TELKOMNIKA Telecommunication, Computing, Electronics and Control
Vol. 18, No. 6, December 2020, pp. 3026~3032
ISSN: 1693-6930, accredited First Grade by Kemenristekdikti, Decree No: 21/E/KPT/2018
DOI: 10.12928/TELKOMNIKA.v18i6.16174

The detection of handguns from live-video in real-time based on deep learning

Mohammed Ghazal, Najwan Waisi, Nawal Abdullah
Department of Computer Engineering Technology, Northern Technical University, Mosul, Iraq

Article Info

Article history:

Received Mar 27, 2020
Revised Jun 12, 2020
Accepted Jul 9, 2020

Keywords:

CNN
Convolutional neural networks
Deep learning
Handgun
MobileNetV3
SSDLite
Weapon

ABSTRACT

Many people have been killed indiscriminately by the use of handguns in different countries. Terrorist acts, online fighting games and mentally disturbed people are considered the common reasons for these crimes. A real-time handguns detection surveillance system is built to overcome these bad acts, based on convolutional neural networks (CNNs). This method is focused on the detection of different weapons, such as (handgun and rifles). The identification of handguns from surveillance cameras and images requires monitoring by human supervisor, that can cause errors. To overcome this issue, the designed detection system sends an alert message to the supervisor when a weapon is detected. In the proposed detection system, a pre-trained deep learning model MobileNetV3-SSDLite is used to perform the handgun detection operation. This model has been selected because it is fast and accurate in infering to integrate network for detecting and classifying weapons in images. The experimental result using global handguns datasets of various weapons showed that the use of MobileNetV3 with SSDLite model both enhance the accuracy level in identifying the real time handguns detection.

Corresponding Author:

Mohammed Talal Ghazal,
Department of Computer Engineering Technology,
Northern Technical University,
Mosul, Iraq.
Email: mohammed.ghazal@ntu.edu.iq

1. INTRODUCTION

Crime rates using handguns have increased in many places in the world [1], especially in the countries that cannot restrict weapons to the state. Daily crimes occur with various weapons in public places and places of worship. The United Nations Office for Drugs and Crime (UNODC) reported the statistics for the number of crimes involving weapons per 100,000 habitants is very high in many countries, for example., 10.08 in Central Iraq, 2.5 in Libya and 2.20 in Syria [2]. In iraq, police tried to impose their control in these places by deploying security men among the crowds and, by installing surveillance cameras. These methods are ineffective and expensive, especially when multiple video streams are present on the camera systems. The suggested method of handgun detection system which is based on the video streams from surveillance cameras can alert the security men if a person with gun is detected in real-time, which will lead to speedup handgun detection before the crime occure.

The proposed system for handgun detection in real-time is based on convolutional neural network (CNN). The CNNs has been rapidly growing in computer vision area during the past few years [3-7]. In computer vision there are three object detection models have been analyzed (faster region-based convolutional neural networks (FR-CNN), region based-fully convolutional networks (R-FCN) and single shot detector (SSD)) [8]. To enhance the speed, accuracy and performance, these models have been combined with different feature extractors such as (VGG, Mobilenet and Resnet-101), for that reason the MobileNetv3 with SSDLite model have been used in the proposed system [9]. The MobileNetv3 is %27 faster than MobileNetv2, while maintaining similar mAP. Moreover, its classification is based on pre-trained dataset on the ImageNet. A widely used dataset is ImageNet, which is an object detection dataset which containes about 1.28 million images with approximately 1,000 object classes [10]. Our approach is to use a global dataset containing images of the most used types of weapons, to be close to real circumstance, the handguns datasets for both the sliding window and region proposals approachs have been selected [11], in addition to, weapon dataset in [12]. These contain images of weapons in a variety of different contexts, situations, and orientations. The datasets are trained on MobileNetv3_SSDLite Convolutional Neural Network CNN.

2. RELATED WORKS

There are two methods for handgun detection system: the first is based on metal detection using X-ray or millimetric wave images, it used in airports to check the passenger luggage's [13]. these methods are unable to non-metallic handgun detection and it is limited to use in specific places. The second is based on RGB images [14, 15], from the recent papers fast region-based convolutional network (FRCNN) and region-based convolutional neural networks (RCNN) based models are applied to automatic handgun detection system in surveillance videos [12], another author use a novel object detection algorithms to detect the visual knife for the given video dataset [16], the foreground segmentation is used for features extraction, the feature detection performed by Features from accelerated segment test (FAST) and the multi-resolution analysis (MRA) is used for classification. A VGGNET19 pre-trained model is used for detecting gun and knife in hands of a person [17]. These systems are accurate, but the handgun detection especially when crimes occur need a faster and more accurate system.

