International Journal of Sociotechnology and Knowledge Development
Volume 12 • Issue 1 • January-March 2020

Automatic Gun Detection Approach for Video Surveillance

Mai Kamal el den Mohamed, Faculty of Computer and Informatics, Cairo, Egypt
Ahmed Taha, Faculty of Computer and Informatics, Cairo, Egypt
Hala H. Zayed, Faculty of Computer and Informatics, Cairo, Egypt

ABSTRACT

The immense crime rates resulting from using pistols have led governments to seek solutions to deal with such terrorist incidents. These incidents have a negative impact on public security and cause panic among citizens. From this point, facing a pandemic of weapon violence has become an important research topic. One way to reduce this kind of violence is to prevent it via remote detection and to give an appropriate response in a short time. Video surveillance is the process of monitoring the behavior of people and objects. Surveillance systems can be employed in security applications as legal evidence. Moreover, it is used widely in suspicious activity detection applications. Intelligent video surveillance systems (IVSSs) are the use of automatic video analytics to enhance the effectiveness of traditional surveillance systems. With the rapid development in Deep Learning (DL), it is now widely used to address the problems existing in traditional detection techniques. In this article, an approach to detect pistols and guns in video surveillance systems is proposed. The presented approach does not need any invasive tools in the weapon detection process. It uses DL in the classification and the detection processes. The proposed approach enhances the obtained results by applying Transfer Learning (TL). It employs two different DL techniques: AlexNet and GoogLeNet. Experimental results verify the adaptability of detecting different types of pistols and guns. The experiments were conducted on a benchmark gun database called Internet Movie Firearms Database (IMFDB). The results obtained suggest that the proposed approach is promising and outperforms its counterparts.

KEYWORDS

Closed Circuit Television, Convolutional Neural Network, Deep Neural Network, Transfer Learning, Video Surveillance Systems

1. INTRODUCTION

The high rate of crime and violence among people is considered the third leading cause of death in 53 countries according to the report of the World Health Organization (WHO) European Region (Sethi et al., 2010). These alarming rates force governments to try to find solutions for such dangerous problems. Video surveillance systems are used for analyzing the objects behavior (Amira & Zagrouba, 2018). It involves object classification to understand the events (normal or abnormal) in videos. Abnormal activity detection plays a crucial role in surveillance applications (Huang et al., 2017; Wang et al., 2018; Cosar et al., 2017; Lloyd et al., 2017; Tripathi et al., 2019). The large-scale presence of surveillance systems is a real source of inspiration for the development of an automated system to detect problems of anti-social behavior such as vandalism, fights, gun killings, etc. In most current surveillance systems, monitoring depends on the existence of a human element. This

DOI: 10.4018/IJSKD.2020010103

makes monitoring a very challenging task. In addition, it is labor-intensive and prone to errors. These traditional systems have many problems such as weak security, low intelligence, high cost, and poor stability. Most of these systems are based on human operators. It is difficult for these operators to watch and analyze all the dangerous situations, especially with the long observation periods and a large number of cameras (Research, 2003). The reports included in (Research, 2003; Cohen et al., 2009; Dadashi, 2008) confirm that the Closed-Circuit Television (CCTV) operator suffers from video blindness after 20 to 40 minutes of active monitoring. In the last two decades, researchers and professionals of the industry have devoted their studies to develop surveillance systems that discover suspicious actions (Zhou & Tan, 2010; Liwei et al., 2010; Kishore et al., 2012; Mandrupkar et al., 2013). Automation is required in complex situations to reduce the workload of the human operator and improve the performance. Hence, surveillance systems still require intervention, improvement, and conversion from traditional surveillance systems to intelligent and smart systems (Shah et al., 2007; Tian et al., 2008). There is no human intervention at all in IVSS. The smart surveillance system automatically triggers an alert if any suspicious action or any illegal activity occurs. Accordingly, the operator focuses his attention only on the video feed and takes the convenient action.

The goal of the proposed approach is to design a system capable of automatically detect the presence of dangerous firearms especially, guns and pistols in real-time in the CCTV images. The proposed approach uses the Convolutional Neural Network (CNN) trained to determine the presence of the guns. CNN is a DL algorithm (Abdelouahab et al., 2018). DL is a subfield of machine learning. It is a technique that educates computers to perform what humans do naturally. Recently, with the emergence and successful deployment of DL techniques in image classification, researchers have emigrated from traditional techniques to DL techniques. DL has recently enriched its high ability in detection and classification. It has the ability to detect the dominant features automatically rather not manually (Tiwari & Verma, 2015; Halima & Hosam, 2016; Tiwari & Verma, 2015; Sheen et al., 2001; Xue et al., 2002; Li et al., 2008). This is the main reason prompted us to use it in our proposed approach. Nevertheless, DL suffers from two drawbacks: first, it requires very large datasets. Second, it needs high-performance computing resources. In order to overcome these two constraints. TL through fine-tuning is employed in the proposed approach. It is the improvement of learning in a new task through the transfer of knowledge from a learned task. TL means re-utilizing the knowledge learned from one problem to another one (Torrey & Shavlik, 2009). Network weights are initialized randomly if a network training is from scratch. However, the weights are initially set to the weights of the pre-trained network if fine-tuning is used. TL technique seeks to save time and get better performance. Figure 1 explains how TL improves the training performance rate.

In the proposed approach, DL has been employed to provide a greater level of performance than other traditional techniques (Mery et al., 2013; Blum et al., 2004; Upadhyay & Rana, 2014; Glowacz et al., 2015; Darker et al., 2007; Blechko et al., 2009; Darker et al., 2008; Arslan et al., 2015). The proposed approach allows the detection in noisy images with a low-quality resolution. Applying CNN in detecting firearms achieves efficient feature extraction results and accurate classification results. This increases the robustness of the presented approach. The presented approach uses two different pre-training networks (AlexNet and GoogLeNet). To avoid overfitting and to accelerate the process of training, the proposed approach uses TL. This training style demonstrates the ability of the TL in achieving tremendous results.

The overall organization of this paper is presented as follows: under section 2, the related work is explored. Section 3 describes the methodology of the proposed approach. The fourth section provides a detailed picture of the experimental results. The conclusion is approached in the last section.

2. RELATED WORK

Nowadays, automatic visual surveillance is an elementary need for security. Today, CCTV is employed as a monitoring and surveillance tool for fighting crimes. CCTV footage\films recently grow to be critical evidence in courts. All weapons, including firearms, pose very serious intimidation and risks to

Figure 1. The training performance with and without transfer learning (Torrey & Shavlik, 2009)

