2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2) | 978-1-6654-1285-8/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICoDT252288.2021.9441523

# Application of Deep Learning for Weapons Detection in Surveillance Videos

Tufail Sajjad Shah Hashmi, Nazeef Ul Haq, Muhammad Moazam Fraz, Muhammad Shahzad
*School of Electrical Engineering and Computer Science,*
*National University of Sciences and Technology (NUST),*
Islamabad, Pakistan
{thashmi.mscs18seecs, nhaq.mscs18seecs, moazam.fraz, muhammad.shehzad}@seecs.edu.pk

***Abstract*—Weapon detection is a very serious and intense issue as far as the security and safety of the public in general, no doubt it's a hard and difficult task furthermore, its troublesome when you need to do it automatically or with some of the AI model. Different object detection models are available but in case of weapons detection it is difficult to detect the weapons of distinctive size and shapes along with the different colors of the background. Currently, a great deal of Convolutional Neural Network (CNN) based deep learning approaches are proposed for the recognition and classification in real-time. In this paper, we have done the comparative analysis of the two versions which is a state of the art model called YOLOV3 and YOLOV4 for weapons detection. For training purpose, we create weapons dataset and the images are collected from google images along with a portion of different assets. We annotate the images one by one manually in different formats in light of fact that YOLO needs annotation file in text format and some other models need annotation file in XML format. We trained both the versions on a large data set of weapons and afterward tested their results for comparative analysis. We explained in the paper that YOLOV4 performs obviously superior to the YOLOV3 in terms of processing time and sensitivity yet we can compare these two in precision metric. The implementation details and trained models are made public at this link:https://cutt.ly/5kBEPhM.**

**Index Terms-**

***Index Terms*—Weapons detection, Object detection, Visual surveillance, YOLO family CNNs**

## I. Introduction

Gun violence is a very serious issue for human rights and freedom in the world. The prime human right is frightened by weapon-related violence. The existences of individuals influenced by weapons on regular routine in the entire world. According to the statistics, the death proportion of individuals because of gun brutality is around 500 every day. More than 44% of assassination involves gun violence worldwide. In the middle of 2012 and 2016 more than 1.4 million deaths were recorded due to firearms violence [1], [28], [32].

Weapons are generally utilized for viciousness than self-defence [2]. This problem requires modern tools and techniques to survive the results of gun violence like the vast majority of the nations utilize the video surveillance system to monitor the people for terrorism and crime [3]. With the deep learning utilization and in specific convolutional neural network (CNN), we can estimate the objects in images by classification and localization called object detection [4]. There are numerous utilization and application of object detection, [34], [26], [27], [28] like face detection [5], [27], [29] pedestrian detection [6], [30], [31] skeleton detection [7]. There are many architectures and algorithms are present like YOLO and its versions [8], [9], [10], [11]. R-CNN and its versions [12], [13], [14], [15]. Girshick et al introduced CNN which is region-based called R-CNN, in which CNN combines with the region-proposals algorithm. The selective search approach is used to extract 2000 regions, then the classification will apply only on the extracted regions rather than the entire image. After that, there are some improvements made by the authors to overcome the limitations of the existing algorithm by giving the entire image as an input to the network instead of region proposals, and then convolutional features map gives the identification of region proposals. Then Faster R-CNN is proposed by Shaoqing Ren et al. by changing the selective search technique to the object detection algorithm [33].

In this paper, we examine the detection of weapons and compare the performance of different models. We inspect the current approaches of these two versions called YOLOV4 and YOLOV3. We have additionally generated a new dataset comprise of 7800 images of weapons that is also a part of this research and exploration. All the images are collected from internet and principal assets are google images, CCTV videos, and movies. After successful collection of these images we manually annotate these images with the help of LabelImg[33] tool. The ground truth of every weapon is available along with the image in the given dataset. The created dataset likewise contain the various formats like YOLO, Faster R-CNN, Pascal VOC along with the rotated bounding boxes.

The important benefactions of this research study are given below.

