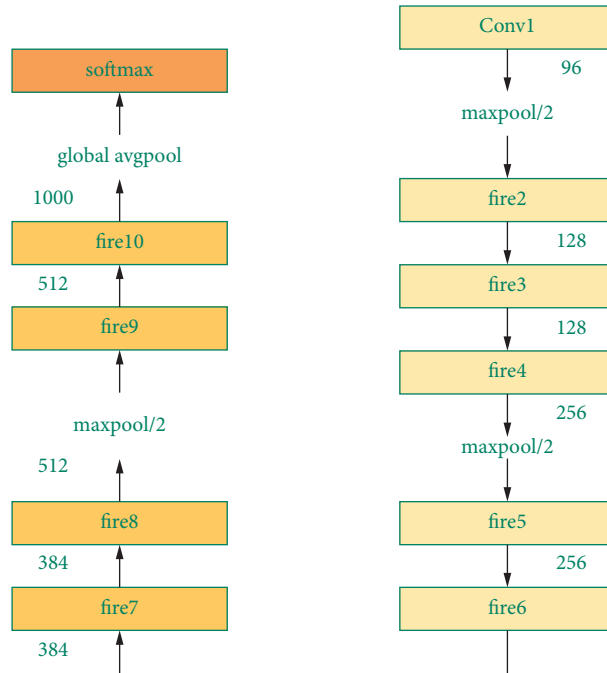


FIGURE 13: SqueezeNet architecture [25].

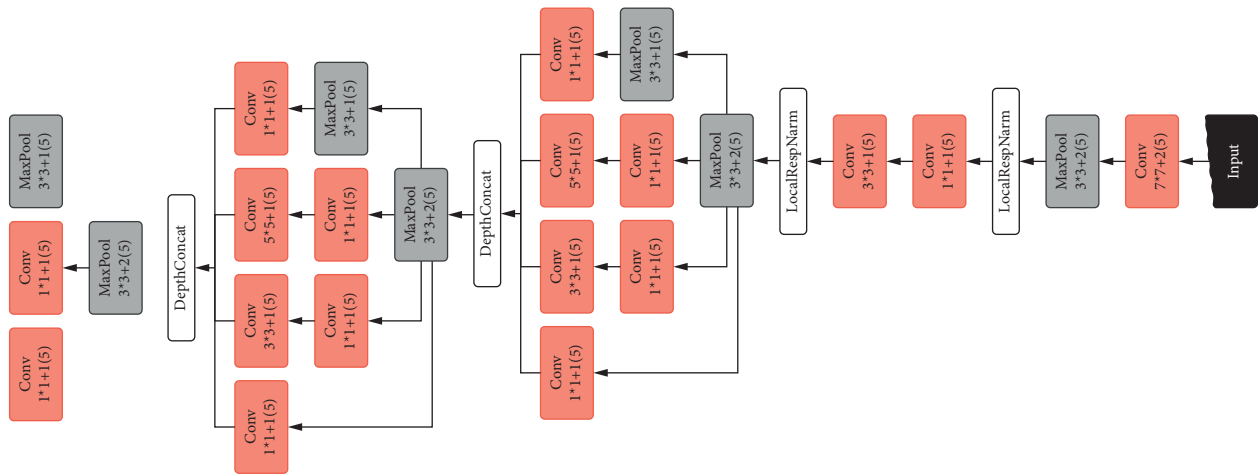


FIGURE 14: GoogleNet architecture [26].

