ISSN: 2414-1895 DOI: 10.6919/ICJE.202112_7(12).0039

Firearms and Knives Detection Model for the Automatic Surveillance

Yuxuan Zhang[1, *], Wangweimin Xing[2, a], Xinyi Fan[3, b], Hanzhong Qiu[4, c], Qisong Lu[5,d]

[1]NanChang University, Nanchang, Jiangxi, China

[2]Zhengzhou Foreign Language High School, Zhengzhou, Henan, China

[3]University of Illinois at Urbana-Champaign, IL, USA

[4]SANLI XUCHENGDA,Shanghai, China

[5]Suzhou dulwich internetional high school, Suzhou, Jiangsu, China

[a]xingwangweimin@gmail.com, [b]fanxinyi6@gmail.com,

[c]2454173341@qq.com, [d]3038190983@qq.com

*Corresponding author:1770040579@qq.com

These authors contributed equally to this work

Abstract

Public safety and the safety of residents are prominent issues around the world. How to efficiently and quickly identify dangerous items such as knives, revolvers, bombs and assault rifles in public has become an issue that many countries attach importance to. Aiming at the problems of low efficiency of traditional manual security inspection and inaccurate classification of items, this paper proposes a method of image recognition for dangerous goods detection based on transfer learning by convolutional neural network. First, use flip, rotation, Gaussian noise and brightness transformation operations to expand and preprocess the acquired images of five types of dangerous goods to obtain a dangerous goods image data set. This paper uses an improved model based on CNN for recognition training, sets a suitable learning rate to speed up the training of the model and enhance the recognition ability. The final recognition accuracy of the built model reaches 72%, and the model has a high accuracy for the recognition of dangerous goods. It can also provide a reference for the precise classification of other items.

Keywords

Security check; item image; convolutional neural network; deep Learning; image classification.

1. Introduction

The globe has witnessed multifarious terrorist and criminal events over the past few decades. A case in point is that on March 16, 2021, 8 people were massacred in a shooting spree in Atlanta, and the incident at Columbine High School with 37 victims on April 20, 19991. The Public Surveillance System has attained much interest to ward off certain activities; an automated weapon detection system for public surveillance using CCTV or other detectors has been pressed for in status quo. According to a U.S. Department of Justice study, about ¾ of the homicide involves handheld firearms from 1994 to 20112. Another examination exhibits that the pistol is the primary weapon used for crimes in diverse scenarios, compared to other weapons like knives and blunt objects[3]. An automated

ISSN: 2414-1895 DOI: 10.6919/ICJE.202112_7(12).0039

weapon detection system could disclose possible danger by recognizing weapons and alarm security departments in time, in this sense preventing criminal activities. This paper will focus on the Deep learning-based model that serves as an automated visual weapon detection system in pictures, including some of the most frequently used and deadly weapons, including knives, pistols, assault rifles, bombs, and so forth.

We initially call for a model of weapon detection to pave the way for the weapon detection surveillance system to work. Object Detection is a domain of Artificial Intelligence regarding Digital Image Processing and Computer Vision, which operates by identifying instances of an object[4], such as intimidating weapons we presented in the following content. Object Detection methods are usually set apart into either Machine Learning-based approaches or Deep learning-based approaches depending upon the complexity of the target class. The Deep Learning-based approach applies Artificial Neural Networks to perform end-to-end object detection without specifying features[5]. The Deep-learning framework stems from Google's Deep Mind's AlphaGo artificial intelligence system in 2016[6]. Being a subfield of machine learning, Deep-learning can be examined in realms of big data, data mining, natural language processing, recommendation systems, semantic image segmentation, and so forth[7]. In addition, the Deep-learning framework has a bearing on neural-network-based learning, which we mainly practiced and discussed convolutional neural networks in this paper.

CNN (Convolutional neural networks) is one of the most symbolic algorithms among Deep learning frameworks and has been extensively practiced in the field of computer vision in the status quo. CNN can straight manage image data as input without handling complex algorithms to detect features, thereupon leave out the tedious operations of data preprocessing procedures. The scheme of a CNN is that when one looks at an image, the alikeness or identical elements, such as a particular pattern, are usually discerned first. Correspondingly, when a person views a portrait image, its identity would be ascertained by appearance traits such as eyes, brows, and mouth, then examines the results successively by probing traits of each section accordingly. CNN's features of local connection, weight sharing, and down-sampling are liable to reduce the fraction of parameters involved and the ramification of the model training trial. In addition, the model performance would be granted with ampler exactness owe to a convolutional layer and a pooling layer. CNN is a feed-forward network, the neurons between adjacent layers in the CNN are partially instead of entirely connected. The sensing region of a neuron stretches from a portion of the upper neural units, and each neuron exerts local sensing to extract basic peculiarities of the input data[9].