- Present the comparison of state of the art algorithm YOLO in term of weapon detection. We also discuss the sensitivity of objects in terms of two different kinds of classes of weapons i-e gun and pistol.
- We also developed a new weapons dataset consist of 7800 images and most of the images contain more than one weapons. Angle oriented bounding boxes are also available in this dataset.
- The dataset is provided in different formats like YOLO, Faster-RCNN, Pascal-VOC. This will help the researchers

978-1-6654-1285-8/20/$31.00 ©2021 IEEE

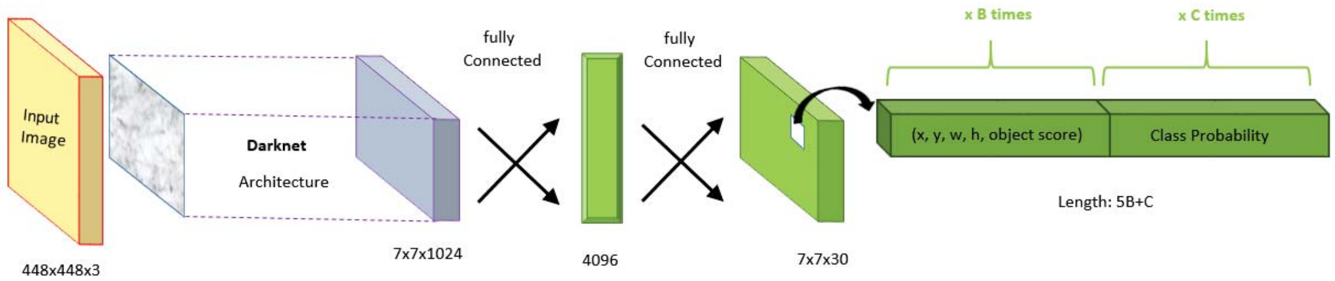


Fig. 1: YOLO Network Architecture to Represent the Implementation of Convolutional Layers

to conduct the fruitful research in future.

In this paper, we measure the performance of two versions of YOLO in terms of weapon detection. We observe the current approaches of the two versions, namely YOLOV4 and YOLOV3.
The rest of the paper is arranged as follows: Section-II explains the associated work about weapon detection. Section-III deliver the comparison between YOLOV4 and YOLOV3 for weapon detection. Section-IV conclude the research and discuss the results.

## II. Related Work

This section of the paper contains the knowledge about the weapons detection problem addressed by Convolutional Neural Networks.
Elmir et al.[16] introduced a model that works in different steps. The initial step is image acquisition and another one is motion detection, the last one is gun detection. They performed experiments on three different models, the first one is the CNN-based model, the second one is the Fast R-CNN-based model, and the third one is the Mobile-Net CNN-based model. For evaluation of these three models, the amount of databases used for the learning phase is 2. For the classification task, the training data set contains 9261 images with 102 classes in which handgun class at 200. For the region task, the training data set contains 3000 images. Test data set for detection and classification contains 608 images with 304 are handguns. They used a sample of 420 images for learning models from a database for the region proposal technique. They tested the first model with 608 images, the second with 200, and the third with 420 images. If we say about the results, they got 55% accuracy on the first model, 80% on the second model, and 90% on the third model.
In current research papers, Region-based Convolutional Neural Networks (RCNN) and Fast Region-based Convolutional Neural Networks (FRCNN) methods are implemented for automated detection of weapons in surveillance videos [17]. The object detection algorithm is also used for the detection of knives in a given videos[18].

### A. Explanation of YOLO

YOLOV4[11]. is the result of improvements in previous versions YOLOV1[8], YOLOV2[9] and YOLOV3[10].