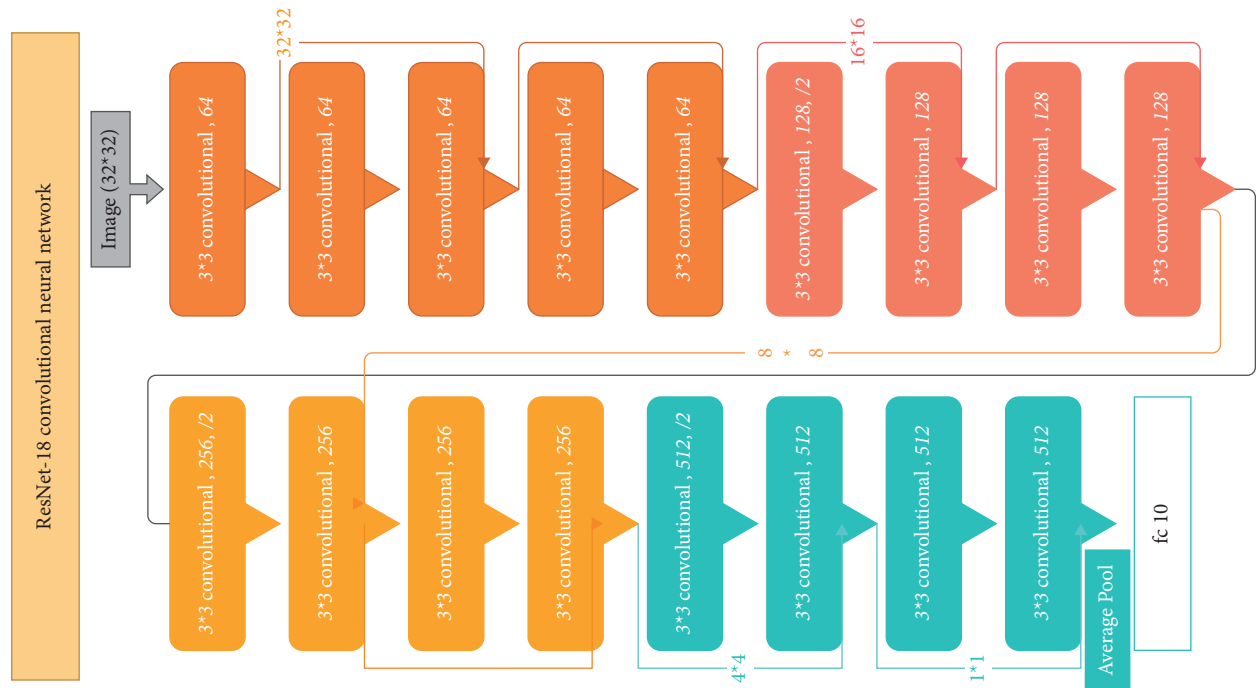


FIGURE 15: ResNet-18 architecture [27].