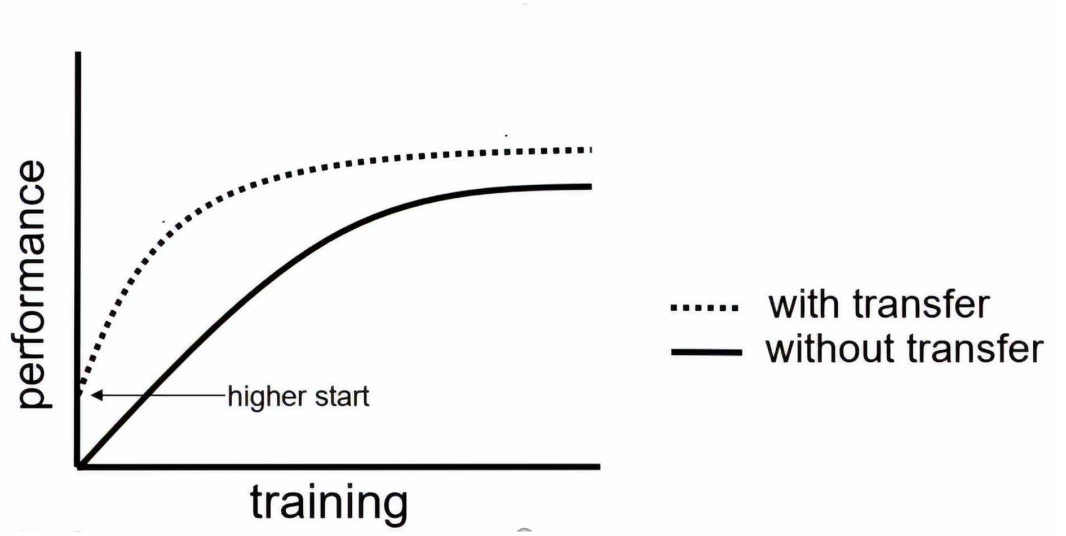
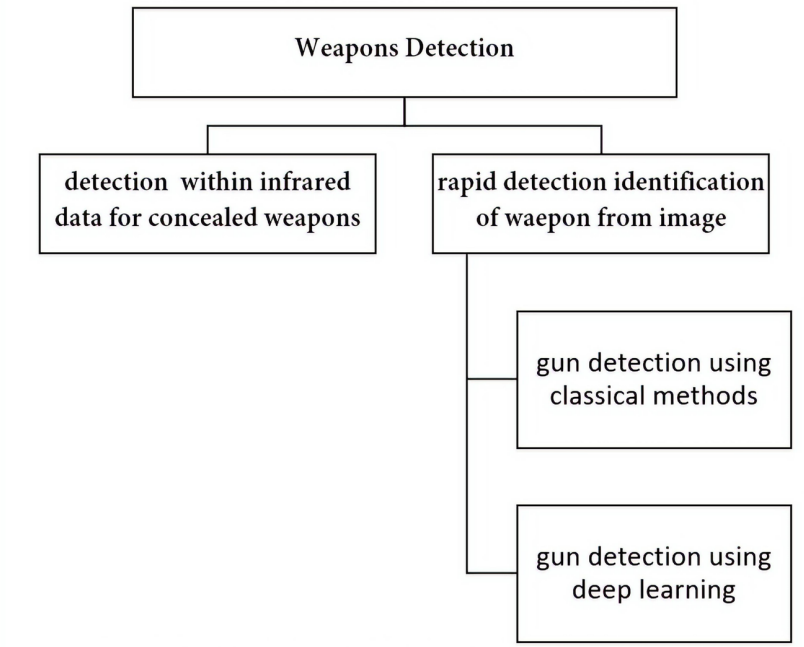


Figure 2. General description of the weapon detection techniques



the safety of the public places. Of all firearms, the handgun is the main weapon employed in different crimes. Therefore, gun detection has high importance for safety concerns. In attempts to combat gun crimes, many researches have emerged. By reviewing the published research in this area, we can classify the work into two main categories (as shown in Figure 2). The first is detecting concealed

weapons through X-ray and infrared machines. The second category is visual-based techniques that detect weapons in video images.

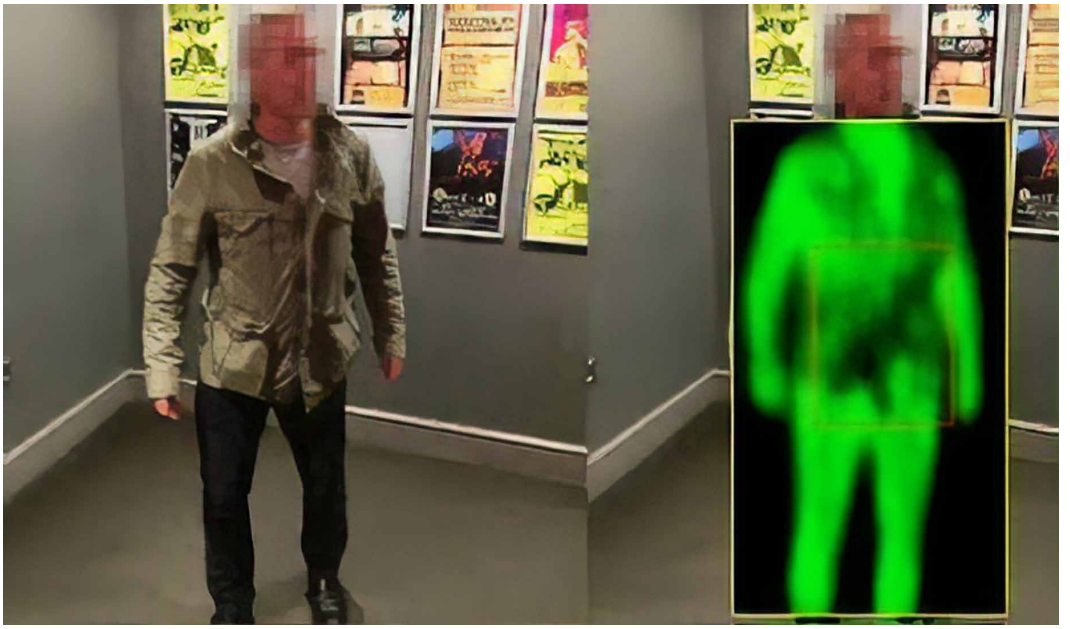
Most of the existing work concentrates on detecting the firearms using technologies like X-ray, millimetric wave and electromagnetic scanning. These methods work to identify when a weapon is found on a person's luggage as shown in Figure 3, or in a person's body as shown in Figure 4. Although there is little research in visual gun detection, there is a lot of research in the field of Concealed Weapon Detection (CWD). (Sheen et al., 2001) introduce a way to detect hidden weapons on a person's body inside airports and safe places. Their method depends on a 3-Dimensional millimeter wave imaging technique. Another CWD method is presented in (Xue et al., 2002). It uses a multi-scale decomposition based fusion method to detect hidden weapons. Also, another method proves the possibility of detecting metal objects like knives and guns by adopting microwave swept frequency radar (Li et al., 2008). Objects can also be identified using X-ray imaging, as interpreted by (Mery et al., 2013). Furthermore, (Blum et al., 2004) recommended a CWD method based on the integration of visual image and infrared (IR) or mm-wave image. The method depends on a multi-resolution mosaic technique. It uses the image mosaic to highlight the concealed weapon of the target image. To raise such a composite image that has a microscopic seam, an image mosaic method is used to combine two or more images. Another CWD method dependent on image fusion is also presented in (Upadhyay & Rana, 2014). The authors employ the fusion of IR and visual image to detect a concealed weapon in a situation and present the overexposed and underexposed area in the image scene. Their methodology consists of implementing a homeomorphic filter to visual and IR images, captured at different exposure condition. Moreover, (Glowacz et al., 2015) suggested a methodology for detecting knives by the baggage scanning system at airports and railway stations. Their method depends on an active appearance model and the Harris corner detector.

Although these techniques achieve high accuracy, they have some major drawbacks in real-world applications. First, CWD system is based only on metal detection and cannot detect non-metallic guns. Second, it is expensive. Finally, it is not accurate in all cases, because it interacts with metallic objects only.

