Weapons Detection in Video Surveillance

PRASHANT BHALLA[1], SACHIN GARG[2]
[1, 2] *Information Technology (MAIT), Maharaja Agrasen Institute of Technology, Delhi, India*

Abstract- The automatic detection of weapons captured by surveillance settings is essential for speeding the otherwise laborious approach. With use of , UCFCrime which is the largest available dataset for automatic visual analysis of anomalies and consists of real-world crime scenes of various categories. In this paper, we introduce HR-Crime, a subset of the UCF-Crime dataset suitable and use a technique for specified object tracking.

I. INTRODUCTION

In the last few years, deep learning techniques and Convolutional Neural Networks (CNNs) have achieved great results in image detection, classification, segmentation and it's being used in several applications. The advancements in technology and the latest innovative detection models such as YOLO,FasterR-CNN,VGG-16 have achieved satisfactory results. The common challenges that are faced while weapon detection is the increase in complexity due to partial or full occlusion of gun deformation and loss of information while transmission. The rate of false-negative and false-positive alsois an issue in weapon detection systems due to such sensitives systems being linked to alarms or such devices. Weapon Detection systems need Real-time processing and fast response times due to their critical nature, so the research has to findand implement techniques that speed processing time weapon detection models

1. Need of study: The main problems such as the high number of false negatives and false positives occur due to challenges such as the similar shape and handling of non- weapon objects which are commonly handheld. Another major challenge is ensuring that the model doesn't fail to detect the weapon and has a very low false negatives rate. The model should also be able to avoid false positives from the background of these images and videos. The already presented models have a very high rate of false negatives when it comes to videos. Suppose if 10 people with weapons are interested to enter a building and out of these 10 only one person succeeded to enter in the building can lead to serious consequences.
2. Goal of the study: To achieve this goal, this study puts forth a model that can take advantage of the latest modelssuch as YOLO which has very fast detection speeds .The main contributions of this work inclued reducing the number off alse positive sand negatives in the domain of Weapon Detection by using Gaussian blur to remove the background and only focusing on the area of Interest and its combine duse with YOLOv5 swith Stochastic gradient descent (SGD)
3. Objective of the study: Weapon Detection systems need Real-time processing and fast response times due to their critical nature, so the research has to find and implementtechniques that speed the processing time of weapon detection model

II. MODELS

A. *YoloV5*

YOLO an acronym for 'You only look once', is an object detection algorithm that divides images into a grid system. Each cell in the grid is responsible for detecting objects within itself.YOLO is one of the most famous object detection algorithms due to its speed and accuracy.YOLOv5 is a family of compound-scaled object detection models trained on the COCO dataset, and includes simple functionality for Test Time Augmentation (TTA), model ensembling, hyperparameter evo-lution, and export to ONNX, CoreML and TFLite.YOLOv5 has four different models including YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x. Generally, YOLOv5 respectively uses the following architecture:-

1. Backbone: Model Backbone is mostly used to extract key features from an input image. CSP(Cross Stage Partial Networks) are used as a

backbone in YOLO v5 to extract rich in useful characteristics from an input image.

2. Neck: The Model Neck is mostly used to create fea- ture pyramids. Feature pyramids aid models in generalizing successfully when it comes to object scaling. It aids in the identification of the same object in various sizes and scales. Feature pyramids are quite beneficial in assisting models to perform effectively on previously unseen data. Other models, such as FPN, BiFPN, and PANet, use various sorts of feature pyramid approaches.
3. PANet is used as a neck in YOLO v5 to get feature pyramids.
4. Head: The model Head is mostly responsible for thefinal detection step. It uses anchor boxes to construct final output vectors with class probabilities, objectness scores, and bounding boxes.

