Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.10 n.47, p. 19-26. May/June, 2024.
DOI: https://doi.org/10.5935/jetia.v10i47.1039

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

WEAPON DETECTION AND CLASSIFICATION USING DEEP LEARNING

Sunil B. Mane[1]

[1]Dept. of Computer Science and Engineering COEP Technological University (COEP Tech) Pune, Maharashtra, India.

[1]http://orcid.org/0000-0002-7111-4908

Email: sunilbmane.comp@coeptech.ac.in

ARTICLE INFO

Article History
Received: January 23th, 2024
Revised: June 03th, 2024
Accepted: June 03th, 2024
Published: July 01th, 2024

Keywords:
Computer Vision,
Weapon Detection,
Deep Learning,
Deep Learning Libraries.

ABSTRACT

High gun-related crime rate poses a great threat to society in the present world. There is a serious need for systems to deal with such gun-related crimes. As CCTVs are installed in almost every part of the city, using the CCTV footage to detect the weapon is the simplest and efficient way to deal with such crimes. Unconcealed weapon detection, in images and videos, can help reduce the number of homicides due to the gun-related violence. In this work, we focus on developing a robust and automatic weapon detection system with ability to classify the detected weapon into different categories. This work provides and extensive survey on already existing weapon detection systems, weapon detection datasets, challenges in weapon detection and deep learning-based object detection technologies. We have developed a new image dataset for weapon detection and classification task. The experimental analysis shows the superiority of the Faster-RCNN models over SSD models for weapon detection systems. The detection results show how the final developed system deals with different challenges related to weapon detection and classification in real world scenarios.

I. INTRODUCTION

The crime rates due to the use of weapons have increased drastically in recent years. The survey [1] shows an average of 32.23% of total homicides by guns and a total of 128k homicides due to guns combined for all the countries. Gun violence is a very concerning issue in every nation, especially in countries where owning a gun is legal. According to a survey [1] average of 10.42 guns are owned per 100 people by the entire world. Despite the efforts taken to control gun violence, the crime rates have not fallen even a bit in the past few years. Crimes involving weapons like knife, hand-held guns, assault rifles, etc includes robbery and are primarily carries out in public places and banks where CCTVs are installed. Computer vision is a prominent domain in Artificial Intelligence as it deals with analysis and extraction of information from images or sequence images. The visual data present is vast, and it is ready to be used for dealing with real world issues. The image or video data generated and upload to the internet today is mind blowing. Forbes 2018 survey [2] tells that every minute 300 hrs of the video is uploaded just on YouTube, which is a very high amount of data for a single video sharing platform. Also, 300 million photos get uploaded on Facebook every day. CCTV and cameras are present everywhere and the data generated from them can be used for solving social issues. Object Detection is Computer Vision technology to detect instances of a real-world object in an image or a video. The problem definition of object detection is to spatially locate the object in the image, known as Object Localization and to predict the class of the object, known as Object Classification. The use of object detection for solving real-world problems has increased tremendously. Object detection has proved tremendous use in the field of Artificial Intelligence where image and video analysis are prominent. Object detection forms the basis for solving more complex problems such as autonomous vehicles, image segmentation, intelligent video surveillance, robotic vision, and many more. Due to the advancements in deep learning, there has been a drastic increase in accuracy as well as the performance of the deep learning models. Application of deep learning models in computer vision has created a pathway for solving many problems related to object detection. With the introduction of Region based Convolutional Neural Network (R-CNN) [3] for feature extraction, the accuracy of the object detection systems has been significantly increased. Convolutional Neural Networks (CNN) are comparatively deeper than traditional feature extraction techniques and can learn multiple and more complex descriptors automatically. Deep convolutional neural networks (DCNN) used in deep learning models have proved way more accurate than hand-made technique in extracting features from images. The feature maps generated by CNNs are more useful for detection of the spatial location of the image while the features generated by the previous generation technique lacked the capability to spatially locate the detected image. The region-based deep CNN models are capable of detecting the object even if a part of the object is visible provided that the visible part is more than some threshold.

