A Deep Learning Based Approach to Detect Suspicious Weapons

Prashant Varshney, Harsh Tyagi, Nikhil Kr. Lohia, Abhishek Kajla and Palak Girdhar

Computer Science & Engineering Department, Bhagwan Parshuram Institute of Technology, Guru Gobind Singh Indraprastha University, Sector 17, Rohini, New Delhi, 110089, India

Abstract

Over the past few decades, the world has witnessed a lot of terrorist and criminal activities. The Public Surveillance System has gained a lot of importance as a response to counter these activities. Various state governments have started to install cameras in their densely populated and important cities to safeguard their citizens. To cover a complete city under a surveillance network of thousands of cameras, hundreds of security personnels are needed to monitor its video feed in real-time. To make this task cost-effective and feasible, one security personnel is monitoring nearly 6 – 8 cameras which usually leads to failure in detection of threats. One slip of concentration can cause damage to many lives. This research paper determines the optimized, efficient & faster way to detect commonly used weapons like AK47, Hand Revolver, Pistol, Combat Knife, Grenade, etc. in a live video feed.

Keywords 1

Object Detection, Neural Network, Deep Learning, Computer Vision, mAP Score

1. Introduction

Object Detection is a field of Artificial Intelligence associated with Digital Image Processing and Computer Vision, which deals with detecting instances of an object like a car, humans, weapons, etc. possessing similar features with the trained object classes [1]. Object Detection methods are generally categorized into either ML-based approach or DL-based approach depending upon the complexity of object class. For Machine Learning based approaches, it is essential to define features beforehand using methods like Haar Cascade, SIFT, etc. which further uses the support vector machine (SVM) technique for detecting object class. The Deep Learning based approaches uses Artificial Neural Networks to do an end-to-end object detection without defining features specifically.

Recent growth in the field of Artificial Intelligence has contributed a lot to solve major crises all over the world. This research paper mainly focuses on different Object Detection models in Deep Learning like RCNN (Region-based Convolutional Neural Network) [2], SSD (Single Shot Detector) [3], and YOLO (You Only Look Once) [4, 5, 6, 7] using which possession of any suspicious weapons are detected in video surveillance. The primary goal of this paper is to analyze the performance of these models and determine the most efficient and reliable model amongst them for surveillance purposes.

The main inspiration for this research paper came from the Mumbai Chhatrapati Shivaji terminus railway station attack where a couple of terrorists entered the railway station with their automatic assault rifles and started indiscriminate shooting which killed 58

ACI'21: Workshop on Advances in Computational Intelligence at ISIC 2021, February 25–27, 2021, Delhi, India

EMAIL: pv03158@gmail.com (P. Varshney); tyagih1699@gmail.com (H. Tyagi); nikhillohia6128@gmail.com (N. Lohia); abhishekkajla5511@gmail.com (A. Kajla); palakgirdhar@bpitindia.com (P. Girdhar)
ORCID: 0000-0003-0497-6214 (P. Varshney); 0000-0001-5919-811X (H. Tyagi); 0000-0002-7338-7619 (N. Lohia); 0000-0002-4042-6001 (P. Girdhar)

CEUR Workshop Proceedings (CEUR-WS.org)

civilians and caused 100 plus casualties [8]. At that time, if there had been any AI-based Weapon Detection Technology that was monitoring the city, then the Police Control Room would have known about them beforehand. We decided to build such a system that might add an extra layer to the security at public places preventing such threatening activities.

2. Related Work

The detection of threatening weapons in the surveillance system is a challenging task to do. To cover a complete city under a surveillance network of thousands of cameras, hundreds of security personnels are needed to monitor its video feed in real-time. To make this task cost-effective and feasible, one security personnel is monitoring nearly 6 – 8 cameras which leads to failure in detection of threats at their initial stage and results in delayed response causing causalities.

According to the research of Velastin et al. of the Queen Mary University of London, carried in 2006 usually after 20 minutes of video monitoring, operators in many instances fail to notice the presence of threatening objects in a video feed [9]. Researching further analyzed that after 12 minutes of monitoring video feed, an operator is likely to fail to notice up to 45% of suspicious activities and after 22 minutes of monitoring, up to 95% of suspicious activities are failed to notice [10]. Thus the most optimal solution to this problem could be to eliminate humans from the equation as much as possible.