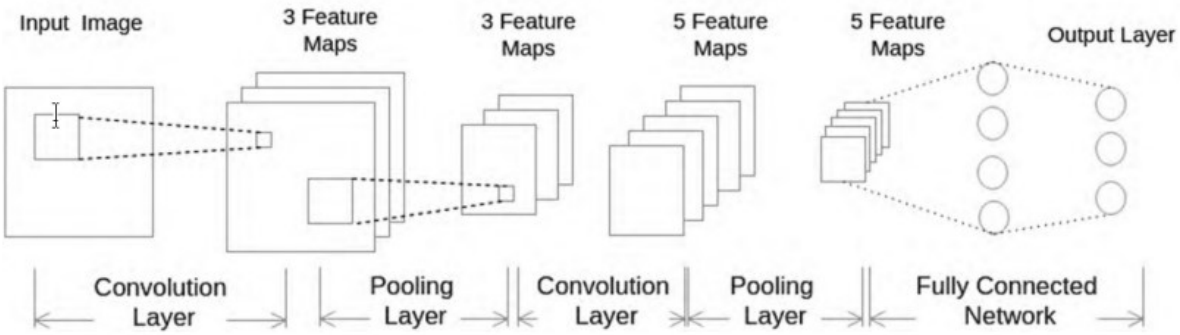

Figure 1 CNN Structure Demonstration

In the course of the training trial, we implemented ADAM(Adaptive Momentum) for parameters adjustment to achieve the desired accuracy. ADAM is a straightforward method for stochastic optimization and craves only first-order gradients with comparatively minor memory requirements. Such a method computes peculiar adaptive learning rates for different parameters from approximations of first and second moments of the gradients. ADAM has been designed intended for ML and DL problems with a chunk of datasets and high-dimensional parameter spaces. The method consolidates the advantages of AdaGrad and the RMSProp, which can be practiced with both sparse gradients and non-stationary objectives[10].

ISSN: 2414-1895 DOI: 10.6919/ICJE.202112_7(12).0039

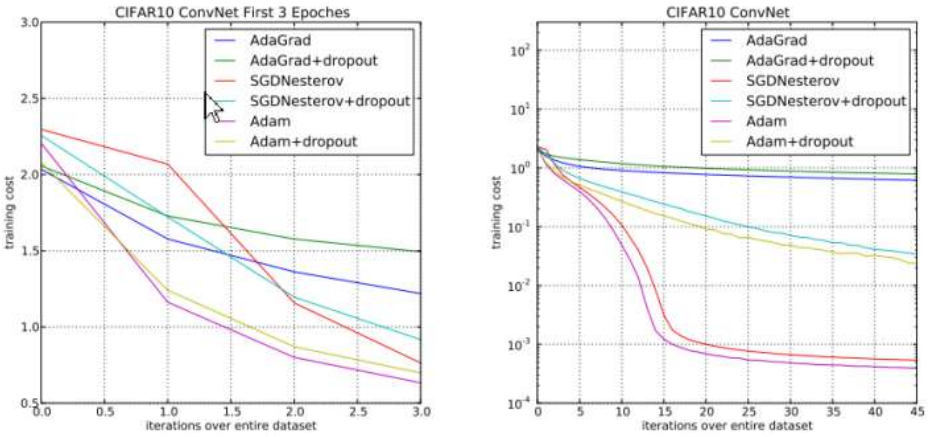

Figure 2 Convolutional neural networks training cost10.

The remaining portion of the paper is constructed as follows: automated surveillance system related work is touched on and introduced in section 3. Then, the methodology, including dataset construction, model training, and testing of our weapon detection model, is explained in section 4 at large, which follows the experiments and results in section 5. Finally, section 6 addressed the conclusion and future work.