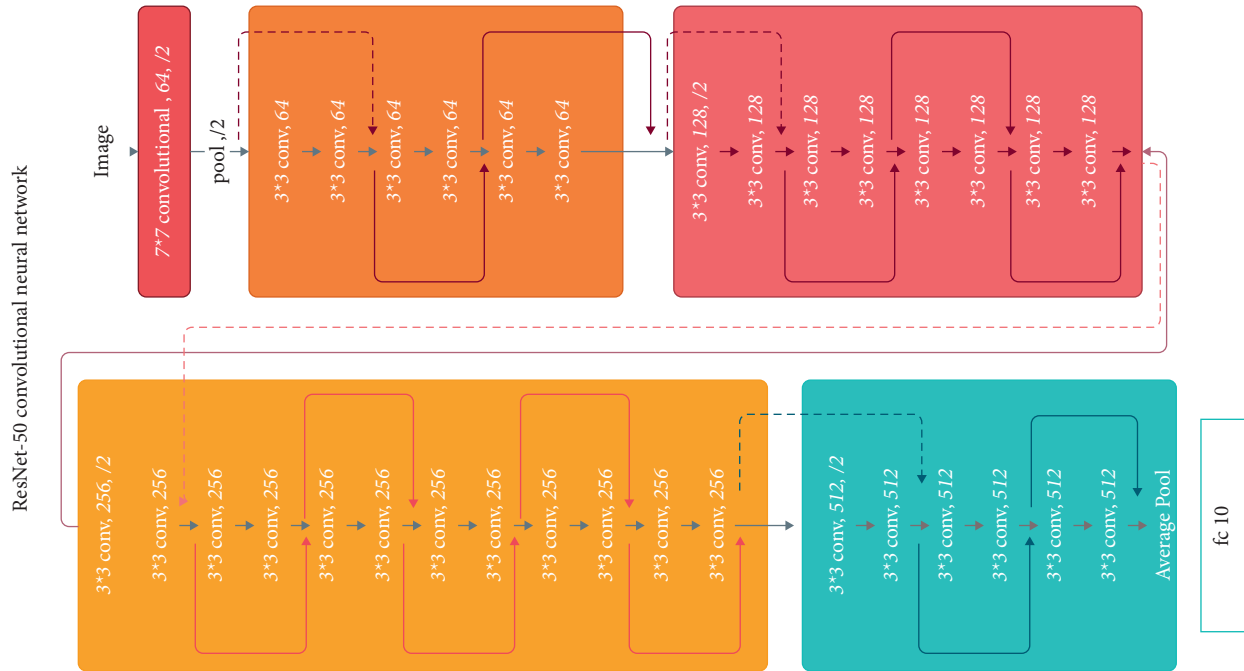


FIGURE 16: ResNet-50 architecture [28].

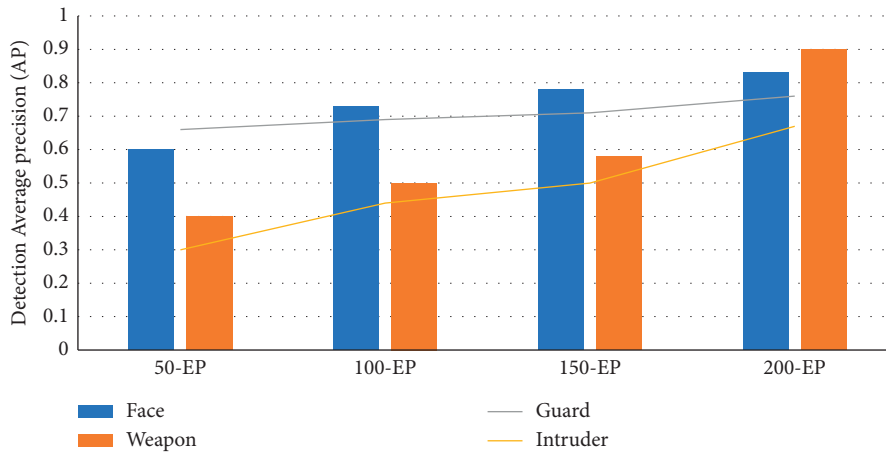


FIGURE 17: Detection average precision.

TABLE 2: Results comparison with existing work.

Method/ref no	Framerate (fps)	Error @ 50%	Memory (MB)
Ref [30]	20.6	0.442	8.5
Ref [29]	19.1	0.422	9.1
Ref [31]	16.2	0.341	10.5
Proposed	15.1	0.311	10.8

