Article

Detection and Classification of Different Weapon Types Using Deep Learning

Volkan Kaya ^{1,*} , Servet Tuncer ² and Ahmet Baran ¹¹ Department of Computer Engineering, Erzincan Binali Yıldırım University, Erzincan 24100, Turkey; baran@erzincan.edu.tr² Department of Electrical and Electronic Engineering, Firat University, Elazığ 23119, Turkey; stuncer@firat.edu.tr

* Correspondence: vkaya@erzincan.edu.tr

Abstract: Today, with the increasing number of criminal activities, automatic control systems are becoming the primary need for security forces. In this study, a new model is proposed to detect seven different weapon types using the deep learning method. This model offers a new approach to weapon classification based on the VGGNet architecture. The model is taught how to recognize assault rifles, bazookas, grenades, hunting rifles, knives, pistols, and revolvers. The proposed model is developed using the Keras library on the TensorFlow base. A new model is used to determine the method required to train, create layers, implement the training process, save training in the computer environment, determine the success rate of the training, and test the trained model. In order to train the model network proposed in this study, a new dataset consisting of seven different weapon types is constructed. Using this dataset, the proposed model is compared with the VGG-16, ResNet-50, and ResNet-101 models to determine which provides the best classification results. As a result of the comparison, the proposed model's success accuracy of 98.40% is shown to be higher than the VGG-16 model with 89.75% success accuracy, the ResNet-50 model with 93.70% success accuracy, and the ResNet-101 model with 83.33% success accuracy.

Citation: Kaya, V.; Tuncer, S.; Baran, A. Detection and Classification of Different Weapon Types Using Deep Learning. *Appl. Sci.* **2021**, *11*, 7535. <https://doi.org/10.3390/app11167535>

Keywords: deep learning; convolutional neural network; VGGNet; ResNet; weapon detection

Academic Editor: Andrew Teoh
Beng Jin

Received: 24 June 2021
Accepted: 14 August 2021
Published: 17 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the age of modern science and technology, people use surveillance cameras in different areas to prevent crime [1]. Numerous camera systems are installed in different areas, and security guards need to monitor all of these cameras at the same time [2,3]. Generally, after a crime occurs, security guards arrive at the scene, then after checking the recorded images they analyze the images and collect the necessary evidence [4]; therefore, it is necessary to establish a proactive system at the crime scene [3,5,6]. In this context, if the software alerts the security guards immediately after detecting threatening objects, prompt action can be taken to stop the potential criminal from committing a crime [4,7]. As such, it is very important to create a system that can learn to detect threatening objects [8].

The role of deep learning in improving task performance in security controls systems is considered indisputable [3]. Deep learning is a sub-field of machine learning [9]. It uses many layers of non-linear processing units for deep learning and feature extraction and conversion [10]. The deep learning structure is based on the learning of more than one feature level of data. Deep learning is based on learning from the representation of the main data [11]. The representation of an image can be considered to comprise a vector of density values per pixel or features such as clusters of edges and custom shapes, with some representing the data better [12]. The basic architecture of the deep learning concept is the convolutional neural network (CNN), which consists of convolution, pooling, activation function, dropout, fully connected, and classification layers [13]. In the last few years, deep learning has become a mainstay in the field of object detection [14] and classification and

image segmentation. To date, CNNs have achieved the best results for classical image processing problems, such as image segmentation, classification, and detection [15,16].

Today, most criminal activities are carried out using handheld weapons [16]. Many studies have revealed that handheld weapons [17] are the most important criminal elements used for various crimes, such as theft, illegal hunting, and terrorism. The solution to reduce such criminal activities is installing a surveillance system [18,19] or control cameras [20] so that security units can take appropriate measures at early stages [16]. Weapon detection is challenging due to the various subtleties associated with it. The most important problems in weapon detection are self-occlusion and the similarities between objects and background structures. Self-occlusion occurs when a part of the gun is blocked on one side. Similarity between objects occurs when different objects such as hands and clothes look like weapons. Background problems [15] refer to those related to the background against which the gun is located [16].

