iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VIII **Month of publication:** August 2021

DOI: https://doi.org/10.22214/ijraset.2021.37512

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

Prediction of Crime Scene Objects Using Deep Learning

Vibhavari B Rao[1], Aditya K[2]
[1, 2] *Department of Information Science and Engineering, MS Ramaiah Institute Of Technology, Bangalore*

Abstract: ***The crime rates today can inevitably put a civilian's life in danger. While consistent efforts are being made to alleviate crime, there is also a dire need to create a smart and proactive surveillance system. Our project implements a smart surveillance system that would alert the authorities in real-time when a crime is being committed. During armed robberies and hostage situations, most often, the police cannot reach the place on time to prevent it from happening, owing to the lag in communication between the informants of the crime scene and the police. We propose an object detection model that implements deep learning algorithms to detect objects of violence such as pistols, knives, rifles from video surveillance footage, and in turn send real-time alerts to the authorities. There are a number of object detection algorithms being developed, each being evaluated under the performance metric mAP. On implementing Faster R-CNN with ResNet 101 architecture we found the mAP score to be about 91%. However, the downside to this is the excessive training and inferencing time it incurs. On the other hand, YOLOv5 architecture resulted in a model that performed very well in terms of speed. Its training speed was found to be 0.012 s / image during training but naturally, the accuracy was not as high as Faster R-CNN. With good computer architecture, it can run at about 40 fps. Thus, there is a tradeoff between speed and accuracy and it's important to strike a balance. We use transfer learning to improve accuracy by training the model on our custom dataset. This project can be deployed on any generic CCTV camera by setting up a live RTSP (real-time streaming protocol) and streaming the footage on a laptop or desktop where the deep learning model is being run.***

Keywords: object detection, smart surveillance system, Faster-RCNN, YOLOv5, ResNet 101, transfer learning

I. INTRODUCTION

Crime is defined as any illegal activity that is undesirable and punishable by law. Among the many types of crime, including cyber crime, burglary, fraud, domestic abuse and victimless crime, the most dangerous is assault which can pose a threat to a civilian's life. Assault can be carried out by use of weapons such as knives, pistols, rifles and other objects of violence. It is important that measures be taken to curb such illegal activities from taking place, especially in public places. Most often, during armed robberies and hostage situations, it is observed that the police cannot reach on time to prevent a crime or robbery from happening. This is because people inside the room most often cannot contact the police. The only way the police would know about this is if bystanders notice something happening, and inform the police with real time updates. Thus, if the law enforcement authorities reached the crime scene after the crime took place, it would slow down the process of finding the criminals and rescue the victims of the crime scene. It is vital that the law enforcement authorities be notified immediately when the crime is taking place. Although surveillance cameras have been put up in most places, real time manual monitoring of the cameras is a cumbersome process. It requires undivided and immense attention. Keeping these existing flaws in mind, we propose an object detection model using the salient features of computer vision which serves as an efficient alert system to notify the police when a crime is happening. Computer vision imparts artificial intelligence to computer systems, enabling them to gather actionable insights from digital documents such as images and videos. One of the widely applied computer vision techniques is object detection which performs both localisation and classification of objects in an image or a video. Our smart surveillance camera uses object detection algorithms to detect weapons of violence such as knives, pistols and rifles and consequently sends alerts to the authorities without any delay. Object detection models typically predict the presence of an object in an image and build a bounding box around it to locate the object in the frame. It returns the coordinates of the upper left corner of the box, the length of the box along the x-axis, and the length of the box along the y-axis. Using these values, we can plot a box around the object to check correctness on test images. There are many factors to take into consideration before choosing an object detection algorithm. Generally, we have to choose between speed and accuracy and strike a balance that works best for the application at hand. It's essential to have great accuracy as we cannot risk a weapon being undetected. This is a huge challenge. However it's a relatively smaller problem if wrong alerts are sent. In other words, false negatives are a far greater problem than false positives.

