WEAPON DETECTION IN REAL TIME USING YOLOv8

Kamlesh A. Kadam
Bachelors of computer engineering
Maulana Mukhtar Ahmad Nadvi Technical Campus
Malegaon, India
ajay1919kadam@gmail.com

Nishant S. Bacchav
Bachelors of computer engineering
Maulana Mukhtar Ahmad Nadvi Technical Campus
Malegaon, India
nishantbacchav2002@gmail.com

Tushar Dilip Ahire
Bachelors of computer engineering
Maulana Mukhtar Ahmad Nadvi Technical Campus
Malegaon, India
tusharahire660@gmail.com

Kamlesh W. More
Bachelors of computer engineering
Maulana Mukhtar Ahmad Nadvi Technical Campus
Malegaon, India
kamleshmore369@gmail.com

Prof. Siddiqui Mohsin
Computer dept
MMANTC
Malegaon, India

Dr.Salman Baig sir
HOD Computer dept
MMANTC
Malegaon, India

Abstract— In today's modern world the security is more essential. To make a nation economically strong, the investors and tourists must provide a safe and secure surrounding. So for monitoring activities and surveillance Closed Circuit Television (CCTV) cameras are used i.e. (robberies or theft) but still this surveillance need the human intervention. For detecting these unauthorized activities using we needed a system. The weapon detection in real time is still a complicated challenge having deep learning algorithm, fast processing hardware, and enhanced CCTV cameras. This work focuses on providing an authorized or safe place CCTV videos as a source to detect suspicious or harmful weapons by applying the YOLOv8 model which uses deep learning and machine learning algorithms.

Two approaches are used i.e. YOLOv4 and YOLOv8 for real time surveillance.

Keywords—knife detection, gun detection, object detection, Machine learning, YOLOv8

I. INTRODUCTION

Suspicious-events or human-beings'-behavior detection is part of a video or image surveillance system and this will have potential and added advantage for surveillance and forensic identification systems. In addition, it helps the manual power investigation time in the detection of criminal activity incidents and makes it an automated one. In today's world pre-suspicious and pre-criminal activity has to be identified as the impact of such incidents leads to damage of assets and threat to the safety of humans. In this work, the authors required the detection of proactive incidents to decrease the cost of security incidents. A method was imposed to carry out suspicious activity behavior identification with YOLOv8 features for a real time suspicious activity detection. A dataset of more than 20 suspicious activity incident cases was validated with an Experiment by using a top-down approach for transferring Features-based knowledge into rules. The results were gathered with the help of the false alarm rate. This concept is applied in many research-based applications related to suspicious-activity detection, pickpocket detection, chain-snatch-event detection, and other social-security detection including criminal-activity detection. Work was carried out to deploy a mechanism using deep learning for real-time suspicious activities like assaults, showing knife, and showing weapons on video surveillance datasets. This would be applied in real-world suspicious activity detection. Work was carried out with a multi-camera dataset to recognize the activities of humans from a dataset with labeled behavior. YOLOv8 methods are used for suspicious-human-activity detection: human detection by background subtraction, feature extraction by CNN.

CCTV cameras play an important role to overcome this problem and are considered to be one of the most important requirements for the security aspect. CCTVs are installed in every public place today and are mainly used for providing safety, crime investigation, and other security measures for detection. CCTV footage is the most important evidence in courts. After a crime is committed, law enforcement agencies arrive at the scene and take the recording of footage with them if we look at the surveillance system of different countries around the world, UK has about 4.5 million cameras, which are used for surveillance. Sweden has about 50000 cameras installed around 2010. The government of Poland was able to reduce drug cases by 60% and street fights by 40% by installing just 450 cameras in the city of Poznan [5]. China has the world's biggest surveillance system and 170 million cameras around the nation, and these are expected to expand three times, through an additional 400 million to be connected by 2020.

In previous years, though having surveillance cameras installed, to use them for security purposes was not an easy and dependable method. A human has to be there all the time to monitor screens. CCTV operator has to monitor 20-

25 screens for 10 hours. He has to look, observe, identify, and control the situation that can be harmful to the individuals and the property. As the number of screens increases, the concentration of the person decreases considerably to monitor each screen with time.