In the year 2001, Paul Viola and Michael Jones proposed the first robust, efficient, and real-time Machine Learning based Object Detection Framework in their paper "Rapid Object Detection using Boosted Cascade of Simple Features" [11,12]. This framework can be trained to discover the variety of objects by taking lots of images that contain the object which we want our classifier to detect (positive images) and the same images but without the object which needs to be detected (negative images), to train the classifier. However, this approach cannot be used to identify the presence of complex objects in different orientations and sizes.

Deep Learning based Object Detection frameworks mainly consist of two types –

1. Region Proposal based framework that includes models like RCNN, FRCNN, and Faster RCNN.
2. Regression-based frameworks that include models like YOLO and SSD.

Region Proposal-based algorithms use a sliding window approach to extract features from the visual data. In the year 2014, Ross Girshick presented RCNN model based on this algorithm, which obtained mAP of 53.3% On the contrary, to the results achieved on the PASCAL VOC dataset, an improvement of 30% was achieved by this model. In this model, the whole image is processed with a Convolution Neural Network to produce a feature map and then a feature vector of fixed-length with a Region of Interest (RoI) pooling layer is extracted from each region proposal.

3. Proposed Methodology

3.1. Flowchart

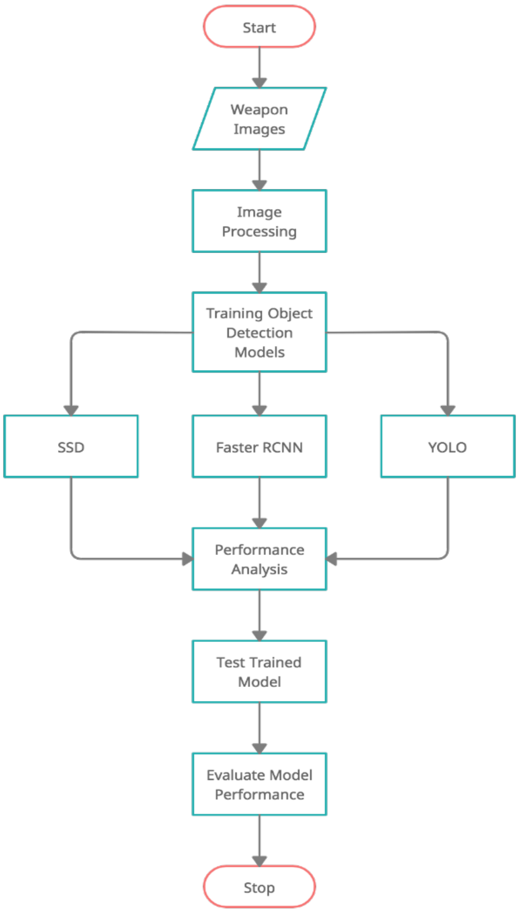

Figure 1: Flowchart of weapon detection system proposed

3.2. Scraping images of commonly used weapons

To build DL-based object detection model, nearly 5000 images of commonly used weapons like AK47, Hand Pistol, Revolver, Shotgun, Combat Knife, etc. were scrapped available in various sizes and orientations, which is later pre-processed for building dataset. These images were gathered from different sources available on the internet using automation scripts.

3.3. Building labelled dataset

After pre-processing and scaling, these scrapped images were labelled using "LabelImg" software. This software helps in labelling class objects by simply marking the object manually in the image using selection tool. The software then creates a text (*.txt) file where the exact 2D coordinates of the bounding box along with the class identifier is stored. These files when combined with images give us a complete labelled dataset which further can be used to train any desired DL-based object detection model.

3.4. Model training

Then proceeding further different Deep Learning frameworks and models such as RCNN, SSD, and YOLO were trained using the above built dataset. The dataset was randomly divided in the ratio of 80 - 20. Using 80 percent of the dataset the model was trained. The remaining 20 percent of the dataset was used for testing purposes. To train these DL-based models Google Collab was used because of its free and powerful GPUs.