2. Related Work

The method of weapons detection stemmed from old object detection techniques tapped into the sliding-window paradigm, which utilizes artificial traits and classifiers on jam-packed image grids to distinguish objects. For example, Viola and Jones11 described in their work that they implemented AdaBoost and the Haar feature to train a set of cascaded classifiers for face detection objectives, attaining adequate accuracy with excellent efficiency12. Another object detection technique is DPM13, mentioned by Felzenszwalb and Girshick, utilizing composites of deformable part models of different scales to represent massively shifting object types12. The current popular methods of object detection primarily tap into Deep learning approaches. Deep Learning-based Object Detection frameworks mainly consist of Region Proposal-based and Regression-based frameworks that involve models such as YOLO and SSD6. Region Proposal-based algorithms adopt a sliding window method to derive features from the visual data. Girshick introduced the RCNN model in 2014. For one thing, the whole image is processed with a Convolution Neural Network to produce a feature map. Then, the fixed-length vector with a Region of Interest pooling layer is extracted from each region proposal6. The research and work of detecting and classifying objects in real-time started after primary advancements in the CCTV field, processing devices, and deep learning frameworks14. Closed Circuit Television technology was first adopted in 1946 in Germany, where these cameras were placed on examining the launch of a rocket named V215. In 1973, Charge-Coupled Device (CCD) was developed and introduced surveillance cameras to the public in 1980. There have been notable technological advancements in the past few years, and advanced models or systems for visual object identification and detection for surveillance, administration, and security were presented14.

In reference to public safety technologies, threatening weapon detection was regularly in conjunction with concealed weapon detection (CWD). Circumstances such as luggage checks and other security schemes at airports and other public areas. CWD was based on imaging techniques approaches like infrared imaging and millimeter-wave for recognizing concealed weapons in checked or carried baggage and other locations. Also include methods such as the fusion-based technique of multi-scale decomposition, which combines visual color picture with infrared picture integration, and an approach based on visual, infrared, or mm-wave pictures using a multi-resolution mosaic technique pointing out the concealed threatening object from the target image14.

Darker first put forward the concept of automated detection of threatening weapons as a component of the MEDUSA project. MEDUSA is designed to detect both guns as objects and people who attempt

ISSN: 2414-1895 DOI: 10.6919/ICJE.202112_7(12).0039

to commit criminal activity via CCTV automatically. This system depends on eliciting the features that CCTV operators could detect as a concealed weapon or suspicious action. Darker and his team also conducted experiments to utilize CCTV as a mechanical sensor for firearm detection17. Marbach introduced a method for automated fire detection based on the temporal variation of fire intensity detected by a visual image sensor. The complete image sequences are examined to decide a candidate flame area. Fire features are extracted from the candidate flame region to decide fire patterns, and an alarm will be triggered if the pattern persists over a certain period18.

Furthermore, Arslan proposed a resolution for threat evaluation using visual hierarchy and conceptual firearms philosophy, generating the sequence of visual hierarchies that have better accuracy and more extensive scale. The threat evaluation equation is improved and adopted high-level information from the conceptual hierarchy. The conceptual philosophy is also refined, and more connections were formed between these two hierarchies19. Barros introduced the FISVER framework for intelligent public safety in video-surveyed vehicles, capable of performing general object detection, including objects such as firearms, knives, and other threatening weapons20.

3. Methodology

3.1 Project Flow

The project process is divided into the following steps:

1) Download data on the website through a crawler tool.
2) explore the data and then process the data in order to make a data set.
3) Construct a convolutional neural network for experiments.
4) analyze and adjust according to the experimental results.

The specific project process is shown in Figure 3 Project Flow Chart.