It is impossible for the person monitoring the screens to keep the same level of attention all the time. The solution to aforementioned problem is to install surveillance cameras with the ability to automatically detect weapons. However, there is not much work done on algorithms for weapon detection in surveillance cameras, and related studies are often considering concealed weapon detection (CWD), mostly using X-rays or milli-meter wave's images employing traditional machine learning techniques. In the past few years, deep learning in particular convolutional neural network (CNN) has given ground breaking results in object categorizing and detection. It has achieved finest results thus far in classical problems of image processing such as grouping, detection and localization. Instead of selecting features manually, CNN automatically learns features from given data.

The main contributions of this work are: presentation of a first detailed and comprehensive work on weapon detection that can achieve detection in videos from real-time CCTV and works well even in low resolution and brightness because most of the work done earlier is on high definition training images but real-time scenario needs real-time training data as well for better results, finding of the most suitable and appropriate CNN based object detector for the application of weapon detection in real-time CCTV video streams, making of a new dataset because real-time detection also needs real-time training data so we made a new database and preprocessed it using different Open CV filters .

Introducing the concept of related confusion classes to reduce false positives and negatives, training and testing of our novel database on the latest state of the deep learning based classification and detection models among them Yolov5 performed best in terms of both speed and accuracy and our selected trained model predict images at almost every orientation, angle, and view, achieving the highest mean average precision of 91.73% along with a F1-score of 91% on Yolov4 and 92.81% of accuracy with YOLOv5.

II. CONTRIBUTION

In previous studies these mechanism is used to detect the suspicious activity using yolov3 and yolo v4 model but some drawbacks occur while using this mechanism such as,

In previous mechanism detect the suspicious activity with low accuracy and speed. Do not detect the suspicious activity, its only detect the weapons because through the weapons persons are perform the suspicious activity. Difficulty to detects the objects in far away and also near because its speed of detection. In these drawbacks, mechanism are generate the wrongly detections of objects. Time is highly consume because its speed of the mechanism.

Now, mechanism is build using yolov8 model and boost up the speed and accuracy in detect the suspicious activity in real time.

III. Literature Survey

This explores the tools, techniques, concepts, and mainly the work which has done before. The limitations and the short introduction of the ideas and experiences learned from some references. These works explore various methods for detecting abnormal or suspicious events in video surveillance. The approaches involve background subtraction, deep learning, motion features, and object tracking. They have a different strengths and limitations, and the choice of method depends on the specific use case and requirements. Overall, the research in this field aims to improve security and safety in various scenarios Real Time Weapon Detection Using Deep Learning Muhammad Tahir Bhatti et al. [1] proposed an algorithm used are VGG16, Inception-V3, Inception-ResnetV2, SSDMobileNetV1, Faster-RCNN Inception-ResnetV2 (FRIRv2), YOLOv3, and YOLOv4. Precision and recall count the most rather than accuracy when object detection is performed so these entire algorithms were tested in terms of them. Yolov4 stands out best amongst all other algorithms and gave a F1-score of 91% along with a mean average precision of 91.73% higher than previously achieved. Suspicious Activity Detection from Video surveillance K. Kranthi Kumar et al. [2] have proposed an algorithm which detect signs of aggression and violence in real time, allowing irregularities to be distinguished from normal patterns by utilizing deep learning models in identification or classification of high movement frames, where we can set off a detection alert in the event of a threat, alerting us to suspicious activity at a specific point in time. Suspicious Activity Detection using Image Processing Phalguni Kadam et al. [3] have presented and demonstrated an automated way of suspicious activity detection using Deep Learning and Image Processing, the proposed work aims to eliminate time and effort wasted on monitoring video surveillance cameras. Predicting human behaviourr is almost impossible. Deep Learning is used to detect suspicious and non-suspicious activity and to warn the user if any suspicious activity is detected. The proposed system strives for the detection of real-world suspicious activities such as burglaries, assaults etc. in surveillance videos. Weapon Detection Using Artificial Intelligence And Deep Learning Mirza Sohail Baig et al. [4] implements automatic gun (or) weapon detection using a convolution neural network (CNN) based SS D and Faster RCNN algorithms. Proposed implementation uses two types of datasets. One dataset, which had pre-labelled images and the other one is a set of images, which were labelled manually. Results are tabulated, both algorithms achieve good accuracy, but their application in real situations can be based on the trade-off between speed and accuracy. While reviewing the literature survey following limitations we have found Limited Data: The literature review relies on existing publications, and there may be gaps in the coverage of recent or specialized research that is not widely published. Outdated Information: The review may not include the most recent developments and studies in the field, particularly if it was conducted some time ago. Varying Quality: The quality of research articles can vary significantly, and the review may include studies of differing methodological rigor. Not Real time Detection: Every literature survey has worked Suspicious activity and non-suspicious activity detection not in real time

