IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Human Detection And Counting In Visual Surveillance

[1] Abhijith A, [2] Adarsh Balan V V, [3] Adarsh K V, [4] Advait Nikesh, [5] Priya V V

[1,2,3,4]Student, [5] Assistant Professor,

[1,2,3,4,5] Department of Computer Science and Engineering,

[1,2,3,4,5]College of Engineering Thalassery, Thalassery, India

Abstract: Dependable individuals checking and human recognition is a significant issue in visual observation. Lately, the field has seen many advances, however, the arrangements have a few limitations: individuals should be moving, the foundation should be straightforward, and the picture goal should be high. We expect to foster a powerful strategy assessing the number of individuals and find each individual in a picture with confounded scenes. The primary focal point of this work is to discover intends to successfully manage the previously mentioned continuous issue utilizing the YOLO calculation to recognize the individual people in a video outline. The dataset is thus used to investigate the individual people and the tally of individual people is shown in the versatile application. Creating to make a productive framework to keep away from the surge in establishments, associations, and surprisingly open spots will help individuals better using time productively.

Index Terms **- Object Detection, YOLO, Firebase.**

I. INTRODUCTION

Coronavirus influenced our everyday life, organizations, upset world exchange, and developments. Medical services laborers and analysts are on the bleeding edges fightingCOVID-19, expecting to slow the spread of the illness and take care of the wiped out. Typical life has quit spilling over a billion groups all throughout the planet. The COVID-19 emergency, accordingly, will feature something that has consistently been valid about ML. In the current emergency, human activity and development will be especially basic in utilizing the force of what ML can do. Exploiting the AI innovation we can recognize the human and check. By utilizing this innovation we can guarantee the social distance among people and control the spreading of COVID 19 up as far as possible.

Individuals tallying might be an essential and testing issue in visual reconnaissance. Programmed observing of the quantity of individuals in open regions is furthermore significant for security control and substantial arranging. As of late, the field has seen many advances, however the arrangements have limitations: individuals should be moving, the foundation should be straightforward, and the picture goal should be high. This undertaking intends to foster a compelling technique for assessing the quantity of individuals and find every person in a low goal picture with convoluted scenes. Number of reconnaissance applications require the location and following of people to ensure security, wellbeing, and site the board. Models remember the assessment of line length for shops, the observing of passage focuses, transport terminals, or train stations. Understanding customer traffic into stores is central to the most fundamental retail measurements. In This ebb and flow pandemic circumstance it might assist us with forestalling the social affair of individuals openly places and we can play it safe to try not to spread of the infections.

A group tallying framework will assist us with deciding the occasions when you need more staff and less. It can likewise assist you with distinguishing your most active days of the week. This sort of arrangement can give you a nitty-gritty investigation of the number of individuals in your store at some random time. In the event that we own various stores or areas, this kind of framework can be utilized to assist you with deciding precisely which area has the most pedestrian activity and its circumstance. You can likewise utilize a group checking framework to recognize which of your representatives is working most proficiently. This sort of arrangement will consider execution throughout any timeframe you need to inspect for a long-haul execution survey.

II. LITERATURE SURVEY

[1] The point of this work is to introduce a technique for assessing the number of individuals in a bunch of situations. One of the foreground detection methods used is the mixture of the Gaussians approach of Stauffer and Grimson, where the authors provide a practical technique of implementing adaptive background subtraction. As indicated by this technique, every pixel can be demonstrated as a weighted blend of Gaussian conveyances. The second frontal area recognition strategy utilized is the pixel layering method. This comprises of coarsely displaying the scene into a bunch of layers that address the foundation of the picture. The probability PA(y) of a pixel y belonging to a layer is computed using non– parametric kernel density estimation with a Gaussian kernel.

[2] Automated Grow Cut (AGC) segmentation algorithm to classify foreground regions efficiently from the labeled seed pixels obtained from the initial depth analysis. This results in the efficient segmentation of human factors without much computation cost as compared to supervised segmentation techniques. RGB-D cameras can capture both color and depth information from an image, where the depth data provides the distance of scene contents from the camera. These camera sensors are low cost and easy to use. The system uses an unsupervised segmentation system based on depth details and grow-cut segmentation. The segmentation process is fast enough to process on intermediate frames and favor further applications like activity recognition or human tracking.