4. Experimental Results

After completion of training with 80 percent of dataset the desired weights were obtained. The remaining 20 percent of the dataset was tested against these obtained weights, giving a complete and detailed analysis of accuracy and precision of these models which was recorded for further comparison and eventually figuring out the best detection model.

4.1. Evaluation Metrics in an Object Detection Model

In computer vision, Mean Average Precision (mAP) is used to evaluate the Object Detection Model [13]. It measures the accuracy by calculating the number of correct predictions that the model made. To find the mAP score of a model, we have to find the value of Intersection of Union (IoU), precision and recall prior.

Object detection models generate predictions in terms of a class label and a bounding box. For every prediction, we will measure IoU by taking the ratio of area of overlap between the predicted bounding box and the ground truth bounding box to the area of union of both bounding boxes. Mathematically,

$$IoU = \frac{\text{Area of Overlap}}{Area\ of\ Union} \qquad (1)$$

We will find the values of Recall and Precision by using this IoU value, for a given threshold. Taking an example, if IoU is set to 0.7 threshold, and the IoU value of 0.8 is achieved for a prediction, which is greater than or equals to the set threshold then the prediction is classified as TP i.e. True Positive otherwise the prediction is classified as FP i.e. False Positive.

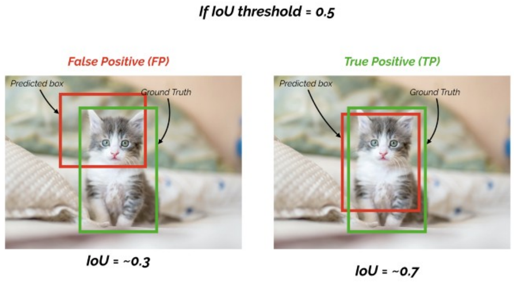

Figure 2: Marking a predicted box as True Positive or False Positive based on IoU value

Precision of an object detection model is the ratio of total number of instances of True Positives to the total number of instances of True Positives and False Positives all together. Mathematically,

$$Precision = \frac{\text{TP}}{(TP + FP)} \qquad (2)$$

where,

TP is true positives i.e. predicted as positive and is true & FP is false positives i.e. predicted as positive and is false.

Recall measures the fraction of relevant predictions that were predicted by the object detection model. It is measured by taking the ratio of total number of instances of True Positives to the total number of instances of True Positives and False Negatives all together. Mathematically,

$$Recall = \frac{\text{TP}}{(TP + FN)} \qquad (3)$$

where,

TP is true positives i.e. predicted as positive and is correct & FN is false negatives i.e. model is failed to predict the presence of the object and object is present there.

The Average Precision (AP) is the area under the Precision vs Recall curve [14]. (Mean Average Precision) mAP is the average of Average Precision (AP).