©IJRASET: All Rights are Reserved

In our model, we aim to optimize the number of true positives and reduce the number of false negatives. Speed is also of the essence as our model should be able to spot the weapons even if it's only visible for a few frames. Since the angle of the weapon may keep changing along with the perpetrator, it's important to have an extensive training data set and apply augmentations to cover all possible angles of a weapon. It was also challenging to build a balanced dataset. We applied and analysed many different object detection algorithms and have attached the results. The best model for our use case was Faster R-CNN. This ensures excellent accuracy and decent training and detection time. Our project thus seeks a solution to detect weapons in real-time, ensuring no lag. It also ensures the least damage by way of quick response. The alerts have to be fast and accurate, minimizing false negatives. This solution can be deployed in surveillance cameras in high security public spaces where weapons are prohibited inside the campus, such as hospitals, banks, educational institutions, train stations and airports.

II. LITERATURE SURVEY

1) *A Robust Object Detector: Application To Detection Of Visual Knives[1]:* The paper proposes a three stage model. Stage one is where the foreground segmentation and identification is performed. In stage two, the model will localize the position of the knife in hand using keypoint detectors. Finally, in stage three, the model utilizes the image data representing the localization centres by feeding it into a classifier that performs the classification task of deciding whether the knife is actually present in the localized image or not. The whole system is implemented using a client and server architecture. The front-end(client) handles segmentation of the images whereas the server will handle the localization and classification. It achieves parallelism by doing most of the calculations in the cloud.
2) *Crime Intention Detection System Using Deep Learning[2]:* The paper proposes a system to detect a gun or knife in the hand of the criminal being pointed at the victim. The system makes use of deep learning models such as VGGNet 19 and GoogleNet. These models are pre-trained on more than a million images, the paper aims to utilise these models to detect violent objects in the image in question with minimum errors.
3) *Uniform and Variational Deep Learning for RGB-D Object Recognition and Person Re-Identification[3]:* The paper proposes an RGB-D object recognition and person re-identification system using **uniform and variational deep learning (UVDL)** method. This system extracts more reliable anthropometric and geometric information, which are robust to different viewpoints, to recognise objects and persons. The existing methods make use of only the visual appearance from RGB images. Two deep convolutional neural networks are used to extract the appearance and the depth features from RGB-D images. A uniform and variational multimodal auto-encoder is designed at the top layer of the neural network to facilitate obtaining a uniform latent variable by projecting it into a shared space. This space contains the entire information of RGB-D images and has both small and large inter-class variation simultaneously. This method allows the model to exploit the relationship between the appearance and depth features. In the end, the system optimizes the two deep convolutional neural networks and the auto-encoder together to bring down the reconstruction error and the discriminative loss to as low as possible.
4) *A Study on CNN Transfer Learning for Image Classification[4]:* The paper proposes the study of a Convolutional Neural Network (CNN) architecture, object detection model, to establish whether the accuracy would improve and if the mode could give better results if new image datasets were used via Transfer Learning, a machine learning technique. Transfer learning approach allows a model to be trained for a particular task and then the same model will be reapplied on another related task, which helps in optimising a scenario by exploiting the model that has been trained on another scenario.
5) *Moving Object Detection Using Deep Learning[5]:* This paper presents a methodology that implements coarse-grained detection as well as fine-grained detection. This paper addresses the problem of noise-induced object fracture during the coarse-grained detection process through the solution of developing a low-complexity connected region detection algorithm to extract moving regions. Deep Convolutional Neural Networks are potentially used in the detection of more precise coordinates and the category of object identification. Initially while coarse-grained detection is used in detecting moving regions, the connected regions detection is performed later. The coordinates of each object are corrected during fine-grained detection, and the class of the object is obtained.

III. PROPOSED MODEL

We explore and analyse two different systems for detection - YOLOv5 and Faster R-CNN. The security cameras deployed in banks and other public places are strategically placed so that they have a good view of the entire room, with minimal blind spots. We propose to stream this video footage in real-time and run the deep learning algorithm on it frame by frame. The most important aspect of this project to explore is the object detection algorithms. We are exploring YOLOv5 for its extraordinary speeds. It trains very fast and it can perform object detection at the rate of 0.02 seconds per frame. But with these speeds, we have to expect less than perfect accuracy.