Video-based weapon detection systems designed to work with surveillance systems are not widely available. The first-time surveillance cameras used to detect weapons was in 2007. (Darker et al., 2007) present a gun detection system as a part of the MEDUSA project in the United Kingdom. The

Figure 3. The weapon detected on a person's luggage

Figure 4. The weapon detected on a person's body



system is considered the foundation stone of developing automated gun crime detection systems. Later, the same team identified some gesture which indicates that an individual person is carrying a concealed firearm (Blechko et al., 2009). The same team also started their experiments to utilize the CCTV to detect the presence of pistols and firearms (Darker et al., 2008). Furthermore, (Arslan et al., 2015) present a solution that integrates two concepts to create a threat assessment system: the visual hierarchy and the theoretical ontology of firearms. (Halima et al., 2016) establish the Bag of Words Surveillance System (BoWSS) algorithm to detect guns. The method begins by extracting features using Scale-Invariant Feature Transform (SIFT), then clustering the attained functions using a k-means algorithm. Later, it employs Support Vector Machine (SVM) for the training. In addition, the authors in (Tiwari & Verma, 2015) implement a visual gun detection approach using SIFT and the Harris interest point detector. Their system segments the recognizable object from an image by a k-means clustering algorithm. Then, Harris interest point detector and Fast Retina Keypoint (FREAK) are exploited to explore the weapon in the segmented images. It handles the object detection challenges like scaling, rotation, and occlusion. It achieves an accuracy equal to 84.26%. Furthermore, the researchers in (Tiwari & Verma, 2015) implement gun detection system based on Speeded Up Robust Features (SURF) interest point detector. An object is marked as a gun if half of the features of weapon descriptor are matched with the SURF features of the object. The accuracy achieved using SURF descriptor was 88.67%. However, all of previously mentioned methods suffer from drawbacks. some of them are mathematically complicated and computationally heavy and thus they are time consuming. Others are not effective for low powered devices. Furthermore, most of them are slow and not good with illumination changes.

The first work that uses DL in detecting firearms is presented by (Olmos, 2018). The work addresses the first solution using the DL. This method uses DL as a classifier on their datasets within a sliding window and region proposals detection-based methods. This method achieved recall equal 100% and precision equal 84.21%. The next work (Castillo et al., 2019) develops an automatic cold steel weapon detection model that used CNN for video surveillance. The next system (Verma & Dhillon, 2017) builds a gun detection system that uses CNN based VGG-16 architecture as a feature extractor. The accuracy of this method was 93%.

3. THE PROPOSED APPROACH

Actually, feature extraction is a critical key factor in producing an effective recognition system. Traditional feature extraction is really a time-consuming and boring mechanism. In addition, it has not the ability to process raw images. However, automatic extraction techniques have the ability to extract the features immediately from raw images. DL is one of the best of these techniques (Liu et al., 2017). It is considered as a landmark in extracting features automatically. DL is featured by its robust capability of feature representation and feature learning compared with the traditional object detection techniques. Figure 5 illustrates how the performance of DL surpasses traditional techniques.

In this context, a DL based approach is proposed to detect pistols and guns in video surveillance systems (VSS). In general, DL techniques attempt to generate better representations and models from large-scale data. There are many architectures of DL. Generally, DL is divided into four categories (Guo et al., 2015; Zhang et al., 2018; Mardi et al., 2019; Lakhtaria & Modi, 2019): CNN, Restricted Boltzmann Machines (RBMs), Autoencoder and Sparse Coding. The categorization of DL methods is shown in Figure 6.

The proposed approach uses CNN as a DL model. The CNN general architecture is shown in Figure7. CNN has attained immense success in image recognition task by automatically generating the hierarchical feature representation of the image from the raw data. It is a special kind of multi-layer-feed-forward neural networks (Mane & Kulkarni, 2017). It is designed to recognize visual patterns straightforwardly from images with minimal preprocessing. It has the ability to learn complex, high-dimensional, and non-linear mappings from an extremely huge number of data (images). CNN architecture involves one input layer and multi-types of hidden layers in addition to one output layer. Generally, CNN has many distinctive characteristics. It is easier to train and it has fewer parameters than other networks. In addition, using CNN leads to savings in memory requirements and computation complexity requirements. Furthermore, CNN gives better performance for applications where the input has a local correlation (e.g., image). Finally, CNN does not only afford the best performance

Figure 5. Diagram explains the performance of DL with respect to other traditional techniques (Torrey & Shavlik)

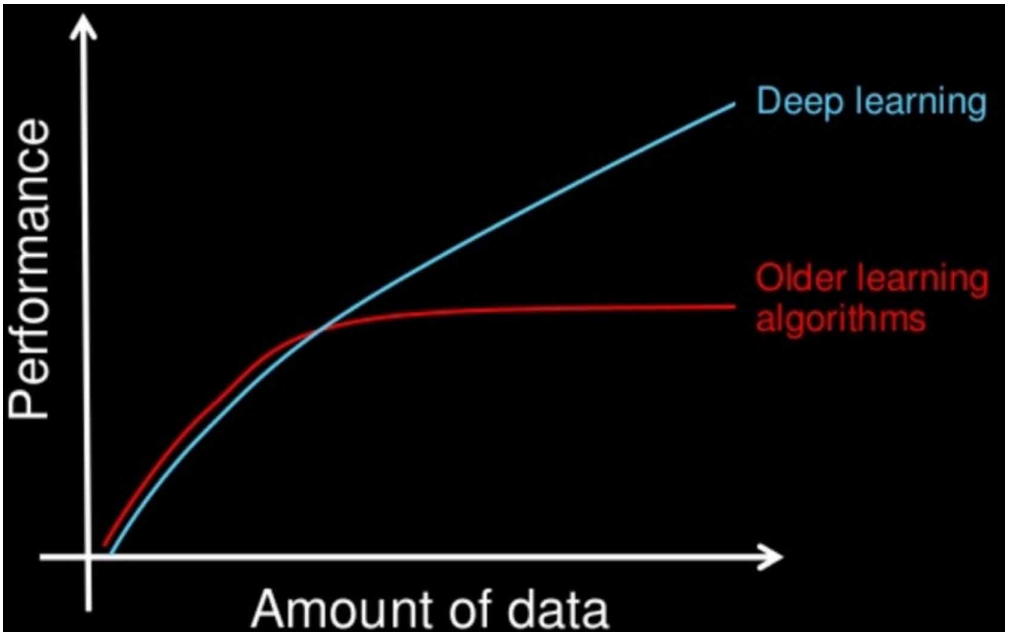
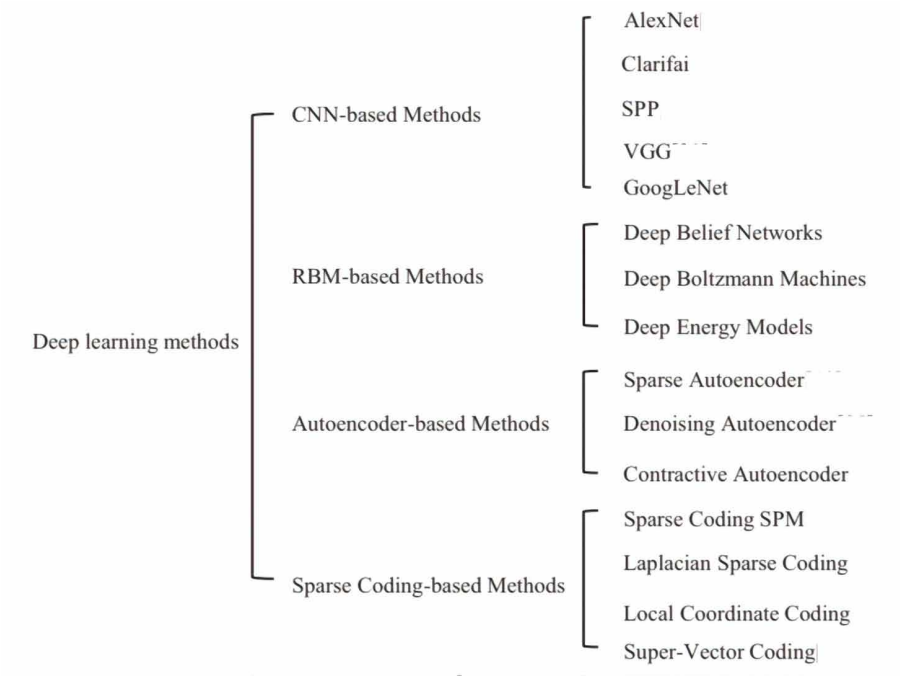