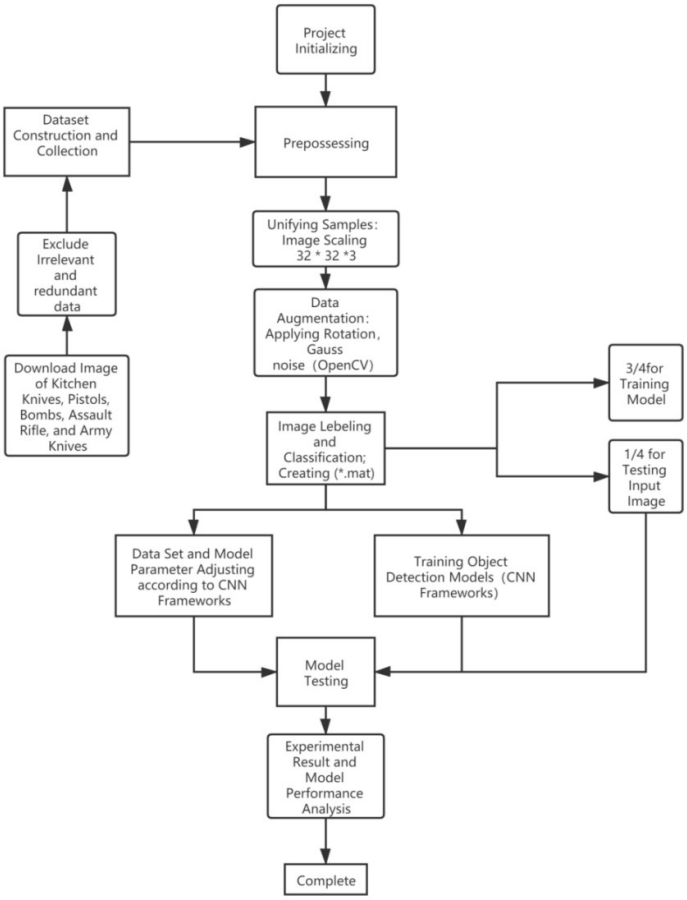

Figure 3 Project Flow Chart

ISSN: 2414-1895 DOI: 10.6919/ICJE.202112_7(12).0039

3.2 The selection and increase of images

To construct the dataset, images of kitchen knives, pistols, bombs, assault rifles, and combat knives, 900 images of each type, were downloaded from different sources available on the internet using automation scripts. Irrelevant or redundant images were erased, and the remaining images were pre-processed using OpenCV following the steps of unifying scaling(32*32*3), RGB to grayscale, and increased the num0ber of the images to around 14000 by applying rotation and Gauss noise.

Figure 4 Data set expansion method

3.3 labeling and classification

Table 1 Data set classification table

Weapon Type	Class identifier Index Assigned	Dataset Size	Image Instance
Bombs	0	6000	
Kitchen Knives	1	6000	
Pistol	2	6000	
Combat Knives	3	6000	
Assault Rifles	4	6000	

ISSN: 2414-1895 DOI: 10.6919/ICJE.202112_7(12).0039

We then operated image labeling and classification using a custom script. This script automatically assigns each image an index according to categories and creating a (*.mat) file where the class identifier (each with index from 0 to 4) is stored. These files give us a complete labeled dataset consisting of 3000 images of each weapon type and 15000 images in total.

3.4 Training and refining of accuracy

Then proceeding Deep Learning frameworks and models of CNN were trained using the pre-processes and labeled dataset. Again, using the above script, We randomly divided the dataset into two segments, 1/5 for testing as the input image and 4/5 for object detection model training. Finally, we operated the object detection model and dataset parameter adjusting according to CNN frameworks parallelly. After completing object detection model training with 4/5 of the dataset, the desired learning weights were obtained. The remaining 1/5 of the dataset was tested against these obtained weights, we then operated a complete and detailed performance analysis of the accuracy and precision of our model, and the experimental results were recorded in spreadsheet.

4. Dataset analysis

The program runs on three devices and the parameters of are shown in the table:

Table 2 Equipment parameter table

Number	*CPU*	*GPU*
1	AND Ryzen7 4800H	RTX 2060
2	intel(R) Core(TM) i7-9750H CPU	GeForce RTX 2080 with Max-Q Design
3	AMD Ryzen9 5900HX with Radeon	GeForce RTX 3080 Laptop GPU

4.1 To obtain training and test data:

In Baidu, the crawler obtained about 1000 photos each in 5 categories (chopper, pistol, submachine gun, bomb, saber). Then we manually clean the data to remove irrelevant photos and use OpenCV to adjust picture parameters (rotate, adjust grayscale, add Gaussian noise). Expand the data sample to about 6000, more than 30,000 in total. And adjust the size of all pictures to 32*32 Integrate five types of data into mat data types and divide them into train and test at a ratio of 4:1. Finally, label the data in one-hot encoding

4.2 Start running the program:

For deep learning or machine learning models, we not only require it to have a good fit (training error) to the training data set. At the same time, it is also hoped that it can have a good fitting result (generalization ability) to the unknown data set (test set), and the resulting test error is called the generalization error.

The most intuitive performance to measure the generalization ability is the over-fitting and under-fitting of the model. Over-fitting and under-fitting are used to describe the two states of the model in the training process.