3. THE PROPOSED MODEL

The main idea of object detection is to recognize the object in the input image and find its location [18]. The designed system focuses on detecting handguns in minimum training time with high accuracy results. A pre-trained model such as GoogleNet, VGGNet-19 and MobileNet have been trained with more than million images to minimize the object detection errors in images [19-21]. Based on fast and accurately training properties, a MobileNetv3 model is used in this method to classify and detect handguns accurately. Experimentally, the system is tested with an image dataset which contains various images of weapons and different position of gunslinger.

The building of mobile models is based on enhancing the efficiency for the building blocks [22]. MobileNetV1 [23] presented depthwise separable convolutions (DSC) as an efficient change for other CNN layers. DSC is utilized to decompose the traditional convolution into depthwise convolution and pointwise convolution, in MobileNet [18]. In Depthwise convolution approach, a single convolutional filter is applied for each input channel, whereas the Pointwise convolution performs a 1*1 convolution to combine those separate channels as shown in Figure 1.

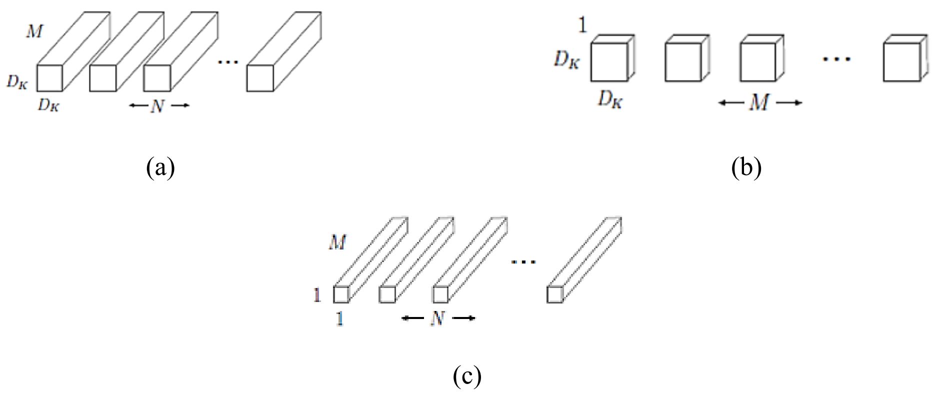

Figure 1. Depth-wise separable convolution DSC in MobileNet;
(a) standard convolution, (b) depthwish convolutional filters, (c) 1x1 convolutional filters called pointwise in the context of depth separable convolution

The novel design for this model leads to reduce both of the complexity of computation and the parameters number. MobileNetV2 [24] presented the linear bottleneck and inverted residual model so as to get a much efficient structure for the CNN layers by utilizing the low rank nature of the problem. The architecture of MobileNetV2 is built based on the structure of inverted residual as shown in Figure 2, where the input and output connections of the residual block are thin bottleneck layers. The intermediate expansion layer utilizes the lightweight depthwise convolutions to filter the features. MnasNet [25] is built based upon the structure of a MobileNetV2 model which presents the lightweight attention modules based on squeeze and excitation into the bottleneck structure. The desgined module is placed in the expansion, followed by the depthwise filters to apply it on the largest representation as shown on Figure 3.

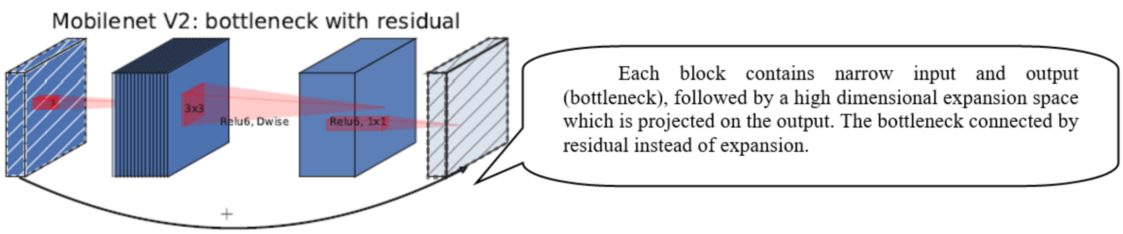

Figure 2. MobileNetV2 layer (inverted residual and linear bottleneck)