Figure 6. Categorization of the DL methods and their representative works (Guo et al., 2015)



compared to other detection algorithms, it even outperforms humans in some cases such as classifying objects into fine-grained categories.

The complete block diagram of the proposed approach that uses CNN to solve the pistol and gun detection problem presents in Figure 8. It consists of six steps. The first step captures the image by a CCTV camera. The second step resizes the RGB image to be suitable to work with CNN. The third step divides the dataset into two groups: learning group with 70% and testing Group with 30%. The fourth one applies the TL technique to avoid work from scratch and accelerates the training process. The next step applies CNN to find the rich features in images. The final step uses these features to train the image classifier.

Based on the steps of the presented algorithm shown above, two different deep learning techniques are employed. The first technique uses AlexNet while the second technique uses GoogLeNet. We

Figure 7. General architecture of CNN

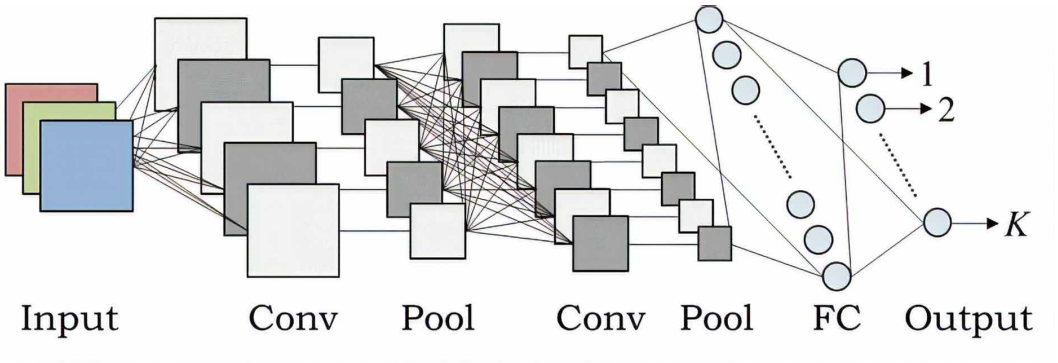
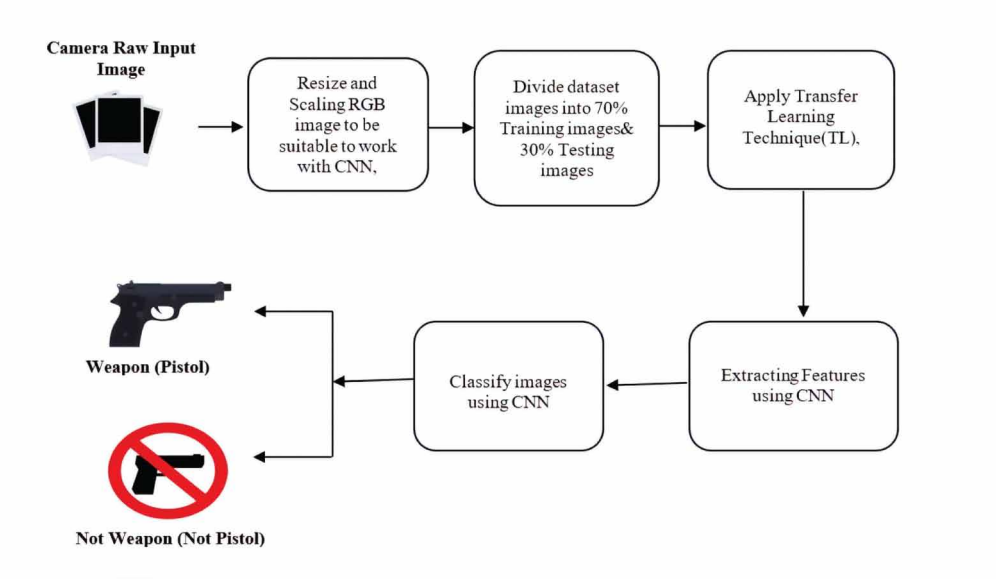


Figure 8. Block diagram of the proposed approach for classifying image pistol/gun or not



applied and tested these steps on the two pre-trained networks and recorded their results in detail. These two pre-trained Networks give the best performance compared to other detection techniques. Each step mentioned in the proposed algorithm will be explained in more details in the next two subsections with AlexNet and GoogLeNet.

One of the limitations of using DL is that it requires a large number of training data. This is because the learning algorithm needs to tune a huge number of parameters. The problem of deep learning is that it starts with a poor initial state and then uses some gradient-based optimization algorithm to converge the network to an optimal solution. To overcome this limitation, the proposed algorithm uses the TL to achieve successful learning in addition to accelerate the overall learning process. TL share the learned weights of another learned network. DL avoids all the drawbacks of the traditional techniques in terms of slowness, illumination changes, and cost time which resulting from being mathematically complicated.

3.1. AlexNet Based Technique for Pistols Detection

AlexNet won Large Scale Visual Recognition Challenge (ILSVRC) 2012, attaining the highest classification performance (Krizhevsky et al., 2012). AlexNet is 8 layers deep with 5 convolutional (conv) layers and 3 fully connected (FC) layers. It is rapid for retraining and classifying new images. It is trained using more than 1000000 images. Moreover, AlexNet classifies images into 1000 object categories. It receives an image as an input and produces a label for the object in the image as an output.

3.1.1. Preprocessing

AlexNet requires the inputs (RGB images) to be of fixed size during both the training and testing processes. In order to achieve this requirement, all images are re-scaled to a fixed image size equal to 227×227. Then, the dataset images are divided into two groups: the first group is "Learning Group" which is used for creating the model in the learning mode. It contains 70% of the dataset (Training images). The second group is "Testing Group" which is used for estimating the classification quality in the testing mode. It contains the remaining 30% of the dataset (Testing images). The total images in the dataset used are 13743 images. Table 1. gives a complete description of the images with respect to both training and testing sets.