In order to prevent overfitting, we use regularization methods to modify the learning algorithm to reduce the generalization error instead of the training error

4.3 L1 regularization

Add an L1 regularization term after the original loss function, that is, the sum of the absolute values of all weights, and then multiply by λ/n. Then the loss function becomes:

Corresponding gradient (derivative):

$$\frac{\partial C}{\partial \omega} = \frac{\partial C_0}{\partial \omega} + \frac{\lambda}{n} sgn(\omega) \tag{1}$$

Which simply takes the sign of each element.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202112_7(12).0039

$$sgn(w) = \begin{cases} 1, w > 0 \\ 0, w = 0 \\ -1. w < 0 \end{cases} \tag{2}$$

When the gradient is descent, the weight update becomes:

$$\omega \to \omega' = \omega - \frac{n\lambda}{n} sgn(w) - \eta \frac{\partial C_0}{\partial \omega} \tag{3}$$

When w=0, |w| is not leadable. So, we can only update w according to the original unregularized method.
When w>0, sgn (w)>0, the updated w becomes smaller when the gradient drops.
When w<0, sgn (w)>0, the updated w becomes larger when the gradient drops.
That is, L1 regularization makes the weight w closer to 0, so that the weight in the network is as 0 as possible, which is equivalent to reducing the network complexity and preventing overfitting.
This is why L1 regularization produces sparser solutions. The sparsity here means that some of the parameters in the optimal value are 0. The sparse nature of L1 regularization has been widely used in feature selection mechanisms to select meaningful features from a subset of available features.

4.4 L2 regularization

L2 regularization is usually called weight decay, which is to add an L2 regularization term after the original loss function, that is, the sum of the squares of all weights w, and then multiply it by λ/2n. Then the loss function becomes:
Corresponding gradient (derivative):

$$C = C_0 + \frac{\lambda}{2n} \cdot \sum w_i^2 \tag{4}$$

It can be found that L2 regularization has no effect on the update of the bias b, but it has an effect on the update of the weight w:

$$\frac{\partial C}{\partial \omega} = \frac{\partial C_0}{\partial \omega} + \frac{\lambda}{n} \omega \tag{5}$$

$$\frac{\partial C}{\partial b} = \frac{\partial C_0}{\partial b} \tag{6}$$

L2 regularization has the effect of making the weight parameter w smaller. The reason it can prevent overfitting is that a smaller weight parameter w means that the complexity of the model is lower, and the fitting of the training data is just right, and the training data will not be overfitted, thereby improving the generalization ability of the model.

Dropout

At the same time, we added the dropout technique, which is equivalent to adding noise in the hidden unit. Dropout refers to randomly deleting a part of hidden units (neurons) with a certain probability (such as 50%) each time during the training process. That is, the activation function of this part of neurons is set to 0 (the output of the activation function is 0), so that these neurons are not calculated.

Dropout can also effectively prevent overfitting

a) Different training models will be produced during the training process, and different training models will also produce different calculation results. As the training continues, the calculation result will fluctuate within a range, but the average value will not change much, so the final training result can be regarded as the average output of different models.
b) It eliminates or weakens the union between neuron nodes and reduces the network's dependence on a single neuron, thereby enhancing the generalization ability.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202112_7(12).0039

Minibatch gradient descent method

The program uses the minibatch gradient descent method.The main problem to be solved is due to the large training data (about 24,000) during actual recognition, and the memory may not be enough when loaded into the computer at a timeTherefore, the training data is divided into multiple copies, and one copy is learned at a time. In this way, the problem of insufficient memory can be solved, and the loss value converges faster.In order to prevent the learning rate from being too large, it will oscillate back and forth when it converges to the global best point. Therefore, the learning rate should be continuously reduced exponentially with the number of training rounds, and the learning step of convergence gradient decline.TensorFlow provides a function to achieve exponential decay learning rate, called tf.train.exponential_decay.The team finally controlled the accuracy rate to 72% by adjusting various parameters.

The network structure parameters are shown below:

Table 3. Data set classification table

Train batch size	16	Test batch size	500
Dropout rate	0.9	Pooling scale	2
Base learning rate	0.001	Decay rate	0.98

5. Conclusion and Future Work

5.1 Conclusion

This research analyzes the automatic identification of dangerous goods, uses the multi-layer feature extraction method of convolutional neural network to extract the characteristics of dangerous goods, and then classifies them according to the characteristics, and finally realizes the identification of dangerous goods and the accuracy of the model. It reached 72%.