*1) Explanation of the algorithm YOLOV1:* YOLO consist of 24 convolutional layers with 2 fully connected layers. Some layers used size 11 of convolutions to decrease the depth of feature map. Another variant utilize 9 convolutional layers called Fast YOLO which affect the accuracy most[23].
YOLO splits the given image into the grid of S x S. A grid cell can correlate with one object and predict the bounding boxes which is in fixed numbers. Every box also assigned a confidence score so that the followings are the details are also predicted for all the bounding boxes (X, Y, W, H, Confidence Score). For the estimation of object classification, the grid cell is concerned with the number of class probabilities from the classes of the model denoted by C. The main theme behind the YOLOV1 to create a single CNN network for the prediction.[24]

$$S \times S \times (B \times 5 + C)$$

S X S denotes the amount of grid cells contain by the given image.
B denotes the bounding boxes contained by each grid cell.
C denotes the amount of classes for training.
When you get the knowledge about the prediction and how they are encoded the remaining part is easy to understand. The structure of the YOLO is similar to the CNN, which contains the convolutional layers and layers of max-pooling along with 2 fully connected layers in the end.[25]

*2) Improved version YOLOV2:* Another challenger for YOLO is introduced, named as SSD[20] (Single Shot Multi-Box Detector). SSD defeat YOLO for real-time object detection in term of accuracy. Hence, an improved version of YOLO is introduced with a lot of improvements to enhance processing time and accuracy.
The first improvement of YOLOV2 over YOLOV1 was the technique of Batch Normalization (BN)[21], presented in 2015. This technique is used for scaling and adjusting the activation to normalize the input layers. It is noticed that mAP is increased by 2% when applying BN to all of the convolutional layers in YOLO.
The second improvement in this version is to apply a High-

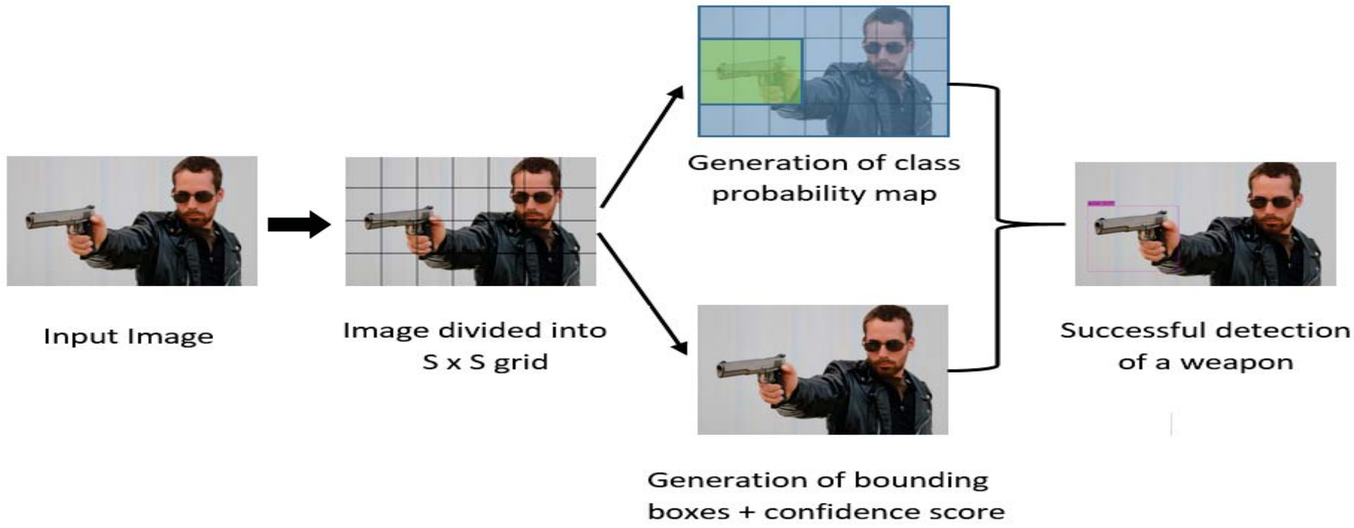

Input Image
Image divided into S x S grid
Generation of class probability map
Generation of bounding boxes + confidence score
Successful detection of a weapon

Fig. 2: YOLO Architecture to Represent the Actions Performed on the Image

Resolution Classifier. Training the model on 224x224 images, for fine-tuning this version also apply 448x448 images in the classification network. So mAP is increased by 4%.

