Real Time Theft Detection Using YOLOv5 Object Detection Model

Varun S
Department of Electronic and Communication Engineering
JSS Science and Technology University
Mysuru, India
varunsg118@gmail.com

V M Bhuvanesh
Department of Electronics and Communication Engineering
JSS Science and Technology University
Mysuru, India
bhuvaneshvm47@gmail.com

***Abstract*—Deep learning is a machine learning approach that trains computers to accomplish things that humans do instinctively. It teaches the compute/AI to visualize the world as humans see it and visualize it. This is done possible by mimicking the human brain along with its complex network of neurons. The concept of neural network is widely considered in such cases where the AI designs its own brain/neural network by taking in examples. While approximate predictions can be done with a single layer neural network, additional hidden layers can lead for higher accuracies and also better optimization. Object detection is a subset of deep learning, which involves the detection of objects within frames. Using the above technologies, the concept of the theft detection can be met. Theft of valuable objects is increasing day by day and is an extremely important problem with respect to the global scale and due to the fact that there are only manual systems present to detect theft, many of the thieves/robbers never get caught and repeat their crimes. Thus, the need for an automated system which detects theft with good accuracy and immediately alerts the owners must be developed. You Only Look Once (YOLO) v5 model, an object detection model that is best suited for our case regarding theft detection is considered here. The YOLO v5 object detection model is then trained using our own dataset. After its training its accuracies and prediction of both objects and theft is done in real time along with an alert being issued to concerned authorities and the owner of those objects/items.**

Keywords—Deep learning, Machine Learning, Object detection, YOLO v5 model, Artificial Intelligence, CNNs.

I. INTRODUCTION

Theft is one of the most popular and oldest criminal activities and it is rising every day at an alarming rate. Theft of items which are high in cost is some of the ever-present problem in the globe. People have suffered from fear and loss as the rate of theft has increased. To combat the global increase in theft, there is a need for a theft deterrent system that is easy to operate, generally free of false alarms, and does not require regular human activity to activate and disarm the system.

Theft rate in India is at a staggering amount of 42.9 thefts per 100,000 population with more than 50% of thefts taken place during midnight. Therefore, more emphasis has to be given to combat thefts during nighttime by developing an intelligent and robust system that can work during both day and nighttime.

The concept of having this unique automated system with high rates of accuracy in detecting thefts is realized with the use of deep learning, machine learning and artificial intelligence. The best way to utilize the above major principles in order to detect theft is done with the help of object detection. Object detection is therefore used as a major concept for real time theft detection. Object detection implies the use of computer vision techniques in order to detect instances of object within frames and videos.

Thus, an efficient object detection model must be chosen which would be trained to give high rates of accuracy. YOLO v5 is one such object detection model, it is a computer vision model which uses a convolution neural network in order or detect objects. In order to realize the concept of theft detection, we must understand the working of the YOLO v5 model. The YOLO v5 model is the fifth generation of the YOLO model, which was released on 18 May 2020.

YOLO is a quick, accurate and precise method for detecting objects within images. The algorithm is based on its name, which is looking only once at an image and detecting all the unique objects within those images/videos including their locations. The working of YOLO is quite complex and innovative, it involves by considering an image and then splitting it into a number of girds. YOLO model then investigates each of the grids and gives out the objects that are present along with their location. In order to find the accuracy, we utilize the test data present, each of the objects that are detected are given a confidence score. Considering another important aspect which is performance, YOLO model is far better as compared to that of other mainstream object detection models. By considering object detection within videos using the YOLO V5 model, it achieves a rate of 150 FPS. The mAP another important parameter which is used to give the accuracy of the model is also very high, at almost twice that of other mainstream models. The base YOLO model is trained using the COCO dataset, in which 330K+ images is used with around 200K+ labeled and 80 object categories. This shows the advantages of the YOLO model, it is a fast, accurate and precise model along with the options are easy customization and user-friendly nature.

Following are the remaining chapters: Section II provides with the explanation and insights of the work's prior research. The suggested objectives and approach is thoroughly explained in Section III. With a comparative analysis, Section IV describes the proposed method's result. Section V wraps up the project and discusses potential future studies.