©IJRASET: All Rights are Reserved

Another system we explore is Faster R-CNN with ResNet architecture that is pre-trained on the ImageNet dataset. This proves to have better accuracy than YOLOv5 with comparable training and testing times. Proposed model architecture constituting YOLOv5 and Faster R-CNN is illustrated as follows

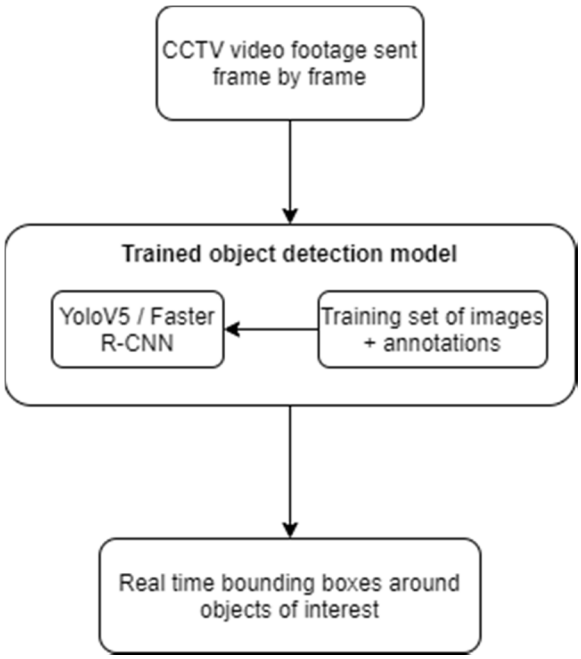


Fig 1 Model overview

A. YOLOv5

YOLOv5 is a single-stage object detector having three important components:

1) Model Backbone
2) Model Neck
3) Model Head

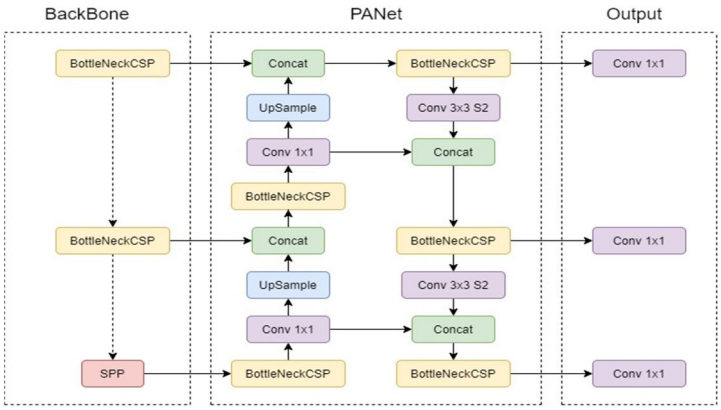


Fig 2 YOLOv5 architecture

Figure 2, shown above, describes the YOLOv5 architecture. The model backbone in the YOLOv5 architecture is used for feature extraction, the first step in object detection. It extracts detailed and important features from the input image. Cross Stage Partial Networks (CSP) are used as a backbone in YOLOv5 to extract rich and informative features from the input image. The next part, model neck, is used to generate feature pyramids (FP). Feature Pyramids are necessary as they assist models to generalize well on object scaling. This means that the model will be able to identify the same object with different scales and sizes. This should help the model perform well on unseen data. The next part, the model head, is used to perform the final detection by the model. Anchor boxes are applied on features which in turn are used to generate output vectors with their associated class probabilities, bounding boxes and objectness scores.

