REAL-WORLD ANOMALY DETECTION USING OBJECT RECOGNITION

Ms. S. Visnu Dharsini, Priyesh Gangrade, Smit Modi, Sarthak Bhatnagar, Ananthu S Nair

Department of Computer Science

SRM Institute of Science and Technology, Chennai, Tamil Nadu – IN 600089

Abstract

The most distinguished application of activity recognition is anomaly detection. Providing security to a private could be a major concern of any society these days thanks to the perpetually increasing actions inflicting threats, ranging from deliberate violence to associate injury caused by associate accident, security system isn't enough because it needs an individual's being to perpetually stay alert and monitor the cameras, that is kind of inefficient, human activity recognition is a vital issue within the pattern recognition field, with applications starting from remote police work to the classification of economic video content. This necessitates the necessity to develop an automatic security system that identifies abnormal activities in the given period and brings immediate facilitate to the victims. This paper aims to supply an associate correct abnormal act recognition system that works in a real time using YOLO-V3. Moreover, smoothness constraints within the loss function to better anomaly detection, anomaly detection considering all anomalies in one cluster and everyone traditional activity in another cluster.

Keyword: - Machine Leaning, activity recognition, you only look once (YoLo-V3), computer vison.

I. Introduction

Human activity recognition is more and more being employed publicly places e.g., streets, intersections, banks, searching malls, etc., and anomaly detection in security systems are one in every one of them to extend public safety. The associate activity recognition system is anticipated to spot the fundamental everyday activities performed by an individual's being. It is difficult to attain high accuracy for recognition of those activities due to the quality and variety of human activities. It's associated with current and open analysis topics in computer vision that features instances of behavioural biometrics, video analysis, animation, and synthesis. Human actions involving gestures and motions of the physical structure area unit understood with the assistance of sensors. The system understands an individual's activity by distinguishing movement and recognizing the pattern of that movement. This is often followed by the development of a complex conceptual abstract model which will acknowledge and classify all human activities. Moreover, activity pattern discovery doesn't need predefined models because it uses solely some low-level device knowledge that area unit captured to seek out the unknown patterns. although the 2 techniques distinction one another, they need a customary target of achieving higher recognition performance for action recognition systems. These techniques are confirmative of every alternative, and that they mix to reinforce the performance by victimization the invention of the activity pattern to outline recognized activity.

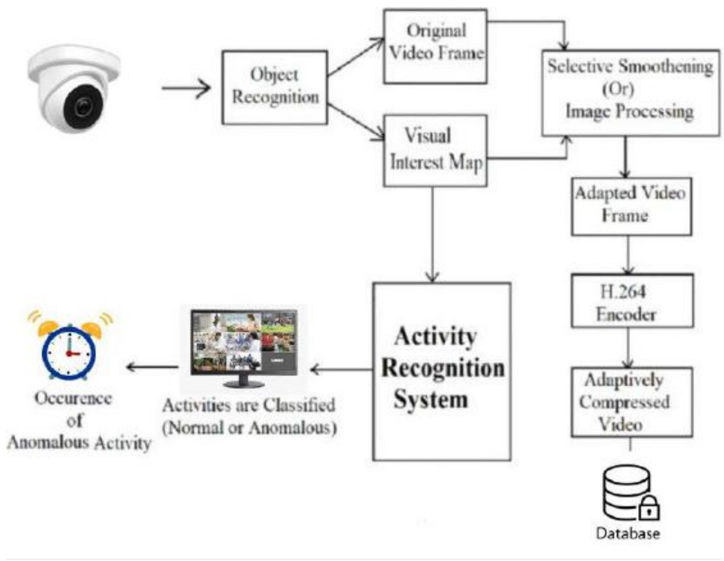

Fig.1 The high-level design of the anomalous human activity recognition system

II. Adaptive Video Compression

There are many approaches that area unit followed for video compression, out of that some ways solely will cut back the scale of the video and thence compress the total video. there'll be a loss of some helpful info once such approaches area unit opted. To avoid loss of serious information from the video, it's thence appropriate to adopt a lot of cheap compression techniques that is that the adaptative video compression compresses insignificant elements of the video, thereby retentive elements that contain the target objects of interest. The video compression commonplace will be controlled. Here, vital and semantically meaningful parts of the video area unit are encoded with higher exactness. this can be achieved by combining low-level options of the video with low machine prices. The insignificant background components are allotted with lower bits within the transmitted illustration. the idea of this methodology is to by selection smoothen individual frames as shown in Fig. 1. Selective smoothening so helps in protective image options of objects that seem to be semantically fascinating, a personality's motion within the case of human action recognition. Every distinct frame in a video is adaptively compressed using the low-level features, eliminating the uninteresting elements, and this result will be used because of the pre-processing state by inserting it into the video coding pipeline. Thus, the adaptative compression technique 1st identifies the target objects then performs selective smoothing wherever the insignificant elements of the image area unit are ironed. The insignificant elements in surveillance work videos area unit largely the backgrounds as they invariably stay constant. Therefore, once some motion is known, it's thought about for recognition and classification. For each movement of a human being, respective humans are captured as the target object. That object is then tracked to acknowledge the activity it performs.

III. Anomaly Detection

The activities of a human being may be broadly classified into normal activities or abnormal activities. a human being’s deviation from normal behavior inflicting hurt to the encircling or to himself is categorized as an abnormal activity. Such behavior is sometimes the result of some mental discomfort. The intensive analysis of human action recognition and its applications have thrown attained anomaly detection. the present approaches for abnormal human action recognition are made supported by the kind and speed of object movements alongside however the objects of interest interact with one another. A study of the various on the market approaches known the drawbacks within the existing activity recognition system, and this stands as a significant motivation to implement an automatic anomaly detection system that works in real-time. OpenCV, Scikit-learn is a number of the libraries alongside YOLOv3 (You Only Look Once, Version 3) that are used for the implementation of the period of time anomaly detection system.

IV. LITERATURE SURVEY

Real-world Anomaly Detection in Surveillance Videos filters by: - Waqas Sultani, Chen Chen, Mubarak Shah

This paper was 14 Feb 2019.They planned a technique to find out anomalies by exploiting each traditional and abnormal video. They additionally think about traditional and abnormal videos as baggage and video segments as instances in multiple instance learning (MIL), and mechanically learn a deep anomaly ranking model that predicts high anomaly scores for abnormal video segments. The algorithmic rule and technology utilized by them are Multiple Instance Learning.