5.2 Future scope

The future work includes reducing the false positives and negatives even more as there is still a need for improvement. We might also try to increase the number of classes or objects in the future but the priority is to further improve precision.

References

[1] Aggarwal, Mayank (March 17, 2021). "Atlanta spa shootings: What we know about 'religious, nerdy' suspect Robert Aaron Long". The Independent. Retrieved March 17, 2021.

[2] Lamb, Gina (April 17, 2008). "Columbine High School". The New York Times. Archived from the original on January 2, 2016. Retrieved February 6, 2016.

[3] M. Planty and J. L. Truman, Firearm Violence, 1993–2011, Washington: U.S. Department of Justice, (2013).

[4] C. A. Perkins, Weapon Use and Violent Crime, 1993–2001. Washington: U.S. Department of Justice, (2003).

[5] Dasiopoulou, Stamatia, et al. "Knowledgeassisted semantic video object detection." IEEE Transactions on Circuits and Systems for Video Technology 15.10 (2005): 1210–1224.

[6] Varshney, P., Tyagi, H., Lohia, N.K., Kajla, A., & Girdhar, P. (2021). A Deep Learning Based Approach to Detect Suspicious Weapons.

[7] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, et al., "Mastering the game of go with deep neural networks and tree search," Nature, vol. 529, pp. 484–489, 2016.

[8] Y.-M. L. Ko-Wei Huang, C.-C. Lin, and Z.-X. Wu, ''A deep learning and image recognition system for image recognition,'' Data Sci. Pattern Recognit., vol. 3, no. 2, pp. 1–23, 2019.

[9] F. Sadoughi, M. Ghaderzadeh, M. Solimany, and R. Fein, "An intelligent system based on back propagation neural network and particle swarm optimization for detection of prostate cancer from benign hyperplasia of prostate," Journal of Health and Medical Informatics, vol. 5(3), pp. 1–5, 2014

[10] Diederik Kingma and JimmyBa. 2015. Adam: A method for stochastic optimization. In International Conference on Learning Representations (ICLR).

[11] P. A. Viola and M. J. Jones. Rapid object detection using a boosted cascade of simple features. In CVPR, pages 511– 518, 2001. 3

[12] Zhang, Shifeng (2018). "Single-Shot Refinement Neural Network for Object Detection". Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition: 4203–4212. arXiv:1711.06897. Bibcode:2017arXiv171106897Z.

[13] P. F. Felzenszwalb, R. B. Girshick, D. A. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. TPAMI, 32(9):1627–1645, 2010.

[14] M. T. Bhatti, M. G. Khan, M. Aslam and M. J. Fiaz, "Weapon Detection in Real-Time CCTV Videos Using Deep Learning," in IEEE Access, vol. 9, pp. 34366-34382, 2021, doi: 10.1109/ACCESS.2021. 3059170.

[15] (2019). From Edison to Internet: A History of Video Surveillance. Accessed: Jun. 13, 2019. [Online].

[16] Darker, I.; Gale, A.; Ward, L.; Blechko, A. Can CCTV Reliably Detect Gun Crime? In Proceedings of the 41st Annual IEEE International Carnahan Conference on Security Technology, Ottawa, ON, USA, 8–11 October 2007; pp. 264–271.

[17] Darker, I.T.; Gale, A.G.; Blechko, A. CCTV as an automated sensor for firearms detection: Human-derived performance as a precursor to automatic recognition. In Proceedings of the International Society for Optical Engineering, 15 September 2008; pp. 11–12.

[18] Marbach, G.; Loepfe, M.; Brupbacher, T. An image processing technique for fire detection in video images. Fire Saf. J. 2006, 41, 285–289.

[19] Arslan, A.N.; Hempelmann, C.F.; Attardo, S.; Blount, G.P.; Sirakov, N.M. Threat assessment using visual hierarchy and conceptual firearms ontology. Opt. Eng. 2015, 54, 105–109.

[20] Barros, H.; Neto, A. FISVER: A Framework for Smart Public Safety in Video-Surveilled Vehicles. In Proceedings of the 3rd Internation Workshop on ADVANCEs in ICT Infrastructures and Services, Miami, FL, USA, 24 August 2014; pp. 221–225.

[21] G. K. Verma and A. Dhillon, ‘‘A handheld gun detection using faster RCNN deep learning,’’ in Proc. 7th Int. Conf. Comput. Commun. Technol.(ICCCT), 2017, pp. 84–88.