4.2. Analysis of Accuracy Precision (AP) and GPU time to process one frame (in ms)

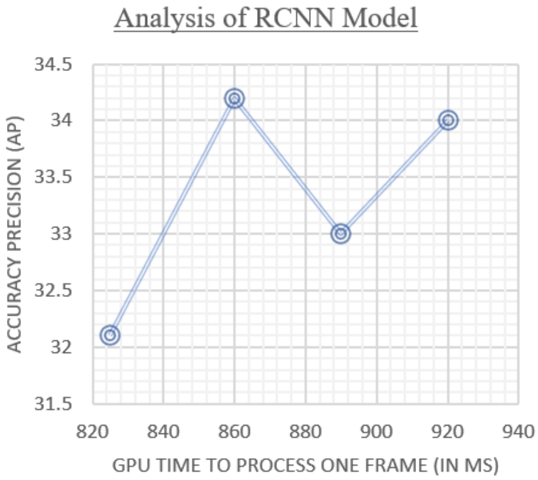

Figure 3: Graph of GPU time vs Accuracy Precision for RCNN Model

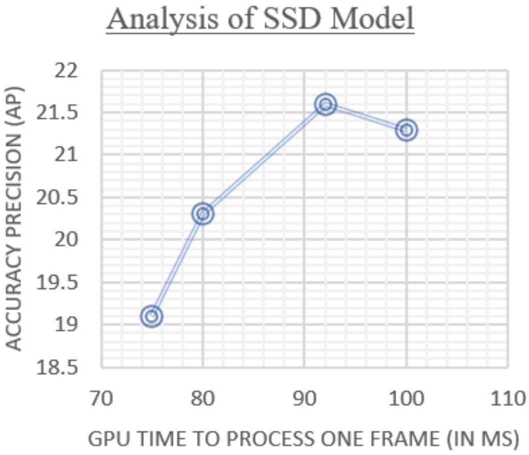

Figure 4: Graph of GPU time vs Accuracy Precision for SSD Model

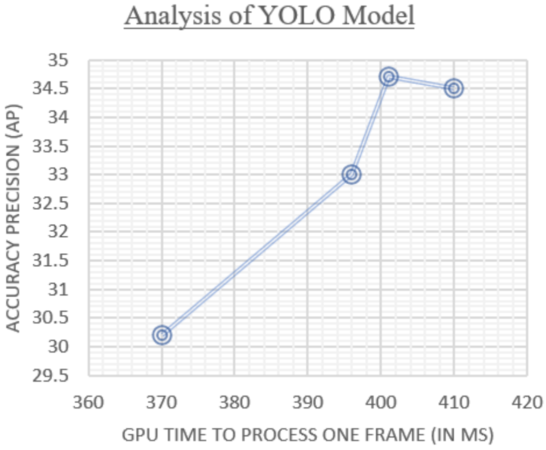

Figure 5: Graph of GPU time vs Accuracy Precision for YOLO Model

Table 1

Comparison of mAP Score and GPU Time of different object detection model

Model	mAP Score	GPU time (in ms)
RCNN	33.325	874
SSD	20.58	86.8
YOLO	33.1	394

5. Conclusion

A comparative study of various DL-based object detectors that uses Artificial Neural Network to classify and localize visual data was conducted. It was tested on a custom dataset of commonly used weapons. SSD amongst all the other DL-based Object Detection Frameworks has the least mAP score of 20, but the computational time to detect the object was the fastest of others. So, it can be a better choice if we need a fast object detector in trade-off to accuracy. YOLO and RCNN provided a similar mAP score of 33 and 34.2 respectively which gives the better accuracy of detecting the object. Although the YOLO trained model is comparatively faster than the RCNN model making it an efficient and reliable object detector.

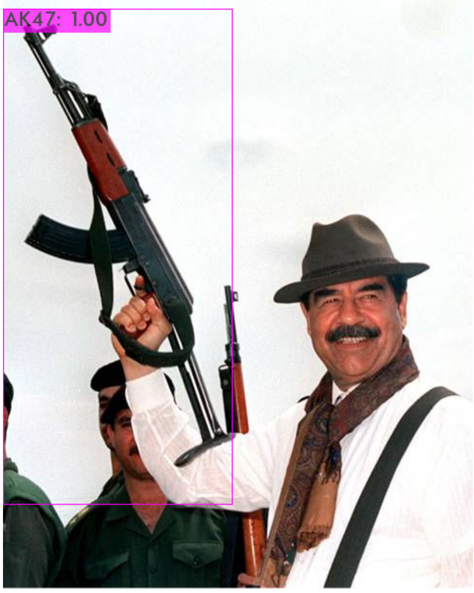

Figure 6: Assault rifle AK47 detected in a captured video

6. Future Scope

Deep Learning based Object Detection framework mainly consists of two types –

1. This research paper is limited to weapons like AK47, Shotguns, 9mm Hand Pistols, Combat Knives, and Hand Grenades which are some of the most commonly used weapons amongst criminals. But there are still a large number of dangerous and illegal weapons whose possession by any individual can cause a problem. So our target will be to add more and more of these labeled datasets into our training part to ensure maximum accuracy.
2. We can enhance this project to a collision detection system where all sorts of vehicle accidents can be traced and reported to the nearby police station to prevent any further damage.
3. Another application can be to use the system for detection of any major blood around the area which will report a nearby hospital to avoid any fatality.
4. Another possible application could be the detection of fire at any place which upon detection can be reported directly to a fire department ensuring that there is minimum damage around that area.
5. One can also monitor through the traffic using this system where the cameras will detect all the vehicles breaking any law and reporting the same to a traffic control department helping them to resolve traffic issues.
6. One of the limitations of our project is that there is no possible solution right now to detect any weapon which is hidden by the criminal in either his pocket or suitcases. We are thinking of a way to overcome this problem and build a better and safer environment for citizens.