The third improvement in this version is to apply convolution with anchor boxes. Fully connected layers are removed by this version and for the prediction of bounding boxes anchor boxes are used. To enhance the output resolution one pooling layer is excluded. The implementation of anchor boxes reduces the mAP by 0.3%, with anchor boxes the recall increase by 7%.

Forth improvement in this version is to apply the dimension cluster. K-means clustering is used to find the correct anchor boxes. For clustering they used IOU score instead of Euclidean distance because smaller boxes generate less error than larger boxes.

The fifth improvement in this version is the direct location prediction. The previous version does not have the restriction on a prediction of the location which leads to the model imbalance at primary iterations.

The sixth improvement in this version is the fine-grained features. To increase the ability to detect small objects, the pass-through layer technique is used by YOLO, in which the low-resolution features are concatenated with the high-resolution features. The mAP increased by 1% by using this technique which looks like identity mapping used in ResNet[22].

*3) Improved version YOLOV3:* The first improvement of this version is the implementation of multi-label classification. To find out the probabilities from the scores soft-max function is used by the preceding versions. Binary cross-entropy loss is used for the classification loss rather than using mean square error like in the preceding versions.

The second improvement of this version is the prediction of different bounding boxes. This version gives the one anchor box for every ground truth object. It neglects the other overlapping bounding box that occurs in the result of defined threshold.

The third improvement of this version is the prediction across scales with the help of feature pyramid networks. This version used the 3 different scales for the prediction of boxes and then extract the features from them.

Forth improvement of this version is the feature extractor called Darknet-53. The number of CNN layers is 53 that follows the skip connection network. It also implements 3x3 and 1x1 convolutional layers and improved the accuracy with smaller floating-point operations.

There are 4 coordinates predicts by the network for each bounding box, tx, ty, tw, th. If the top left corner of the image is denoted by (Cx,Cy), height and width of the bounding box is denoted by (Pw,Ph), then we express the predictions as:

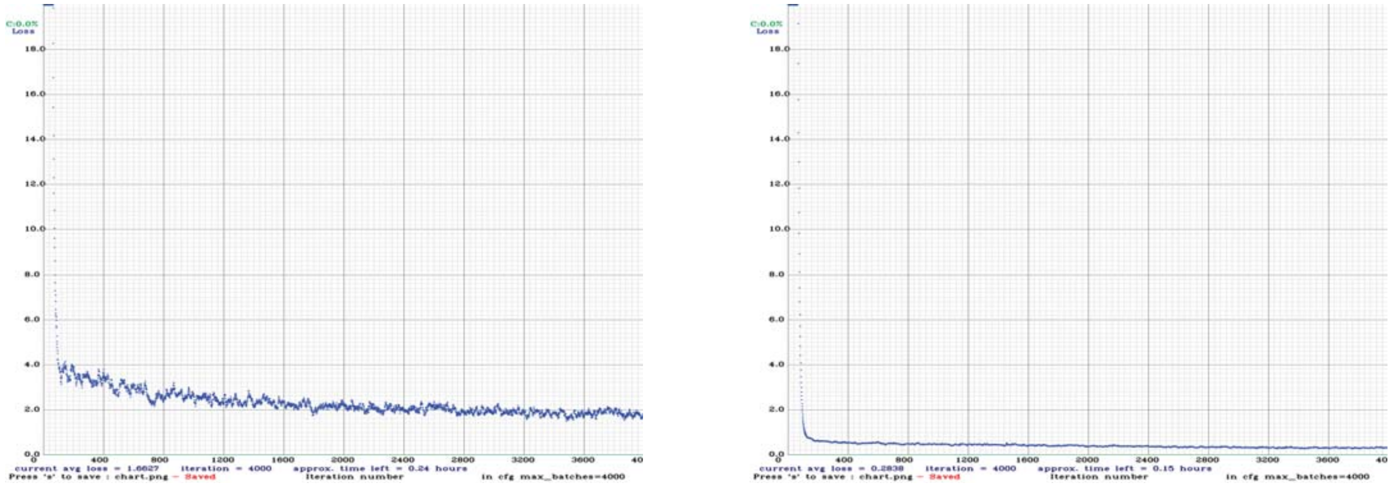
$$b_x = \sigma(t_x) + C_x$$

$$b_y = \sigma(t_y) + C_y$$

$$b_w = P_w e^{t_w}$$

$$b_h = P_h e^{t_h}$$

*4) Improved version YOLOV4:* This version used the cross-stage partial connections (CSP), the latest backbone for CNN to increase the learning ability. Cross mini-Batch Normalization (CmBN) is used to divide the batches into mini-batches. Self Adversarial Training (SAT) presents another data augmentation approach that works in both backward forward stages. The non-monotonic neural activation function is used which is self regularized known as mish-Activation. For augmentation, a technique is used that combines the 4 training images rather than a single image called mosaic data augmentation. A finer regularization technique is used for CNN called drop block regularization. To achieve accuracy and speed for bounding box regression problem CIoUs loss is used.



(a) Validation Accuracy Graph of YOLOV4

(b) Validation Accuracy Graph of YOLOV3

Fig. 3: Validation Accuracy Graphs for YOLOV4 and YOLOV3

## III. FACT-FINDING COMPARISON BETWEEN YOLOV4 AND YOLOV3

In this portion, we will discuss about the training data and testing data and also consider the hardware and software used for experiments. Evaluation of the models is provided for a better understanding of the comparison. The detection results of YOLOV4 and YOLOV3 are shown in Fig. 4 respectively.

TABLE I: STATISTICS ABOUT THE WEAPONS DATASET

| Dataset | Total Images | Gun | Pistol | Total Instances |
|---|---|---|---|---|
| Training | 6240 | 4384 | 2963 | 7347 |
| Testing | 1561 | 1128 | 776 | 1904 |

### A. *Explanation of the Dataset*

To conduct the experiment for the comparison, we create a weapons dataset split into a training set and test set. There are 6240 images in the training set and 7347 instances of labeled weapons in which 4384 instances belong to the gun class and 2963 belong to the pistol class. The test set contains a total of 1561 images and 1904 instances of labeled weapons in which 1128 instances belong to the gun class and 776 belong to the pistol class. This dataset was created after a collection of google images and a lot of other resources like movies and CCTV images. We collect the images of weapons with different color, background, size, and shape to check the model inclusively. After doing the tiring full process of collecting images we manually annotate each image with the help of an annotation tool known as LabelImg[35]. Annotation files are available in different formats.

For YOLO we need a text file that contains the parameters (x, y, w, h, class) "x" and "y" denotes the center points, "w" denotes the width of the box while "h" denotes the height of the box and "class" represents that the object belongs to which class. Another annotation format is Pascal-VOC in the form of the XML file that contains the four points (xmin,ymin,xmax,ymax). Our dataset is also available in the form of oriented bounding boxes which will be very helpful for the study of orientation aware weapon detection. There are two types of classes included in our dataset first class is known as "Gun" and the second is "Pistol". The division of the dataset into the training data and testing data is in the ratio of 80:20.

TABLE II: PERFORMANCE COMPARISON OF YOLO WITH DIFFERENT DATASETS

| Detector | mAP Value | Used Data-set |
|---|---|---|
| YOLOV4 | 84.85 | Weapons Data-set |
| YOLOV3 | 77.30 | Weapons Data-set |
| YOLOV3 | 57.80 | COCO Data-set |
| YOLOV3 | 25.8 | BDD100K Data-set |
| YOLOV4 | 50.1 | BDD100K Data-set |

### B. *Explanation of Software and Hardware tools*

Now we discuss the configuration, in YOLOV4 we used the default configuration, the optimizer used with the momentum of 0.9, the height and width both are set to 512. The learning rate is set to 0.001, and decay is set to 0.0005. The size of the batch is set to 64, the amount of max batches are equal to 4000. The threshold for overlapping the anchor box with the ground truth is set to be 0.7 in this case, if the value of IOU is greater than 0.7 then the box is selected otherwise it will be neglected. In YOLOV3 we also use the default configuration momentum and the learning rate is also the same as in YOLOV4, but height and width are set to 416 and all the other configurations are the same as in YOLOV4.