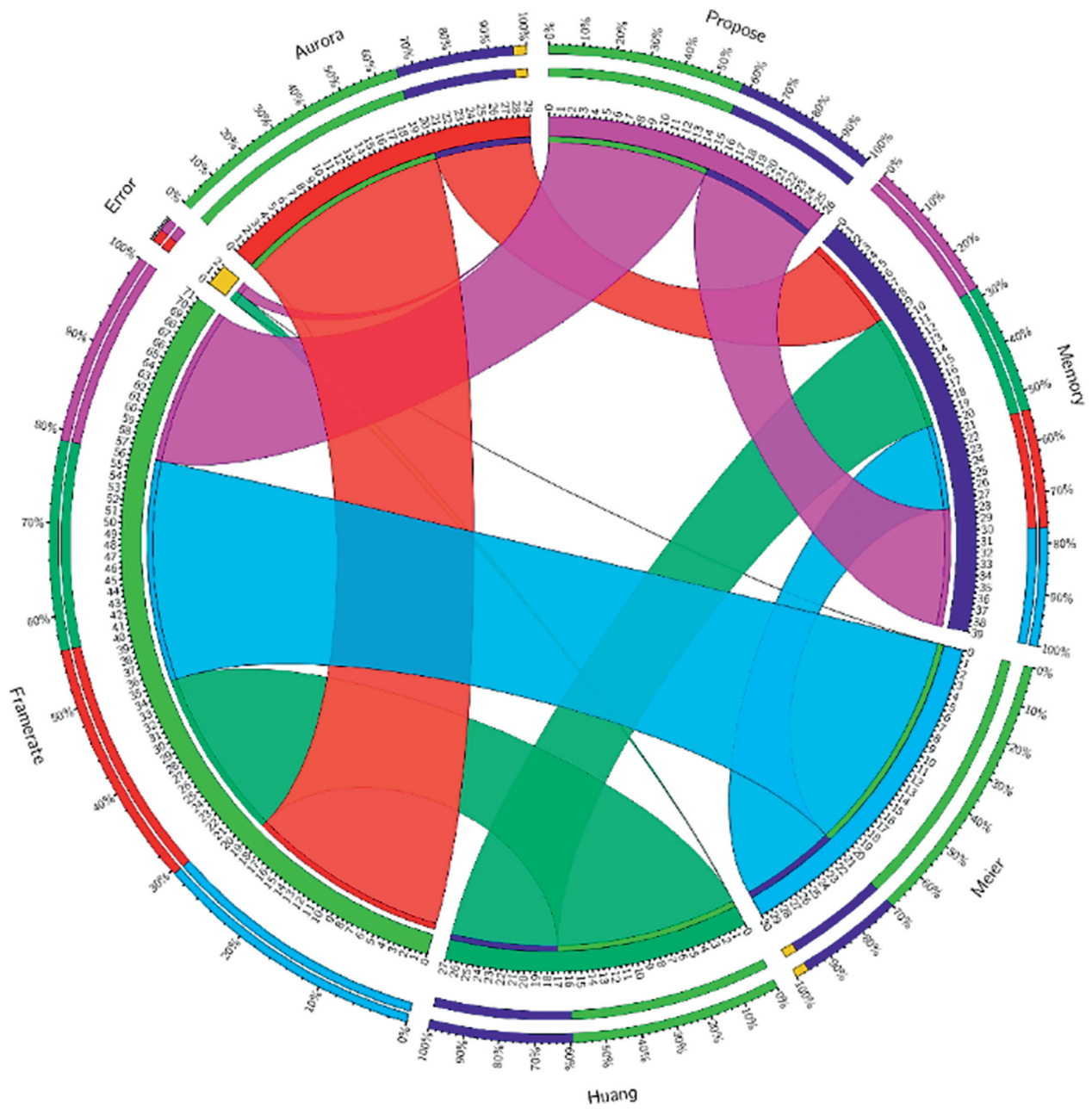


FIGURE 18: Circos graph results comparison.

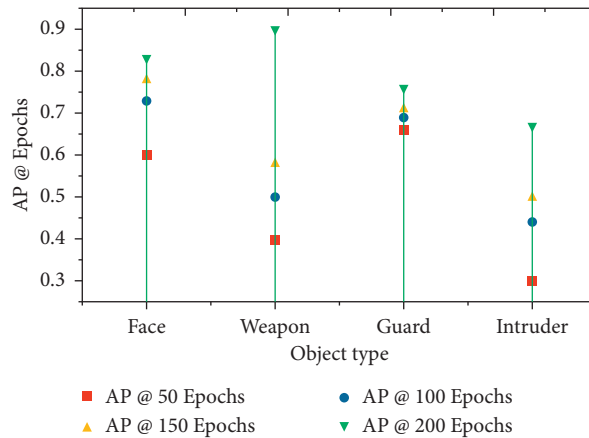


FIGURE 19: Average precision detection accuracy.