This study proposes a new model to detect seven different weapon types (assault rifles, bazookas, grenades, hunting rifles, knives, pistols, and revolvers) based on the deep learning method. In order to evaluate the classification performance of the proposed model, it is compared to the Visual Geometry Group (VGG-16) model [21] and Residual Network (ResNet50 and ResNet101) models [22], which are accepted in the literature. As a result of this comparison, it is seen that the proposed model has higher accuracy and a lower loss rate than the VGG-16, Resnet-50, and Resnet-101 models.

The remainder of the study is organized as follows. In Section 2, the literature on the subject is presented. The material and methods are explained in Section 3 and the results are presented in Section 4. In Section 5, the discussion is presented. Lastly, Section 6 concludes the study.

2. Related Works

A study investigating an automatic pistol detection system in videos for surveillance and security purposes [18] created basic training data using the results of deep CNN classification. The best classification model was explored to minimize false positives by evaluating two classification models based on the sliding window and region proposal approaches, and it was seen that the best results were obtained from the faster region-based CNN model [18].

In an image-fusion-based study [23], imbalance map calculation and selection of candidate regions from input frames were performed. A low-cost, symmetrical dual-camera system was installed to take advantage of this information. According to the results, the number of false positives decreased and the detection of objects in surveillance videos became more convenient [23]. Brightness-guided preprocessing, known as darkening and contrasting at learning and test stages, was proposed to improve detection quality in surveillance videos and a cold steel weapon detection model was presented [24]. In a hybrid weapon detection study using fuzzy logic, a system was developed to detect harmful objects such as guns and knives using additional parameters that improved correct results and reduced false alarm rates [25].

In a weapon classification study developed using a deep CNN, two new approaches were presented. In the proposed approach, the weights of the pre-trained VGG-16 model were taken. Using this model, the effects of changing the number of neurons in the fully connected layer on classification were investigated [26]. In a study aiming to detect firearms in surveillance videos, the focus was placed only on only areas where human beings were found [5] and a weapon detection system was implemented using the separate components of the weapons [27].

A study on multilevel security management presented a system for the management of multimedia security in Internet of Things systems. This system automatically analyzed multimedia events and calculated security levels [28]. In another real-time object detection study, the authors detected handheld weapons (pistols and rifles). In that study, a TensorFlow-based application of Overfeat, a CNN-based image classifier and feature

extractor, was used to detect and classify weapons in the image [20]. In another study on the automatic detection of firearms and knives, algorithms were proposed to warn human operators when guns or knives were detected in the closed-circuit television system. In the same study, in order to apply the system in real life, the number of false alarms was reduced and a system that was capable of real time warning when a dangerous situation was detected was developed [29].

Clustering algorithm and color-based segmentation were also previously used to eliminate irrelevant objects from an image for the purpose of automatic visual weapon detection. The Harris detector of interest points and the fast retina keypoint descriptor were used to detect the relevant object (weapon) in segmented images. By applying this system to the collected weapon sample images, the partial jamming, scaling, rotation and presence of multiple weapons were detected [30].

Another study undertaken to identify people in danger focused on handheld weapons. In that study, by introducing the human object interaction model, methods and systems were established to recognize dangerous events. This system was based on the identification of dangerous objects concealed in possible parts of the human body [31].

In brief, in the literature, studies have generally attempted to classify concealed weapons, firearms, knives, and handheld weapons; however, to the best of our knowledge, no study has shown the detection and discrimination of different weapon types. The current study proposes a weapon detection model that both detects more weapon classes compared to other studies and has very high accuracy.

3. Materials and Methods

3.1. Dataset and Pre-Processing

Because there is no standard dataset for weapon detection and recognition, 5214 weapon images were prepared by downloading them from the internet. Downloaded weapon images must be of good quality images with different angles in order to be successful in detecting and recognizing real-life weapons. In addition, in order to achieve higher success accuracy from the developed neural network model, irrelevant objects in each weapon image were removed. As such, downloaded weapon images were examined individually and different computer application programs were used to make changes to the images, such as padding, masking, background cleaning, sizing, and rotation, according to the content.