II. RELEATED WORK

The theft detection process makes use of the machine learning-specific YOLO (You Only Look Once) object detection technique. Security and monitoring tools, including facial recognition, anomaly detection, and intrusion detection, are included. Various CCTV cameras use a theft detection system built on machine learning. Cameras for surveillance to identify theft. It is used in many applications to monitor theft in homes, malls, museums, and other places [1].

For the purpose of object detection the high efficiency technique like artificial intelligence and machine learning has been used in order to get better results with the use of OpenCV. Techniques including object detection, recognition, and segmentation are used in image identification. In order to process the data at high the speed and to maintain the quality of the result OpenCV and techniques like machine learning and artificial intelligence has been adapted. It is crucial to use computer vision and machine learning to identify and categorize various activities. For instance, monitoring the direction the driver is looking while operating the car, his speed, the direction he is going, where the people around him are, etc. Thus, the primary objective is to prevent accidents by increasing efficiency [2].

Deep learning provides detailed accomplishment for the detection of object for surveillance system. This technique provides improved quality of image resolution which helps in easy detection of object. Since the movie offers a lot of information with significant disparities and challenges, the job is really difficult. All surveillance systems still require human supervision. Dramatic efficiency benefits result from recent developments in the use of OpenCV, which plays a major role in monitoring of the video input. Here, theft detection using CCTV is combined with theft tracking. Here the movement of the thieve or the burglars is to be detected or identified without use of sensors or by image processing from the CCTV input or the real time video from it. From the real time input from CCTV the analysis of the movement of thief or burglars is being detected and security officers can be alerted about the suspicious person breaking into building, giving them opportunity to stop it [3].

Providing the surety and keep watch on has become very difficult in the present day world. Activities like threat and violence which has made critical necessity for the effective tape supervision systems in order to make sure the owner of household members to know about ongoing thefts occurring on the spot. In the past years due to lack of the information or knowledge and the production of less number or scarcity of the apparatus the rate of theft have been increased over the years. Legacy systems are unable to alert the home's owner of a theft in real time or recognize faces that are partially or completely masked. During the night time, the vision cannot be captured properly, which makes difficult to find invader in the dark. The main problem with this type of installation is that it requires the availability of a homeowner or family member for round-the-clock or manual video surveillance, which is virtually impossible [7].

Here they have proposed a new version of YOLO that is YOLO-FIR for the detection of infrared image along with the YOLOv5 in order to overcome the difficulties faced in identifying infrared images in object detection, it was introduced to overcome the problems like low rate of resolution at long distance also low identification and low energy. In this paper they increased the accuracy of the small objects with this technique by increasing the resolution to detect the object. Through which they have concluded, the efficiency for object detection performance in been increased which the help of YOLO-FIR compared other YOLOs, it can also be used for real time video application for detecting of objects [12].

In this paper, the effectiveness and accuracy of YOLO V3 and V4 in detecting firearms in a real-world setting has been compared. The focus is in the identification of weapons/firearms since it is a critical concern from the perspective of public safety, several AI models and deep learning techniques like CNN (conventional neural network) are integrated [13].

III. METHODLOGY

This chapter contains the objectives of the project, a block diagram of the same, and the methodology. Fig 1 gives the approach towards the YOLO v5 model.

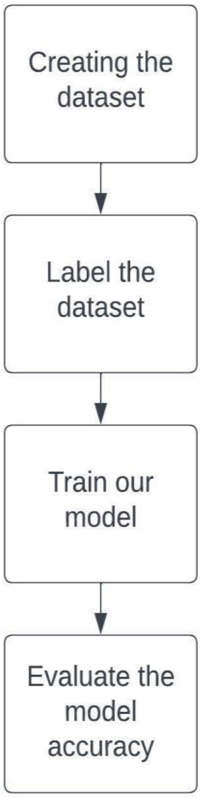

Fig. 1. Block Diagram of the Proposed approach in detail

These are the general objectives that was considered and executed:

- Creating the dataset
- Label our dataset
- Training of the model using our dataset
- Performance evaluation of the model
- Detection of the Object/Person in real time
- Theft detection based on object displacement and issuance of alert

From the above objectives let us deep dive into each of the objectives and try to analyze the main and important parts that were considered in each of them.