In this research work, the focus is on building a weapon detection system which would have capabilities to classify the detected weapon based on the predefined weapon categories. The task of classification is only possible if labels are provided with the bounding box annotations in the dataset used for training. We found that the dataset required for developing a weapon detection and classification dataset was not freely available. Hence, we also focus on developing an image dataset for weapon detection and classification dataset, with bounding box annotations as well as the weapon class labels. In Section 2, we have provided survey of the existing state of weapon detection systems along with its advantages and disadvantages and weapon detection datasets and their limitations. In Section 3, we provide the steps taken while preparing our own weapon detection and classification dataset and the details of the final dataset. In Section 6 we describe the implementation framework used in this work and evaluation results of all the models used in this experimental analysis. In Section 8 we show the detection results of the final system developed in this work and see how well the system addresses the different challenges in weapon detection systems. Finally, we conclude our work in Section 8.5 and provide some directions for future work related to the domain of weapon detection and classification using deep learning.

II. PREVIOUS WORK

Several weapon detection systems have been implemented in the last decade using various methods and algorithms. In current section will be discussing their assets and liabilities with respect to various aspects which includes the nature of the problem statement, approach and algorithm used to train and deploy the model, the dataset used for training and testing, the accuracy and efficiency of the system in terms of multiple metrics [4]. Used state-of-the-art Faster R-CNN [5] model on Internet Movie Firearm Database (IMFDB) using VGGNet-16 [6] architecture without the use of GPUs. The model was pre-trained on ImageNet [7] dataset and they used Stochastic Gradient Descent (SGD) for fitting the model and updating the weights. They talked about challenges in Firearm detection in terms of occlusion, inter-class variation, and noises in the images of the gun. The model used for detection achieved 93% accuracy with Boosted Tree classification which outperformed KNN with 91.5% accuracy and SVM classification with 92.6% accuracy. The trained model was only capable of detecting a weapon but not classifying a weapon into different weapon category. Talked about challenges such as occlusion and dataset creation, especially for gun detection. They compared two detection algorithms: Sliding window approach and region-based approach [8]. The dataset used contained 3000 images of handheld guns in varying context. Similar to [4], they used Faster R-CNN [5] model based on VGGNet16 [6] architecture which was pre-trained on ImageNet dataset. The model achieved 100% recall and 84.21% precision. Also, the system uses SVM classifier to fire the alarm based on the input from the detection algorithm which gave five successive True Positives within a 0.2s interval in 27 out of 30 different scenes. The dataset used for training purpose was not benchmarked and hence the comparative study of used models to the state-of-the-art models is not possible. Akcay et al. [9] compared different frameworks including Sliding Window based CNN (SW-CNN) [10], R-CNN [3], Fast R-CNN [11], R-FCN [12] and YOLOv2 [13] using architectures: AlexNet [14], VGGNet-16 [6] and ResNet [15] with different pipelining, for detection of guns, laptops and other items in X-ray images. The results they obtained while experimenting clearly shows the superiority of Faster R-CNN based ResNets in object detection with 98.6% accuracy. CNN feature extraction proved to be superior then Bag of VisualWords (BoVW) with entire network achieving 99.6% True positive, 99.4 Accuracy and 0.011 False positives. The trained model was only capable of detecting weapons in X-ray images that have only one channel (greyscale) and hence cannot be used for detection of weapons in normal images.

For [16] compared VGGNet-16 [6] and GoogleNetOverFeat [17] for detecting handguns in an input image or video file. The dataset used for training and evaluation purpose was IMFDB. OverFeat was used with three different sets of hyper-parameters. OverFeat with 30% confidence threshold and 0.0003 learning rate outperform other mentioned models and achieved an excellent training accuracy of 93% and test accuracy of 89%. The experimentation showed how hyper-parameter tuning can improve the accuracy and performance of the deep learning models. Despite achieving high accuracy rates, the model lacked the capability classify the detected gun to be introduced in real life system as the speed of the detection was very slow (1.3s per classification). Several weapon detection systems were developed using deep learning techniques, but these weapon detection systems lack some of the required features such as classifying the weapons detected into various categories based on the type of the gun, such as handguns, knives, assault rifles, etc.