Anomaly Detection in Video Sequence with Appearance-Motion Correspondence filters by Trong Nguyen Nguyen, Jean Meunier

This paper was revised on 17 Aug 2019.The given paper proposed a deep convolutional neural network that addresses the downside of anomaly detection in surveillance by learning a correspondence between common object appearances (e.g., pedestrian, tree, etc.) and their associated emotions. The algorithmic rule and technology utilized by them are a convolutional neural network (CNN).

Deep anomaly detection through visual attention in surveillance videos filters by: - Nasaruddin Nasaruddin, Kahlil Muchtar, Afdhal Afdhal & Alvin Prayuda Juniarta Dwiyantoro.

This paper was published on 16 October 2020.This paper used a technique for learning anomaly behavior within the video by finding an attention region from spatiotemporal information, in distinction to the full-frame learning. They additionally used a similar algorithmic rule and technology utilized by them is a convolutional neural network (CNN).

Video anomaly detection method based on future frame prediction and attention mechanism filters by: Chenxu Wang, Yanxin Yao , Han Yao.

This paper was published on 17 March 2021.The given paper proposes a video anomaly detection algorithm based on the future frame prediction using Generative Adversarial Network (GAN) and attention mechanism. Limitations and technical gap of this paper is that this paper needs to provide different types of data continuously to check if it works accurately or not.

Activity recognition and anomaly detection in smart homes filters by: Labiba Gillani Fahad, Syed Fahad Tahir.

This paper was revised on 12 November 2020.They planned the approach that acknowledges the activities performed in a very sensible home and separates the conventional from the abnormal activities. The algorithmic rule and technology utilized by them is water autoencoder.

Contextual Multi-Scale Region Convolutional 3D Network for Activity Detection by: Yancheng Bai, Huijuan Xu, Kate Saenko, Bernard Ghanem.

This paper was published on 28 January 2018. They proposed the contextual multi-scale region convolutional 3D network (CMSRC3D) for activity detection. Limitations and technical gap from this paper is that it complicates the vision-based detection systems and the issue of objects appearing in images with different pixel-wise.

Anomalies Detection and Tracking Using Siamese Neural Networks by: Yan, Weiqi

This paper was published on 15 May 2020. The proposed model is based on the single-target tracking network Siamese-RPN, which assists multi-target tracking through a cyclic structure. Limitations and technical gap from this paper is that the distribution of training samples is imbalanced, positive samples are far less than negative samples, leading to ineffective training of the Siamese network.

Machine Learning for Anomaly Detection: A Systematic Review

By: Ali Bou Nassif, Manar Abu Talib Qassim Nasir Fatima Mohamad Dakalbab.

This paper was revised on 25 May 2021.The given paper conducted a Systematic Literature Review (SLR) which analyses ML models that detect anomalies in their application. Limitations and technical gap from this paper is that an SLR's quality depends on what has been published in the literature.

Real-Time Anomaly Recognition Through CCTV Using Neural Networks

By: Virender Singha, Swati Singha, Dr. Pooja Guptaa.

This paper was published on 1 July 2020. The given paper proposed an idea to reduce the wastage of time and labour, they are utilizing deep learning algorithms for Automating Threat Recognition System. Limitations and technical gap from this paper is that it is difficult to detect small objects along with that it only predicts a label, not a segmentation box.

Anomaly Event Detection in Security Surveillance Using Two-Stream Based Model by: Wangli Hao,1 Ruixian Zhang,1 Shancang Li,2 Junyu Li,1 Fuzhong Li ,1 Shanshan Zhao,2 and Wuping Zhang1.

This paper was revised on 03 Aug 2020. The paper proposed a novel two-stream convolutional networks model for anomaly detection in surveillance videos. The algorithm used in the given paper is RGB and Flow two-stream networks. Limitations and technical gap from this paper is that it is non-useful for object specification and recognition of colors along with that it is difficult to determine a specific colour.

Proposed Methodology

Recent surveillance cameras can record the video if motion is detected but in old surveillance cameras continuously records the video regardless of motion detected. This will improve system efficiency by reducing processing, searching time and storage required to save the recorded videos. The protective services and authorities often fail to respond efficiently in crime incidents, because they follow reactive approach. In reactive approach authorities depends on witness report or closed-circuit television (CCTV) footage for analysing about the crime after it had occurred. In most of the cases when an incident was occurred, investigators visit the site of the incident, manually retrieve the footage from camera, and then try to locate the appropriate footage either by watching the full length of the video or by Processing it by using advanced algorithms. An efficient crime prediction analysis system for smart home is required to enable the robust security management, thus minimizing the crime incidents and losses. In this paper a framework for real time crime analysis and prediction in smart home using webcam is implemented. This framework has three main steps they are:

- Intelligent Motion detection
- Object detection
- Face recognition

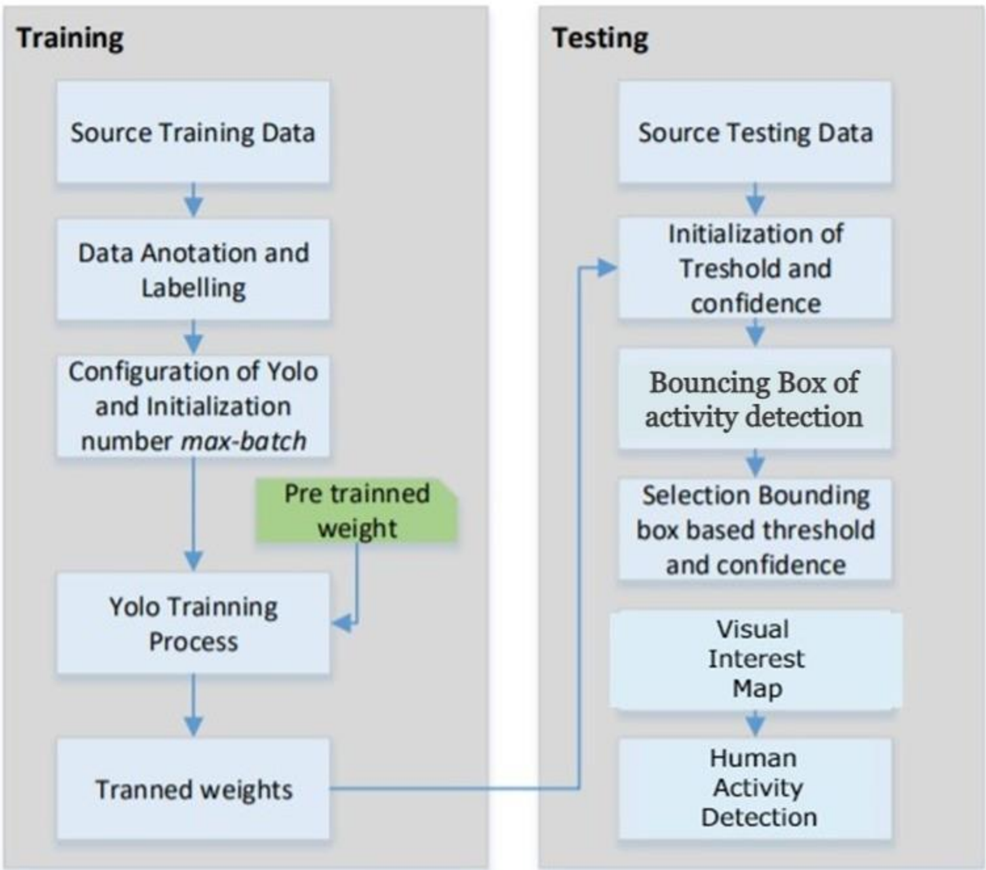

Fig.2 Proposed Method

V. Yolo Architecture

The YOLOv3 algorithm first separates an image into a grid. Each grid cell predicts some number of boundary boxes (sometimes referred to as anchor boxes) around objects that score highly with the aforementioned predefined classes.

Each boundary box has a respective confidence score of how accurate it assumes that prediction should be and detects only one object per bounding box. The boundary boxes are generated by clustering the dimensions of the ground truth boxes from the original dataset to find the most common shapes and sizes.

Other comparable algorithms that can carry out the same objective are R-CNN (Region-based Convolutional Neural Networks made in 2015) and Fast R-CNN (R-CNN improvement developed in 2017), and Mask R-CNN.

However, unlike systems like R-CNN and Fast R-CNN, YOLO is trained to do classification and bounding box regression at the same time.

VI. Recent Updates in YOLO Algorithm

There are major differences between YOLOv3, and older versions occur in terms of speed, precision, and specificity of classes. YOLOv2 and YOLOv3 are worlds apart in terms of accuracy, speed, and architecture. YOLOv2 came out in 2016, two years before YOLO v3.

The following sections will give you an overview of what's new in YOLOv3.

VII. Speed

YOLOv2 was using Darknet-19 as its backbone feature extractor, while YOLOv3 now uses Darknet-53. Darknet-53 is a backbone also made by the YOLO creators Joseph Redmon and Ali Farhadi.

Darknet-53 has 53 convolutional layers instead of the previous 19, making it more powerful than Darknet-19 and more efficient than competing backbones (ResNet-101 or ResNet-152).

VIII. Specificity of Classes

The new YOLOv3 uses independent logistic classifiers and binary cross-entropy loss for the class predictions during training. These edits make it possible to use complex datasets such as Microsoft's Open Images Dataset (OID) for YOLOv3 model training. OID contains dozens of overlapping labels, such as "man" and "person" for images in the dataset.

YOLO v3 uses a multilabel approach which allows classes to be more specific and be multiple for individual bounding boxes. Meanwhile, YOLOv2 used a softmax, which is a mathematical function that converts a vector of numbers into a vector of probabilities, where the probabilities of each value are proportional to the relative scale of each value in the vector.

Using a softmax makes it so that each bounding box can only belong to one class, which is sometimes not the case, especially with datasets like OID.

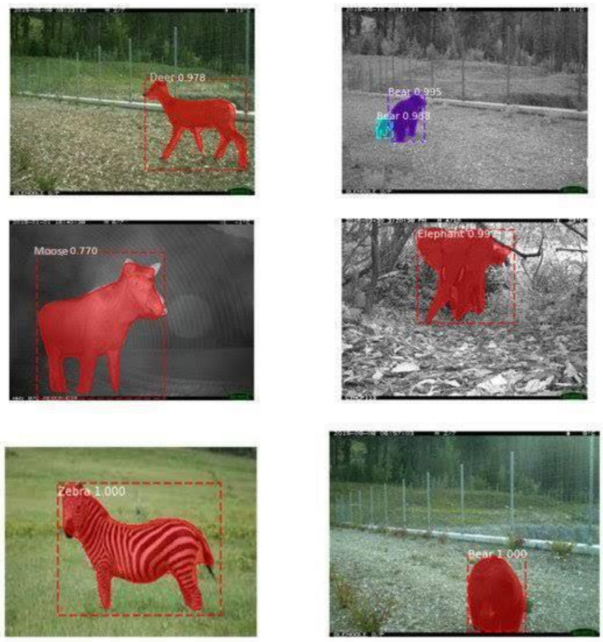

Fig 3: Object Detection to recognize animals with YOLO in a farming application.

IX. Precision for Small Objects

The chart below (taken and modified from the YOLOv3 paper) shows the average precision (AP) of detecting small, medium, and large images with various algorithms and backbones. The higher the AP, the more accurate it is for that variable.

The precision for small objects in YOLOv2 was incomparable to other algorithms because of how inaccurate YOLO was at detecting small objects. With an AP of 5.0, it paled compared to other algorithms like RetinaNet (21.8) or SSD513 (10.2), which had the second-lowest AP for small objects.

Table: YOLOv3 comparison for different object sizes showing the average precision (AP) for AP-S (small object size), AP-M (medium object size), AP-L (large object size)

	backbone	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Two-stage methods							
Faster R-CNN+++	ResNet-101-C4	34.9	55.7	37.4	15.6	38.7	50.9
Faster R-CNN w FPN	ResNet-101-FPN	36.2	59.1	39.0	18.2	39.0	48.2
Faster R-CNN by G-RMI	Inception-ResNet-v2	34.7	55.5	36.7	13.5	38.1	52.0
Faster R-CNN w TDM	Inception-ResNet-v2-TDM	36.8	57.7	39.2	16.2	39.8	**52.1**
One-stage methods							
YOLOv2	DarkNet-19	21.6	44.0	19.2	5.0	22.4	35.5
SSD513	ResNet-101-SSD	31.2	50.4	33.3	10.2	34.5	49.8
DSSD513	ResNet-101-DSSD	33.2	53.3	35.2	13.0	35.4	51.1
RetinaNet	ResNet-101-FPN	39.1	59.1	42.3	21.8	42.7	50.2
RetinaNet	ResNeXt-101-FPN	**40.8**	**61.1**	**44.1**	**24.1**	**44.2**	51.2
YOLOv3 608×608	Darknet-53	33.0	57.9	34.4	18.3	35.4	41.9

YOLOv3 increased the AP for small objects by 13.3, which is a massive advance from YOLOv2. However, the average precision (AP) for all objects (small, medium, large) is still less than RetinaNet.