Configuration of the computer is given below:

- CPU: Intel Core i7-9700
- Graphics Card: Nvidia Ge Force 1080 TI

- RAM: 62 GB RAM
- Operating System: Windows 10

### C. Performance Evaluation

For the comparative analysis of the two versions, we utilize these parameters (PRECISION, RECALL, F1 SCORE, QUALITY, mAP) and description of these parameters are given below:

$$PRECISION = \frac{TP}{TP + FP}$$

$$RECALL = \frac{TP}{TP + FN}$$

$$F1 \quad SCORE = 2*Precision*Recall(Precision+Recall)$$

$$QUALITY = \frac{TP}{TP + FP + FN}$$

where TP (TRUE POSITIVES) shows the number of weapons detected by the model. FP (FALSE POSITIVES) shows the amount of non-weapon objects that are detected falsely as a weapon. FN (FALSE NEGATIVES) shows the amount of weapons that the model was unable to recognize them as a weapon.

TABLE III: Evaluation Metrics of YOLOV4 and YOLOV3 on Test Data

| Measure | YOLOV4 | YOLOV3 |
|---|---|---|
| TP (TRUE POSITIVES) | 1491 | 1343 |
| FP (FALSE POSITIVES) | 256 | 257 |
| FN (FALSE NEGATIVES) | 413 | 561 |
| PRECISION | 85% | 84% |
| RECALL | 78% | 71% |
| F1 SCORE | 82% | 77% |
| QUALITY | 69% | 62% |
| mAP | 84.85% | 77.30% |

### D. Comparison between YOLOV4 and YOLOV3

In this section, we will compare both the versions on the basis of given metrics. Table II shows that both versions have a high value of precision and this high value of precision specifies, when the model detects the object as a weapon then the chance is very high that the object is a weapon. So, the capacity of the model to detect the true weapon is very high. The probability to detect the non-weapon object as a weapon is very small. You can see the clear difference in the evaluation metrics. This evaluation shows that both versions have a high precision rate(85% for V4 and 84% for V3). But when inspecting the recall, we noticed that YOLOV4 recall is greater than YOLOV3 (78% for V4 vs 71% for V3). The recall shows the capacity of the model to detect all the instances of the weapon in the given image. If the value of False Negatives is high it's possible that the model misses some of the instances. F1 Score gives the general idea about the robustness of the model so there is also a difference between the F1 Scores of two versions. You can also see the big difference between mAP of both the versions.

## IV. Conclusion

In this research paper, Comparative analysis have been made for the two versions of the state of the art object detection algorithm known as YOLOV4 and YOLOV3. We have done a fact-finding comparative analysis for a weapons detection task. We take the beginning from the outline of both the versions, take a look at the architecture and improvements of the preceding versions. From that point onward, we made a comparative analysis with the assistance of an independent self-made weapons dataset. Dataset was divided into the training set and testing set, both the versions trained on that dataset and furthermore measure the performance on a given dataset. The performance is estimated on the basis of given parameters e.g Precision, Recall, F1 Score, Quality, mAP, and so on. We have demonstrated that YOLOV4 performance is obviously superior to YOLOV3 and highlight the things behind the improvement. This comparison gives the researchers a super arrangement to see things profoundly and give the information that how the little changes give better outcomes. For future work we will build the measure to increase the images in our dataset and furthermore increment the measure of classes to extend the detection of weapons.

## References

[1] https://www.amnesty.org/en/what-we-do/arms-control/gun-violence/

[2] https://www.hsph.harvard.edu/hicrc/firearms-research/gun-threats-and-self-defense-gun-use-2/

[3] https://www.ifsecglobal.com/video-surveillance/role-cctv-cameras-public-privacy-protection/

[4] Z. Zhao, P. Zheng, S. Xu and X. Wu, ”Object Detection With Deep Learning: A Review,” in IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 11, pp. 3212-3232, Nov. 2019, doi: 10.1109/TNNLS.2018.2876865.