[3]This paper focuses on developing an efficient method for estimating the number of individuals and locating each individual during a low-resolution image with complicated scenes. First, post-processing steps are performed on background-subtracted results to get the number of individuals during a scene, second, an Expectation-Maximization (EM)-based method has been developed to locate individuals during a low-resolution scene. A replacement cluster model is employed to represent everyone within the scene. Next is, the count of individuals is employed as a priori for locating individuals supported feature points. Hence, the methods for estimating the number of individuals and for locating individuals are connected.

[4] To help within the rescue effort, indoor counting solutions can provide emergency personnel with the quantity of people who have evacuated the building, and from which floors. Life count implements a completely unique two-stage neural network-based algorithm to accurately count the amount of individuals passing through a hallway. A novel CSI-based human counting algorithm called Life count is developed that can count the number of individuals passing through a corridor with higher accuracy than existing approaches from literature life count utilizes a single TX/RX link and can count multiple people crossing the sensing region simultaneously.

[5] CNN and Motion-based model for human detection in the infrared video feed. It can be listed as 3 steps called preprocessing, human detection, and human tracking. Noise cancellation filters, histogram equalization on raw image, CNN-based human detection on an enhanced image. The human discovery pace of the general framework can be upgraded utilizing a versatile movement model dependent on KCF Tracker. Depth-wise separable is a combination of two convolution steps called depth-wise convolution and pointwise convolution. Noise filters and histogram equalization were used in the preprocessing stage. As a result, we could enhance the detection rate by 5.1 than an anon- prepossessed step. As the subsequent advance, we applied human identification utilizing AI, which is the most ideal approach to tackle object location issues.

[6] Exceptional individuals checking technique upheld head discovery and following to measure the number of individuals who move under an overhead camera. There fundamental four pieces of the proposed framework are frontal area extraction, head discovery, head following, and intersection line judgment. The Proposed method utilizes an efficient foreground extraction method to urge foreground regions of moving people, and a few morphological operations are performed to optimize the foreground regions. It takes advantage of the LBP highlight-based Adaboost classifier for head identification inside the streamlined forefront areas. After head recognition is played out, the competitor head object is followed by the neighborhood head following strategies upheld by the Mean Shift calculation.

[7] An automatic people counting system will count multiple folks that interact within the region of interest, by using only one camera. The algorithm uses the Viola-Jones method of face recognition to detect people. Employing a single overhead mounted camera, the system counts the number of people going into an observed area. Counting is performed by analyzing the image and detecting faces in it.

[8] This paper introduces a singular multi-person tracking system for crowd counting in surveillance environments. The proposed system consists of 4 modules and they're: people detection, head-torso template extraction, tracking and crowd cluster analysis. Firstly, the system extracts human silhouettes using inverse transform, also a median filter reducing the value of computing and handling various complex monitoring situations. Secondly, people are detected by their head torso to be less varied and hardly occluded. Thirdly, every-one is tracked through consecutive frames using the Kalman filter techniques with Jaccard similarity and normalized cross-correlation. Finally, the template marking is employed for crowd counting having cues localization and clustered using Gaussian mapping for normal/abnormal events detection.

[9] Histograms of oriented gradients (HOG) are proven to be ready to significantly outperform existing feature sets for human detection. Extracting the FASTHOG human features on the foreground image, then training the positive and negative samples by SVM linear classifier. The moving targets as noise and eliminate them to urge a background image, the first 11 step is to get a background-removal image. It solves the problem of missing detection caused by HOG without rotation variance when the direction or action of the human body changes drastically. It cancels the noise of background interference by the background subtraction, hence the accuracy is improved.

[10] Indoor people counting is important for many applications such as crowd control and smart building. They consider people counting in a quasi-static scenario and propose a non-intrusive training-free method using the Channel State Information (CSI) on a single pair of commercial Wi-Fi devices. A training-free approach for passive crowd counting using a single pair of commodity Wi-Fi devices. First, calculate the breathing rate of the multi-personalities. The next stage is the people counting. In the stage of people counting is the conversion of Breathing Rates to People Counting and the second one is Extracting Breathing Traces, the third is Trace Concatenating. Iterative powerful programming gives the breathing follow to each time window and decides the inhabitance level dependent on the following number. Experiments were conducted using a pair of commodity WIFI devices, one as Tx and the other as Rx.