X. High-Level Design of the Model

The system design of the planned model shown in Fig. one describes that the videos captured from the p surveillance cameras are reborn into frames for visual perception. Once the target object is known, a visible interest map is generated, that is employed to coach the activity recognition system. at the same time, the frames are by selection ironed leading to adaptive compression. Thus, the associate degree intermediate step of adaptive video compression is pipelined into the activity recognition system aspiring to enhance the system's performance.

XI. Model Detection

YOLOv3 has wonderful detection effects within the field of object detection. YOLO algorithmic program improves the speed of detection as a result of it will predict objects in a period of time. YOLO may be a predictive technique that has correct results with the minimal background errors. The algorithmic program has wonderful learning capabilities that alter it to find out the representations of objects and apply them in object detection.

Motion detection is the mechanism by which a change in the location of an object relative to its background or a change in the background relative to an object is detected. The key applications of motion detection are the detection of unauthorized entry and the detection of a moving object that allows a camera to record subsequent events. A simple motion detection algorithm compares the current image to a reference image and simply counts the number of different pixels. Due to factors such as changing lighting, camera flicker, and CCD dark currents, images will naturally differ, pre-processing is useful to minimize the number of false positive output. For detecting the moving objects in video, background subtraction model is used.

XII. Face Detection

If motion is detected, then next step is to detect face in live stream. Face detection is performed by using Haar feature based cascade classifier which is an effective detector of objects. It is an approach based on machine learning. Lot of Positive and negative images are used to train the cascade function, then it is used for comparing with other images for object detection. There are huge individual XML files with lot of features, each xml files have a specific use case feature. Here for face detection, haarcascade_frontalface_default.xml is used which has features for detecting the front face. This xml has values which is obtained when training with lot of positive and negative images for detecting the front face. Face detection model is designed using OpenCV which is most familiar way to detect the face.

XIII. Implementing Object Detection using YOLOv3

We have used YOLOv3 to implement the object recognition & the algorithm automatically identifies the category of interest and their individual Anchor box and maintains a counter. we've got used the pre-trained YOLOV3 model, which is capable of detecting “person” as a category, and that we count the number of individuals by maintaining a counter. Time period video frames had given as input to our model & our model detected the anomaly counter.

Fig.4: Anchor Box Detection

XIV. Object detection and Recognition

Object Recognition is one of the computer vision techniques for identifying instances of objects in images or videos. The primary objective of object detection is to replicate the human intelligence of detecting object in a video or image to computers. The use cases of object detection are infinite some of them are monitoring objects, video surveillance, pedestrian identification, identification of anomalies, people counting, self-driving cars or face detection and so on. Object detection model is designed based on YOLOV3 and Darknet for custom data. Custom data here considered are Knife and Gun. Following are procedures involved in designing the object detection model are:

1. Set up YOLO V3 on windows. Install all dependencies they are Visual Studio 2019, CUDA>=10.0, cuDNN>=7.0, CMake>=3.12, OpenCV>=2.4. Ensure to add OpenCV, CUDA, cuDNN directory in environmental variables. Then clone the darknet directory from https://github.com/AlexeyAB/darknet. Set up the config file with the CUDA version installed cuDNN directory in environmental variables. Then clone the darknet directory from https://github.com/AlexeyAB/darknet. Set up the config file with the CUDA version installed and then build the solution using visual studio 2019 which will generate darknet.exe file. Copy cuDNN64_7.dll, OpenCV ffmpeg420_64.dll, OpenCV_world420.dll file to darknet bin folder.

2. Using Image annotation tool Labelimg which is a powerful tool used for image annotation and labelling. Using this tool labelImg and annotation and labelling is done and save the file generated for each custom image in txt file which contains annotation and labelled values for each image. Define class file which has names of the objects that should be detected.

3. Prepare config file for custom data by modifying the yolo config file in darknet. Create object name folder for training and object data folder which has train data path, validation data path, classes, names of custom object file and path for storing the trained data. Download pre trained CNN weights for YOLO.

4. Train using darknet for custom data using pre trained weights.

5. Using an object detector model coded using yolo v3 and OpenCV for detect the objects in real time which is able to detect the knife and gun.

Once the image frames have been loaded, parameters for nms_thresh and iou_thresh, we are able to use the YOLO algorithmic program to discover objects within the image. We tend to discover the objects using the detect_objects(m, resized_image, iou_thresh, nms_thresh)function from the utils module. This function takes in the model to come to the resized image, and therefore the NMS and IOU thresholds, and returns the Anchor boxes of the objects found.

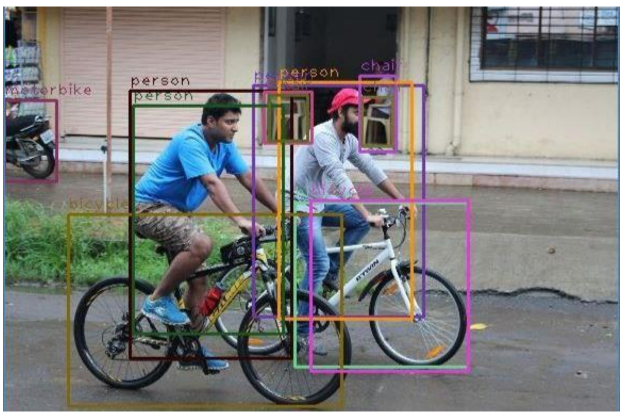

Fig5. Object Detection

XV. Input Image Processing

The input image is passed through a CNN network which is a series of Convolution and Detection layers, followed by two fully connected layers to get the output. The input image in this case is divided into 19*19 grids cell, with 5 predefined anchor boxes for each grid cell, totally we will have 1805 anchor boxes and each anchor box will have 85 predicted elements by the network as the output.