©IJRASET: All Rights are Reserved

B. Faster R-CNN with ResNet 101

The main breakthrough of the faster r-cnn model is the replacement of the slow selective search algorithm implemented in the fast r-cnn models. It introduces the concept of region proposal network (RPN). In the previous feature map, RPN uses a 3x3 sliding window that moves across the feature map. For each sliding window location, *k* number of fixed ratio anchor boxes are generated. For each of these boxes, the 4 coordinate points are produced along with a softmax probability score of this box containing an object. Finally, if an anchor box has a probability score greater than a specified threshold, its corresponding coordinates are passed as a region proposal to a Fast R-CNN network. This comprises a Faster R-CNN model.

In our model, we use the ResNet architecture as the base Convolutional Neural Network. The advantage of using ResNet over VGG is that the former is bigger. It is observed that more layers in the CNN enhances its capacity to learn better.

The ResNet architecture we've implemented is ResNet 101. ResNet-101 is a convolutional neural network that is 101 layers deep and pretrained on over a million ImageNet images. We use ResNet as a pretrained model and repurpose it in our project. We import the weights from ResNet and use a transfer learning approach to fine tune the model. Repurposing a pretrained model for feature extraction holds a lot of importance when we are using a small dataset bearing high similarity to the original dataset. Thus, we fine tune the softmax layer that is used to give respective probabilities of a pistol and other weapon classes. This dramatically enhances the time and performance of our model.

A fixed feature extraction mechanism is the name for this method. We solely retrain the new output layer we created, leaving the weights of the other layers alone. Figure 2, shows the proposed Faster R-CNN architecture with fine tuning of the softmax layer.