III. DATASET PREPARATION

For training a deep learning model for detection of weapon, image dataset is required. Existing image datasets for weapon detection, cannot be used for classification of the weapon into various categories. For such classification, image dataset must have bounding box annotations with proper class labels. These labels are then used by the deep learning model to correctly predict the class of the weapon detected in the image. To overcome the accuracy-related issues, many image conditions were taken into consideration while preparing the dataset. The image conditions considered are: – Illumination conditions: day, night – Blur and Motion – Different viewpoints – Different colors – Different environments: war, city, movies, cartoons – Different conditional situation – firing, holding, – Various types of weapons in same category. (For e.g. in Handguns – automatic, folding, revolver, self-loading and in Assault-rifles – ak47, m4, g36) We have compiled an image dataset with images containing the instances of single or multiple weapons of various weapon categories. Figure 1 shows the working of the data collection and data preparation steps used for compiling the weapon dataset. The steps in the weapon dataset creation are as follows:

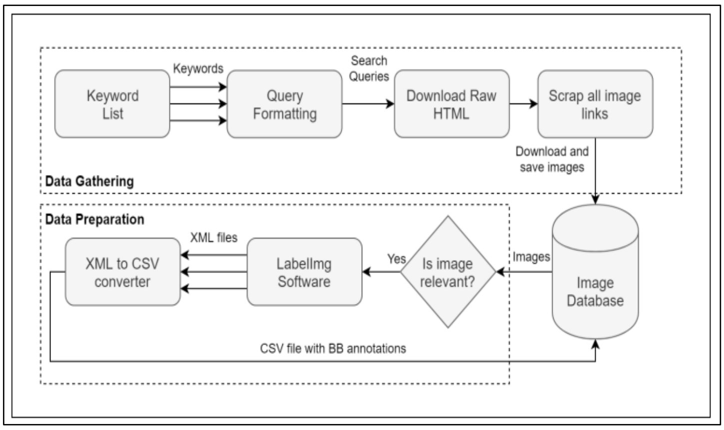

Figure 1: Workflow of weapon detection and classification. dataset creation.
Source: Author, (2024).

IV. DATA COLLECTION USING WEB SCRAPING

We have used a pre-existing module to scrap the images from internet. The working of the script is as follows:

a. We input search keywords in the form of a list and specify the number of images to be downloaded per keyword.
b. A search query is created for 'Google Images' using the keyword specifies in the script and the search query is fired onto the internet. A raw HTML file is returned, which is downloaded on a temporary basis.
c. Image links are extracted from the raw HTML file and the links are used to download the images from the internet. The raw HTML is deleted.
d. These downloaded images are stored in their native shapes using the correct file extension.
e. Steps b to e are repeated for each keyword in the keyword list provided in the script.

V. MANUAL REMOVAL OF IRRELEVANT IMAGES

After scrapping the images, we found that there were a large number of irrelevant images in the dataset. We could have kept them in the dataset as the presence of images, with no instance of objects, does not affect the working or the accuracy of the deep learning models. But to make the dataset compact, we have removed the images with no instance any of the weapon categories. Before removal, the number of images in the dataset was 15,473, which reduced to 8,843 after pruning the irrelevant images from the dataset.

V.1 BOUNDING BOX ANNOTATIONS

For detection of the weapon in the image, deep learning models need to the have the bounding box highlighting the presence of weapon instance in the image. Deep learning models use the region of an image inside the bounding box to learn the features of the object. For annotating the images in our dataset, we have used 'Labelling' [18]. Labelling is image annotation tool written in Python and has an easy graphical interface for bounding box annotations. The output of the annotations is saved as an XML files which are in PASCAL VOC format. The weapon categories for class labels taken into consideration are Assault-rifle, Handgun, Knife, Shotgun, Sniper-rifle.