XVI. Face Recognition

The next step after object detection is face recognition. Face recognition model is built using a Face recognition function defined by adam geitgey. Face recognition package is downloaded using pip command. Then built a model to recognize the face in real time. Humans are capable to identify the person easily and quickly, but computers cannot. In order to make computers to do that following procedures are involved they are: find face in image, analyze facial features, compare against known face and then prediction. The first step is finding the face which involves, convert the RGB image to gray image then divide the image into 16*16 pixel each. For each pixel calculate the gradients point in each major direction replace that square in the image with the arrow directions that were the strongest. Using HOG Face is detected for given image. Face landmark estimation algorithm is used for locating the 68 face landmarks on given image, condition is that eyes and nose should be visible in image. A pre trained convolution neural Network “Open Face” is used to generate 128 measurements for each face. By calculating the Euclidean distance between the image encodings, comparing the distance face prediction is performed.

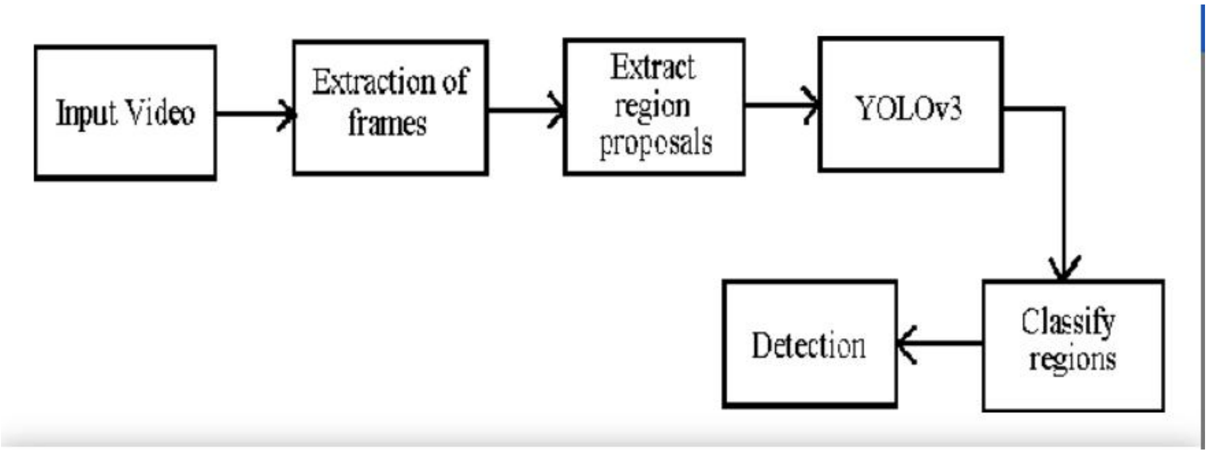

Fig.6 Basic block diagram of suspicious activity detection.

XVII. Project Architecture

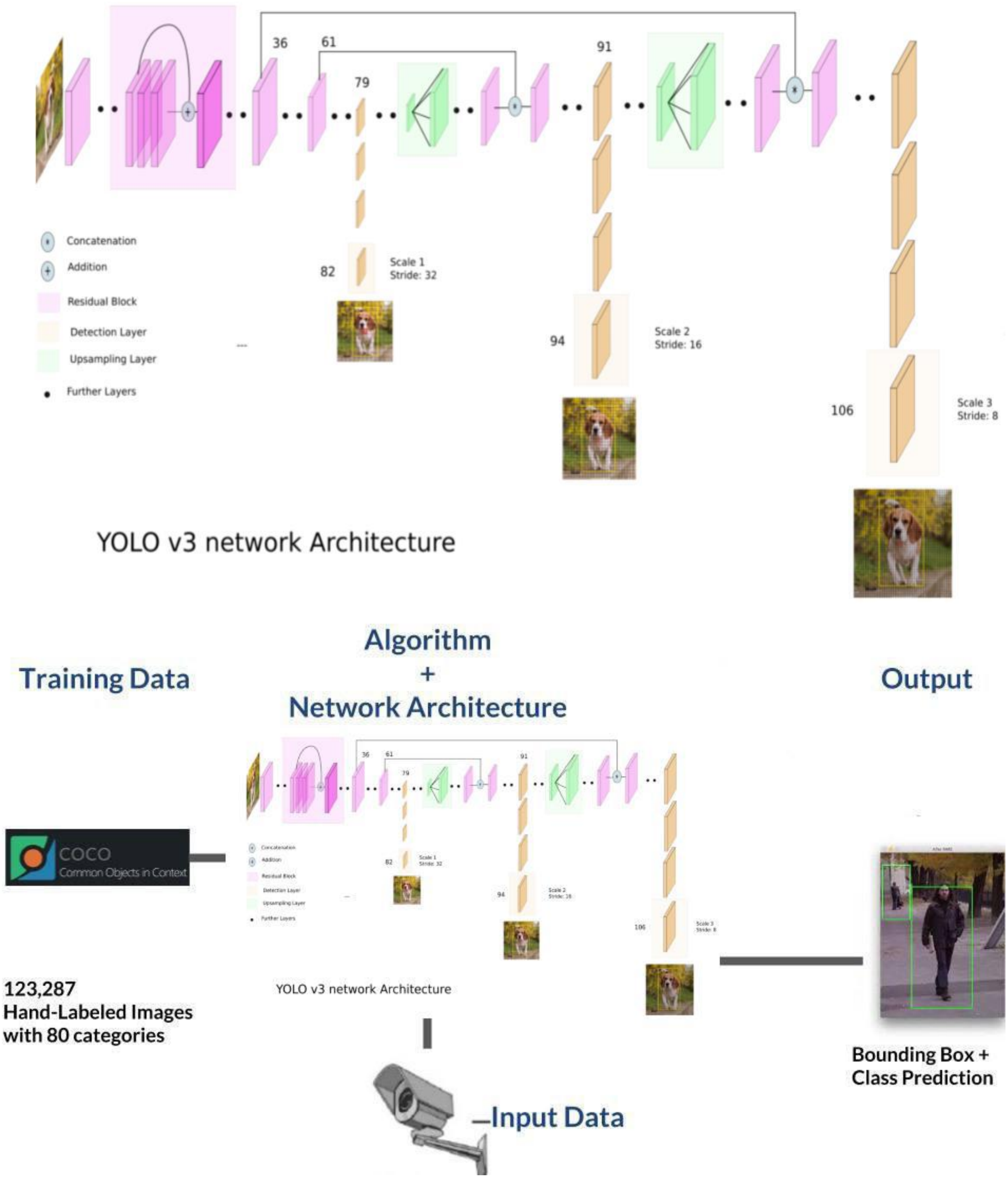

Fig.7 YOLO v3 Network Architecture.