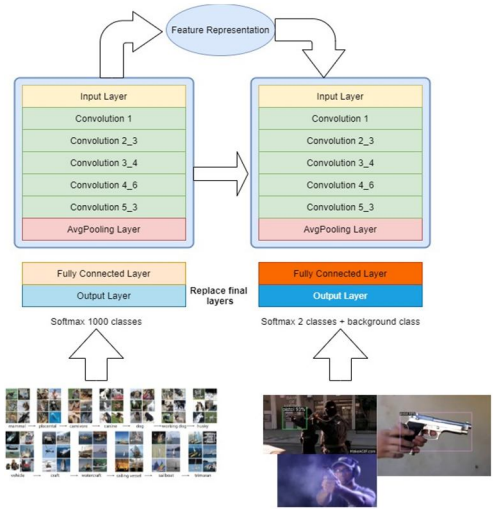


Fig 3 Architecture of Faster R-CNN with fine tuning

C. Sequence diagram

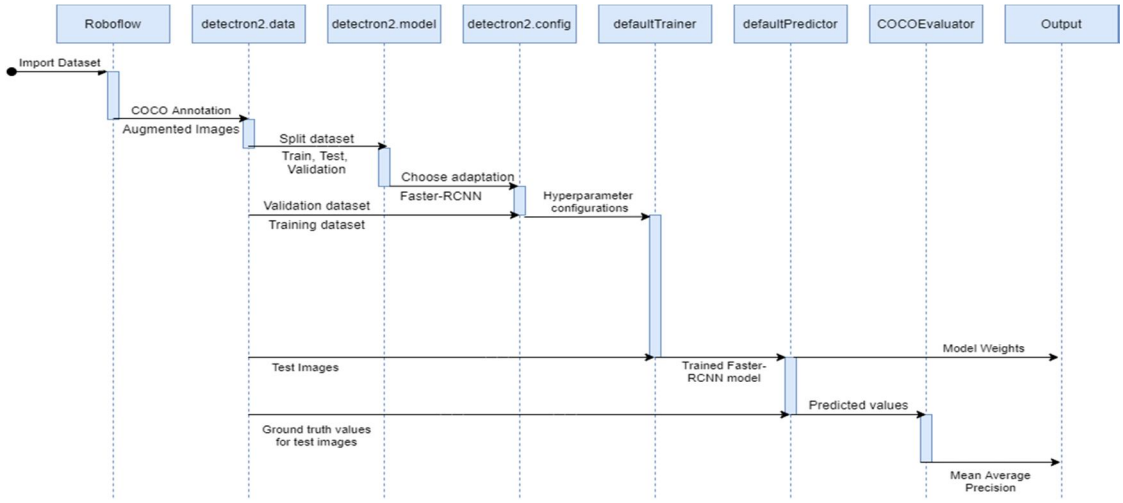


Fig 4 Sequence diagram implementing Faster R-CNN

©IJRASET: All Rights are Reserved

The first step in this project is to import the dataset - images along with bounding boxes, and the respective class of the image. The bounding box is represented by four values - the x and y coordinates present in the top left corner of the bounding box, the height and width of the box. These are imported into Roboflow. Roboflow is a platform that allows us to use this dataset in the COCO annotated format. We also augment the images in cases where the training images are less in number. Augmentation allows the model to train on the same but augmented images, and this gives improved results. The next step is to split the dataset into train, test and validation data. These are used in different stages of the project. Next, we use the detectron2 to choose our model. In this project we have used the Faster R-CNN model. This is followed by configuring the hyper parameters for our model. The model is now ready to be trained. Detectron2 is used as a platform where state of the art computer vision models can be deployed.We use the default trainer from detectron2 trainer which then trains the model using the train and validation images. The trained model is then passed to the default predictor along with the test images to obtain the predictions made by our model. We can then view the predictions by drawing the bounding boxes on the test images along with their confidence value using the Visualizer class of the detectron library. The predicted values are compared with the true values by the COCO Evaluator. The metric used is called mean average precision. Throughout this process, the hyperparameters are adjusted accordingly to get the best accuracy.

