Artificial Intelligence and Deep Learning for Weapon Identification in Security Systems

Makkena Brahmaiah¹, Srinivasa Rao Madala², Ch Mastan Chowdary³

¹M.Tech., Scholar, Department of Computer Science and Engineering
²Professor & HOD, Department of Computer Science and Engineering,
³Asst. Professor, Department of Computer Science and Engineering,
PACE Institute of Technology and Sciences

Abstract. As crime rates rise at large events and possibly lonely places, security is always a top concern in every field. A wide range of issues may be solved with the use of computer vision, including anomalous detection and monitoring. Intelligence monitoring is becoming more dependent on video surveillance systems that can recognise and analyse scene and anomaly occurrences. Using SSD and Faster RCNN techniques, this paper provides automated gun (or weapon) identification. Use of two different kinds of datasets is included in the proposed approach. As opposed to the first dataset, the second one comprises pictures that have been manually tagged. However, the trade-off between speed and precision in real-world situations determines whether or not each method will be useful.

Keywords: Firearm identification, computer vision, faster RCNN, SSD, CCTV, and Artificial Intelligence are some of the key terms (AI).

1. Introduction
WMD detection, also called as "anomaly" or "anomoly", is the identification of unusual, unexpected occurrences or things that are not considered to be part of the normal course of a pattern or dataset, and thus vary from the existing patterns of behaviour. In biology, an abnormality is a pattern that does not fit into a typical pattern. Anomalous events are thus modified as a consequence. It is possible to identify objects by utilising feature extraction and learning methods [6]. Detecting and categorising firearms correctly is the goal of the suggested implementation. There's also the issue of accuracy, since a false warning may lead to undesirable outcomes [11] [12]. A delicate balance has to be struck between accuracy and speed while choosing the appropriate technique. The deep learning-based weapons detection method is shown in Figure 1 below. Frames are retrieved from the video input. Use of the frame differencing technique and generation of the bounding box are required before an item can be detected. It uses a dataset that has been created, trained, and supplied to the object detection algorithm to find objects. If a gun had to be detected, the application dictated which detection technique to use (SSD or RCNN). RCNN and Single Shot Detection (SSD) are machine learning models that are used to address the detection issue [2][9][15].

2. Literature Review
All locations are thus faced with the task of minimising potentially life-threatening situations and maintaining high levels of security. As a consequence, researchers have utilised object detection to monitor a variety of activities and behaviours. This kind of technology is able to extract low-level information, such as features engineering or object tracking, recognise unusual human behaviour, or
even find and detect weapons. Systems at the highest levels of government are intended to make judgments about anomalous occurrences. Object-centered and integrated methods, which may be divided into two groups, are used to identify anomalous occurrences. CNN spatial-temporal system lowers processing costs because of its restricted usage. According to researchers at [14], surveillance video of complex circumstances may be used to detect and uncover aberrant behaviour. Objects may be captured in both time and frequency domains by using a spatial-temporal convolution layer, which allows for the extraction of both object presence and movement information contained in continuous frames. There are spatial-temporal convolution layers that can be utilised to decrease local noise and improve detection accuracy, but only for moving pixels. To identify anomalous behaviours, researchers created a multi-instance learning graph-based model that highlights positive occurrences by training coarse filters using kernel-SVM classifiers and anchor dictionary learning, an enhanced dictionary learning technique. By comparing the cost and time of SRC with those of other techniques, such as utilising anomalous information and reducing SRC's time and cost, normality may be determined.

Three phases of object detection are used by Hu et al. [15] in their object detection method, which contributes to the identification of various objects in traffic scenarios. In the beginning, it uses three different types of objects: cars, bicycles, and traffic signs to identify, recognise, and track items in motion. A learning-based detection system that combines a dense feature extractor and trimodal class detection is used to identify all items. Over and beyond, the detectors extract dense features that they then share with each other so that they may be more efficient overall. An object subcategorization technique that uses intraclass variance is given.

This means that when knives or weapons are spotted in a CCTV image, the security guard or operator is alerted. In order to reduce false alarms, the algorithm has a 94.93 percent specificity and an 81.18 percent sensitivity for knife detection. There is 96.69 percent fire alarm specificity, whereas the fire alarm sensitivities for different items in the movie are 34.98 percent. Histogram of Directed Tracklets (HDT) is a method that identifies abnormal situations in complex pictures. Researchers Mousavi and colleagues reported their findings in [17]. There are now descriptors based on long-range motion projections called tracklets as opposed to traditional techniques that use optical flow, which only evaluate edge features from two subsequent frames. Video sequences of spatial-temporal cuboids have been statistically gathered on the tracklets that go through them.