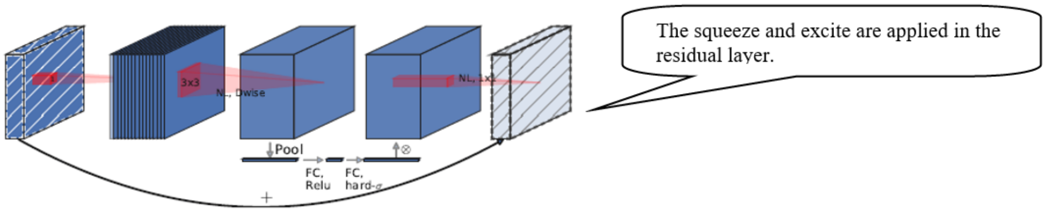

Figure 3. MobileNetV2 with squeeze-and-excite

To build more efficint model, MobileNetV3 is combined to improved these layers with modified swish nonlinearities. it is replaced the sigmoid activation function by hard sigmoid because its easier to calculate and more accurate. MobileNetV3 has two models, which are MobileNetV3-Large and MobileNetV3-Small. The MobileNetV3-Large has a higher resource than the MobileNetV3-Small, the specification for both networks are shown in Tables 1 and 2.

Table 1. Specification for the mobileNet V3-large

Input	Operator	exp size	#out	SE	NL	s
224^2 x 3	conv2d	-	16	-	HS	2
112^2 x16	bnec, 3x3	16	16	-	RE	1
112^2 x 16	bnec, 3x3	64	24	-	RE	2
56^2 x 24	bnec, 3x3	72	24	-	RE	1
56^2 x 24	bnec, 5x5	72	40	✓	RE	2
28^2 x 40	bnec, 5x5	120	40	✓	RE	1
28^2 x 40	bnec, 5x5	120	40	✓	RE	1
28^2 x 40	bnec, 3x3	240	80	-	HS	2
14^2 x 80	bnec, 3x3	200	80	-	HS	1
14^2 x 80	bnec, 3x3	184	80	-	HS	1
14^2 x 80	bnec, 3x3	184	80	-	HS	1
14^2 x 80	bnec, 3x3	480	112	✓	HS	1
14^2 x 112	bnec, 3x3	672	112	✓	HS	1
14^2 x 112	bnec, 5x5	672	160	✓	HS	1
14^2 x 112	bnec, 5x5	672	160	✓	HS	2
7^2 x 160	bnec, 5x5	960	160	✓	HS	1
7^2 x 160	conv2d, 1x1	-	960	-	HS	1
7^2 x 960	Pool, 7x7	-	-	-	HS	-
1^2 x 960	conv2d 1x1, nbn	-	1280	-	HS	1
1^2 x 1280	conv2d 1x1, nbn	-	k	-		-

Table 2. Specification for MobileNetV3-small

Input	Operator	exp size	*#out*	SE	NL
224^2 x 3	conv2d, 3x3	-	16	-	HS
112^2 x24	bnec, 3x3	16	16	✓	RE
56^2 x 24	bnec, 3x3	72	24	-	RE
28^2 x 24	bnec, 3x3	88	24	-	RE
28^2 x 40	bnec, 5x5	96	40	✓	HS
14^2 x 40	bnec, 5x5	240	40	✓	HS
14^2 x 40	bnec, 5x5	240	40	✓	HS
14^2 x 40	bnec, 5x5	120	48	✓	HS
14^2 x 48	bnec, 5x5	144	48	✓	HS
14^2 x 96	bnec, 5x5	288	96	✓	HS
7^2 x 96	bnec, 5x5	576	96	✓	HS
7^2 x 96	bnec, 5x5	576	96	✓	HS
7^2 x 96	conv2d, 1x1	-	576	✓	HS
7^2 x 576	Pool, 7x7	-	-	-	HS
1^2 x 576	conv2d, 1x1	-	1280	-	HS
1^2 x 1280	conv2d, 1x1	-	k	-	HS

4. EXPERIMENTAL MODEL TRAINING

The experiments have been accomplished by using a workstation equipped with an Intel Core I7-7700 CPU @ 3.60GHZ, NVIDIA Graphics Processing Unit (GPU) of 3840 CUDA cores and 16GB DDR4 RAM 2400 MHz, with Ubuntu 16.04 operating system. To evaluate the proposed model, three datasets are used for the learning phase:

- The handgun dataset of sliding window approach, which consists of 102 classes with around 9261 images [11].
- The handgun dataset of region proposals approach, which consists of 608 images and includes 304 images of handguns [11].
- Weapon dataset, which contains 3000 images of guns taken in a variety of contexts [12].