II. EXISTING SYSTEM

Usually, strategies decide the quantity of individuals and their areas at the same time. Detection-based methods determine the number of people by identifying individuals within the scene. These techniques decide the quantity of individuals and their areas all the while. However, the low-resolution images from the cameras always make this work. Occlusion could also be a serious problem for background segmentation techniques. From the machine vision perspective, it's hard to differentiate an object sort of a person's because of its sizable amount of possible appearances. Interesting progress is being made employing a local-based approach for human detection. Nowadays there are only methods to count and detect humans and there is no app to visible these counts to the people by using an application.

IV. PROPOSED SYSTEM

People counting and human detection is a crucial problem in visual surveillance. In recent years, the sector has seen many advances, but the solutions have restrictions: people must be moving, the background must be simple, and the image resolution must be high. So we proposed a system that can be used for both human detection and people counting mechanisms. With the help of the camera, the system will capture video footage. This real-time video footage will go through frame separation by using the frames we will detect the human count. The count of the people in various locations can be viewed by using another application. And, the number of people is used as a prior for locating individuals based on feature points. Henceforth, the strategies for assessing the quantity of individuals and for finding people are associated. This proposed system may help to save our time and helps to avoid gathering of people in public shops and etc.

4.1 Objectives

1. To develop a system that identifies humans from a video frame.
2. To count the humans when the humans are identified.
3. To develop a real time human Counting system which shows the count in an Android Application.

4.2 Overall Algorithm

1. Real time video footage is used as input which is acquired using a CCTV/webcam.
2. The obtained visuals are converted into image frames with a time slice.
3. The frames generated are arranged in a sequence according to time.
4. Image is converted to an N X N array of grid.
5. Each grid is responsible for predicting whether an object is present or not.
6. Apply Bounding box regression.
7. Apply non-max suppression to remove weak predictions.
8. Filtering humans from detected objects.
9. Count of humans in each frame is stored and can be viewed by the user through an Android Application.