‘Creating the dataset’ implies that a dataset is required in order to implement the project. We have considered our very own dataset, containing videos and images. We have complied our own 7 videos depicting theft and normal situations. Each video being approx. 2 minutes in length. From this, we have considered 5 videos for training purpose and 2 videos for final validation and accuracy check. From these 5 videos, we have split the video into image frames. This has allowed us to create a large dataset of around 700+ images. These images are then used to train our model using YOLO v5. The final validation images or the testing images consisted of around 200+ images. This gave us an accurate idea of the results, since having a large dataset fine-tunes during the testing procedure.

The next step in order to achieve our results is to train our AI model in order to detect the objects. Since our main goal is to detect theft, we have considered 3 main objects that will be detected by our YOLO v5 model i.e., Person (Thief), Bag, and a Book. For the AI to differentiate between a person, bag, and book, annotation of the dataset must be done. All the 700+ images in our train set have been annotated. Annotation basically involves adding labels or additional information to the data points in the dataset to help categorize or understand them better. These labels can be in the form of tags, keywords, or any other metadata that provide context or meaning to the data.

After the annotation and training of the model is done, we must analyze the results and our model performance. This is usually done by seeing the prediction results, mAP (Mean Average Precision) etc. This is further explained in the results and discussion’s part. This concludes the first half of the project.

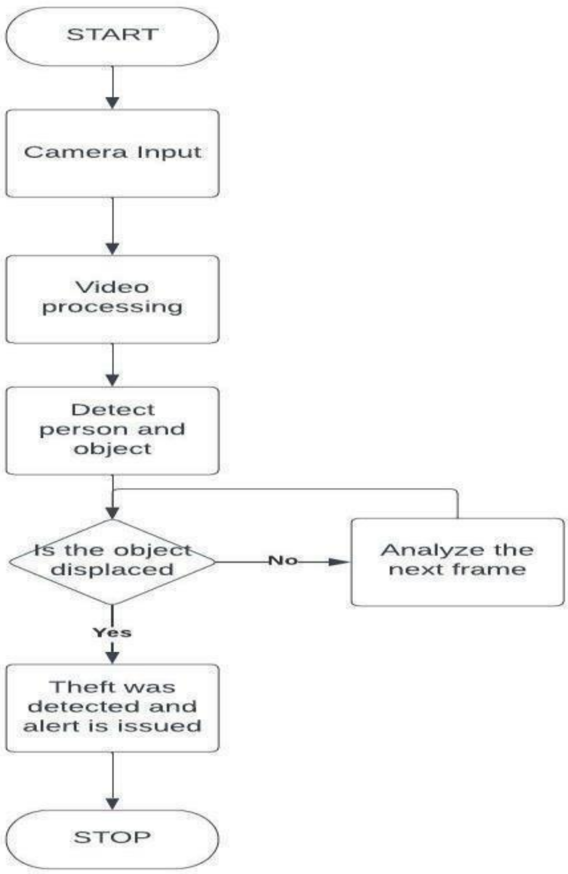

Fig. 2. Block Diagram real time theft detection

In the second half, we must consider how to achieve real time theft detection. This is done using various libraries present in python. The most important one being that of OpenCV, (Open Source Computer Vision) is an open-source library of computer vision and image processing functions. These functions are mainly used for tasks such as video and image analysis, face and object recognition and detection and much more.

The Real Time scenario can be explained using the following flow chart i.e., Fig 2.

Camera input is taken directly from CCTV cameras and fed to a central hub for processing. Each frame of the video is analyzed by our model. From this we get the various objects which are present and their coordinates are noted. Theft is detected when a person disturbs the objects by a large displacement from their initial value of resting position. These is achieved by using the OpenCV library. If the object was indeed displaced, then we arrive at a conclusion that a theft has happened. The police officials and the owner are informed about the same in real-time via email. This was done with the help of a library called smtplib.

Python's smtplib module allows you to send email messages over the Simple Mail Transfer Protocol (SMTP). Python programs may use it to send email to any email address on the internet.

Depending on the interaction between the person(thief) and the object displacement, Theft is detected.

IV. Results and Discussions

This chapter contains results that were obtained after training our own model and analyzing the results obtained. We also deep dive into the results obtained and try to break down the important parameters.