Figure 4 shows some samples of handgun dataset. The proposed model trained with caffe deep learning library to approach handgun detection in live videos with near to real time. Figure 5 illustrates the handguns detection system structure. At first, the program initializes the trained list of class labels by MobileNetV3- SSDLite to detect handguns. After that, a set of bounding box colors are generated for each class before loading the proposed model. Then, begining to input images and resize them to the size of 300x300 pixels, and then convert them to blobs. Finaly, passing the blobs through the CNN network to obtain detection and predications of handguns. Each time the handgun detector finds a weapon, the CNN model returns boundary box around the object (handgun or any other types of defined weapons in the datasets), with determined centroid, the system sends an alert message to the supervisor.

Figure 4. Samples of weapon dataset

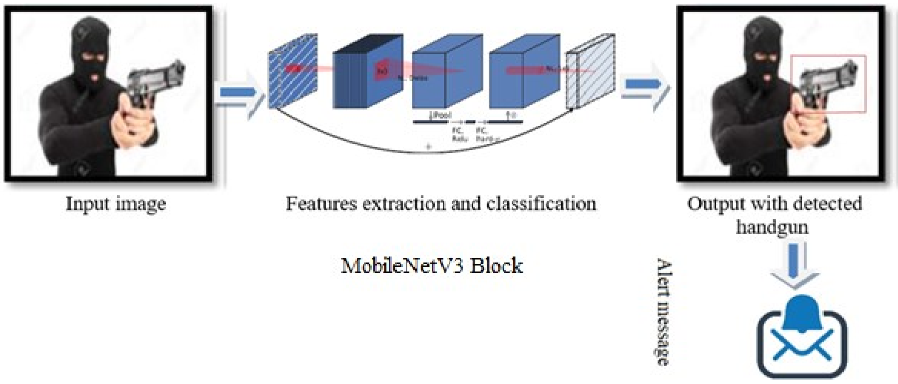

Figure 5. Handgun detection system structure

5. RESULTS

The findings of this method suggest that MobileNetV3 is suitable for handgun detection as it delivers a perfect correspondence between prediction speed and accuracy. The proposed model achieved 96% training accuracy rate while the result of previous networks: MobileNetV2-SSD and GoogleNet inception were 89% and 87% respectively. The speed achieved by the experiment using MobileNetV3-SSDLite is ~20–23 FPS, whereas MobileNetV2-SSD has achieved ~15–18 FPS. In terms of prediction, MobileNetV3-SSDLite has the best performance 0.942 for feature extracter while MobileNetV2-SSD has 0.816. The experiments have been performed in a real-time detection in live videos and images as shown in Figure 6. The system is provided with SMS feature, which allow the system to sends an alert message to the supervisor whenever a handgun is detected.

Figure 6. Example of handgun detection system

6. CONCLUSION

The proposed Handgun Detection System is really an applicable system which helps to detect crimes and gunslinger. The use of MobileNetV3 pre-trained model is much efficient because it is faster than other models, such as MobileNetV2 and GoogleNet. Moreover, the security alarm property increases system effectiveness. The proposed system can be used in various applications, such as, real time detection of guns in supermarkets which is monitored by cameras, and also checking whether the uploaded videos to youtube and other social media networks include scenes with guns.

ACKNOWLEDGEMENTS

This research is funded by cooperation between Deep Learning team in Northern Technical University NTU in Iraq. Website: For Northern Technical University https://www.ntu.edu.iq