V. SYSTEM ARCHITECTURE

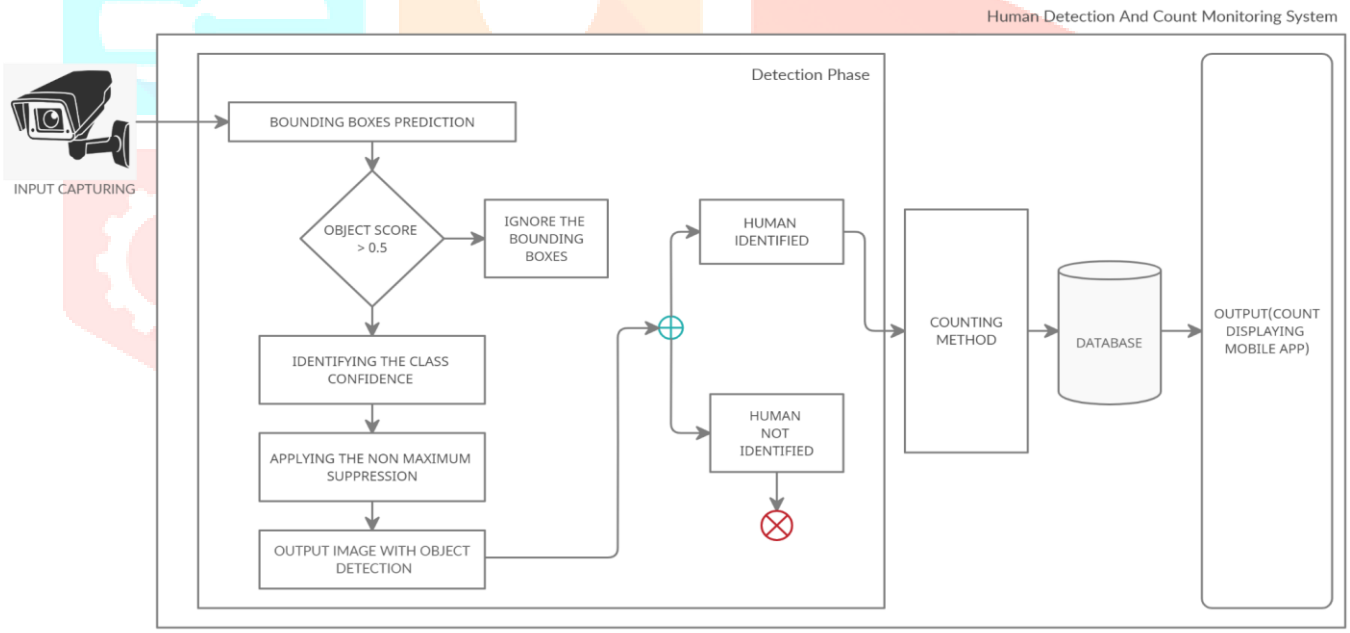

Figure 1: Architecture Diagram

5.1 Image Acquisition

The first phase of our system is image acquisition. It's defined as the action of capturing a picture from a source, usually a hardware-based source for processing. After the image has been obtained, the various routines of processing are applied to the image to perform the different vision tasks. Performing image acquisition in image processing is the first step in the workflow sequence because, without an image, no processing is possible. There are various ways to acquire images with the help of a camera or scanner. Acquired images should retain all the features.

5.2 Frame Separation

Figure 2: Frame Separation

In our proposed system, input is the real-time video footage which is captured by Closed-Circuit Television (CCTV) or any type of video capturing devices. But in order to process and find out each of the individual persons we require images rather than video footage .So we need to convert the video into images or different video frames. Then the collected images are transferred to the image sequence.

5.3 Image Separation

A picture grouping is a progression of successive still pictures that address casings of liveliness. Normally, the pictures are saved inside one envelope and are named with an augmentations document name to save the sequential request. Picture succession or video combination is the utilization of incredible pragmatic significance.

VI. YOLO ALGORITHM

YOLO is a calculation that utilizes neural organizations to give ongoing item discovery. This calculation is mainstream because of its speed and precision. It has been used in different applications to recognize traffic lights, individuals, stopping meters, and creatures. YOLO is a contraction for the term You Only Look Once. This is a calculation that distinguishes and perceives different items during a picture (continuously). Item discovery in YOLO is finished as a relapse issue and gives the classification probabilities of the recognized pictures. YOLO calculation utilizes convolutional neural organizations (CNN) to distinguish objects progressively. As the name proposes, the calculation requires just one forward spread through a neural organization to identify objects. This implies that the forecast inside the whole picture is finished during a solitary calculation run. The CNN is utilized to foresee different class probabilities and jumping boxes all the while.
YOLO algorithm works using the subsequent three techniques:

- Residual Blocks
- Bounding Box Regression
- Intersection over Union (IOU)

6.1 Residual Blocks

First, the image is divided into various grids. Each grid has a dimension of n x n. The input image is divided into grids. In the image above, grid cells are of equal dimension. Every grid cell is responsible to detect objects that appear inside them. For example, if an object center appears within a particular grid cell, then this cell is going to be liable for detecting it.

6.2 Bounding Box Regression

A bounding box is an outline that focus an object in an image. Every bounding box of a image will have these attributes:
- Width (bw)
- Height (bh)
- Class (for example, person, car, traffic signal, etc.)- This is often represented by the letter c.
- Bounding box center (bx , by)

The following image shows an example of a bounding box. The bounding box is represented by a yellow outline.