XVIII. EXPERIMENTATION AND RESULTS

A. Implementation Details

We use YOLOv3 to implement object recognition and the algorithm automatically identifies the desired class and anchor box respectively and maintains a counter. We use the pre-trained YOLOV3 model, which is capable of detecting "people" as a class and we count the number of people by creating a counter. Real-time video frames have been provided as input to our model, and our model detects an anomaly counter.

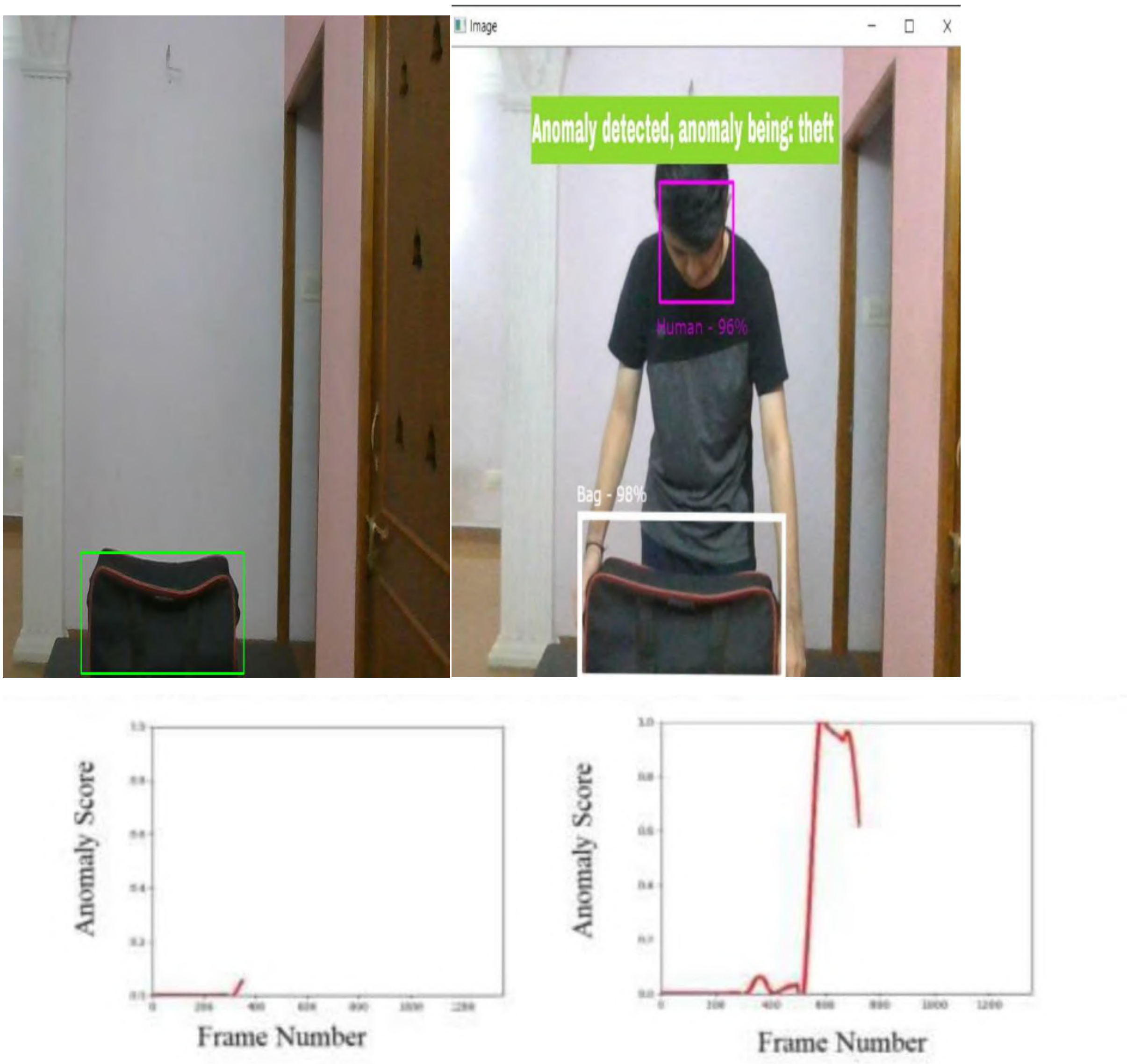

Fig.8: Anomaly detected, anomaly being: theft

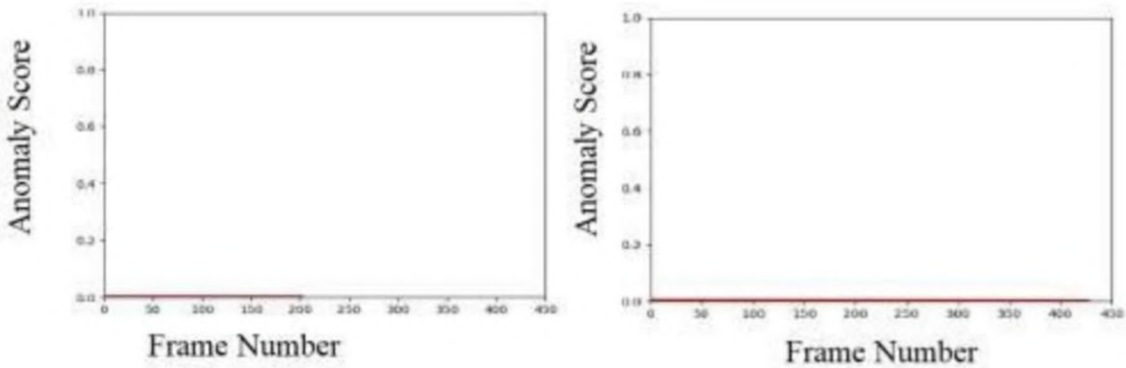

Fig.9: Test result for a video with no anomaly

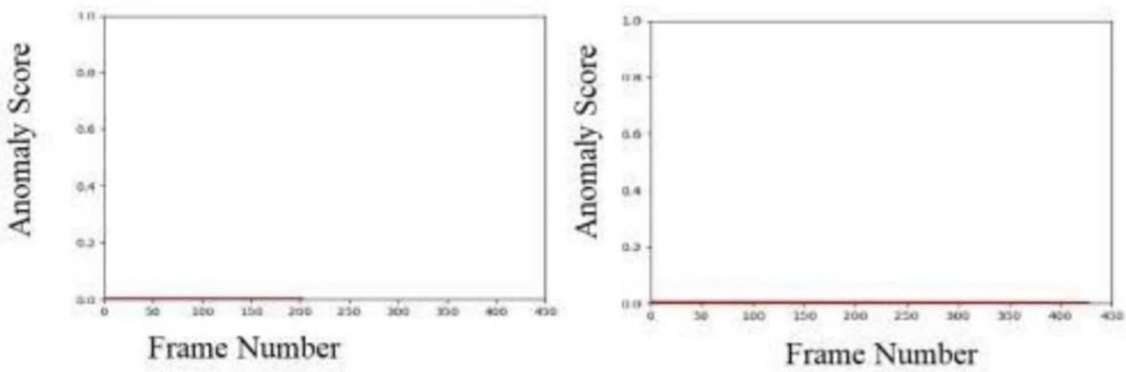

Fig.10: Test result for a video with no anomaly