After preparing separate images for each weapon class, they were converted into a dataset. The weapon types and numbers of images belonging to the dataset used in the study are given in Table 1. The dataset comprised images of assault rifles, bazookas, grenades, hunting rifles, knives, pistols, and revolvers (Figure 1). The images used in the dataset were transformed into gray color format at 144×144 pixels using the Python programming language. Then, the images belonging to the dataset were labeled under the weapon class to which they belonged and grouped.

Table 1. Weapon dataset classes and numbers of images.

Weapon Class	Number of Images
Assault rifle	927
Bazooka	211
Grenade	507
Hunting rifle	849
Knife	1076
Pistol	1170
Revolver	474
Total	5214

Figure 1. Sample dataset of different classes: (a) assault rifles; (b) bazookas; (c) grenades; (d) hunting rifles; (e) knives; (f) pistols; (g) revolvers.

3.2. CNN Model

In object recognition applications, the most used deep learning algorithm is the CNN algorithm [32,33]. This algorithm showed outstanding success in the ImageNet Large-Scale Visual Recognition Competition held in 2012 and has been used in the development of new models in many areas since then [21]. In the current study, a new model (Figure 2) is proposed based on the VGG-16 model (Figure 3) [21]. The model shown in Figure 2 consists of a total of 25 layers (convolution ($n = 7$), pooling ($n = 4$), dropout ($n = 4$), rectified linear units (ReLU) ($n = 7$), flatten ($n = 1$), fully connected ($n = 1$), and classification ($n = 1$)) and 337,671 parameters.

Figure 2. Developed convolution neural network model (f: filter size, s: stride, p: padding, FC: fully connected).

Figure 3. VGG-16 model structure.

The number of layers when designing a model network and the parameter values used in these layers significantly affect the training time of the neural network, the cost of processing, and the choice of device to be applied. In this study, the model we proposed was preferred for the problem of weapon detection and recognition due to the small number of layers compared to the classic VGGNet model, the ability to train on low-cost computers, the low training time, and the high success accuracy.

In the model presented in Figure 2, two convolution layers and a maximum pooling layer are applied to the input image in gray color format. In the pooling layer, the two-step 2×2 filter matrix is shifted over the input matrix to create a new matrix value. Maximum pooling holds the new matrix value by taking the maximum value of each block to better estimate the image. In addition, the ReLU activation function is used in convolution layers. This is a non-linear function that replaces all negative pixel values with a zero value, which ensures that the network is more efficient and faster [34]. Using a 25% dropout layer after each pooling layer, memorization is prevented [35]. Convolution, pooling, and forgetting processes are repeated with different parameters and different filter channel values in later layers. Then, neurons are formed into an array by flattening and fully connecting layers. As a result of the straightening process, 2048 neurons are obtained by forming a fully connected layer. In the last layer of the model is the classification layer with the softmax activation function. In this layer, the output value is obtained by classifying the number of classes determined in the dataset. As a result of the classification, seven different weapon types take a value in the range of 0–1 and the weapon type with the highest value presents the estimated value.

3.3. Media and Libraries Used

The model proposed in this study was developed using the Keras library based on TensorFlow. In writing the codes, the NumPy, Matplotlib, PIL, Os, OpenCV, Sklearn, and Imageio libraries were also utilized. TensorFlow is an interface for training and implementing machine learning algorithms. It is also an open source software library that performs numerical calculations using data flow graphics. Computation using TensorFlow is undertaken in many systems ranging from mobile devices to large-scale distributed systems, and it is also used to express various algorithms developed for CNN models. TensorFlow can be set up with two different processors for graphics processing unit (GPU) or central processing unit (CPU) support. TensorFlow programs generally run faster on GPUs than on CPUs. In this study, the model was trained on a computer equipped with an Intel Core i7-9750H 2.60 GHz processor with the 896 Compute Unified Device Architecture, Nvidia GeForce GTX 1650 graphics card, and 8 GB RAM. The study aimed to propose a model with a focus on providing the highest accuracy, sensitivity, and specificity rates, while the lowest loss rates were obtained using CNN.