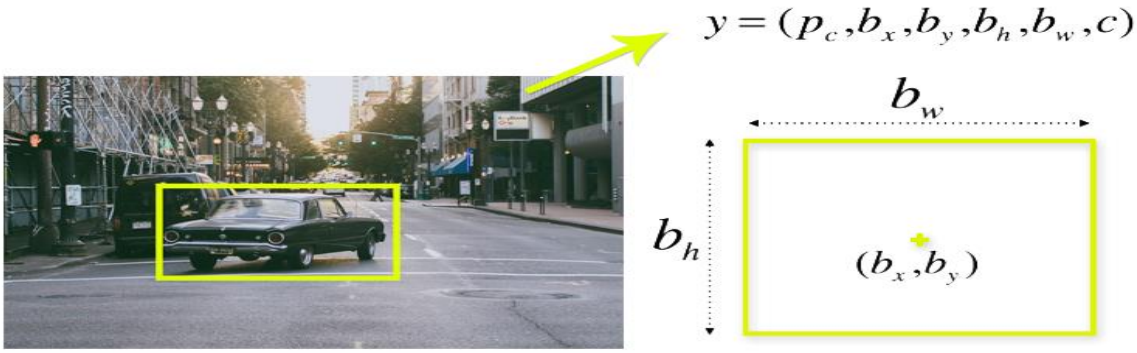

Figure 4: Bounding Box

YOLO uses one bounding box regression to predict the peak, width, center, and sophistication of objects. The image above represents the probability of an object appearing within the bounding box.

6.3 Intersection Over Union (IOU)

Convergence over Union (IOU) is a wonder in object identification that depicts how boxes cross-over. YOLO utilizes IOU to give a yield box that encompasses the items impeccably. Every matrix cell is answerable for anticipating the bouncing boxes and their certainty scores. The IOU is equivalent to 1 if the anticipated jumping box is equivalent to the genuine box. This instrument dispenses with jumping boxes that are not equivalent to the genuine box. The accompanying picture gives a straightforward illustration of how IOU functions. In the picture above, there are two jumping boxes, one in green and the other in blue. The blue box is the anticipated box while the green box is the genuine box. YOLO guarantees that the two bouncing boxes are equivalent.

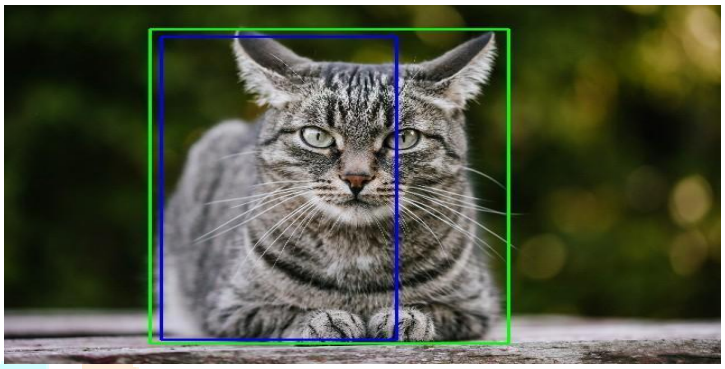

Figure 5: Intersection over Union

VII. DATABASE

Firebase real-time database is a fully-managed NoSQL document database for mobile and web app development. It's intended to effectively store and sync application information at a worldwide scale, and it's presently accessible in beta.
The below figure shows the flow of storing data inside a firebase real-time database. The mainly real-time database follows the structure of JSON. In this project, all the organization's details that as the name of the organization, address of the organization, the image of the organization, and the varying count are stored inside the main directory called customers. And inside that, each organization has its own unique value.