V.2 FINAL DATASET DESCRIPTION

The final image dataset compiled is now ready to be used for weapon detection and classification task. The image dataset can be used for detecting and classifying five different kinds of weapon categories: Assault rifles, Handguns, Knives, Shotguns and Sniper-rifles. The image dataset contains around 8.8k image files with 11.5k instances of weapons. Table 1 describes the actual number of images and weapon instances present in our dataset.

Figure 2: Some examples of the weapon detection and classification dataset.
Source: Author, (2024)

Table 1: Image Dataset description for weapon detection and classification.

Category	**No. Of Images**	**No. Of Instances**
Assault-rifle	1649	1826
Handgun	2866	3591
Knife	1640	2201
Shortgun	976	1610
Sniper-rifle	1712	2224
Total	8843	11452

Source: Author, (2024).

The Image dataset is uploaded on Google drive and is made open-source. The annotations are also uploaded along with the images. The dataset and the annotations are made freely available to anyone who wants to work in the field of weapon detection and classification.

VI. RESULTS AND DISCUSSIONS

We have used four different deep learning models for weapon detection and classification task trained and evaluated on our own dataset. The models used for experimental analysis of deep learning techniques are as follows: (1) Faster RCNN with Inceptionv2 (2) Faster RCNN with Resnet50 (3) SSD with Inceptionv2 (4) SSD with Resnet50 the image dataset was divided in two parts - 80% was used for training and 20% was used for evaluation. The training was performed for exactly 300k steps for each model inorder to compare them on equal grounds. Data Augmentation was used to boost the performance of each deep learning model used in this study. The technique of data augmentation was Horizontal Flip with Random 50% probability. Learning rate was kept constant with initial value at 0.0002 and the IoU threshold was set to 0.6. Training was performed on GPU - NVIDIA Quadro K620 with 2 GB of memory. The batch size was kept constant at 1 due to the low memory of the GPU.

VI.1 COMPARISON OF DEEP LEARNING MODELS

There are three types of losses calculated while training deep learning models:

1. Localization loss is the loss while localizing the object in the given images. The localization loss signifies how good the model detects the spatial location of the relevant objects in the image. Figure 3 shows the localization loss while training all four models used in this work. The graph clearly shows the superiority of Faster RCNN models over SSD models for localization task. The RPN used in Faster RCNN are very powerful for detection the spatial location of the object in the image.

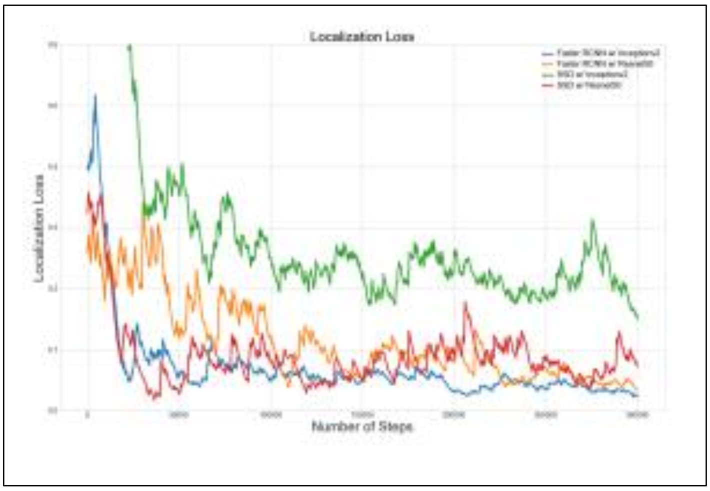

Figure 3: Localization Loss of the models while training.
Source: Author, (2024).