To train the model, around 723 images were used, where each single image was annotated to 3 objects being Person, Bag and Book. The model was trained using YOLOv5 custom dataset option, also the model was trained for 100 epochs in order to get the highest possible accuracy and less false detections. The data used to train our model is defined as the training data and we use this in order or ensure the model predicts accurately. Test data as the name suggests is mainly used to assess the performance of the trained model. From this test data we get the final results regarding important machine learning parameters like accuracy, precession etc.

From Fig 3, we can find the various graphs which are related to the results of the trained model. The model is trained for 100 epochs and this can be seen in the x-axis. The y-axis indicates the various parameters that are relevant to our models accuracy, precision etc. From the figure, Precision being an important parameter is used to find the ratio of positive samples correctly identified to the total number of positive samples identified which includes both positive and negative while the recall is percentage of positive samples that were accurately identified as positive to the total number of positive samples.

Mean average precision metric referred to as mAp, is used to evaluate YOLOv5 object detection models extensively. The mAP assesses a model's accuracy in localizing and recognizing items within an image. As a result, a high mAP suggests that the model is effective at both detecting and categorizing items in the picture, whereas a low mAP implies that the model is less effective. The mAP is typically employed as a measure of the model's performance in YOLOv5 and is frequently presented as a percentage.

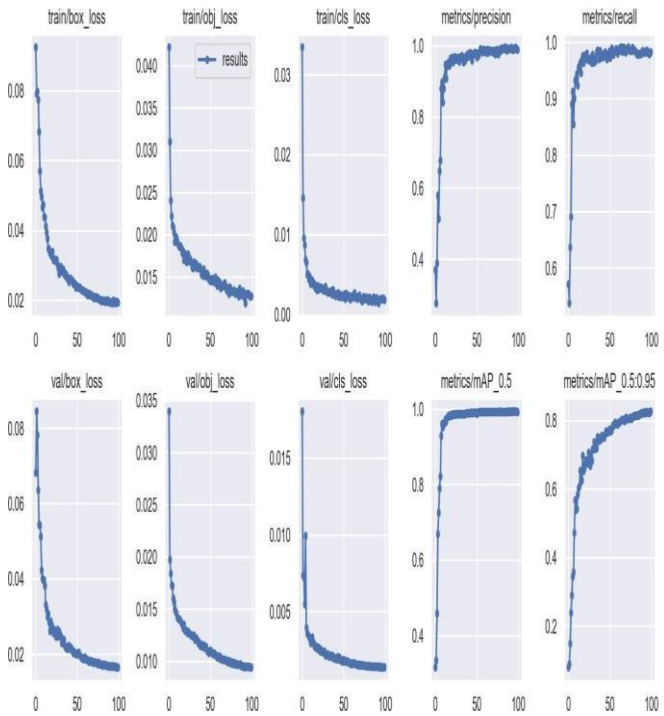

Fig. 3. Model training results

Epoch means the amount of times or passes the training data goes through the machine learning algorithm which causes the model to be trained accurately. Higher the value of Epoch, the more finetuned and precise our model will become. If the dataset is larger, we usually divide it into small batches. Our trained model can be seen to achieve satisfactory results after 50+ epochs. This can be seen through the Precession and recall curves. At 100 epochs we can see that the parameter mAP is almost 90%+ which in turn shows that model is now able to accurately detect the images/video.

The training of the model with 700+ image dataset took almost up to 3.5 hours to reach 100 epochs. This is to conclude that a faster and powerful system will require less time to be taken.

A confusion matrix is an important result that is widely used in model training analysis which implies how well a classification method works. A confusion matrix is used to summarizes a classification algorithm's performance. Here we test data by how much percentage has been identified by itself. Considering Fig 4, we can see the different objects that we have considered and the model prediction results.

From fig 5, we can see our model trying to identify the different objects. This is taken from the final trained weights i.e., after 100 epochs. It also important to note the object detection accuracy and the bounding box accuracy is very high and in most cases it is at 100%. This indicates our model is extremely accurate, which is a must needed during theft detection.