IV. METHODOLOGY

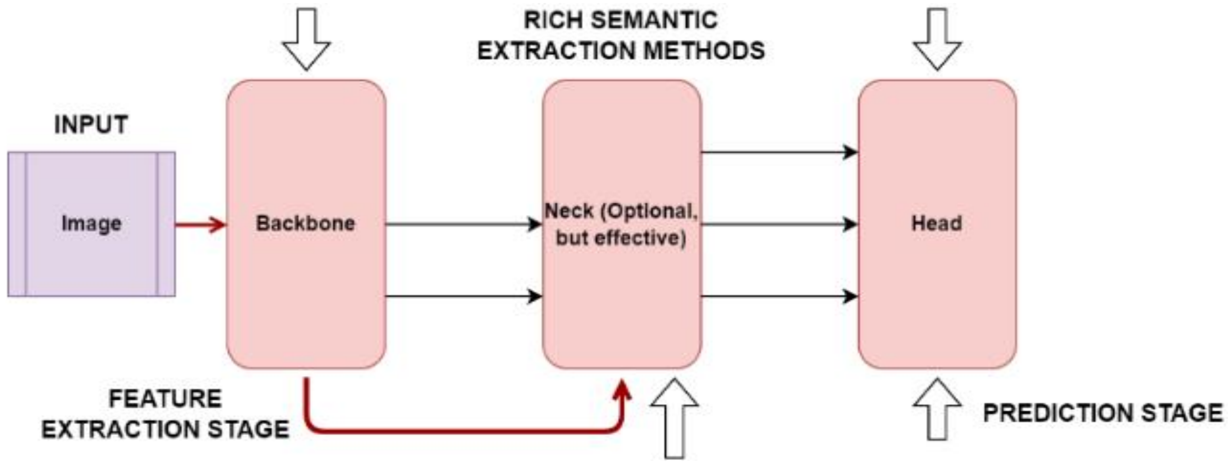

Fig 3.1 YOLOv8 Model Architecture

In this model the detail explanation about the work which has been done, the methods to object recognition, the classification and the detection approach such as Sliding window/Classification Models and Region proposal/Object Detection Models and training mechanism.

System Model Architecture:

The YOLOv8 model are many parts of this model The input layer is training image's they detect to the backbone all images are many of frames in YOLOv8 Each bounding box prediction includes the coordinates of the bounding box, the confidence score, and the class probabilities. Pass the preprocessed images through the YOLOv8 model to perform object detection The Detection Neck and Detection Head together can be called as the Object Detector.

Input (One and Two Stage Detector)

It processes input images in a single pass, efficiently detecting objects in real-time by dividing the all-image grid and predicting bounding boxes and class objects within each grid cell.

Backbone (One and Two Stage Detector)

The backbone in a one-stage detector is a neural network architecture that serves as the foundation for the YOLOv4 object detection model. It extracts features from the input image, while incorporating design to improve accuracy and speed.

Neck (One and Two Stage Detector)

The neck module helps improve object detection accuracy by fusing information from various scales and hierarchies, aiding in the precise localization and classification of objects in a single pass.

Dense Prediction (One and Two Stage Detector)

It processes the entire image for object detection in one pass, optimizing for real-time inference.

The YOLOv4 first proposes regions of interest and then refines these proposals for detection, which often yields higher accuracy at the cost of slightly reduced speed.

Sparse Prediction (Two Stage Detector)

Sparse Prediction in a two-stage detector framework involves an initial region proposal network (RPN) stage, where potential object locations are identified, followed by a refinement stage that focuses on fine-grained object detection within these proposed regions.

This approach enhances accuracy by reducing false positives, albeit at a computational cost compared to one-stage detectors.

3.1 Object Recognition

As the name suggests, it is the process of predicting the real class or category of an image to which it belongs by making probability high only for that particular class. YOLOv8's are used to efficiently perform this process. Many states of the art Classification and Detection algorithms uses YOLOv8 as a backend to perform their tasks. Fig. 1 depicts that classification and localization come under the category of recognition and combined classification and localization is performed to do object detection. Let us have a brief overview of the object classification, localization, and detection.