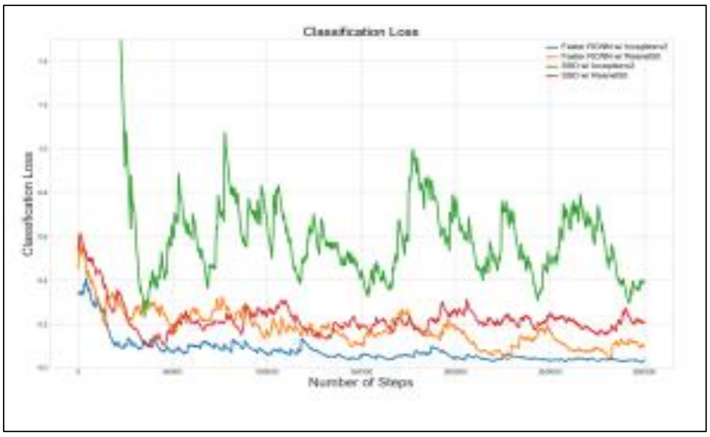

Figure 4: Classification Loss of the models while training.
Source: Author, (2024).

Figure 5: Total Loss of the models while training.
Source: Author, (2024).

2. Classification loss is the loss while classifying the detected object into correct category. The classification loss signifies how accurately the model classifies the object detected in the image. Figure 3 exhibits the classification loss while training all four models used in this work. Faster RCNN w/Inceptionv2 outperforms all other models in the classification task while SSD w/Inceptionv2 shows poor performance.

3. Total loss is the combined loss for localization and classification of objects in the input images. The total loss gives the overall performance of the model for detection and classification task. Figure 3 presents the total loss while training all four models used in this work. We see significant performance difference between Faster RCNN and SSD models. The loss for Faster RCNN models model stabilizes after 250k steps and remains constant afterwards but SSD models fail to do so.

For measuring how good the model works, we need to evaluate the model. The models were evaluated using the evaluation metrics of mAP used by COCO competition. The models were evaluated at 200k and 300k steps to compare the performance of the models. Table 2 shows the COCO metrics obtained by all four models used in this experimental analysis.

The results clearly states substantial accuracy difference between Faster RCNN and SSD models. Faster RCNN w/Inceptionv2 achieves highest mean Average Precision with a score of 0.662, followed by Faster RCNN w/Inceptionv2 with a mAP score of 0.573. SSD w/Inceptionv2 has worst performance of all with overall mAP of 0.238. Figure 6 shows the comparison between the mAP scores of all the four models used in this experimental analysis.

Table 2: Evaluation metrics obtained by all four models after training for 300k steps.

Evaluation Metrices	Faster RCNN w/Inception V2	Faster RCNN w/Resnet50	SSD w/Inception V2	SSD w/Resnet50
mAP	0.662	0.573	0.238	0.333
mAP@.50 IOU	0.837	0.720	0.454	0.484
mAP@.75 IOU	0.742	0.652	0.221	0.379
mAP (small)	0.151	0.000	0.000	0.101
mAP (medium)	0.293	0.189	0.012	0.093
mAP (large)	0.689	0.601	0.266	0.352
AR@1	0.605	0.557	0.286	0.465
AR@10	0.780	0.640	0.405	0.632
AR@100	0.790	0.000	0.444	0.656
AR@1 (small)	0.420	0.000	0.000	0.360
AR@10 (medium)	0.564	0.191	0.045	0.292
AR@100 (large)	0.806	0.671	0.485	0.680

Source: Author, (2024).

VI.2 DISCUSSIONS

Our inferences from the training and evaluation results are as follows:

– Faster RCNN models are better than SSD at detecting weapons in small regions.

– Training time for Faster RCNN and SSD models vary significantly even with the use of same backbone DCNNs.

– SSD models shows significant speed up on Faster RCNN models while detection.

– Faster RCNN w/Inceptionv2 outperforms every other model in our study, with high margin for weapon detection and classification task.