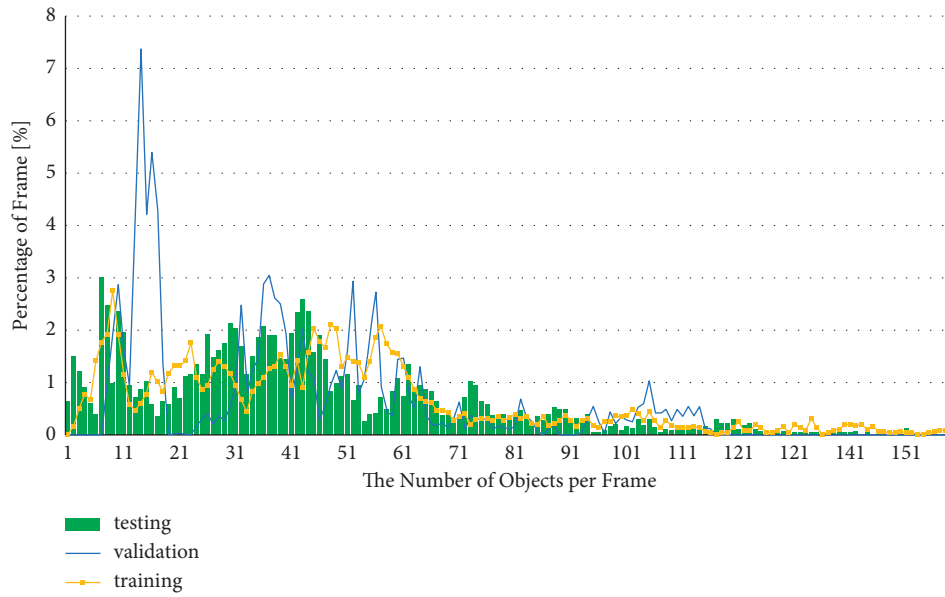


FIGURE 20: Number of objects per frame vs percentage of frames.

TABLE 3: Detailed results of training and testing for various object detectors.

Object detector	CNN feature extractors	No. of training images	Train-set AP	Time to train in hours	No. of testing images	Test-set AP
SqueezeNet-based FasterRCNN	Face	1805	0.22	13.86	695	0.23
	Weapon	1700	0.21	12.95	550	0.20
	Guard	1600	0.28	11.80	500	0.25
	Intruder	1550	0.20	11.56	500	0.18
GoogLeNet-based FasterRCNN	Face	1805	0.96	25.08	695	0.97
	Weapon	1700	0.94	23.33	550	0.96
	Guard	1600	0.97	22.01	500	0.95
	Intruder	1550	0.80	21.64	500	0.79
ResNet-18-based FasterRCNN	Face	1805	0.97	17.01	695	0.97
	Weapon	1700	0.95	15.23	550	0.94
	Guard	1600	0.96	14.85	500	0.93
	Intruder	1550	0.79	14.66	500	0.83
ResNet-50-based FasterRCNN	Face	1805	0.97	47.01	695	0.98
	Weapon	1700	0.96	45.20	550	0.97
	Guard	1600	0.97	44.09	500	0.96
	Intruder	1550	0.81	42.77	500	0.85

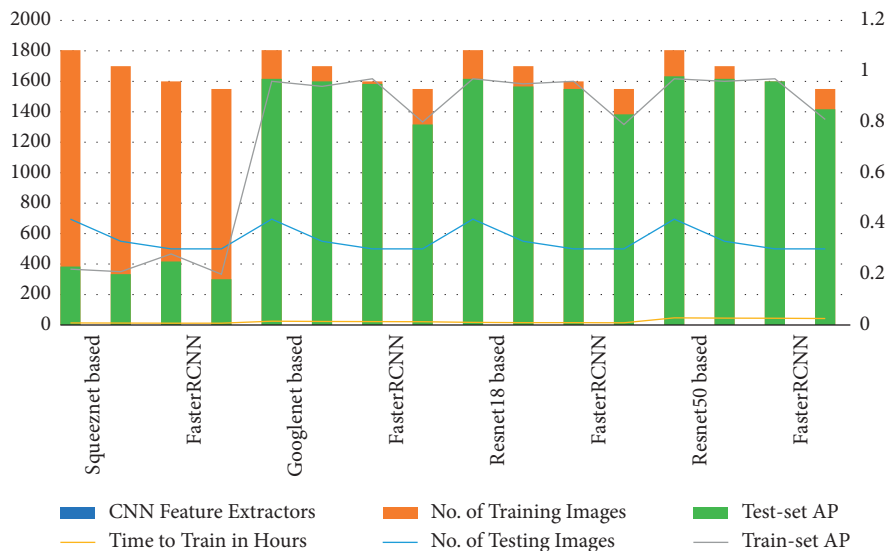


FIGURE 21: Train and test set average precisions (APs).