REFERENCES

[1] Sturup J., *et al.*, "Increased gun violence among young males in Sweden: a descriptive national survey and international compa," *European Journal on Criminal Policy and Research*,vol. 25, no. 4, pp. 365-378, May 2018.
[2] United Nations Office on Drugs and Crime (UNODC), "Global study on homicide 2019," Data: UNODC Homicide Statistics, 2019.
[3] Chen W., *et al.*, "All you need is a few shifts: Designing efficient convolutional neural networks for image classification". *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 2019.
[4] Liu L., *et al.*, "Deep learning for generic object detection: A survey". *International Journal of Computer Vision*, vol. 128, pp. 261–318, September 2018.
[5] Zoph B, *et al.*, "Learning transferable architectures for scalable image recognition". *Proceedings of the IEEE conference on computer vision and pattern recognition*, June 2018.
[6] Li Z., Gavrilyuk K., Gavves E., Jain M., Snoek C. G., "Video LSTM convolves, attends and flows for action recognition*," Computer Vision and Image Understanding*, vol. 166, pp. 41-50, January 2018.
[7] Martínez F., Hernánde C., Martínez F. E., "Evaluation of deep neural network architectures in the identification of bone fissures," *TELKOMNIKA Telecommunication Computing Electronics and Control*, vol. 18, no. 2, pp. 807-814, April 2020.
[8] Heredia A, Barros-Gavilanes G, editors. " Video processing inside embedded devices using ssd-mobilenet to count mobility actors," *2019 IEEE Colombian Conference on Applications in Computational Intelligence (ColCACI)*, June 2019.
[9] Howard A., *et al.*, "Searching for mobilenetv3", *Proceedings of the IEEE International Conference on Computer Vision*, November 2019.
[10] Simonyan K, Zisserman A. "Very deep convolutional networks for large-scale image recognition". *arXiv preprint arXiv*, September 2014.
[11] Elmir Y., Laouar S. A., Hamdaoui L.," Deep learning for automatic detection of handguns in video sequences," *3rd edition of the National Study Day on Research on Computer Sciences (JERI 2019),* Saida, Algeria, vol. 2351, pp. 1-10, April 27, 2019. [Online]. Available: http://ceur-ws.org/Vol-2351/paper_69.pdf.
[12] Olmos R., Tabik S., Herrera F., "Automatic handgun detection alarm in videos using deep learning," *Neurocomputing*, vol. 275, pp. 66-72, February 2018.
[13] Flitton G., Breckon T. P., Megherbi N., "A comparison of 3D interest point descriptors with application to airport baggage object detection in complex CT imagery," *Pattern Recognition*, vol. 46, no. 9, pp. 2420-2436, September 2013.
[14] Tiwari R. K., Verma G. K., "A computer vision based framework for visual gun detection using harris interest point detector," *Procedia Computer Science*, vol. 54, pp. 703-712, 2015.
[15] Tiwari R. K., Verma G. K., "A computer vision based framework for visual gun detection using SURF". *2015 International Conference on Electrical, Electronics, Signals, Communication and Optimization (EESCO)*, January 2015.
[16] Buckchash H., *et al.*, "A robust object detector: application to detection of visual knives," *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, July 2017.
[17] Navalgund U. V., Priyadharshini K., "Crime intention detection system using deep learning," *2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET)*, December 2018.
[18] Feng X., *et al.*, "Computer vision algorithms and hardware implementations: A survey," *Integration*, vol. 69, pp. 3019-320, November 2019.
[19] Masud M., *et al.,* "Deep learning-based intelligent face recognition in IoT-cloud environment," *Computer Communications*, vol. 152, pp. 215-222, February 2020.
[20] Kurdthongmee W., "A comparative study of the effectiveness of using popular DNN object detection algorithms for pith detection in cross-sectional images of parawood," *Heliyon*, vol. 6, no. 2, February 2020.
[21] Attamimi M., Mardiyanto R., Irfansyah A., "Inclined image recognition for aerial mapping using deep learning and tree based models," *Telecommunication, Computing, Electronics and Control*, vol. 16, no. 6, pp. 3034-3044, December 2018.
[22] Shadeed G. A., Tawfeeq M. A., Mahmoud S. M., "Deep learning model for thorax diseases detection". *Telecommunication, Computing, Electronics and Control*, vol. 18, no. 1, pp. 441-449, February 2020.
[23] Howard AG, et al., "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv*, April 2017.
[24] Sandler M., *et al.*, "Mobilenetv2: inverted residuals and linear bottlenecks," *Proceedings of the IEEE conference on computer vision and pattern recognition*, June 2018.
[25] Tan M., *et al.*, "Mnasnet: Platform-aware neural architecture search for mobile," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 2019.

BIOGRAPHIES OF AUTHORS

Mohammed Talal Ghazal obtained his M.Sc. degree from Computer Engineering Technology, Northern Technical University, Mosul, Iraq in 2016. His M.Sc. thesis entitled: "Wheelchair Robot Control Using EOG signals". His research interests include the design of face recognition algorithms, deep learning CNNs models and object detection.

Najwan Zuhair Waisi obtained her M.Sc. degree from Computer Science Department, Mousl University, Mosul, Iraq in 2014. Her M.Sc. thesis entitled: "Design and Implementation of Client Honeypot" and her current research interests include the deep learning CNNs models and object detection.

Nawal Younis Abdullah obtained her M.Sc. degree from Computer Engineering Technology, Northern Technical University, Mosul, Iraq in 2014. Her M.Sc. thesis entitled: "FPGA Based Video Scene Boundaries Detection Using Enhanced Sobel Filter" and her current research interests include the deep learning CNNs models and object detection.