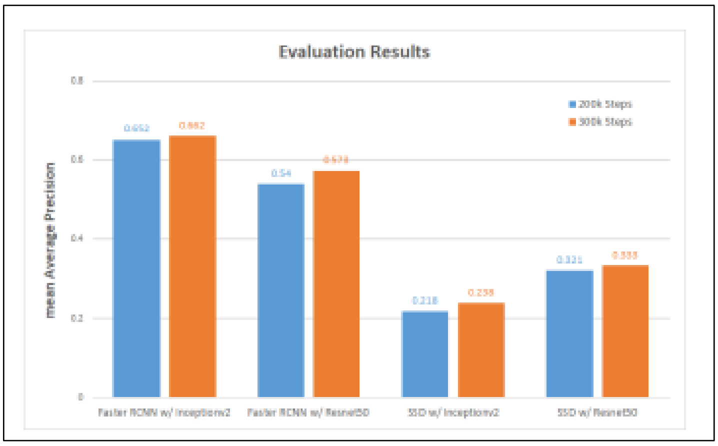

Figure 6: Comparison of mean Average Precision secured by four models.
Source: Author, (2024).

VI.3 DETECTION RESULTS

In this section, we have shown the results of detection and classification of the weapons using some real-life images. The Faster RCNN w/Inceptionv2 was used for the detection as it showed highest accuracy in the model comparison phase. The model was trained for 1 million steps with same configuration. The images chosen for the detection results contain various environments and various contextual situations. The detection results show how well our model tackles different challenges while detecting and classifying weapons.

Positive and Negative detection:

Some of the positive and negative results are shown in this section. Figure 7 shows the correct classification of gun with perfect bounding box prediction. Figure 8 shows that our trained model fails at certain situations to detect the weapon

X-ray images:

The trained model detects the weapon using the features of the weapon and not the color. Hence our model is able to detect weapons and classify them even in the X-ray images. This provides an extra use-case for our model. Figure 9 shows the weapon detection in the X-ray images.

Occlusion:

The major issue with weapon detection is occlusion of the weapon with the holder's hand or body. This issue is easily handled by our model. The model is able to detect the weapon in the image, even with as low as 30-40 percent visibility of the weapon in the image or video frame.

Illumination

The model is also capable of detecting weapon in lowlight or at night. The illumination is not a problem for the model until and unless the border edges are somewhat visible in the image. Figure 11 shows the detection of weapon in low-light situations.

View-point and Angles

The weapon should be detected from any angle and view-point if it appears in the image. Our model is capable of detecting gun in any view-point. This is only possible because the images of weapons with different angles and view-points are present in the image dataset we prepared in the earlier stage. Figure 12 shows the weapon detection in different angles and viewpoints.

Figure 7: Positive examples of the weapon detection.
Source: Author, (2024).

Figure 8: Some Positive examples of the weapon detection.
Source: Authors, (2024).

Figure 9: Weapon detection in X-ray images.
Source: Author, (2024).

Figure 10: Solving the issue of occlusion and scale.
Source: Author, (2024).

Figure 11: Low illumination condition.
Source: Author, (2024).

Figure 12: Handling different viewpoints.
Source: Author, (2024).

VII. CONCLUSIONS

The proposed work was mainly focused on creation of a new image dataset for weapon detection and classification and implementation of weapon detection system with ability to classify the weapon based on the weapon type, using Deep learning techniques. We have created an image dataset with more than 11.5k instances of weapon. The weapons in the images have bounding box annotations with five different categories of weapons. The dataset was then used for training the state-of-the-art deep learning models used for salient object detection problems. We have used four different models for training and evaluated them on the test set using different evaluation metrics. Our experimental results shows that Faster R-CNN w/Inceptionv2 outperformed all the other models used in this work and achieved mean Average Precision of 0.69 and mean Average Recall of 0.62. The detection speed of Faster-RCNN is not quick, but this can be fastened by using high performance GPUs. Even though SSD models are incredibly fast at detection task, lack the real-life use as the accuracy of the detection boxes and classification is not acceptable. Due to the lack of computing power, we were not able to perform comparison of modern DCNNs such as Inception-ResNets or SE-ResNets.