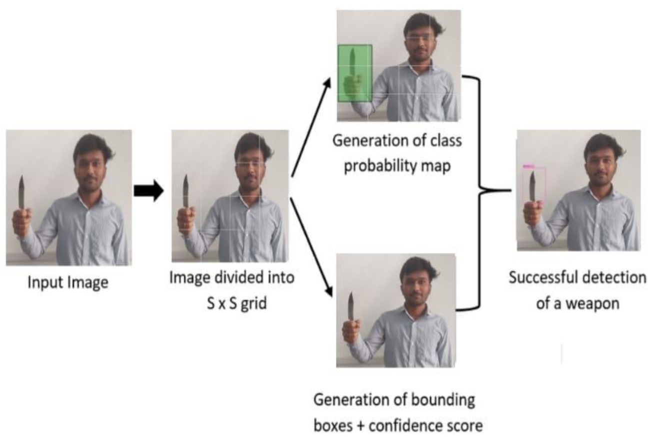

Fig 3.4 Object Recognition to detection

The classification model takes an image and slide the kernel/filter over the whole image to get the feature maps. From the feature extracted, it then predicts the label based on the probability.

3.1.2 Object Localization

This method outputs the actual location of an object in an image by giving the associated height and width along with its coordinates.

3.1.3 Object Detection

his task uses the properties of the aforementioned algorithms. The detection algorithm tells us the bounding box having x and y coordinates with associated width and height along with the class label. Non-max suppression is used to output the box with our desired threshold. This process gives the following results altogether:

- Bounding Box
- Probability

In past object detection was very limited because of less data and low processing power of computers but with the passage of time the computing power of computers increased and world moved from CPUs to Graphic Processing Units (GPU). GPU's were firstly made for increasing the graphic quality of the systems and for gaming but later GPUs were used extensively for deep learning. In ImageNet, competitions started and contained about 1000 classes. This was the evolution of machine learning and deep learning. In the beginning, the models were not very deep, means there were not many layers as they are now in an algorithm. Because of the aforementioned developments, in 2012 A. Hrushevsky presented a model called Alex Net trained on ImageNet and got the first position in that competition. This was the beginning of object detection in deep learning. It gave a way to researchers and then every year the algorithms and models keep on coming. All these algorithms contain layers that work on the principle of the model YOLOv8.

3.2 Classification and Detection Approach

There are many ways to generate region proposals, but the simplest way of generating them is by using the sliding window approach. The sliding window method is slow because filter slides over the entire frame and has limitations, which were tackled by the region proposal approach, so we have the following two approaches used in our work for both classification and detection models are:

- Sliding window/Classification Models
- Region proposal/Object Detection Models

3.2.1 Sliding Window/Classification Models

In the method to the sliding window, a box or window is moved over a picture to select an area and use the object recognition model to identify each frame patch covered by the window. It is an exhaustive search over the whole picture for objects. Not only do we need to search in the picture for all feasible places, we also need to search on distinct scales. This is because models are usually trained on a particular range. The outcomes are in tens of thousands (104) of picture spots being classified. The sliding window method is computationally very costly because of the search with various aspect ratios and especially for each pixel of an image if the stride or step value is less.

Fig 5.1 knife Detection in Real time YOLOv4

In Fig 5.1 shows that the model which is taken for working on to implement weapon detection of the knife with the person hand accurately with good accuracy score of the detection of knife is 84%

Fig 5.2 knife Detection in Real time YOLOv8

In Fig 5.2 shows that the model which is taken for working on to implement detect the suspicious activity of the knife with the person hand accurately with good accuracy score of the detection of knife is 97%

3.3.2 Region Proposal/Object Detection Models

This technique takes an image as the bounding boxes of input and output proposals related to all areas in a picture most probable to be the object. These regional proposals may be noisy; coinciding not containing the object flawlessly, but there is a proposal among these region proposals related to the original target object. As this method takes a picture as the bounding boxes of input and output related to all patches in a picture most probable to be a category, so it proposes a region with the maximum score as the location of an object. Instead of considering all possible regions of the input frame as possibilities, this method uses detection proposal techniques to select regions. Region-based CNNs (R-CNN) was the first detection model to introduce CNNs under this approach. The selective search method of this approach produces 2000 boxes having maximum likelihood. Selective search is a widely used proposal generation method because it is very fast having a good recall value. It is dependent on the hierarchical calculation of desired areas established on the compatibility of color, texture, size, and shape. Yolo series is among the state of the art object detection models. Unlike the other region proposal-based methods it divides the input image into a grid and then simultaneously predicts the probability and bounding boxes for an object with a center falling into a grid cell.