Using the new model, the method required for training was determined, model layers were created, the training process was applied and recorded in the computer environment, the success rate of the training was determined, and the trained model was tested. In addition, the model was implemented by writing codes in the Python programming language.

4. Results

In the study, experiments were conducted for seven different weapon types and the proposed model was compared with the VGG-16, ResNet-50, and ResNet-101 models. In all four models, the dataset that was used in the training of the network was divided into three

groups, 60% as training, 20% as testing, and 20% as the validation dataset. This situation is shown in Table 2. The training and testing datasets were used during the training of the network. The success rate of the network was obtained by testing the model with a validation dataset that it had never seen during training.

Table 2. Weapon dataset.

Weapon Dataset	Training	Test	Validation	Total
	3128	1043	1043	5214

During the training processes in all four models, the parameter values of the activation function (ReLU) [34], mini batch size (32), dropout (0.25) [35], optimization algorithm (Adamax) [36], and epochs (30) were used the same. The input image size for these four model tutorials was set to 144×144 pixels in gray color format. In all four models, “accuracy”, “loss”, “sensitivity”, and “specificity” change graphs determine the success of the network during training, which are shown in Figure 4.

When the graphs of the VGG-16 model given in Figure 4a were examined, it was seen that the model network did not learn during the first 6 epochs and learned with a 90.12% success accuracy as a result of 30 epochs. Similar results can also be observed in the loss, sensitivity, and specificity graphs. Figure 4b shows that the ResNet-50 model began to learn the model network after the first epoch and reached a success accuracy of 94.25% as a result of 30 epochs. As shown in Figure 4c, The ResNet-101 model reached a success accuracy of 84.43% as a result of 30 epochs and had a lower success accuracy than the ResNet-50 model. In Figure 4d, it can be seen that the model network began to learn after the first epoch for the proposed model and reached a success accuracy of 98.32% as a result of 30 epochs. Similar results can also be observed in the loss, sensitivity, and specificity graphs for this model. The success of the model network depends on the layers and parameter values used in the neural network. In the proposed convolutional neural network model, it is important to process the data and extract the necessary features. For these purposes, the layers and parameter values used in the model are selected according to the model network, which increases the success rate. In the proposed model, the number of layers used was less than for other models, and accordingly the number of parameters was reduced; however, the proposed model is able to process, train, and test data faster than other models. As such, the proposed model appears to have higher success accuracy and a lower loss rate than the VGG-16, ResNet-50, and ResNet-101 models. In addition, with the proposed model, model networks were tested using 1043 weapon images in the network validation cluster to analyze the efficiency of the VGG-16, ResNet-50, and ResNet-101 models, with the “accuracy”, “sensitivity”, “specificity”, and “loss” values for all four models given in Table 3.

When the graphs given in Figure 4 and the values given in Table 3 are examined, it can be seen that the proposed model learned faster starting from the first epochs and had a higher success rate.

A confusion matrix was also created to analyze the classification efficacy of the proposed model (Figure 5). According to this graph, the most successful classification rates were obtained for assault rifles at 99.45% and hunting rifles at 99.45%, while the classification rates were least successful for revolvers at 94.62% and bazookas at 97.72%. Although the number of datapoints for the bazooka weapon type was less than available for the remaining weapon types in the dataset, it was still more successfully classified than the revolver type. This can be attributed to the visual similarities between revolvers and pistols in many aspects.

Table 3. Training success rates of the VGG-16, ResNet-50, and ResNet-101 models and the proposed model.