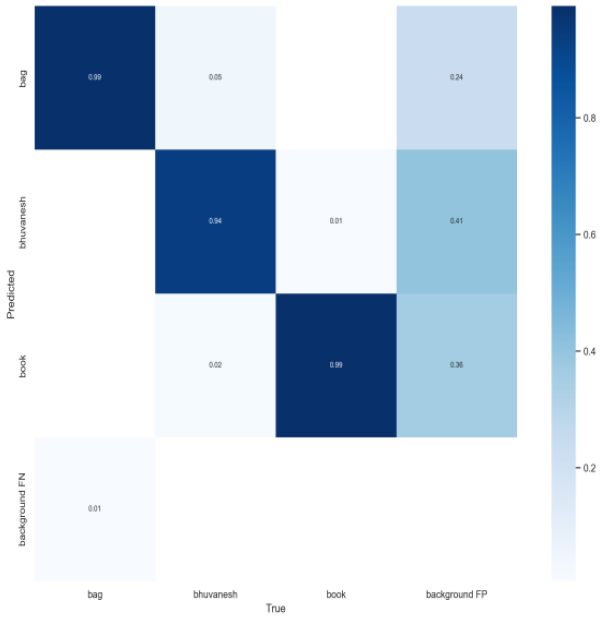

Fig. 4. Confusion matrix

Further, real time results are obtained via OpenCV since the displacement of objects is the main parameter that is considered to detect real time theft. Required code was written to use the functions of OpenCV in order to detect object displacement. Issuing of email alert was done by considering the functions from the library smptlib. A dummy email id was used in order to send the alert to the owner's email id and the official police email id.

Fig. 5. Object Detection

All the initial values that are required for the working is given within the code. This may include the owners email id, Police email id etc.

Taking Fig 6 into consideration, the various objects which are being tracked include the thief, bag, and the book. The above snippet is also taken at the time when the object Book is displaced, the thief has displaced the book from its stationary position which causes a log to be displayed in the python console. The log indicates the following 'Object 1. BOOK is displaced by a large factor'. From our earlier consideration, we can conclude that a theft has taken place and as a result of this necessary alerts are sent to the owner and the police. Fig 6 indicates the E-mail alert sent to the owner. The contents of the E-mail, title, recipients are already given earlier and modification is possible.

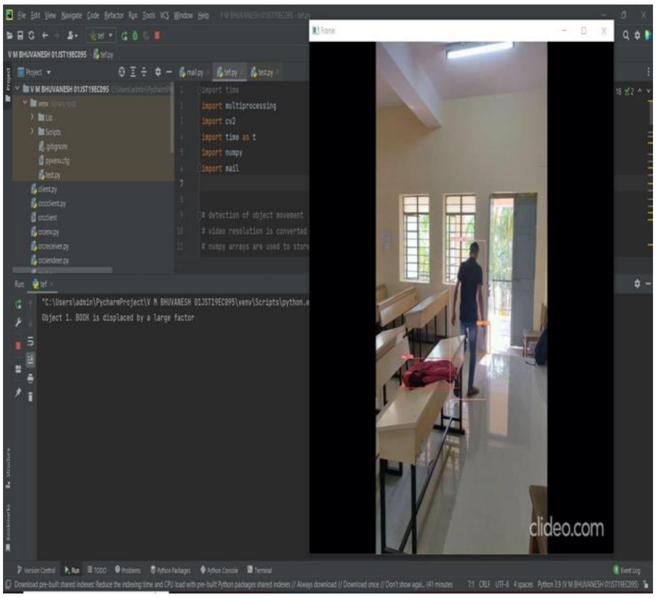

Fig. 6. Real time result

In our case, the id from which the alert is sent is a Gmail id. Other email id from different services can also be used to achieve the same result.