Table 1. Total images divided into two classes

Type	Total Images		
	Total No. of Images	Training Images No.	Test Images No.
PE	4099	2869	1229
NE	9644	6750	2893

3.1.2. Transfer Learning

TL is the task of using the knowledge equipped by a pre-trained network to learn new patterns in new data. TL improves the training performance rate (Sharma & Kumar, 2019). This performance rate can never be reached if the training begins from scratch. TL is employed in the proposed technique because of its tremendous ability to reduce the learning time and to enhance results.

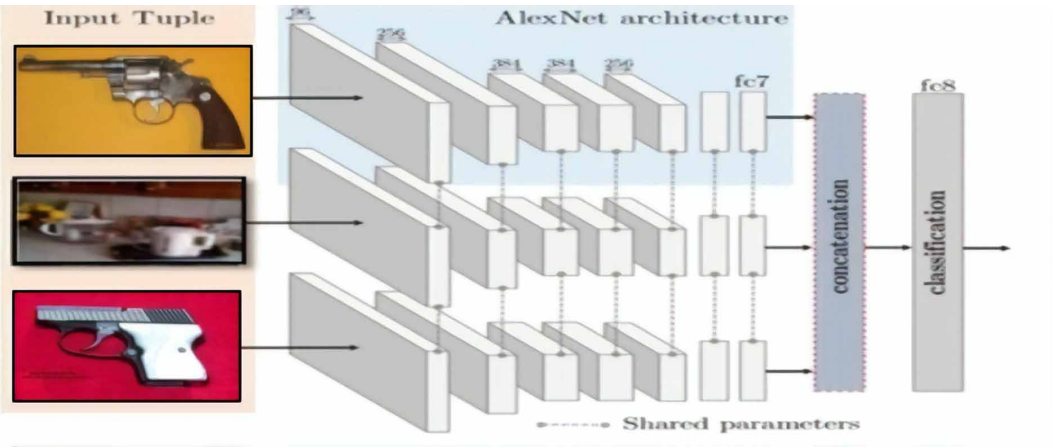
3.1.3. Extracting Features and Classification

In this step, pre-trained networks are used to extract rich features and to accomplish the classification process. Figure 9. shows image features extracted from the AlexNet pre-trained network. AlexNet starts with convolutions, then max pooling. There are two types of pooling, average pooling, and max pooling. Max pooling is the best because it accelerates the learning process and it is more effective in computation. After that, it attaches ReLU activation function after every convolutional and fully-connected layer. ReLU activation function is very fast and is easier in computation than other activation functions. Also, it is characterized by its ability to vanish gradient problem. The final result of AlexNet is the class label that describes the result of the classification process: positive or negative.

3.2. GoogLeNet Based Technique for Pistols Detection

GoogLeNet won ILSVRC 2014. It is 22 layers deep (Szegedy et al., 2015). GoogLeNet has the ability to attain great accuracy using limited computational cost. GoogLeNet is smaller and faster than VGG networks.

Figure 9. The training process of the first proposed technique with negative and positive dataset images



3.2.1. Preprocessing

GoogLeNet requires the inputs (RGB images) to be of fixed size during both the training and testing processes. In order to achieve this requirement, all images are resized to a fixed image size to be 224×224. Then, the dataset images are divided as shown in Table 1 into two groups: the first group is "Learning Group" which is used for creating the model in the learning mode. It contains 70% of the dataset (Training images). The second group is "Testing Group" which is used for estimating the classification quality in the testing mode. It contains the remaining 30% of the dataset (Testing images). The total images in the dataset used are 13743 images. Learning group contains 2869 positive images and 6750 negative images. On the other side, the testing group contains 1229 positive images and 2893 negative images.

3.2.2. Transfer Learning

TL is the task of using the knowledge equipped by a pre-trained network to learn new patterns in new data. The goal of transfer learning is to improve learning in the target task by leveraging knowledge from the source task. It saves time and facilitates the learning process. TL technique helps us to use fine-tune existing networks that are trained on a large dataset. Then, to continue the training on our smaller dataset, TL is employed in the proposed technique.

3.2.3. Extracting Features and Classification

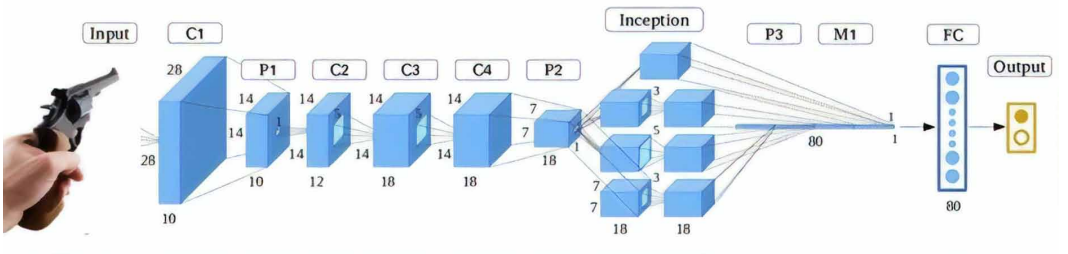
GoogLeNet module depends on several very small convolutions in order to reduce the number of parameters. In the proposed technique, pre-trained networks are used to extract rich features and also to complete the classification process. Figure10 describes the feature extraction process using GoogLeNet.

4. EXPERIMENTAL RESULTS

Performance evaluation is a very crucial task at the end of the development process. It is significant to show that the proposed approach achieves an acceptable level of performance and it achieves a significant improvement over existing algorithms. The proposed approach is implemented using the Matlab 2017b. The implementation is running under 64-bit Windows 10 operating system on a computer with Intel Core i7 processor and 16GB RAM.

The Internet Movie Firearms Database (IMFDb) is an online database of firearms launched in May 2007 by "Bunni". It contains different scenes for individuals carrying pistols or guns collected from certain movies, television series, video games, and anime. IMFDb contains many categories, one of these categories is gun (http://www.imfdb.org/wiki/Category:Gun). Gun has 27 subcategories such as Assault Rifle, Flare Gun, Fictional Firearm, Machine Gun, Revolver, Rifle, Pistol, etc. This is considered a benchmark gun database.

Figure 10. The training process of the first proposed system with negative and positive dataset images



In order to evaluate the proposed approach, positive and negative data are used. We limited the problem of weapons detection in our proposed approach into only two classes: positive class and negative class. The positive class contains 4099 images downloaded from the pistol category in the gun category of IMFDB database (http://www.imfdb.org/wiki/Category:Pistol). Figure 11 shows some samples from this positive dataset. The negative class contains 9644 images collected from the Internet (http://kt.agh.edu.pl/~matiolanski/KnivesImagesDatabase/) and (https://sci2s.ugr.es/weapons-detection#TestSet). Figure 12, shows some samples from this negative dataset images.