[5] Rein-Lien Hsu, M. Abdel-Mottaleb and A. K. Jain, ”Face detection in color images,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pp. 696-706, May 2002, doi: 10.1109/34.1000242.

[6] M. Szarvas, A. Yoshizawa, M. Yamamoto and J. Ogata, ”Pedestrian detection with convolutional neural networks,” IEEE Proceedings. Intelligent Vehicles Symposium, 2005., Las Vegas, NV, USA, 2005, pp. 224-229, doi: 10.1109/IVS.2005.1505106.

[7] X. Bai, Xinggang Wang, L. J. Latecki, W. Liu and Z. Tu, ”Active skeleton for non-rigid object detection,” 2009 IEEE 12th International Conference on Computer Vision, Kyoto, 2009, pp. 575-582, doi: 10.1109/ICCV.2009.5459188.

[8] J. Redmon, S. K. Divvala, R. B. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016, pp. 779–788, 2016.

[9] J. Redmon and A. Farhadi, “YOLO9000: better, faster, stronger,” in 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017, pp. 6517–6525, 2017.

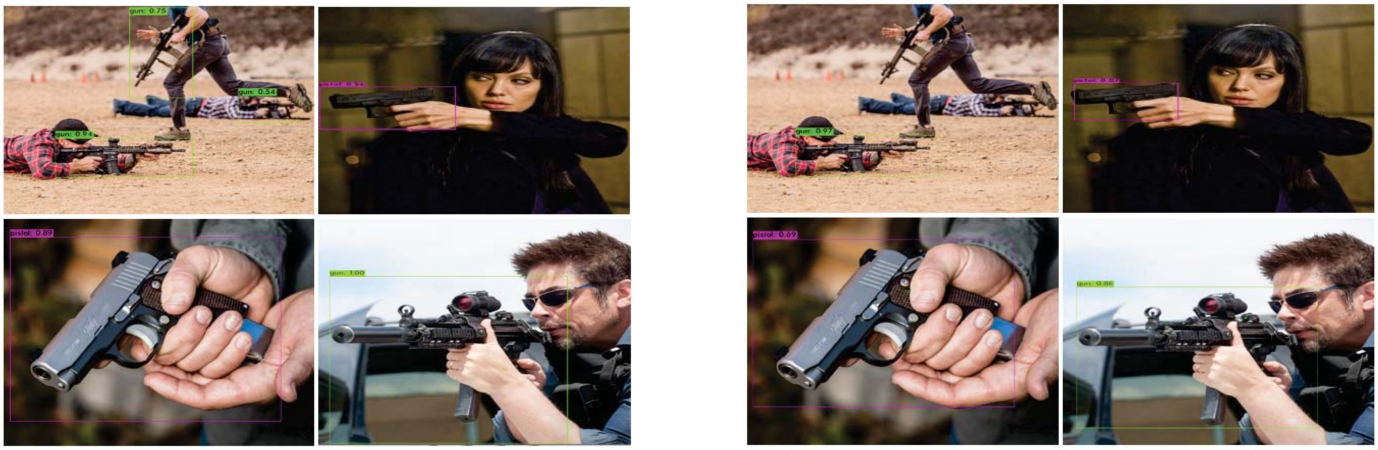
[10] J. Redmon and A. Farhadi, “Yolov3: An incremental improvement,” CoRR, vol. abs/1804.02767, 2018.

[11] Alexey Bochkovskiy et al., YOLOv4: Optimal Speed and Accuracy of Object Detection, Apr 2020.

[12] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 580–587, 2014.

[13] R. Girshick, “Fast R-CNN,” in Proceedings of the IEEE International Conference on Computer Vision, 2015.

[14] R. Girshick, “Fast R-CNN,” in Proceedings of the IEEE International Conference on Computer Vision, 2015.