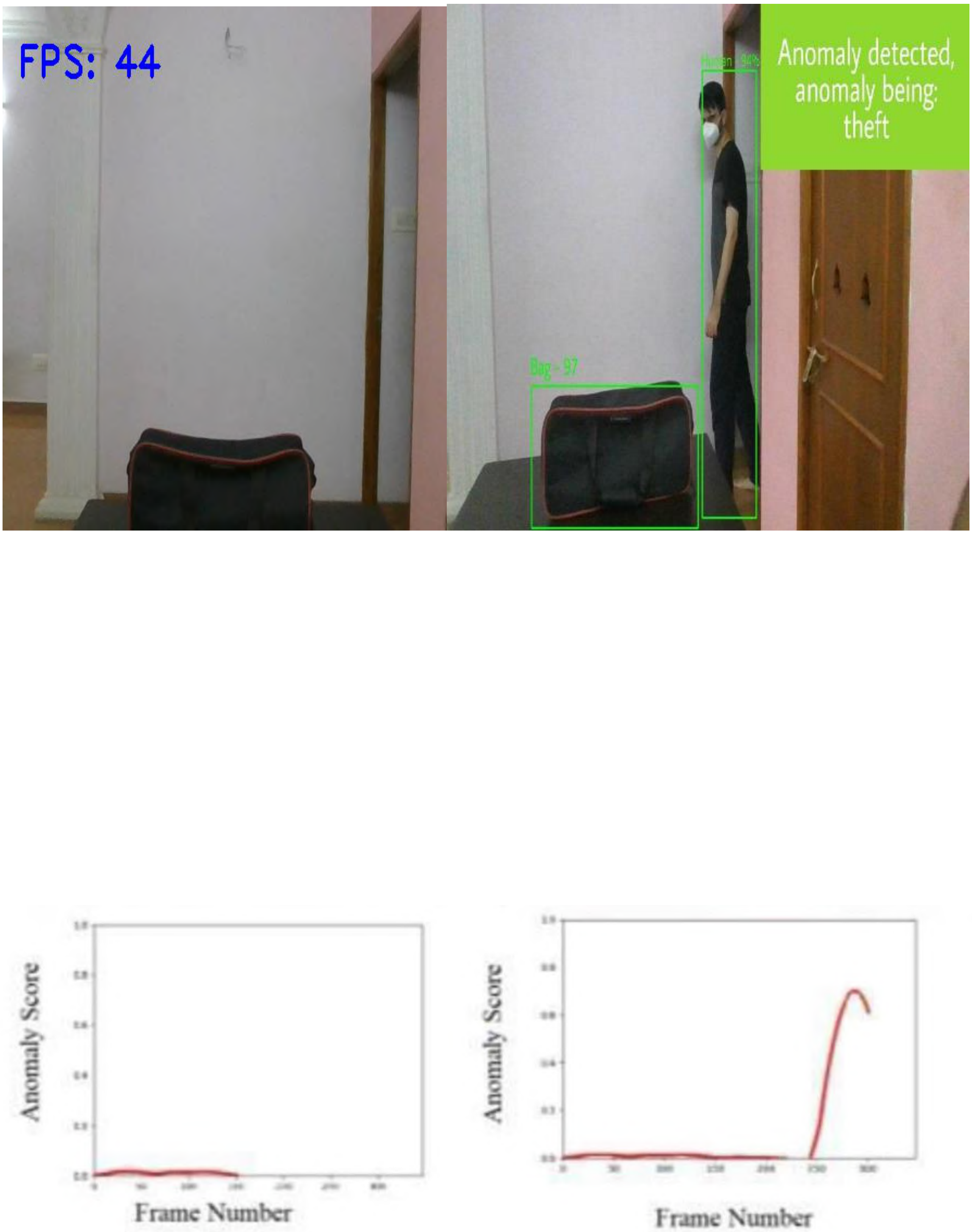

Fig11.: Anomaly detected, anomaly being theft

B. Pre-Processing Data

Suppose this detection system is installed in the house so first, we will scan the house owner's face then our model will get trained on the faces it collected after this our detection system will start video stream this stream will act as input to our model. When any person comes near to the system it will run a similarity check with the face in the input stream and the trained faces, if the similarity accuracy is greater than 85-90 percent then our model will keep on running and if the similarity is less than 80 percent the model will start the lockdown procedure in which alarms will on and the registered members will get triggered.