4.1 Training Mechanism

The general methodology used in training and optimization. It starts with defining a problem, finding the required dataset, applying pre-processing methods, and then finally training and evaluating the dataset. If the evaluation is correct then we save those weights as a classifier but if it's incorrect then comes the process of backpropagation algorithm along with the gradient descent algorithm. In backpropagation, weights are optimized by subtracting the partial derivative of cost function J(O) with a multiplier of the learning rate alpha α from the old or previous weight value. Gradient descent is the main weight optimization algorithm. It is used as a base in all optimizers used for the modeling and it helps in converging the model and reaching the minima where we get the best and desired weights values.

4.2 Confusion Object Inclusion

We have formulated the problem to reduce the number of false positives and negatives by adding relevant confusion object. The weapon category includes all the handheld weapons such as, pistol, revolver, shotgun and other than weapon includes the objects that can most be confused with pistol classes e.g., mobile, metal detector, selfie stick, purse, etc. By understanding the differences between classification and detection algorithms.

4.3 Summary

In this chapter YOLOv4 is a one-stage object detection model that builds off of the original YOLO models. Modern object detectors are usually composed of two components backbone and a head. Then also discuss the output how to work that.
The knife object can detect and the show the accuracy is 66.33% and person accuracy can
show the 84.34%.

In this chapter YOLOv8 is a one-stage object detection model that builds off of the original YOLO models. Modern object detectors are usually composed of two components backbone and a head. Then also discuss the output how to work that.
The knife object can detect and the show the accuracy is 96.00 % and person accuracy can
show the 98.33%.

5.1 Conclusion

Implementing real-time suspicious activity detection using a trained YOLOv8 model is a complexbut powerful solution for enhancing security and safety in various environments. In conclusion, using a YOLOv8-trained model for real-time suspicious activity detection is a powerful application of computer vision and deep learning. This approach can enhance security and safety in various environments by identifying and alerting to potential threats. The success of such a system relies on high-quality training data, an effective detection logic, and compliance with legal and ethical considerations. Regular maintenance and updates are essential to keep the system accurate and relevant over time. For both monitoring and control purposes, this work has presenteda novel automatic weapon detection system in real time. This work will indeed help in improvingthe security, law and order situation for the betterment and safety of humanity, especially for the countries who had suffered a lot with these kind of violent activities.
This will bring a positive impact on the economy by attracting investors and tourists, as securityand safety are their primary needs. We have focused on detecting the weapon in live CCTV streams and at the same time reduced the false negatives and positives. To achieve high precisionand recall we constructed a new training database for the real-time scenario, then trained, and evaluated it on the latest state-of-the-art deep learning models using two approaches, i.e. sliding window/classification and region proposal/object detection

6.2 Future Work

As we conclude this chapter, we cast our eyes toward the future a future that brims with possibilities and opportunities.

- Advanced Algorithms: Exploring and developing more sophisticated algorithms, including deep learning models, to enhance detection accuracy and reduce false positives.
- Real-world Integration: Implementing the system in various real-world environments and assessing its performance in diverse contexts, such as urban surveillance, industrial settings, or public transportation.

REFERENCES

1 Muhammad Tahir Bhatti, Muhammad Gufran Khan, (Senior Member, Imee), Masood Aslam, And Muhammad Junaid Fiaz "Weapon Detection in Real-Time CCTV Videos Using Deep Learning" Volume 9, 2021

2 K. Kranthi Kumar 1, B. Hema Kumari, T. Saikumar, U. Sridhar, G. Srinivas, G. Sai Karan Reddy “Suspicious Activity Detection from Video surveillance” International Journal of Research Publication and Reviews Vol 3, Issue 6, pp 2373-2377, June 2022

3 Phalguni Kadam, Shweta Gawande, Akshita Thorat, Rohini Mule” Suspicious Activity Detection using Image Processing” Journal of Science and Technology ISSN: 2456-5660 Volume 06, Special Issue 01, August 2021

4 Mirza Sohail Baig, Dr. Pathan Ahmed Khan “Weapon Detection Using Artificial Intelligence and Deep Learning for Security Applications” International Journal for Advanced Research in Science and Technology