5. Conclusions

This research investigates a deep learning-based security and surveillance method for an organization, institution, or any other official procession. The surveillance system directly synchronizes the real-time data with the control room. The air surveillance system validates individuals and the community's protection by notifying the concerned departments about vulnerabilities to lay the foundation. Multiple experiments are carried out to validate the working system. Multiple experiments contained different kinds of CNNs feature to know their capabilities to detect various objects. Four well-known CNNs, namely, SqueezeNet, GoogleNet, ResNet-18, and ResNet-50, which are feature-based FasterRCNNs, were evaluated and reported. There are manifold advantages of an air surveillance system over the existing surveillance system used in our country. Firstly, the air surveillance system's pace will be optimum as they are aerial vehicles capable of roaming at any desired location. Secondly, air surveillance will be abandoned as these (UAV's) undergo surveillance irrespective of getting overtired or exhausted, contrary to those humans who need rest. Thirdly, this latest system will be systematic and least prone to errors, having not as many defaults as human calculations. Human beings are more liable to cause errors than machines. Finally, the air surveillance system is much cheaper than existing surveillance systems because the cost to deploy this system is comparatively less, and it does not involve assigning a task to human resources during surveillance because they are incorporated with sensors and are made intelligent. This system can also be utilized in monitoring the disaster-stricken area to assist while taking management initiatives afterwards. In the future, this work can be extended using the ensemble of other deep learning models, such as VGG19, InceptionV3, and Xception, along with the extension of the data set with fire and smoke categories.

Data Availability

The data used in the experiments will be available upon request.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors thank the University of Engineering and Technology Taxila and Higher Education Commission Pakistan for supporting this work under NRPU Project no. 6338.

References

- [1] D. He, S. Chan, and M. Guizani, "Drone-assisted public safety networks: the security aspect," *IEEE Communications Magazine*, vol. 55, no. 8, pp. 218–223, 2017.
- [2] S. Berrahal, J. H. Kim, S. Rekhis, N. Boudriga, D. Wilkins, and J. Acevedo, "Border surveillance monitoring using quadcopter UAV-aided wireless sensor networks," *Journal of Communications Software and Systems*, vol. 12, no. 1, pp. 67–82, 2016.
- [3] Z. Zaheer, A. Usmani, E. Khan, and M. A. Qadeer, "Aerial surveillance system using UAV," in *Proceedings of the 2016 Thirteenth International Conference on Wireless and Optical Communications Networks (WOCN)*, pp. 1–7.
- [4] J. S. Gadda and R. D. Patil, "Quad copter (UAVs) for border security with GUI system," *International Journal of Engineering Research and Technology*, vol. 2, no. 12, pp. 620–624, 2013.
- [5] J.-N. Lee and K.-C. Kwak, "A trends analysis of image processing in unmanned aerial vehicle," *World Academy of Science, Engineering and Technology*, vol. 8, no. 2, 2014.
- [6] T. Krajník, M. Nitsche, S. Pedre, L. Přeučil, and M. E. Mejail, "A simple visual navigation system for an UAV," in *Proceedings of the International Multi-Conference on Systems, Signals & Devices*, pp. 1–6, Chemnitz, Germany..
- [7] M. R. Haque, M. Muhammad, D. Swarnaker, and M. Arifuzzaman, "Autonomous quadcopter for product home delivery," in *Proceedings of the 2014 International Conference on Electrical Engineering and Information & Communication Technology*, pp. 1–5, 2014.
- [8] G. Ding, Q. Wu, L. Zhang, Y. Lin, T. A. Tsiftsis, and Y.-D. Yao, "An amateur drone surveillance system based on the cognitive Internet of Things," *IEEE Communications Magazine*, vol. 56, no. 1, pp. 29–35, 2018.
- [9] G. Amato, P. Barsocchi, F. Falchi et al., "Towards multimodal surveillance for smart building security," *Proceedings*, vol. 2, no. 2, 2018.
- [10] B. Alford, E. Curran, and S. Cole, "Determining the value of UAVs in Iraq," *The Journal of Conventional Weapons Destruction*, vol. 22, no. 1, 2018.
- [11] P. S. Bangare, A. Pote, S. L. Bangare, P. Kurhekar, and D. Patil, "The online home security system: ways to protect home from intruders & thefts," *International Journal of Innovative Technology and Exploring Engineering*, vol. 2, no. 3, pp. 2278–3075, 2013.
- [12] M. S. Munagekar, "Smart Surveillance system for theft detection using image processing," *International Research Journal of Engineering and Technology (IRJET)*, vol. 5, no. 8, pp. 232–234, 2018.
- [13] A. Alshammari and D. B. Rawat, "Intelligent multicamera video surveillance system for smart city applications," in *Proceedings of the 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 0317–0323, IEEE, uary, 2019.
- [14] P. Bhagyalakshmi, P. Indhumathi, and L. R. Bhavadharini, "Real time video surveillance for automated weapon detection," *International Journal of Trend in Scientific Research and Development (IJTSRD)*, vol. 3, no. 3, 2019.
- [15] A. Mahalanobis, J. L. Cannon, S. R. Stanfill, R. R. Muise, and M. A. Shah, "Network video image processing for security, surveillance, and situational awareness," in *Proceedings of the SPIE - The International Society for Optical Engineering*, pp. 1–8, FL, USA, August 2004.
- [16] A. Reghunath, S. V. Nair, and J. Shah, "Deep learning based customized model for features extraction," in *Proceedings of the 2019 International Conference on Communication and Electronics Systems (ICCES)*, pp. 1406–1411, 2019.
- [17] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1, pp. 37–52, 1987.