The following metrics are widely used to evaluate this kind of applications. The metrics includes True Positive Rate (TPR), True Negative Rate (TNR), Positive Predictive Value (PPV), False Omission Rate (FOR), recall, precision, and accuracy. These measurements are calculated by the following equations (Vajhala et al., 2016):

Figure 11. Sample images from IMFDB database

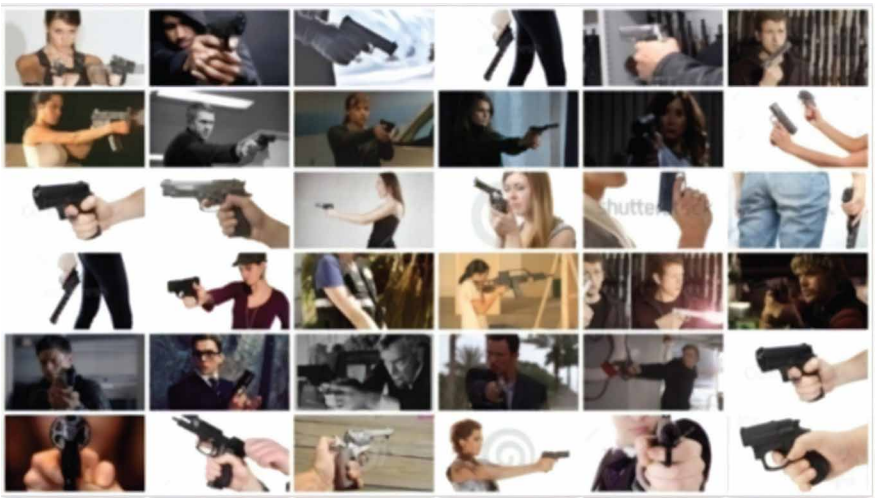
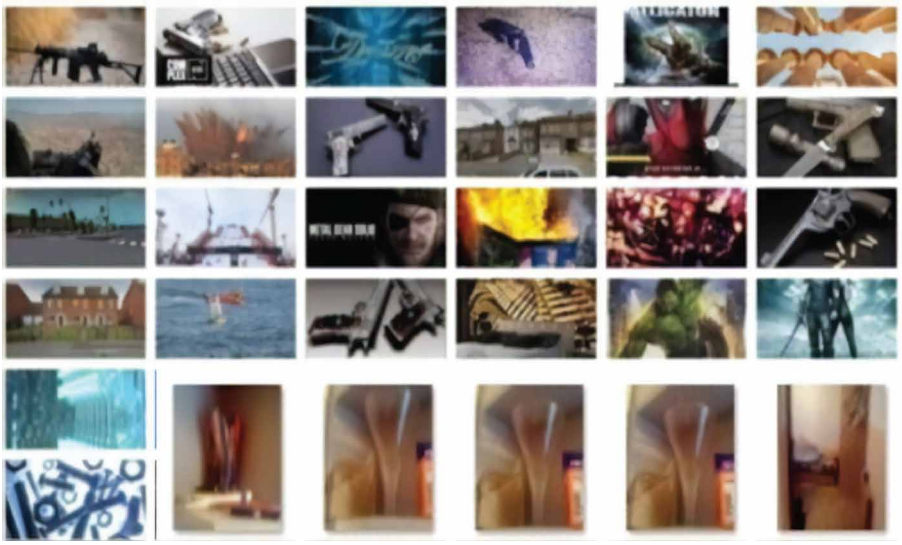


Figure 12. Sample images from negative images



$$\text{TruePositiveRate } \left(TPR\right) = \frac{TruePositives}{TotalPositives} \tag{1}$$

$$\text{TrueNegativeRate } \left(TNR\right) = \frac{TrueNegatives}{TotalNegatives} \tag{2}$$

$$\text{PositivePredictiveValue } \left(PPV\right) = \frac{TruePositives}{TruePositives + FlasePositives} \tag{3}$$

$$\text{FlaseOmissionRate } \left(FOR\right) = \frac{TrueNegatives}{TrueNegatives + FlaseNegatives} \tag{4}$$

$$\text{Recall } \left(\text{Sensitivity}\right) = \frac{TruePositives}{TotalPositives} \tag{5}$$

$$\text{Specifcity} = \frac{TrueNegatives}{FalsePositives + TrueNegatives} \tag{6}$$

$$\text{Accuracy} = \frac{TruePositives + TrueNegatives}{TotalPositives + TotalNegatives} \tag{7}$$

The previous metrics depends on four values: True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN). A True positive means that the object is positive (gun) and the classifier output is also positive (gun). A False positive means that the object is negative (not-gun) and the classifier output are positive (gun). A True negative means that the object is negative (not-gun) and the classifier output is negative (not-gun). A False negative means that the object is positive (gun) but the classifier output is negative (not-gun).

The evaluation of the presented approach is performed into two parts. The first part evaluates the technique that uses AlexNet while the second part evaluates the technique that uses GoogleNet. Depending on the previous four outcomes, Table 2 shows the result of the confusion matrix of the first technique that employs AlexNet as a pre-trained network to detect pistols. The first column in the confusion matrix represents pistols while the second column represents negative images (not-pistols). The first row represents pistols, and the second row represents not-pistols. The green cells represent correctly classified images and red cells represent false classified images. From this confusion matrix, the accuracy of Alex Net classifier is 99.2%.

The performance of the proposed technique is tested against varying conditions such as occlusion, assorted background with guns, etc. The performance of the technique is evaluated with the previous

Table 2. Confusion matrix for weapon detection using AlexNet

Actual Class			
		Pistol (P)	**Not a Pistol (NP)**
Predicted Class	**Pistol (P)**	0.9902	0.0098
	Not a pistol (NP)	0.0059	0.9941

seven parameters: TPR, TNR, PPV, FOR, recall, precision, and accuracy according to the Equations 1-7, respectively as shown in Table 3. The results of the AlexNet based technique proved its ability to detect pistols in different conditions like occlusion, and varied background of the gun. This classifier produces a low number of false negatives equal to 17. Furthermore, the true negative rate achieves low result 29.7%, while precision and accuracy achieve respectively 99.5% and 99.22%. These results prove the efficiency of the proposed technique. This is due to the good initialization of the weights done by TL. In fact, TL decreases the time of learning process and achieves very high results.

Another group of experiments is conducted to evaluate GoogLeNet based technique. We considered also two classes and trained the classification model on the same database used in evaluating the first AlexNet based technique. Table 4 shows the confusion matrix of the second technique that employs GoogLeNet as a pre-trained network to detect pistols. The green cells in this matrix represent the correct classified images and red cells represent false classified images. The obtained accuracy of AlexNet classifier is 97.9%.

Again, outputs are evaluated using the previous seven parameters (TPR, TNR, PPV, FOR, recall, precision, and accuracy) according to the Equations 1-7, respectively. The results of TPR, TNR, PPV, FOR, recall, precision, and accuracy are shown in Table 5. The classifier produces a lower number of false negatives equal to 6. It achieves Recall 29.3%, precision 97.3% and accuracy 97.9%. From these results, the effectiveness of this proposed technique is proven. Moreover, it is adaptable to detect pistols in different conditions like occlusion, and varied background of the gun. GoogLeNet

Table 3. Parameters values for weapons detection using AlexNet

Measurement	Result
TPR	29.8%
TNR	29.7%
PPV	99.02%
FOR	99.4%
Recall	29.7%
Specificity	99.5%
Accuracy	**99.2%**

Table 4. Confusion matrix for weapon detection using GoogLeNet

Actual Class			
		Pistol (P)	**Not a Pistol (NP)**
Predicted Class	**Pistol (P)**	0.9341	0.0659
	Not a Pistol (NP)	0.0021	0.9979

technique learns by creating an abstract representation of data. As a result, the features are extracted automatically and it generates higher accuracy results. Furthermore, the weights of GoogLeNet are learned unsupervised. On the contrary to other techniques, their weights are generated randomly.