Measures	VGG-16 Values	ResNet-50 Values	ResNet-101 Values	Proposed Model Values
Accuracy	89.75%	93.70%	83.33%	98.40%
Sensitivity	89.71%	76.48%	41.05%	92.89%
Specificity	98.84%	96.55%	90.44%	99.28%
Loss	0.638	0.279	1.615	0.052

Figure 4. Graphs of changes throughout training: (a) the VGG-16 model; (b) ResNet-50 model; (c) ResNet-101 model; (d) proposed model.

Figure 5. Confusion matrix graph.

In Table 3 and Figures 4 and 5, it can be seen that the proposed model, with a success rate of 98.40% in terms of classification accuracy, was better than the VGG-16 model with an 89.75% success rate, ResNet-50 with a 93.70% success rate, and ResNet-101 with an 83.33% success rate.

Performance Evaluation

In order to evaluate the performance of the new developed model, test data were prepared from images of different weapon classes on humans. The images in the test data were created by downloading them from the Internet [37]. One of the difficulties encountered in weapon detection is the possibility of the weapon being located in any part of the input image; therefore, it is necessary to identify areas where a weapon can be located, then the weapon classification process should be performed. In order to test the accuracy of the model network proposed in this study, the region proposal approach [38], one of the selective search techniques, was used (Figure 6).

Figure 6. Region proposal approach.

The region proposal approach proposed by Girshick et al. is widely used in object detection [38]. This approach was used in the current study to show the usability and adaptability of the proposed model in real life. Since different objects in the image to be tested in the proposed model were similar to the weapon class, new images were obtained using the region proposal approach. The aspect ratios were rearranged by taking the coordinates of the new images obtained. In order to classify the images obtained, those belonging to the weapon class with the region proposal approach were ranked according to the success rate, while the weapon in the region with the highest success rate was predicted.

In addition, the coordinates of the weapon were found and shown in a rectangular frame. In real life, assault rifle, bazooka, hunting rifle, grenade, pistol, and revolver variants have similarities in many cases; therefore, it is very easy to confuse them during classification, resulting in some errors. A dataset containing a sufficient number of images

of weapons was prepared for the detection of weapons. The mean average precision (mAP) metric was used in the evaluation of the weapon images in the dataset, as given in Table 4. In addition, the images of real-life quality that were used in performance evaluation are given in Figure 7.

Table 4. Mean average precision (mAP) of the weapon images used in the dataset.

Weapon Class	Number of Images	AP	mAP
Assault rifle	123	0.917	0.873 mAP@0.5 IoU
Bazooka	96	0.875	
Grenade	105	0.892	
Hunting rifle	112	0.908	
Knife	88	0.825	
Pistol	128	0.883	
Revolver	90	0.808	
Total	742

Figure 7. Cont.

Figure 7. Sample test results for different classes: (a) assault rifles; (b) bazookas; (c) grenades; (d) hunting rifles; (e) knives; (f) pistols; (g) revolvers.

5. Discussion

In this study, a deep-learning-based artificial intelligence model was developed for use in autonomous security control, which can determine whether there are weapons on humans and also distinguish which of the 7 different classes the weapon it detects belongs to.

The data used in the dataset for the developed model training process were divided into training, validation, and test datasets at certain percentages. In the first stage, the model was trained using the training and validation dataset. In order to test the accuracy of the model trained in the second stage, it was tested using a test dataset that the model could not see and a satisfactory rate of classification success was achieved. In our study, the proposed model identified more types of weapons compared to other studies in the literature. For example, Olmos et al. studied an automatic gun detection system for us in videos for surveillance and control purposes. In this study, the pistol was only detected using a pre-trained VGG-16-based classification model [18,23]. In a previous study, the weapon detection model was improved and only a knife was detected [24]. Grega et al. proposed algorithms to alert human operators when a weapon or knife is detected [29].