Currently, there are two types of detection methods based on deep learning: 1-stage detector and 2-stage detector. Firstly, 2- stage detector in which regional proposal and classification are performed sequentially. The faster R-CNN and mask R-CNN Menelaos-NT Research Report template by Zhouyan Qiu, University of Vigo correspond to the kind of 2-stage detector. In contrast to 2- stage detector, in the 1-stage detector, a regional proposal and classification are performed simultaneously. In other words, it is a method of solving classification and localization problemsat the same time. YOLO , TPH-YOLOv5 , SSD , SSD MobileNet , Focal Loss , and RefineDet ; are representative algorithm of 1-stage detector. While it was popular in thepast, Fast R-CNN has an inefficient problem in learning and execution speed because the candidate area generation module is performed in a separate module independently of CNN . The YOLO is a famous object detection algorithm with several versions. It is easy to implement and can train the entire image immediately. For this reason, YOLO has developed gradually . In 2020, the fifth version of YOLO was released. Compared to fast R-CNN, speed and accuracy have increased. Since YOLO does not apply a separate network for extracting candidate regions, it shows better performance in terms of processing time than Fast R-CNN . Because Fast R-CNN was the combining hand-crafted and deep convolutional features method is used, there are limitations in detecting objects or humans . The basic structure of the previous YOLOv5 is largely divided into the backbone network part, the neck part, and the head part, as shown in Figure 1 . Backbone is a convolutional neural network formed by aggregating image features in various particle sizes. Neck is a series of layersthat mix and combine image features to deliver prior to prediction, and Head consumes features from Neck (PAnet) and takes box and class prediction steps. The biggest feature of YOLOv5 is that it has Focus and CSP (cross-stage partial connections) layer. The focus layer was created to reduce layers, parameters, FLOPS, and CUDA memory and improve forward and backward speed while minimizing the impact of mAP. Three layers were used in YOLOv3 [31], but in the previous YOLOv5, it was changed to one layer . The CSP layer extends to shallow information in the focus layer to maximize functionality, while the feature extraction module is iteratedto extract detailed information and functions more thoroughly

III. DATASET

We have used dataset and combined them into two cate- gories Anoma- lous and non-anomalous.The dataset used is DaSic weapon-detection Dataset.The datasets included in this section have been designed for the object detection task based on Deep Learning architectures with a CNN backbone. The selected images contain weapons and objects but also consider an enriched context of different background objects as well as the way objects are handled. After the training stage on these datasets, the detection models must locate and distinguish between weapons and different common objects present in the background or handled similarly. The datasets also attached the annotation files in Pascal VOC format with the region of the target objects in xml files.

1. Handgun detection - The Pistol detection dataset contains 3000 images of short guns with rich context in the background. The images selected from the internet contain one or more handguns in diverse situations including video surveillance contexts.
2. Knife detection - The Knife detection dataset contains 2078 images where at least one knife

appears.The selected images were download from Internet, and some frames were extracted from Youtube videos or surveillance videos. The dataset take into account: cold steel weapon of diverse types, shape, colors, size, and made of different materials knives located at different distances near and far from the camera knives occluded partially by the hand objects that can be handled in the same way as knives images captured in indoor and outdoor scenarios

3. Weapons and similar handled object detection The Sohas weapon detection dataset is formed by weapons and small objects that are handled in a similar way. It includes six different objects such as pistol, knife, bill, purse, smartphone and card.

IV. METHODOLOGY

In this section, the different steps involved in the proposed method are detailed, the specific procedures or techniques used to identify, select, process, and analyze information about analomous behaviour with help of Human-Pose Estimation andObject Detection