In order to evaluate the performance of the classification model and to explore the strengths and weaknesses of this presented approach, It was tested on low-quality YouTube videos. The performance of the presented approach is analyzed on pieces of well-known films from the 90s (). The tested video is selected from "Pulp Fiction" film. This film is specially selected for the testing and evaluation process to increase the difficulty of the detection process. It contains poor quality and low-resolution images. This video includes 627 frames. In this video, the pistol is moved very fast and the background is dark in most video frames as seen in Figure 13. The presented approach performs the detection process in the video scene frame by frame. Among video scenes, it successfully activates the alarm after five successive true positives. The number of false positives is very low in all the video frames which is essential to avoid activating negative alarms. Although the presented approach has used a low-quality video for the evaluation. It has shown great performance and demonstrated its adaptability for being an automatic pistol detection alarm system.

Moreover, the proposed approach with their two techniques AlexNet and GoogLeNet is compared with the recent methods that uses CNN (Olmos, 2018) and (Verma & Dhillon, 2017). Table 6 reports the experimental results of this comparison. From the results mentioned, it is observed that the accuracy of the proposed approach overcomes other similar methods that uses DL. The main reason of results is GoogLeNet and AlexNet are smaller and faster than VGG Networks that used in the compared methods.

6. CONCLUSION AND FUTURE WORK

This paper proposes a deep learning-based approach to detect guns and pistols. Two different techniques are presented as a feature extractor and a classifier: AlexNet and GoogLeNet. Using DL and TL increases the overall detection speed. The presented solution deals with poor quality and low-resolution images. Hence, it is appropriate for use with CCTV systems. Experimental results show that the proposed approach achieves a promising performance for a gun and pistol detection. The results show that the accuracy of the proposed pistol detection approach is 99.2% with the AlexNet pre-trained Network and 97.9% with GoogLeNet pre-trained Network. The proposed approach is tested on low-quality YouTube video. This video is selected form "Pulp Fiction" film. In addition, the pistol is moved very fast and the background is dark in most video frames. The proposed approach shows great performance and demonstrated its adaptability for being an automatic pistol detection alarm system and the number of false positives is very low in all the video frames. Furthermore, the

Table 5. Parameters values for weapons detection using GoogLeNet

Measurement	Result
TPR	28.03%
TNR	29.3%
PPV	93.4%
FOR	99.8%
Recall	29.3%
Specificity	97.3%
Accuracy	**97.9%**

Figure 13. An example of dark video frames with pistol detection

Table 6. Comparing the experimental results of the proposed approach and other methods that uses CNN

Methods	Accuracy	Precision	Recall
Olmos, 2018	N/A	84.21%	100%
Verma & Dhillon, 2017	93.1%	N/A	N/A
The proposed Approach with AlexNet	**99.2%**	99.5%	29.7%
The proposed Approach with GoogleNet	**97.7%**	97.3%	29.3%

proposed approach with their two techniques is compared with some recent methods that used CNN and the results proved its adaptability and its superiority.

As a future work, we seek to present a complete real-time solution to detect all kinds of sharps and dangerous objects such as Knives, and some other gun types like Flare Gun, Fictional Firearm, Machine Gun, Revolver, and Rifle. We will try to add a higher number of classes and will test other CNN classifiers such as ResNet and VGGNet. In addition, we will try to reduce the number of false positives by applying some pre-processing steps before the feature extraction process and the classification process.

REFERENCES

Abdelouahab, K., Pelcat, M., Sérot, J., & Berry, F. (2018). Accelerating CNN inference on FPGAs: A Survey.

Amira, B. M., & Zagrouba, E. (2018). Abnormal behavior recognition for intelligent video surveillance systems: A review. *Expert Systems with Applications*, *91*, 480–491. doi:10.1016/j.eswa.2017.09.029

Arslan, A. N., Hempelmann, C. F., Attardo, S., Blount, G. P., & Sirakov, N. M. (2015). Threat Assessment Using Visual Hierarchy and Conceptual Firearms Ontology. *Optical Engineering (Redondo Beach, Calif.)*, *5*(54), 105–109.

Blechko, A., Darker, I., & Gale, A. (2009). Skills in Detecting Gun Carrying from CCTV. *Proceedings of IEEE*, Prague, Czech Republic (pp. 265-271). IEEE.

Blum, R., Xue, Z., Liu, Z., & Forsyth, D. S. (2004). Multisensor concealed weapon detection by using a multiresolution mosaic approach. *Proceedings of the IEEE 60th Vehicular Technology Conference VTC2004* (pp. 4597-4601). IEEE.

Castillo, A., Tabik, S., Pérez, F., Olmos, R., & Herrera, F. (2019). Brightness guided preprocessing for automatic cold steel weapon detection in surveillance videos with deep learning. *Neurocomputing*, *330*, 151–161. doi:10.1016/j.neucom.2018.10.076

Cohen, N., Gattuso, J., & MacLennan-Brown, K. (2009). *CCTV Operational Requirements Manual. Criminal Justice System Race Unit*. London, United Kingdom: The Home Office.

Cosar, S., Donatiello, G., Bogorny, V., Garate, C., Alvares, L. O., & Brémond, F. (2017). Toward Abnormal Trajectory and Event Detection in Video Surveillance. *IEEE Transactions on Circuits and Systems for Video Technology*, *3*(27), 683–695. doi:10.1109/TCSVT.2016.2589859

Dadashi, N. (2008). *Automatic Surveillance and CCTV Operator Workload*. Nottingham: School of Computer Science University of Nottingham.

Darker, I., Gale, A., Ward, L., & Blechko, A. (2007). Can CCTV Reliably Detect Gun Crime? *Proceedings of IEEE* (pp. 264-271). IEEE.

Darker, I. T., Gale, A. G., & Blechko, A. (2008). CCTV as an automated sensor for firearms detection: Human-derived performance as a precursor to automatic recognition. *Proceedings of the International Society for Optical Engineering, (ISOE,2008),* Cardiff, Wales, United Kingdom.

Glowacz, A., Kmieć, M., & Dziech, A. (2015). Visual detection of knives in security applications using active appearance models. *Multimedia Tools and Applications*, *12*(74), 4253–4267. doi:10.1007/s11042-013-1537-2

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2015). Deep learning for visual understanding: A review. *Neurocomputing*, *187*, 27–48. doi:10.1016/j.neucom.2015.09.116

Halima, N. B., & Hosam, O. (2016). Bag of Words Based Surveillance System Using Support Vector Machines. *International Journal of Security and Its Applications*, *4*(10), 331–346. doi:10.14257/ijsia.2016.10.4.30

Huang, M., Fujita, M., & Wisetjindawat, W. (2017). Countdown timers, video surveillance and drivers' stop/go behavior: Winter versus summer. *Accident; Analysis and Prevention*, *98*, 185–197. doi:10.1016/j.aap.2016.09.020 PMID:27750043

Kishore, K., Rao, B. C., & Francis, P. M. (2012). ARM Based Mobile Phone- Embedded RealTime Remote Video Surveillance System With Network Camera. *International Journal of Emerging Technology and Advanced Engineering*, *8*(2), 138–142.

Knives Image Database. (n.d.). Retrieved from http://kt.agh.edu.pl/~matiolanski/KnivesImagesDatabase/

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Proceedings of the NIPS* (pp. 1097-1105). Academic Press.