D. Class Diagram

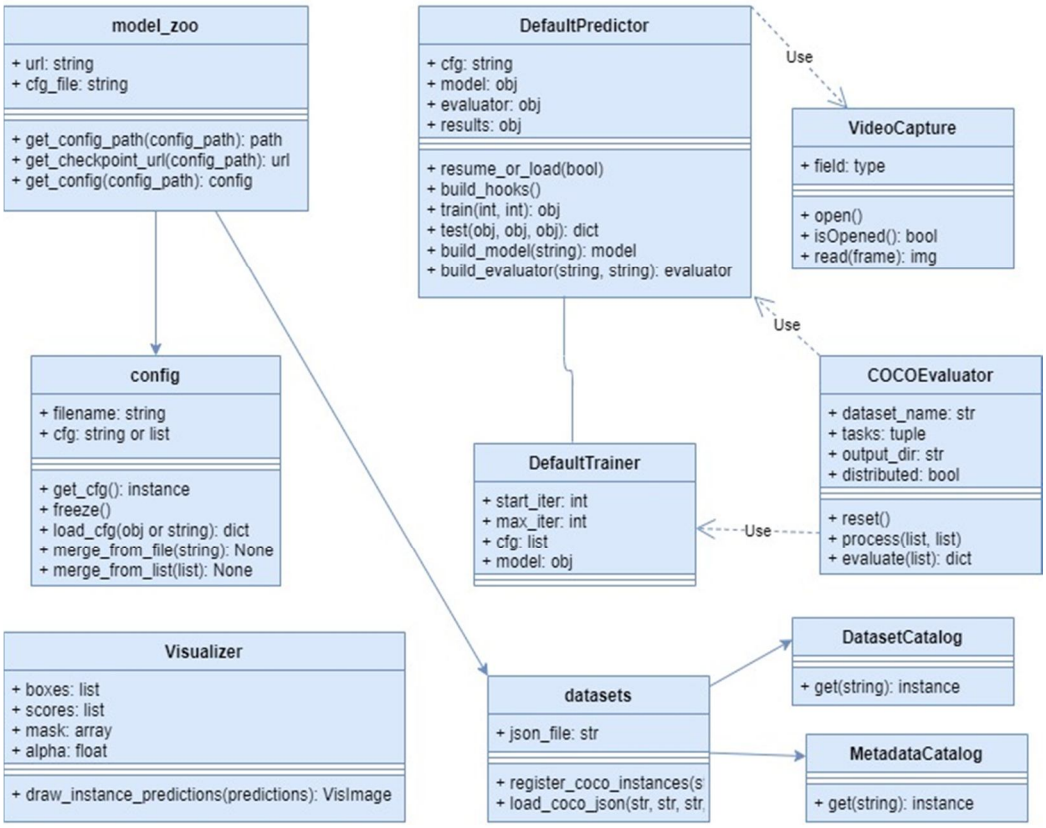


Fig 5 Class diagram

IV. EXPERIMENTAL SETUP AND RESULT ANALYSIS

A. Dataset and Training

The training, testing and validation dataset for the YOLO v5 and Faster R-CNN models were sourced from the public research weapon dataset prepared by the University of Granada research group. The created datasets were annotated and augmented as part of the preprocessing. The dataset was split as 70% for training, 20% for testing and 10% for validation. Out of 737 images 516 images were the train set, 147 images constituted the test set while 74 images constituted the validation set. The model was trained to detect pistols, rifles and knives. The training epochs was set to 3000

©IJRASET: All Rights are Reserved

B. Preprocessing and Augmentations

It is observed that a model trained on a well augmented balanced dataset performs better than that trained on a dataset that is not augmented. Image data augmentation can be considered as a unique approach to transform a small dataset into a large one that contains different versions and variations of the images in the dataset. This will not only enable the deep learning model to learn the dataset better but also enhances its generalisation capability since it is exposed to a variety of data . The augmentations we explored in this project that made our model more skilled are:

1) Auto-Orient: Applied
2) Resize: Stretch to 416x416
3) Grayscale: Applied
4) Auto-Adjust Contrast: Using Adaptive Equalization
5) Outputs per training example: 3
6) Rotation: Between -15° and +15°
7) Shear: ±15° Horizontal, ±15° Vertical



Fig 6 Augmented train set

C. True Positive Results

After the Faster-RCNN model was inferenced on a video footage, the true positive results are shown below. It is evident that the accuracy on small objects and blurry objects is exceptionally good. This is because the mAP scores for Faster R-CNN is found to be 77% for small objects.



Figure 7 True positives from different set of test images

©IJRASET: All Rights are Reserved

D. False Positive Results



Figure 8 Test images that were predicted wrongly to be true

In Fig 8, we see that a watch is being wrongly predicted as a pistol. However, the true positive object does not go undetected in the picture.

E. True Negative



Figure 9 Images that were predicted correctly to be negative

F. Evaluation Metrics

1) *AP (Average Precision):* Common object detectors Faster R-CNN, YOLOv5 and SSD use Average precision as a popular metric in measuring the accuracy. It calculates the average precision for recall values ranging from 0 to 1.
2) *IoU (Intersection over union):* The degree of overlap between bound figures is determined by IOU. The focus is mainly on the extent of overlap between the predicted bounding box and the ground truth bounding box. It is important to predefine a threshold criterion of IOU which is used to indicate if a prediction is a true positive or a false positive.
3) *COCO mAP:* The calculation in COCO mAP uses a 101-point interpolated AP definition. AP is computed by taking the average of numerous IoU for COCO dataset. The average AP for IoU with a step size of 0.05 is AP@[.5:.95]. All categories are averaged -AP@[.50:.05:.95] starts at 0.5 and goes up to 0.95 with a 0.05 step size.

G. Evaluation of Faster R-CNN

```
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.587
 Average Precision  (AP) @[ IoU=0.50      | area=   all | maxDets=100 ] = 0.912
 Average Precision  (AP) @[ IoU=0.75      | area=   all | maxDets=100 ] = 0.552
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.744
 Average Precision  (AP) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.471
 Average Precision  (AP) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.608
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=  1 ] = 0.606
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets= 10 ] = 0.694
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=   all | maxDets=100 ] = 0.694
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= small | maxDets=100 ] = 0.767
 Average Recall     (AR) @[ IoU=0.50:0.95 | area=medium | maxDets=100 ] = 0.591
 Average Recall     (AR) @[ IoU=0.50:0.95 | area= large | maxDets=100 ] = 0.687
```