[22] Dhiraj, Deepak Kumar Jain, An evaluation of deep learning based object detection strategies for threat object detection in baggage security imagery, Pattern Recognition Letters, Volume 120, 2019, Pages 112-119, ISSN 0167-8655,

[23] K. He, X. Zhang, S. Ren, and J. Sun, ‘‘Deep residual learning for image recognition,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778.

[24] M. Nakib, R. T. Khan, M. S. Hasan, and J. Uddin, ‘‘Crime scene prediction by detecting threatening objects using convolutional neural network,’’in Proc. Int. Conf. Comput., Commun., Chem., Mater. Electron. Eng.(IC4ME2), Feb. 2018, pp. 1–4.

[25] J. O. Laguna, A. G. Olaya, and D. Borrajo, ‘‘A dynamic sliding window approach for activity recognition,’’ in Proc. Int. Conf. User Modeling,

[26] M. Grega, A. Matiolański, P. Guzik, and M. Leszczuk, ‘‘Automated detection of firearms and knives in a CCTV image,’’ Sensors, vol. 16, no. 1,p. 47, Jan. 2016.

[27] G. Flitton, T. P. Breckon, and N. Megherbi, ‘‘A comparison of 3D interest point descriptors with application to airport baggage object detection in complex CT imagery,’’ Pattern Recognit., vol. 46, no. 9, pp. 2420–2436, Sep. 2013.

[28] R. Gesick, C. Saritac, and C.-C. Hung, ‘‘Automatic image analysis process for the detection of concealed weapons,’’ in Proc. 5th Annu. Workshop Cyber Security. Inf. Intell. Res. Cyber Security. Inf. Intell. Challenges Strategies(CSIIRW), 2009, p. 20.

[29] R. K. Tiwari and G. K. Verma, ‘‘A computer vision based framework for visual gun detection using Harris interest point detector,’’ Procedia Comput. Sci., vol. 54, pp. 703–712, Aug. 2015.

[30] A. C. Sankaranarayanan, A. Veeraraghavan, and R. Chellappa, ‘‘Object detection, tracking and recognition for multiple smart cameras,’’ Proc.IEEE, vol. 96, no. 10, pp. 1606–1624, Oct. 2008.

[31] Z. Xiao, X. Lu, J. Yan, L. Wu, and L. Ren, ‘‘Automatic detection of concealed pistols using passive millimeter wave imaging,’’ in Proc. IEEE Int. Conf. Imag. Syst. Techn. (IST), Sep. 2015, pp. 1–4.

[32] S. Zhang, C. Wang, S.-C. Chan, X. Wei, and C.-H. Ho, ‘‘New object detection, tracking, and recognition approaches for video surveillance over camera network,’’ IEEE Sensors J., vol. 15, no. 5, pp. 2679–2691, May 2015.

[33] H. Jain, A. Vikram, Mohana, A. Kashyap and A. Jain, "Weapon Detection using Artificial Intelligence and Deep Learning for Security Applications," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020, pp. 193-198, doi: 10.1109/ICESC48915. 2020.9155832.

[34] T. S. S. Hashmi, N. U. Haq, M. M. Fraz, and M. Shahzad, "Application of Deep Learning for Weapons Detection in Surveillance Videos," 2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2), 2021, pp. 1-6, doi: 10.1109/ICoDT252288.2021.9441523.

[35] Wu, R., Yan, S., Shan, Y., Dang, Q., & Sun, G. (2015). Deep Image: Scaling up Image Recognition. ArXiv, abs/1501.02876.

[36] Krizhevsky, A., Sutskever, I., & Hinton, G.E. (2012). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60, 84 - 90.

[37] Jose L. Salazar González, Carlos Zaccaro, Juan A. Álvarez-García, Luis M. Soria Morillo, Fernando Sancho Caparrini, Real-time gun detection in CCTV: An open problem, Neural Networks, Volume 132, 2020, Pages 297-308, ISSN 0893-6080

[38] Zhang Q, Zhang R M, Chen B. Research review of image recognition technology based on deep learning[J]. Journal of the Hebei Academy of Sciences, 2019, 36(3): 28-36.

[39] Yuanpan Zheng , Guangyang Li and Ye Li. Survey of application of deep learning in image recognition[J]. Computer Engineering and Applications, 2019, 55(12): 20–36.