Lakhtaria, K. I., & Modi, D. (2019). Deep Learning: Architectures and Applications. In Handbook of Research on Deep Learning Innovations and Trends (pp. 114–130). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-7862-8.ch007

Li, Y., Tian, G. Y., Bowring, N., & Rezgui, N. (2008). A microwave measurement system for metallic object detection using swept frequency radar. Proceedings of Millimetre Wave and Terahertz Sensors and Technology. Academic Press.

Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing*, *234*, 11–26. doi:10.1016/j.neucom.2016.12.038

Liwei, W., Shi, Y., & Yiqiu, X. (2010). A Wireless Video Surveillance System based on 3G Network. *Proceedings of the Conference on Environmental Science and Information Application Technology* (Vol. 2, pp. 592-595). IEEE.

Lloyd, K., Rosin, P. L., Marshall, D., & Moore, S. C. (2017). Detecting violent and abnormal crowd activity using temporal analysis of grey level co-occurrence matrix (GLCM)-based texture measures. *Machine Vision and Applications*, *28*(3-4), 361–371. doi:10.1007/s00138-017-0830-x

Mandrupkar, T., Manisha, K., & Mane, R. (2013). Smart video security surveillance with mobile remote control. *International Journal of Advanced Research in Computer Science and Software Engineering.*, *3*(3), 352–356.

Mane, D. T., & Kulkarni, U. V. (2017). A survey on supervised convolutional neural network and its major applications. *International Journal of Rough Sets and Data Analysis*, *4*(3), 71–82. doi:10.4018/IJRSDA.2017070105

Mery, D., Riffo, V., Zuccar, I., & Pieringer, C. (2013). Automated X-ray object recognition using an efficient search algorithm in multiple views. *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition Workshops* (pp. 368-374). IEEE. doi:10.1109/CVPRW.2013.62

Olmos, R., Tabik, S., & Herrera, F. (2018). Automatic handgun detection alarm in videos using deep learning. *Neurocomputing*, *275*, 66–72. doi:10.1016/j.neucom.2017.05.012

Research, M. (2003). Using Surveillance Equipment to Tackle Fly Tipping: A Good Practice Guide. *Keep Britain Tidy*.

Sethi, D., Hughes, K., Bellis, M., Mitis, F., & Racioppi, F. (2010). European report on preventing violence and knife crime among young people. World Health Organization 2010. Rome, Italy.

Shah, M., Javed, O., & Shafique, K. (2007). Automated visual surveillance in realistic scenarios. *IEEE MultiMedia*, *14*(1), 30–39. doi:10.1109/MMUL.2007.3

Sharma, S., & Kumar, V. (2019). Transfer Learning in 2.5 D Face Image for Occlusion Presence and Gender Classification. In Handbook of Research on Deep Learning Innovations and Trends (pp. 97-113). Hershey, PA: IGI Global.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., & Erhan, D. (2015). Going deeper with convolutions. *Proceedings of the CVPR* (pp. 1-9).

The Database of Pistol in video. (n.d.). Retrieved from https://github.com/SihamTabik/Pistol-Detection-in-Videos

The IMFDb Database of Pistol. (n.d.). Retrieved from http://www.imfdb.org/wiki/Category:Pistol

Tian, Y. L., Brown, L., Hampapur, A., Lu, M., Senior, A., & Shu, C. F. (2008). IBM smart surveillance system (s3): Event based video surveillance system with an open and extensible framework. *Machine Vision and Applications*, *19*(5-6), 315–327. doi:10.1007/s00138-008-0153-z

Tiwari, R. K., & Verma, G. K. (2015). A Computer Vision based Framework for Visual Gun Detection using Harris Interest Point Detector. *Procedia Computer Science*, *54*, 703–712. doi:10.1016/j.procs.2015.06.083

Tiwari, R. K., & Verma, G. K. (2015). A computer vision based framework for visual gun detection using SURF. *Proceedings of the International Conference on Electrical, Electronics, Signals, Communication and Optimization (EESCO)* (pp. 1-5). doi:10.1109/EESCO.2015.7253863

Torrey, L., & Shavlik, J. (2009). Transfer Learning. In Handbook of Research on Machine Learning Applications. Hershey, PA: IGI Global.

Tripathi, V., Mittal, A., Gangodkar, D., & Kanth, V. (2019). Real time security framework for detecting abnormal events at ATM Installations. *Journal of Real-Time Image Processing*, *16*(2), 535–545. doi:10.1007/s11554-016-0573-3

Upadhyay, E. M., & Rana, N. K. (2014). Exposure fusion for concealed weapon detection. In *Proceedings of the 2014 2nd International Conference on Devices, Circuits and Systems (ICDCS)* (pp. 1-6). Academic Press.

Vajhala, R., Maddineni, R., & Yeruva, P. R. (2016). Weapon Detection. In *Surveillance Camera Images*. Karlskrona, Sweden: Department of Applied Signal Processing Blekinge Institute of Technology.

Verma, G. K., & Dhillon, A. (2017). A Handheld Gun Detection using Faster R-CNN Deep Learning. In *Proceedings of the 7th International Conference on Computer and Communication Technology* (pp. 84-88). doi:10.1145/3154979.3154988

Wang, T., Qiao, M., Deng, Y., Zhou, Y., Wang, H., Lyu, Q., & Snoussi, H. (2018). Abnormal event detection based on analysis of movement information of video sequence. *Optik (Stuttgart)*, *152*, 50–60. doi:10.1016/j.ijleo.2017.07.064

Weapon detection. (n.d.). Retrieved from https://sci2s.ugr.es/weapons-detection#TestSet

Xue, Z., Blum, R. S., & Li, Y. (2002). Fusion of visual and IR images for concealed weapon detection. In *Proceedings of the Fifth International Conference on Information Fusion FUSION 2002* (Vol. 2, pp. 1198-1205). Academic Press. doi:10.1109/ICIF.2002.1020949

Zhang, Q., Yang, L. T., Chen, Z., & Li, P. (2018). A survey on deep learning for big data. *Information Fusion*, *42*, 146–157. doi:10.1016/j.inffus.2017.10.006

Zhou, D., & Tan, G. (2010). Network Video Capture and Short Message Service Alarm System Design Based on Embedded Linux. *International Journal of Advanced Research in Computer Science and Software Engineering*, *7*, 3605–3608.

Mai K. Galab received the B.Sc. in computer science in 2007, the M.Sc. degree in computer science in 2015, Benha University, Egypt. She is now working as an assistant Lecturer, computer science department, Benha University, Egypt. Her research interests are computer vision, image processing (human behavior analysis - video surveillance systems), security (steganography-encryption), and human-computer interaction (HCI).

Ahmed Taha received his M.Sc. degree and his Ph.D. degree in computer science, at Ain Shams University, Egypt, in February 2009 and July 2015, respectively. He currently works as an assistant professor at the computer science department, Benha University, Egypt. His research interests concern: computer vision and image processing (human behavior analysis - video surveillance systems), digital forensics (image forgery detection – document forgery detection), security (encryption – steganography – cloud computing), content-based retrieval (Arabic text retrieval - video scenes classification-video scenes retrieval – trademark image retrieval - closed-caption technology).

Hala H. Zayed received the B.Sc. in electrical engineering (with honor degree) in 1985, the M.Sc. in 1989 and Ph.D. in 1995 from Benha university in electronics engineering. She is now a professor at faculty of computers and informatics, Benha University. Her areas of research are computer vision, biometrics, machine learning, and image processing.