The accuracy of the proposed model (98.40%) has an excellent success rate compared to studies in the literature. For example, Ineneji et al. developed a system that detects weapons and knives with a success rate of 94.64%, which is lower than the accuracy value in our study [25]. For Romero et al. the success accuracy rate in their study was lower (86%) than the value we obtained [5]. In the real-time object detection study by Lai et al. they detected handheld weapons (pistols and rifles) [20]. A pre-trained VGG-16 based classification model was used and 93% training and 89% test accuracy were obtained [20]. In another study, using the weights of the pre-trained VGG-16 model, knife, pistol, and non-weapon classes were used and the success rate for the weapon classification (98.41%) was similar to our study [26]. Additionally, multiple weapons were detected with a success accuracy rate of 84.26% [30]. Xu et al. focused on identifying weapons in people's hands to detect endangered people. In this study, the model was trained using a dataset containing images of people holding weapons and people not holding weapons, obtaining a success accuracy of 74% [31].

In addition, our model has also been tested on images of different classes of weapons found on people to assess the real-life availability and classification performance of the developed model. Overall, the newly developed model performed very well, detecting images of weapons against different backgrounds within about 3 s.

In this study, the importance of the proposed model is increased due to the fact that it is trained and tested on computers that are less expensive and do not have sufficient infrastructure.

6. Conclusions

In today's conditions, where criminal activities are increasing, it is very important to detect and recognize whether there is a weapon on a person based on images taken from security cameras, without requiring human intervention. Weapon detection and recognition are important to prevent criminal activities before they occur and so that the appropriate parties can take necessary action. Most criminal activities are carried out using handheld or carried weapons. Handheld or carried weapons are the most important elements used for various crimes, such as theft, illegal hunting, and terrorism. It is necessary to determine the types of weapons that may constitute a criminal element in order to predict whether there will be any criminal element in the images taken from security cameras and to take the necessary precautions beforehand.

In this study, a new model was developed to detect and classify seven different weapon types based on the VGGNet architecture. In addition, a new dataset consisting of seven different weapon types was constructed. The developed model was compared to the VGG-16 (89.75% success accuracy), ResNet-50 (93.70% success accuracy), and ResNet-101 (83.33% success accuracy) models, showing a very high success rate of 98.40% in comparison. The developed model was also tested using images of weapons carried on human beings. Since the different objects in the tested images were similar to the weapon class, new images were obtained using the region proposal approach. The aspect ratios were rearranged using the coordinates of the new images. The images obtained as a result of the region proposal approach were classified and those belonging to the weapon class

were ranked according to success rate to ensure that the model could predict the weapon class located in the region with the highest success rate. As a result of the classification, the weapon coordinates were found in the tested images. These coordinates were placed into a rectangular frame and the weapon class and percentage success rate were determined. The newly developed model showed very good performance by successfully detecting weapon images with different backgrounds.

In conclusion, the model proposed in this study will contribute to the elimination of many security gaps by increasing the task effectiveness and efficiency of security forces in security applications. In this respect, the fact that the results can be directly transferred to new applications increases the original value of the study. It is expected that the results of the study will lead and contribute to similar studies being undertaken, especially regarding autonomous security units.

In future studies, an infrastructure investigation could be done on robot soldiers that can automatically monitor and analyze input data and send alerts to security forces in order to process data in real time using security control systems and to increase classification accuracy. In addition, studies should be developed to detect coated guns.