Ji and his colleagues developed a surveillance video system that can automatically detect human activities using convolutional neural networks (CNNs) and deep learning (DL). For the 3D CNN model for classification, regularisation of outputs using high-level characteristics and the integration of data from a variety of various models are thus needed to achieve success.

Pang et al. [19] demonstrated real-time concealed object identification under human clothes. In order to get passive millimetre wave imaging, a small sample of human bones was subjected to the YOLO algorithm. So, we've compared SSD-VGG16 with YOLOv3-13 and YOLOv3-53. Using 36 frames per second and 95 percent mean average precision, the accuracy of weapon detection was determined. Researchers at Warsi A et al. were able to detect handguns with fewer false negatives and positives using Faster Region-Based CNN (RCNN) and YOLO V3 algorithm [20]. After capturing real-time images and incorporating them into the ImageNet dataset, the YOLO V3 algorithm was trained. If four different movies were used, it was discovered that YOLO V3 was faster in real-time settings than FasterRCNN when using four different movies.

3. Implementation

3.1 An overview of the tools and resources utilised for implementation

To learn more about our products and services please visit our website or contact us by phone or email at [email protected] or [email protected].
Objects having common labels are stored in the COCO Dataset.
- NVIDIA GeForce 820M GPU
- GeForce is a brand of graphics processing units developed by NVIDIA.

3.2 Data set Specifications

Case-I
- Intel i5 7th generation (4 cores)
- 2.5 GHz clock speed
- NVIDIA GeForce 820M graphics card
- 29.97 frames per second in the input
- 0.20 frames per second in the output.
- .mov video format, 4.14 MB video size, COCO and self-created picture dataset

Case II
- Intel i5 7th Generation (4 cores)
- 2.5 GHz • NVIDIA GeForce 820M
- 200-300 KB input picture size
- 0.6 seconds for SSD and 1.7 seconds for RCNN
- JPG image format
- COCO and self-created image dataset

For execution, hypotheses and limitations were made
- Camera line-of-sight: the gun is completely or partly visible to the camera.
- As a result, the ammunition can be easily seen in the background.
- When detecting ammunition, a high-performance GPU was utilised to minimise latency.
- There isn’t a completely automated procedure in place for this process. Every gun detection warning will be double-checked by someone in authority.

Fig 1: Methodology
3.3 FASTER R-CNN
Figures 3 and 4 show CNN layers and a faster RCNN architecture, respectively. It has two networks, one for generating region recommendations and the other for detecting objects. It employs a selective search strategy to generate region proposals. The RPN network ranks anchors or region boxes.

![Fig 3. Layers in CNN Architecture [5]](image)
As a result, Fatkun Batch Image Downloader (chrome extension) is used to download large numbers of Google Images. The pictures are then labelled. In all, 80 percent of the pictures were utilised for training, while 20 percent were used for testing purposes. On this data, the Single Shot Detector (SSD) model was trained using 2669 iterations/steps, which ensured the loss was less than 0.05, thereby improving precision and accuracy. Photographs from tests and training are shown in Figure 5. As seen in Figure 6, the picture has been labelled. The python xml to csv.py command in Anaconda Prompt converts XML data into a CSV file. There are produced CSV files for both test and training datasets in Figures 7 and 8.
Fig. 6. Image along with its label

Table 1. CSV file of testing dataset
Table 2. CSV file of Training dataset

3.3 SSD (Single Shot Detector)

In terms of precision and performance detection, the SSD algorithm has achieved new heights. By removing the requirement for a region proposal network, SSD speeds up the procedure. SSD uses a combination of technology, including default boxes and multi-scale characteristics, to compensate for the loss of precision. These enhancements allow SSD to match the accuracy of the Faster R-CNN utilising lower resolution images, greatly increasing speed. On the COCO dataset, the average scoring is around 74 percent MAP and 59 frames per second.

4. Results

According to Table 1, accuracy for the pre-labeled dataset (AK47) is the greatest on average, whereas the Colt M1911, Smith & Wesson Model 10, UZI Model, and Remington Model range from 76 to 91%. R-CNN is quicker, with an average accuracy of 84.6 percent and a frame speed of 1.606 seconds, compared to the other two algorithms mentioned above. The pre-labeled dataset outperformed the self-created dataset in terms of accuracy since it was trained on millions of images. SSD. SSD's performance analysis is shown in Table 2.