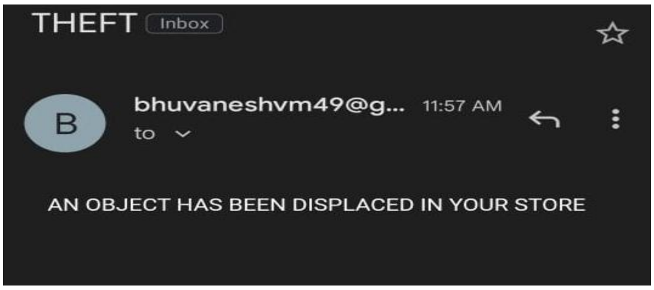

Fig. 7. Email alert

V. Conclusion

This final part includes the work's conclusion and possible future uses for real time theft detection. This project allows us to demonstrate that an effective and accurate real time theft detection model can be created with the help of the YOLOv5 model. The YOLOv5 object detection model has successfully helped in providing an effective and compelling solution to address the concept of theft and enhancing the security and development of security systems. The proposed system was able to accurately detect people and objects in real-time video footage and analyze their movements to identify potential theft incidents. We have achieved high rates of accurate detection in terms of object classification and identification by our trained model along with an effective real time alert system via email. The object classification accuracy was greater than 90% after training for 100 epochs. The model that we had developed and trained was tested on datasets involving real-world and real-time scenarios and achieved high accuracy and detection rates. Furthermore, the system was able to generate alerts and notifications in real-time, enabling quick response to potential theft incidents. Overall, the proposed theft detection system using YOLOv5 shows great promise in enhancing security and reducing loss due to theft in various settings, such as retail stores and public spaces.

The future scope of this project includes more objects being tracked, increasing the dataset so that there are no false instances being reported, implementing daytime theft protection and also alerting via SMS and phone calls. Further updated versions of YOLO model can also be used in the future since it will provide faster training of the model, more accuracy and precision and faster processing. Multi angle camera inputs along with a centralized server can be implemented if provided with adequate resources. The project can also be scaled to so that it can be implemented in both small scale commercial stores upto big supermarkets. It would replace the traditional system of using cameras after theft has taken place.

References

[1] Ayushi Sharma, Jyotsna Pathak, Mukshan Prakash, J N Singh (2021). Object detection using Open CV and Python. IEEE | doi: 10.1109/ICAC3N53548.2021.9725638

[2] Sebin Mangalan, Rutvik Kakadiya, Reuel Lemos, Meghana Pillai (2019). AI Based Automatic Robbery/Theft Detection using Smart Surveillance in Banks. IEEE Xplore ISBN: 978-1-7281-0167-5

[3] Swati Khadse, Babli Nandanwar, Ashwini Kirme, Taiksha Bagde, Vijaya Kamble (2022).Machine Learning Based Theft Detection Using YOLO Object detection.
Doi https://doi.org/10.32628/IJSRST

[4] M. S. Munagekar (2018). Smart Surveillance system for theft detection using image processing. International Research Journal of Engineering and Technology.

[5] K. Piniarski and P. Pawlowski. Efficient pedestrian detection with enhanced object segmentation in far IR night vision in Proc. Signal Process, Algorithms, Archit, Arrangements, Appl. (SPA), Aug 2017 pp. 160_165.

[6] H. Yang and X. A. Wang. Cascade classifier for face detection. Journal of Algorithms & Computational Technology, 2016.

[7] S. Pandya, H. Ghayvat, K. Kotecha, M. H. Yepand P. Gope(2018). Smart Home Anti-Theft System: A Novel Approach for Near Real-Time Monitoring, Smart Home Security and Large Video Data Handling for Wellness Protocol.doi:10.20944/preprints201807.207.v1(2018).

[8] Ren, Z., Yang, S., Zou, F., Yang, F., Luan, C., and Li, K. (2017). A face tracking framework based on convolutional neural networks and Kalman filter. Proceedings of the 8th IEEE International Conference on Software Engineering and Services Science, pp. 410-413.

[9] J. Wu, C. Geyer, and J. Rehg, "Real-time human detection using contour cues," in Robotics and Automation (ICRA), 2011 IEEE International Conference on, May, pp. 860–867.

[10] Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7263-7271). http://openaccess.thecvf.com/menu.py

[11] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934, 2020.

[12] Shasha Li, Yongjun Li, Yao Li, Mengjun Li, and Xiaorong Xu (2021). YOLO-FIRI: Improved YOLOv5 for Infrared Image Object Detection. Digital Object Identifier 10.1109/ACCESS.2021.3120870.

[13] Hashmi, Tufail Sajjad Shah, et al. "Application of deep learning for weapons detection in surveillance videos." 2021 international conference on digital futures and transformative technologies (ICoDT2). IEEE, 2021.