7. Acknowledgement

The Graphics Processing Unit (GTX 1050 Ti) used in this research paper is provided by the Computer Science & Engineering Department, Bhagwan Parshuram Institute of Technology, Delhi, India.

8. References

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

[2] Ross, Girshick (2014). "Rich feature hierarchies for accurate object detection and semantic segmentation" (PDF). *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE: 580–587. arXiv:1311.2524. doi:10.1109/CVPR.2014.81. ISBN 978-1-4799-5118-5. S2CID 215827080

[3] Liu, Wei (October 2016). "SSD: Single shot multibox detector". *Computer Vision – ECCV 2016*. *European Conference on Computer Vision*. Lecture Notes in Computer Science. **9905**. pp. 21–37. arXiv:1512.02325. doi:10.1007/978-3-319-46448-0_2. ISBN 978-3-319-46447-3. S2CID 2141740.

[4] Redmon, Joseph (2016). "You only look once: Unified, real-time object detection". *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. arXiv:1506.02640. Bibcode: 2015arXiv150602640R.

[5] Redmon, Joseph (2017). "YOLO9000: better, faster, stronger". arXiv:1612.08242 [cs.CV].

[6] Redmon, Joseph (2018). "Yolov3: An incremental improvement". arXiv:1804.02767 [cs.CV.

[7] 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.

[8] "26/11 Mumbai Terror Attacks Aftermath: Security Audits Carried Out On 227 Non-Major Seaports Till Date". *NDTV*. Press Trust of India. 26 November 2017. Retrieved 7 December 2017.

[9] S. A. Velastin, B. A. Boghossian, M. A. Vicencio-Silva, A motion-based image processing system for detecting potentially dangerous situations in underground railway stations, Transportation ResearchPart C: Emerging Technologies 14 (2) (2006) 96–113.

[10] T. Ainsworth, Buyer beware, Security Oz 19 (2002) 18–26.

[11] Rapid object detection using a boosted cascade of simple features

[12] Viola, Jones: Robust Real-time Object Detection, IJCV 2001 (pages 1,3).

[13] Hughes, G. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE transactions on information theory*, *14*(1), 55-63.

[14] Buckland, M., & Gey, F. (1994). The relationship between recall and precision. *Journal of the American society for information science*, *45*(1), 12-19.

[15] Ahmad Salihi Ben Musa, Sanjay Kumar Singh, Prateek Agrawal, *"Suspicious Human Activity Recognition for Video Surveillance System"*, International Conference on Control, Instrumentation, Communication & Computational Technologies ICCICCT-2014, IEEEXplore.

[16] Nivid Limbasiya, Prateek Agrawal, "Bidirectional Long Short Term Memory Based Spatio-Temporal in Community Question Answering", A book on Deep learning based approaches for sentiment analysis, pp. 291-310, Jan 2020, Springer.

[17] Prateek Agrawal, Deepak Chaudhary, Vishu Madaan, Anatoliy Zabrovskiy, Radu Prodan, Dragi Kimovski, Christian Timmerer, "Automated Bank Cheque Verification Using Image Processing and Deep Learning Methods", Multimedia tools and applications (MTAP), 80(1), pp. 1-32. https://doi.org/10.1007/s11042-020-09818-1.

[18] Prateek Agrawal, Deepak Chaudhary, Vishu Madaan, Anatoliy Zabrovskiy, Radu Prodan, Dragi Kimovski, Christian Timmerer, "Automated Bank Cheque Verification Using Image Processing and Deep Learning Methods", Multimedia tools and applications (MTAP), 80(1), pp. 1-32. https://doi.org/10.1007/s11042-020-09818-1.

[19] Neha Bhadwal, Prateek Agrawal, Vishu Madaan, Awadhesh Shukla, Anuj Kakran, "Smart Border Surveillance System using Wireless Sensor Network and Computer Vision", International Conference on Automation, Computational and Technology Management (ICACTM'19), pp. 183-190, IEEEXplore.