- [18] S. C. Ng, "Principal component analysis to reduce dimension on digital image," *Procedia Computer Science*, vol. 111, pp. 113–119, 2017.
- [19] W. Y. Lu and M. Yang, "Face detection based on viola-jones algorithm applying composite features," in *Proceedings of the 2019 International Conference on Robots & Intelligent System (ICRIS)*, pp. 82–85, Haikou, China, 2019.
- [20] S. Ren, K. He, R. Girshick, J. Sun, and R.-C. N. N. Faster, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.
- [21] T. Liu and T. Stathaki, "Faster R-CNN for robust pedestrian detection using semantic segmentation network," *Frontiers in Neurorobotics*, vol. 12, no. 64, 2018, in English.
- [22] R. Ezhilarasi and P. Varalakshmi, "Tumor detection in the brain using faster R-CNN," in *Proceedings of the 2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pp. 388–392, Palladam, India, ust, 2018.
- [23] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, 2018.
- [24] Matlab, "Getting started with R-CNN, fast R-CNN, and faster R-CNN," 2021, <https://www.mathworks.com/help/vision/ug/getting-started-with-r-cnn-fast-r-cnn-and-faster-r-cnn.html>.
- [25] O. Russakovsky, J. Deng, H. Su et al., "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.
- [26] C. Szegedy, Y. Jia, W. Liu et al., "Going deeper with convolutions," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, 2015.
- [27] A. S. Das, "CNN architectures: LeNet, VGG, GoogLeNet, R. A. More," 2019, Accessed: Mar. 20, <https://medium.com/sidereal/cnns-architectures-lenet-alexnet-vgggooglenet-resnet-and-more-666091488df5>.
- [28] O. A. Resnet50, <https://www.mathworks.com/help/deeplearning/ref/resnet50.html>.
- [29] P. Angelov, P. Sadeghi-Tehran, and C. Clarke, "AURORA: autonomous real-time on-board video analytics," *Neural Computing & Applications*, vol. 28, no. 5, pp. 855–865, 2017.
- [30] D. Meier, R. Brockers, L. Matthies, R. Siegart, and S. Weiss, "Detection and characterization of moving objects with aerial vehicles using inertial-optical flow," in *Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2473–2480, Hamburg, Germany, ember.
- [31] C. Huang, P. Chen, X. Yang, and K. T. Cheng, "REDBEE: a visual-inertial drone system for real-time moving object detection," in *Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1725–1731, ember.