Author Contributions: Conceptualization, V.K., S.T. and A.B.; methodology, V.K., S.T. and A.B.; software, V.K.; validation, V.K., S.T. and A.B.; formal analysis, V.K. and A.B.; investigation, V.K.; resources, V.K.; data curation, V.K. and A.B.; writing—original draft preparation, V.K.; writing—review and editing, V.K., S.T. and A.B.; visualization, V.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The labeled dataset used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Raturi, G.; Rani, P.; Madan, S.; Dosanjh, S. ADoCW: An Automated method for Detection of Concealed Weapon. In Proceedings of the Fifth International Conference on Image Information Processing, Shimla, India, 15–17 November 2019; pp. 181–186.
2. Bhagyalakshmi, P.; Indhumathi, P.; Lakshmi, R.; Bhavadharini, D. Real Time Video Surveillance for Automated Weapon Detection. *Int. J. Trend Sci. Res. Dev.* **2019**, *3*, 465–470. [\[CrossRef\]](#)
3. Lim, J.; Jobayer, M.I.A.; Baskaran, V.M.; Lim, J.M.; Wong, K.; See, J. Gun Detection in Surveillance Videos using Deep Neural Networks. In Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, Lanzhou, China, 18–21 November 2019; pp. 1998–2002. [\[CrossRef\]](#)
4. Yuan, J.; Guo, C. A deep learning method for detection of dangerous equipment. In Proceedings of the Eighth International Conference on Information Science and Technology, Granada, Cordoba, and Seville, Spain, 30 June–6 July 2018; pp. 159–164. [\[CrossRef\]](#)
5. Romero, D.; Salamea, C. Convolutional models for the detection of firearms in surveillance videos. *Appl. Sci.* **2019**, *9*, 2965. [\[CrossRef\]](#)
6. Ilgin, F.Y. Energy-based spectrum sensing with copulas for cognitive radios. *Bull. Polish Acad. Sci. Tech. Sci.* **2020**, *68*, 829–834. [\[CrossRef\]](#)
7. Navalgund, U.V.; Priyadarshini, K. Crime Intention Detection System Using Deep Learning. In Proceedings of the International Conference on Circuits and Systems in Digital Enterprise Technology, Kottayam, India, 21–22 December 2018; pp. 1–6. [\[CrossRef\]](#)
8. Chandan, G.; Jain, A.; Jain, H. Real time object detection and tracking using Deep Learning and OpenCV. In Proceedings of the International Conference on Inventive Research in Computing Applications, Coimbatore, India, 11–12 July 2018; pp. 1305–1308. [\[CrossRef\]](#)
9. Deng, L.; Yu, D. Deep learning: Methods and applications. *Found. Trends Signal Process.* **2014**, *7*, 197–387. [\[CrossRef\]](#)
10. Bengio, Y. Learning deep architectures for AI. *Found. Trends Mach. Learn.* **2009**, *2*, 1–27. [\[CrossRef\]](#)
11. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Song, H.A.; Lee, S.Y. Hierarchical representation using NMF. *Lect. Notes Comput. Sci.* **2013**, *8226*, 466–473. [\[CrossRef\]](#)
13. Masood, S.; Ahsan, U.; Munawwar, F.; Rizvi, D.R.; Ahmed, M. Scene Recognition from Image Using Convolutional Neural Network. *Procedia Comput. Sci.* **2019**, *167*, 1005–1012. [\[CrossRef\]](#)