(a)

AP	AP50	AP75	APs	APm	APl
58.704	91.223	55.190	74.422	47.136	60.846

(b)

Figure 10 (a) and (b) Evaluation results on bbox represented by average precision

©IJRASET: All Rights are Reserved

As seen in Figure 11, On AP50 gives an accuracy of 91.223 with IOU threshold 50%.This is considered a good value considering we are giving more importance to the correct object being detected rather than the bounding box overlapping degree. This basically means that the model can detect an object with 91.223 % accuracy by drawing a bounding box which overlaps with 50% of the ground truth box. The average inference time is 0.192346 s / img.

H. Evaluation of YOLOv5

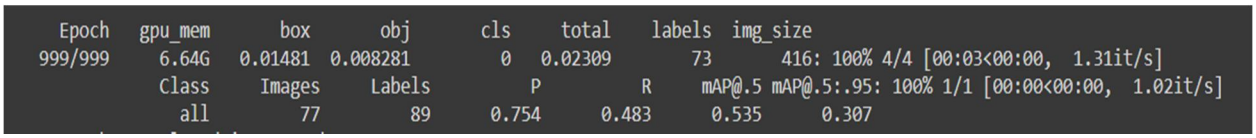


Figure 11 Results obtained using YOLOv5

On training the model for 1000 epochs, we observe from the figure above that Yolov5's accuracy is approximated to be 53.5% at AP50 which is not a great accuracy compared to Faster R-CNN which had about 91.223%. However the inference time is found to be 0.012 s / img which makes it very fast for detecting objects and sending alerts.

Model	Accuracy	Inference speed
FasterRCNN	91.2%	0.192 s/img
YOLOv5	53.5%	0.012 s/img