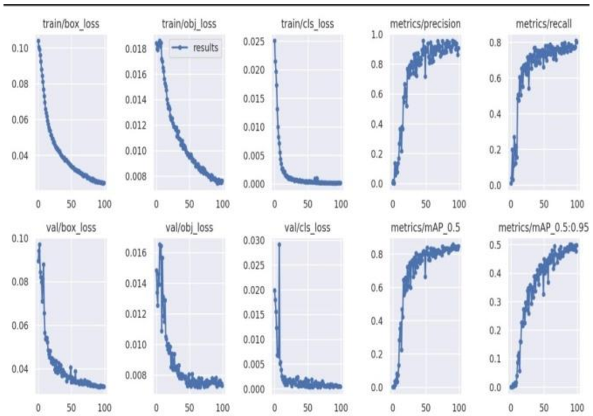
A. Yolov5

YOLO-v5 as a learner to detect weapon (as objects) with their centroids and sizes. The YOLO-v5 DL-based object detection method is used in this paper to construct a model that can make the clustering algorithm free of initialization. This model finds automatically the appropriate clustering initializa-tion parameters such as the number of weapon, possible cen- troids and weapon sizes. The input of the proposed model is a transformed data in 2D feature space. The proposed solution assumes that a high-dimensional data-set can be represented in a lower dimension space, pairwise distance is preserved among weapon. Given a transformed data-set of n points X = x1, . . . , xn, where xi R2 , the clustering algorithms, such as k-means, partition the data-set into k weapon C1, C2, . . . , Ck, where each weapon Cj is represented by a weapon center cj . Let k denotes the correct number of weapon in a data-set. To build accurate clustering models 6 that can achieve high clustering accuracy, k and other clustering parameters (e.g. no. of weapons and weapon sizes) are required as input. Getting a good estimate of these parameters and especially k is not a straightforward task. Several metrics were presented previously to determine the number of weapon. However, computing these metrics requires a high computation and resources overhead. The input of the proposed solution is an image representing the weaponed data. Thus, with large data- set, the computation overhead of the proposed solution will remain the same, contrarily to the existing techniques that require an exponential computation complexity in function of the data size The YOLO-v5 model was trained using a generated data-set. This data-set consists of a large set ofimages provided by DaSIC.

3.3 Model Implementation The YOLO-v5 model imple-mentation of was used to build our proposed learner model. To train this model, the SGD optimizer was used with an initial value of 102 and a batch of size 16. All the other parameters are the standard parameters of the Yolo-V5 code. In the training phase, the Yolo-V5 model learns to recognize all the weapon with 20 epochs. At the inference phase, once the model detects the bounding boxes, the center for each weapon is computed and this value will be the initial value for the initial weapon. As Yolo-V5 being very efficient and lightweight, it can detect quickly objects (weapon) and canbe implemented on GPU or CPU (after being trained). Let us indicate that the euclidean distance between two points p and q in the d space can be computed according to the following equation:
The lower is d (p, q), the closer are the points p and q. When d is equal to 0, p and q are collocated. These results show that the proposed method provides better approximation of the correct centroids and consequently this will result in increasing the accuracy of the clustering algorithm, in addition to reducing the memory consumption and the number of iterations for the clustering algorithms to converge.Theclustering accuracy rate (AR) can be computed as follows:
where n(ck) is the number of data points that were correctly included in cluster k, and n is the total number of data points. The higher is the AR, the better is the clustering correctness.

V. RESULT

We propose an anomaly detection algorithm that relies on object detection(that solely focus on weapon detection like knife and gun in the frame).The advantages of the proposed solution is that it is: lightweight, fast, and robust .The proposed algorithm works on both fine-grained anomaly detection, where the goal is to detect variations of a single action (e.g., appearance of weapon), as well as a new coarse grained anomaly detection setting, where the goal is to distinguish between normal and abnormal actions.



REFERENCES

[1] Dongkuan Xu and Yingjie Tian. A comprehensive survey of clustering algorithms. Annals of Data Science, 2(2):165–193, 2015.

[2] Christophe Guyeux, Stéphane Chrétien, Gaby Bou Tayeh, Jacques Demerjian, and Jacques Bahi. Introducing and comparing recent clustering methods for massive data management in the internet of things. Journal of Sensor and Actuator Networks, 8(4):56 (25), dec 2019.

[3] Kristina P Sinaga and Miin-Shen Yang. Unsupervised k-means clustering algorithm. IEEE Access, 8:80716–80727, 2020.

[4] Adil Fahad, Najlaa Alshatri, Zahir Tari, Abdullah Alamri, Ibrahim Khalil, Albert Y Zomaya, Sebti Foufou, and Abdelaziz Bouras. A survey of clustering algorithms for big data: Taxonomy and empirical analysis. IEEE transactions on emerging topics in computing, 2(3):267–279, 2014.

[5] Maria Camila N Barioni, Humberto Razente, Alessandra MR Marcelino, Agma JM Traina, and Caetano Traina Jr. Open issues for partitioning clustering methods: an overview. Wiley Interdisciplinary Reviews: Data Mining and KnowledgeDiscovery, 4(3):161–177, 2014.

[6] Mohammad Alhawarat and M Hegazi. Revisiting k- means and topic modeling, a comparison study to cluster arabic documents. IEEE Access, 6:42740–42749, 2018.

[7] Yinfeng Meng, Jiye Liang, Fuyuan Cao, and Yijun He. A new distance with derivative information for func- tional k-means clustering algorithm. Information Sciences, 463:166–185, 2018.

[8] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the em algorithm. Journal of the Royal Statistical Society : Series B, 39(1):1–38, 1977.