Seventy-three percent accuracy and 0.736 frames per second are achieved by using SSDs. There were five distinct types of firearms that Model was trained with: the AK47, Smith & Wesson Model 10, Colt M1911 UZI-style rifle and Remington-style rifle. SSD and RCNN Inception V2 models were used to train the guns, and the results were impressive. The SSD model took 12 hours longer to train than the RCNN model, but it was less accurate. Overall, the quicker algorithm was 10.8 percent more accurate than the slower method (SSD). SSD is quicker than Faster R-CNN by around 0.7 seconds in terms of performance. In SSD and Faster R-CNN models, pre-labeled datasets like AK47 guns offer better accuracy compared to self-created images.
5. Conclusions
Using a pre-labeled, self-created picture dataset, the SSD and Faster RCNN algorithms are used to identify weapons (guns). But its usage in real time necessitates a tradeoff between accuracy and speed. With 0.736 frames per second, the SSD algorithm is quicker. As opposed to SSD, RCNN is much slower, with frame rates of 1.606s each frame. With an accuracy of 84.6 percent, a faster RCNN is more accurate. When compared to RCNN, SSD has a much lower accuracy of 73.8 percent, which is disappointing. As a result of its greater speed, SSD allowed for real-time detection, while RCNNs with better accuracy were quicker. GPUs and high-end DSP and FPGA packages may also be utilised to train for bigger datasets [16] [17].

References
[1] Madala, S. R., Rajavarman, V. N., & Vivek, T. V. S. (2018). Analysis of Different Pattern Evaluation Procedures for Big Data Visualization in Data Analysis. In Data Engineering and Intelligent Computing (pp. 453-461). Springer, Singapore.
[2] Madala, S. R., & Rajavarman, V. N. (2018). Efficient Outline Computation for Multi View Data Visualization on Big Data. International Journal of Pure and Applied Mathematics, 119(7), 745-755.
[3] Vivek, T. V. S., Rajavarman, V. N., & Madala, S. R. (2020). Advanced graphical-based security approach to handle hard AI problems based on visual security. International Journal of Intelligent Enterprise, 7(1-3), 250-266.
[4] H R Rohit et al., " A Review of Artificial Intelligence Methods for Data Science and Data Analytics: Applicat ions and Research Challenges," 2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), 2018.
[5] Abhiraj Biswas et al., “ Classification of Object s in Video Records using Neural Network Framework,” International conference on Smart Systems and Inventive Technology, 2018.
[6] Pallavi Raj et al., “Simulation and Performance Analysis of Feature Extraction and Matching Algorithms for Image Processing/Applicant ions’,” IEEE International Conference on Intelligent Sustainable Systems, 2019.
[7] Mohana et al., “Simulation of Object Detect ion Algorithms for Video Surveillance/Applicant ions”, International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), 2018.
[8] Yojan Chitkara et al., “Background Modelling techniques for foreground detect ion and Tracking using Gaussian Mixture model” International Conference on Computing Methodologies and Communication, 2019.
[9] Rubner et al., “A metric for distributions with applications to imagedat abases”, International Conference on Computer Vision, 2016. [10] N. Jain et al., “Performance Analysis of Object Detect ion and Tracking Algorithms for Traffic Surveillance Applicat ions using Neural Networks,” 2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), 2019.
[10] Kishor Kumar Reddy C, Anisha P R, Aporvka K, “Early Prediction of Pneumonia using Convolutional Neural Network and X-Ray Images”, Smart Innovation, Systems and Technologies, 2021
[11] A. Glowaczelet al., “Visual Detect ion of Knives in Securit y Applications using Active Appearance Model”, Multimedia Tools Applicat ions, 2015.
[12] S. P ankanti et al., “Robust abandoned object detect ion using regionlevel analysis,” International Conference on Image Processing, 2011.
[13] Ayush Jain et al., “Survey on Edge Computing - Key Technology in Retail Industry,” International Conference on Intelligent Computing and Control Systems, 2019.
[14] Mohana et al., Performance Evaluat ion of Background ModelingMet hod s for Object Detect ion and Tracking,” International Conference on Inventive Systems and Control, 2020.
[15] J. Wang et al., “Detecting static object s in busy scenes”, Technical Report TR99-1730, Department of Computer Science, Cornell University, 2014.
[16] V. P. Korakoppa et al., “An area efficient FPGA implementation of moving object detect ion and face detect ion using adaptive threshold method,” International Conference on Recent Trends in Electronics, Information & Communication Technology, 2017.
[17] S. K. Mankani et al., “Real-time implementat ion of object detect ion and tracking on DSP for video surveillance applications,” International Conference on Recent Trends in Electronics, Information & Communication Technology, 2016.