Table 1 Performance analysis of Faster R-CNN and YOLOv5

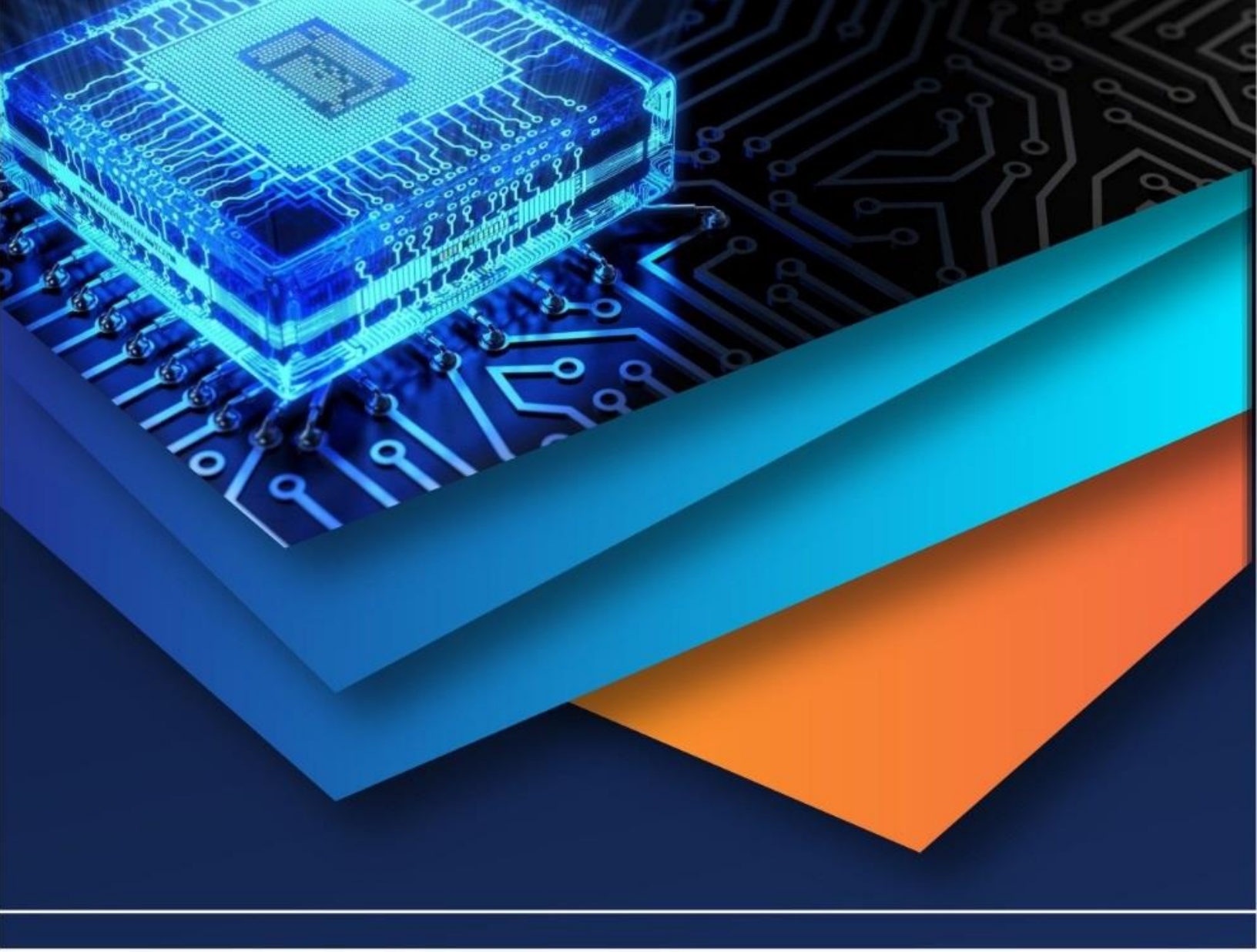
V. CONCLUSION

For the implementation of this project we used the existing technology and libraries available in the world of Image processing and Artificial Intelligence. The suitable models we considered after initial analysis were YOLOv5 and Faster R-CNN with ResNet 101 architecture. An appropriate dataset was compiled using augmentations which eventually gave us a balanced dataset. Since we had two models available, we were able to compare the performance of these models on our dataset. Faster R-CNN gave the best prediction accuracy between the two models, although it was considerably slower than YOLOv5. The Faster R-CNN model included a fine tuning approach for the best optimization of the model's performance. The lower computation time of YOLOv5 might be desirable for real-time use cases. It is also to be noted that the accuracy we received through Faster R-CNN is not a fixed accuracy that will always be obtained. It primarily depends on the training set size and the variety of training images that is fed to the network. The future, evolved implementation of the project would be to deploy an end-to-end model where the surveillance camera directly sends the footage, in real-time to a cloud database. The software implementing the object detection model will extract the video from the database, read each frame of the video and if the positive class (one of the crime scene objects-guns,pistols) is detected, it would create an alert and send it to the local police authorities and notify the owner of the property.

REFERENCES

[1] A ROBUST OBJECT DETECTOR: APPLICATION TO DETECTION OF VISUAL KNIVES, Himanshu Buckchash, Balasubramanian Raman, Department of Computer Science and Engineering, Indian Institute of Technology Roorkee.
[2] CRIME INTENTION DETECTION SYSTEM USING DEEP LEARNING, Umadevi V Navalgund, Priyadharshini. K, Computer Science and Engineering, KLE Technological University, Hubballi, India.
[3] Uniform and Variational Deep Learning for RGB-D Object Recognition and Person Re-Identification, Liangliang Ren, Jiwen Lu Jie Zhou, Senior Member, IEEE and Jianjiang Feng, Member, IEEE.
[4] A Study on CNN Transfer Learning for Image Classification, Mahbub Hussain, Jordan J. Bird, and Diego R. Faria, School of Engineering and Applied Science Aston University, Birmingham, B4 7ET, UK.
[5] Moving Object Detection Using Deep Learning, Haidi Zhu, Xin Yan, Hongying Tang, Yuchao Chang, Baoqing Li and Xiaobing Yuan.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)