The deep learning models trained on our image dataset overcomes many challenges in the field of weapon detection such as illumination, blur, motion, and occlusion. The models are also capable of detecting weapons in X-ray images; hence the system can be used for detecting concealed weapons in baggage at airports, railway stations, malls, etc. SSD w/ Resnet50 can be used for real-time weapon detection in CCTV cameras as the detection speed of this model is high. The research work done in this thesis can be extended in many different ways:

1. More complex models such as Faster R-CNN w/ Inception-Resnetv2 can be trained for more accurate detection and classification of weapons.
2. The dataset can be used to train small and light deep learning models, which can be used for detecting weapons on mobile or embedded systems
3. Our image dataset can be extended by adding new categories of weapons for detection of new types of weapons such as missiles, grenades, etc.

VIII. AUTHOR'S CONTRIBUTION

Conceptualization: Sunil B. Mane.
Methodology: Sunil B. Mane.
Investigation: Sunil B. Mane.
Discussion of results: Sunil B. Mane.
Writing – Original Draft: Sunil B. Mane.
Writing – Review and Editing: Sunil B. Mane.
Resources, Supervision: Sunil B. Mane.
Approval of the final text: Sunil B. Mane.

IX. REFERENCES

[1] "Estimated number of civilian guns per capita by country," Wikipedia, 29-Apr-2019. [Online]. Available: https://en.wikipedia.org/wiki/Estimated number of civilian guns per capita by country. [Accessed: 03-May-2019].

[2] B. Marr, "How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read," Forbes, 11-Mar-2019. [Online]. Available: https://www.forbes.com/sites/bernardmarr/2018/05/21/howmuch-data-do-we-create-every-day-the-mind-blowingstats-everyone-should-read/#2ecf668960ba. [Accessed: 03-May-2019].

[3] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014.

[4] G. K. Verma and A. Dhillon, "A Handheld Gun Detection using Faster R-CNN Deep Learning," Proceedings of the 7th International Conference on Computer and Communication Technology - ICCCT-2017, 2017.

[5] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, 2017.

[6] Simonyan K., Zisserman A., "Very deep convolutional networks for large scale image recognition," ICLR, 2015.

[7] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009.

[8] R. Olmos, S. Tabik, and F. Herrera, "Automatic handgun detection alarm in videos using deep learning," Neurocomputing, vol. 275, pp. 66–72, 2018.

[9] S. Akcay, M. E. Kundegorski, C. G. Willcocks, and T. P. Breckon, "Using Deep Convolutional Neural Network Architectures for Object Classification and Detection Within X-Ray Baggage Security Imagery," IEEE Transactions on Information Forensics and Security, vol. 13, no. 9, pp. 2203–2215, 2018.

[10] H. Nakahara, H. Yonekawa and S. Sato, "An object detector based on multiscale sliding window search using a fully pipelined binarized CNN on an FPGA," 2017 International Conference on Field Programmable Technology (ICFPT), Melbourne, VIC, 2017, pp. 168-175.

[11] R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), 2015.

[12] Dai, J.; Li, Y.; He, K. & Sun, J. Lee, D. D.; Sugiyama, M.; Luxburg, U. V.; Guyon, I. & Garnett, R. (Eds.), "R-FCN: Object Detection via Region-based Fully Convolutional Networks Advances", Neural Information Processing Systems (NIPS) 29, 2016, 379-387.

[13] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[14] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[16] J. Lai, "Developing a Real-Time Gun Detection Classifier," 2017.

[17] Sermanet, Pierre, David Eigen, Xiang Zhang, MichaA˜ l Mathieu, Rob Fergus and Yann LeCun. "OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks," CoRR, 2014, abs/1312.6229.

[18] Tzutalin "LabelImg". Git code (2015). https://github.com/tzutalin/labelImg