(a) Weapons detection using YOLOV4

(b) Weapons detection using YOLOV3

Fig. 4: Comparison between YOLOV4 and YOLOV3

[15] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards RealTime Object Detection with," IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 2017.

[16] Elmir, Youssef, Sid Ahmed Laouar, and Larbi Hamdaoui. "Deep Learning for Automatic Detection of Handguns in Video Sequences." JERI. 2019.

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

[18] 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.

[19] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," International Conference on Learning Representations (ICRL), 2015.

[20] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in Lecture Notes in Computer Science (including subseries Lecture Notes in ArtificialIntelligence and Lecture Notes in Bioinformatics), 2016.

[21] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," CoRR, vol. abs/1502.03167, 2015.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Arxiv.Org, 2015.

[23] M. Haroon, M. Shahzad, M.M. Fraz , "Multi-sized Object Detection Using Spaceborne Optical Imagery", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (JSTARS), Vol. 0 , No. 0, PP. , Jun, 2020.

[24] I Khurram, M.M. Fraz , M Shahzad, NM Rajpoot , "Dense-CaptionNet: A Sentence Generation Architecture for Fine-Grained Description of Image Semantics", Cognitive Computing, Vol. 12 , No. 2, PP. 1-31, Mar, 2020.

[25] S. B. Ahmed, S. F. Ali, J. Ahmad, M. Adnan, M. M. Fraz , "On the Frontiers of Pose Invariant Face Recognition: A Review", Artificial Intelligence Review, Vol. 2019 , No. 1, PP. 1-64, Jul, 2019.

[26] R M S Bashir, M Shahzad, M M Fraz , "VR-PROUD: Vehicle Re-identification using PROgressive Unsupervised Deep architecture", Pattern Recognition, Vol. 90 , No. 1, PP. 52-65, Jan, 2019.

[27] N. Pervaiz, M. M. Fraz, M. Shahzad , "Person Re-Identification Using Hybrid Representation Reinforced by Metric Learning", IEEE Access, Vol. 7 , No. 1, Dec, 2018

[28] N. Perwaiz, M.M. Fraz, M. Shahzad , "Smart Visual Surveillance: Proactive Person Re-identification instead of Impulsive Person Search", Proceedings of the 23rd IEEE International Multitopic Conference (INMIC 2020), Nov, 2020, Bahawalpur

[29] N. Pervaiz, M. M. Fraz, M. Shahzad , "Hierarchical Refined Local Associations for Robust Person Re-Identification", Proceedings of the International Conference on Robotics and Automation in Industry (ICRAI), Oct, 2019, Islamabad , Pakistan.

[30] S. Batool, M. Z. Ali ; M. Shahzad and M. M. Fraz , "End to End Person Re-Identification for Automated Visual Surveillance", Proceedings of the International Conference on Image Processing, Applications and Systems (IPAS), Dec, 2018, Sophia Antipolis , France.

[31] W. Anser, M M Fraz, M Shahzad , , "Two Stream Deep CNN-RNN Attentive Pooling Architecture for Video-based Person Re-identification", Proceedings of the 23rd Iberoamerican Congress on Pattern Recognition, Nov, 2018, Madrid , Spain.

[32] R M S Bashir, M Shahzad, M M Fraz , , "DUPL-VR: Deep Unsupervised Progressive Learning for Vehicle Re-Identification", Proceedings of the 13th International Symposium on Visual Computing, Nov, 2018, Las Vegas , United States.

[33] I Khurram, M M Fraz, M Shahzad , , "Detailed Sentence Generation Architecture for Image Semantics Description", Proceedings of the 13th International Symposium on Visual Computing, Nov, 2018, Las Vegas , United States.

[34] SKJ Rizwi, MA Azad, M.M. Fraz , "Spectrum of Advancements and Developments in Multidisciplinary Domains for Generative Adversarial Networks (GANs)", Archives of Computational Methods in Engineering, Vol. 2021 , No. 1, Apr, 2021

[35] "LabelImg,"https://github.com/tzutalin/labelImg.