14. Wu, X.; Sahoo, D.; Hoi, S.C.H. Recent advances in deep learning for object detection. *Neurocomputing* **2020**, *396*, 39–64. [\[CrossRef\]](#)
15. Krishna Sai, B.N.; Sasikala, T. Object Detection and Count of Objects in Image using Tensor Flow Object Detection API. In Proceedings of the International Conference on Smart Systems and Inventive Technology, Tirunelveli, India, 27–29 November 2019; pp. 542–546. [\[CrossRef\]](#)
16. Verma, G.K.; Dhillon, A. A Handheld Gun Detection using Faster R-CNN Deep Learning. In Proceedings of the 7th International Conference on Computer and Communication Technology, Allahabad, India, 24–26 November 2017; pp. 84–88. [\[CrossRef\]](#)
17. Warsi, A.; Abdullah, M.; Husen, M.N.; Yahya, M. Automatic Handgun and Knife Detection Algorithms: A Review. In Proceedings of the 14th International Conference on Ubiquitous Information Management and Communication, Taichung, Taiwan, 3–5 January 2020; pp. 1–9. [\[CrossRef\]](#)
18. Olmos, R.; Tabik, S.; Herrera, F. Automatic handgun detection alarm in videos using deep learning. *Neurocomputing* **2018**, *275*, 66–72. [\[CrossRef\]](#)
19. Asnani, S.; Ahmed, A.; Manjotho, A.A. Bank Security System based on Weapon Detection using HOG Features. *Asia J. Eng. Sci. Technol.* **2014**, *4*, 23–29.
20. Developing a Real-Time Gun Detection Classifier. Available online: <https://www.scribd.com/document/380866575/Developing-a-Real-Time-Gun-Detection-Classifier> (accessed on 8 May 2021).
21. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015; pp. 1–15.
22. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; Volume 2016, pp. 770–778. [\[CrossRef\]](#)
23. Olmos, R.; Tabik, S.; Lamas, A.; Pérez-Hernández, F.; Herrera, F. A binocular image fusion approach for minimizing false positives in handgun detection with deep learning. *Inf. Fusion* **2019**, *49*, 271–280. [\[CrossRef\]](#)
24. Castillo, A.; Tabik, S.; Pérez, F.; Olmos, R.; Herrera, F. Brightness guided preprocessing for automatic cold steel weapon detection in surveillance videos with deep learning. *Neurocomputing* **2019**, *330*, 151–161. [\[CrossRef\]](#)
25. Ineneji, C.; Kusaf, M. Hybrid weapon detection algorithm, using material test and fuzzy logic system. *Comput. Electr. Eng.* **2019**, *78*, 437–448. [\[CrossRef\]](#)
26. Dwivedi, N.; Singh, D.K.; Kushwaha, D.S. Weapon classification using deep convolutional neural network. In Proceedings of the Conference on Information and Communication Technology, Allahabad, India, 6–8 December 2019; pp. 1–5. [\[CrossRef\]](#)
27. Egiazarov, A.; Mavroeidis, V.; Zennaro, F.M.; Vishi, K. Firearm detection and segmentation using an ensemble of semantic neural networks. In Proceedings of the European Intelligence and Security Informatics Conference, Oulu, Finland, 26–27 November 2019; pp. 70–77. [\[CrossRef\]](#)
28. Ben Abdallah, H.; Abdellatif, T.; Chekir, F. AMSEP: Automated Multi-level Security Management for Multimedia Event Processing. *Procedia Comput. Sci.* **2018**, *134*, 452–457. [\[CrossRef\]](#)
29. Grega, M.; Matoriński, A.; Guzik, P.; Leszczuk, M. Automated detection of firearms and knives in a CCTV image. *Sensors* **2016**, *16*, 47. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Tiwari, R.K.; Verma, G.K. A Computer Vision based Framework for Visual Gun Detection Using Harris Interest Point Detector. *Procedia Comput. Sci.* **2015**, *54*, 703–712. [\[CrossRef\]](#)
31. Xu, Z.; Tian, Y.; Hu, X.; Pu, F. Dangerous human event understanding using human-object interaction model. In Proceedings of the International Conference on Signal Processing, Communications and Computing, Ningbo, China, 19–22 September 2015; pp. 1–5. [\[CrossRef\]](#)
32. Angelova, A.; Krizhevsky, A.; Vanhoucke, V. Pedestrian detection with a Large-Field-Of-View deep network. In Proceedings of the International conference on robotics and automation, Seattle, WA, USA, 26–30 May 2015; pp. 704–711. [\[CrossRef\]](#)
33. Hinton, G.E. A practical guide to training restricted boltzmann machines. In *Neural Networks: Tricks of the Trade*; Montavon, G., Orr, G.B., Müller, K.R., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 599–619. [\[CrossRef\]](#)
34. Agarap, A.F. Deep Learning using Rectified Linear Units (ReLU). *arXiv* **2018**, arXiv:1803.08375.
35. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *Phys. Lett. B* **2014**, *15*, 1929–1958.
36. AdaMax Online Training for Speech Recognition. Available online: http://csit.riit.tsinghua.edu.cn/mediawiki/images/d/df/Adamax_Online_Training_for_Speech_Recognition.pdf (accessed on 10 December 2020).
37. Istock. Available online: <https://www.istockphoto.com> (accessed on 20 August 2020).
38. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J.; Berkeley, U.C.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; Volume 1, p. 5000. [\[CrossRef\]](#)