C. Comparison With Other Model

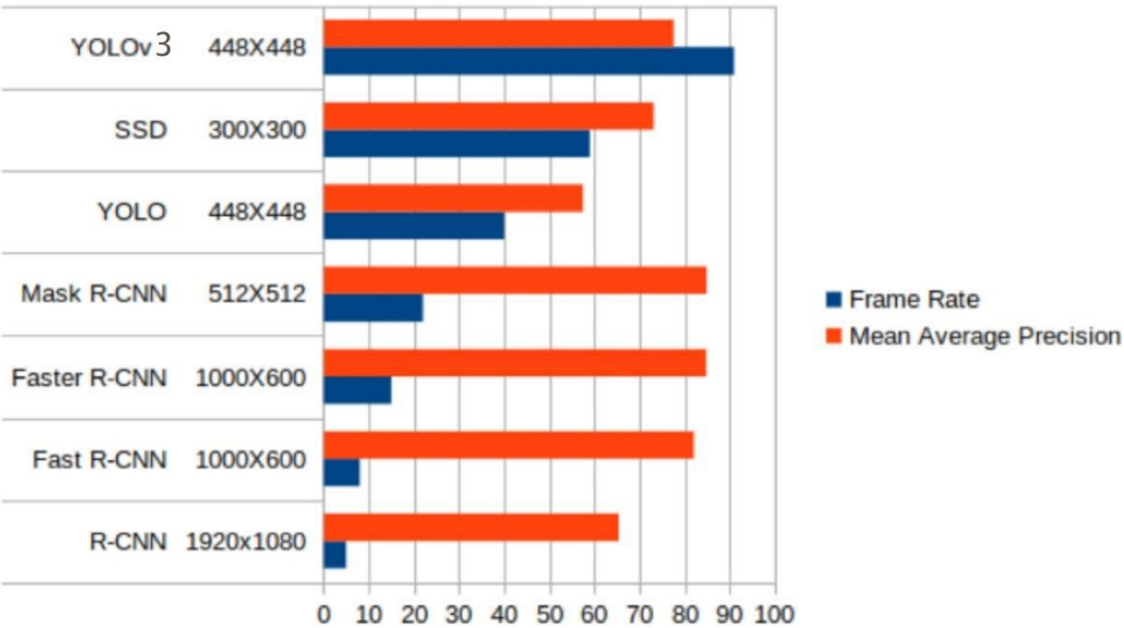

Fig.12.: Anomaly detected, anomaly being theft

Prior detection systems repurpose classifiers or localizers to perform detection. They apply the model to a picture at multiple locations and scales. High grading regions of the image are thought of as detections. We use a completely totally different approach. we have a tendency to apply one neural network to the total image. This network divides the image into regions and predicts bounding boxes and chances for every region. These bounding boxes are weighted by the expected changes. Our model has many benefits over classifier-based systems. It looks at the whole image at test time so its predictions are informed by global context in the image. It conjointly makes predictions with one network evaluation, not like systems like R-CNN that need thousands for one image. This makes it extraordinarily quick, quite 1000x quicker than R-CNN and 100x quicker than quick R-CNN. See our paper for additional details on the total system.

D. False alarm rate

In real-world Anomaly Detection, A robust anomaly detection should have a low false alarm rates as compared to normal situation. Therefore, we evaluate the performance of our approach and the other normal methods then false alarm rates of different approaches at 50% threshold. Our approach has a much lower false alarm rate than other methods, indicating a more robust anomaly detection system.

E. Accuracy

The accuracy of the YOLO algorithm, there are 2 accuracy calculations, namely Detection Accuracy (% DO) and Recognition Accuracy (% TL). Detection accuracy is the value of how accurate the YOLO algorithm is in detecting a object. This accuracy can be searched by looking at how accurate the YOLO algorithm is in making boxes of all objects in the image/video. Recognition accuracy is the value of how accurate the YOLO algorithm is in recognizing a object. This accuracy can be searched by looking at how accurate the YOLO algorithm is in giving the right labels and according to the type of object.The detection accuracy and recognition accuracy formula are:

$Detection\ Accuracy$ (%DO) = $Number\ of\ objects\ detected\ Total\ number\ of\ objects$ $*$100 (%) (1)

$Recognition\ Accuracy$ (%TL) = $The\ number\ of\ correct\ labels\ Total\ number\ of\ labels$ $*$100 (%) (2)

The calculation of accuracy is divided into two, namely calculating accuracy based on the recognized human beings and the calculation of accuracy based on their anomalous Activities. We have been carried out with data without pre-processing; the accuracy obtained is shown

in Table

Table: Accuracy of a dataset with pre-processing.

Dataset with pre-processing	Detection Accuracy (% DO)	Recognition Accuracy (% TL)
Without distraction	98.2	88
Brightness (+25)	97.6	88

Brightness (+50)	95.9	80
Brightness (+75)	94.1	86
Brightness (-25)	98.2	88
Brightness (-50)	97.7	84
Brightness (-75)	97.0	84

XIX. Conclusion and Future Work

There is a rising demand for the installation of surveillance cameras all public places as well as streets due to the increase in crime rates. Responsibility to curb criminal activities doesn't finish with the installation of CCTV cameras; there should be a mechanism to provide immediate facilitate to the victims of crime at the side of immediate action against the criminals. this will happen solely with constant and careful watching of the video police work that eventually needs a manpower. Thus, developing a real-time automated system to recognize abnormal human activities can be an extended to several public places, specific to the application environment like schools, colleges, airports, bus stops, hospitals and railway stations supported their specific requirements. Thus, victimization the intermediate results of adjustive video compression an correct, time anomaly detection system is enforced. The planned technique outperforms different existing systems in terms of accuracy and timeliness.

REFERENCE

[1] Tran D, Bourdev L, Fergus R, Torresani L, Paluri M. Learning Spatiotemporal Features with 3D Convolutional Networks. Computer Vision and Pattern Recognition (CVPR); 2017.

[2] R. T. Ionescu, S. Smeureanu, B. Alexe et al., "Unmasking the abnormal events in video", Proceedings of the IEEE International Conference on Computer Vision[C], pp. 2914-2922, 2017.

[3] Liu B, Yu X, Zhang P, Yu A, Fu Q, Wei X. Supervised Deep Feature Extraction for Hyperspectral Image Classification IEEE Transactions on Geoscience and Remote Sensing (TGRS). IEEE;
2017.

[4] Sultani W, Chen C, Shah M, Real-world anomaly detection in surveillance videos. In: CVPR, 2018.

[5] Fa L, Song Y, Shu X, Global and Local C3D Ensemble System for First Person Interactive Action Recognition. In: International Conference on Multimedia Modeling, 2018.

[6] Ali Bou Nassif, Manar Abu Talib Qassim Nasir Fatima Mohamad Dakalbab. Machine Learning for Anomaly Detection: A Systematic Review. In: IEEE; 2021

[7] Chenxu Wang, Yanxin Yao , Han Yao. Video anomaly detection method based on future frame prediction and attention mechanism. In: IEEE; 2021

[8] Zhou S,Shen W, Zeng D, Fang M, Wei Y, Zhang Z. Spatial–temporal convolutional neural networks for anomaly detection and localization in crowded scenes. Signal Proc Image Commun. 2016;47:358–68