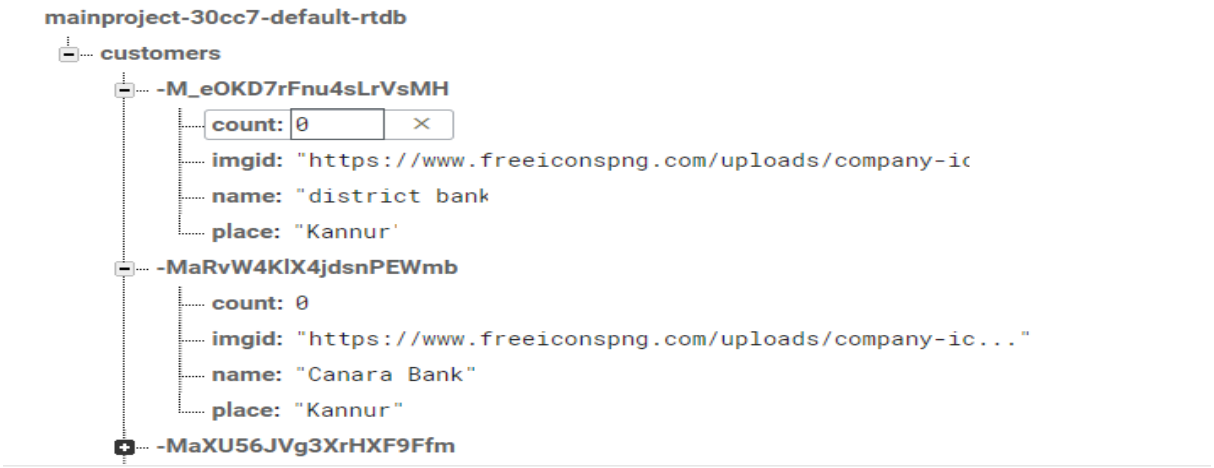

Figure 6: Database

VIII. MOBILE APPLICATION

For showing the count monitoring details we developed a mobile application that uses flutter as a framework. Our system aims to provide the information of people in an organization which is under CCTV surveillance. We can have the count of people in different organizations in our android application. So we can know the status of whether the organization is crowded or not in prior.

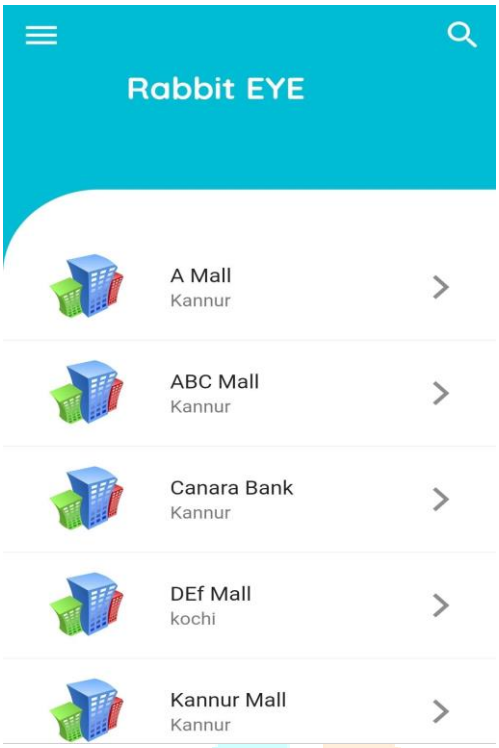

Figure 7:Home Screen

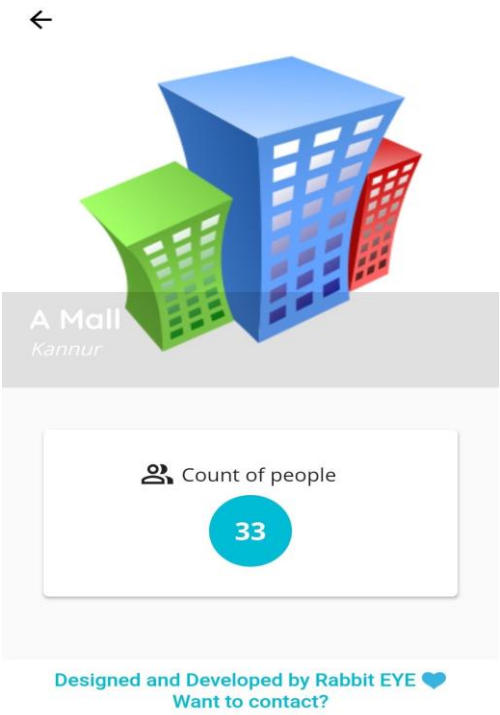

Figure 8: Details Screen

The above figure displays the count of people. First page user can see a nice looking splash screen that gives the information about the project name and a brand logo. After five seconds it automatically redirect to the homes page and shows the list-out various organizations or public places that are connected through our proposed system. In the Detailed page users can see the organization details and the real-time count of people currently present at the organization. With this real time information obtained we can choose our own time which is more comfortable to visit organizations such as Banks, Supermarkets and Public distribution systems.

IX. RESULT AND DISCUSSION

After the model is built, Next step is to test and analyzed its performance. The model trained on coco dataset. Our expected outcome is to detect humans from video frames and images. We have tested the model by giving inputs as videos as well as images. The output we obtained is given in the figure below.

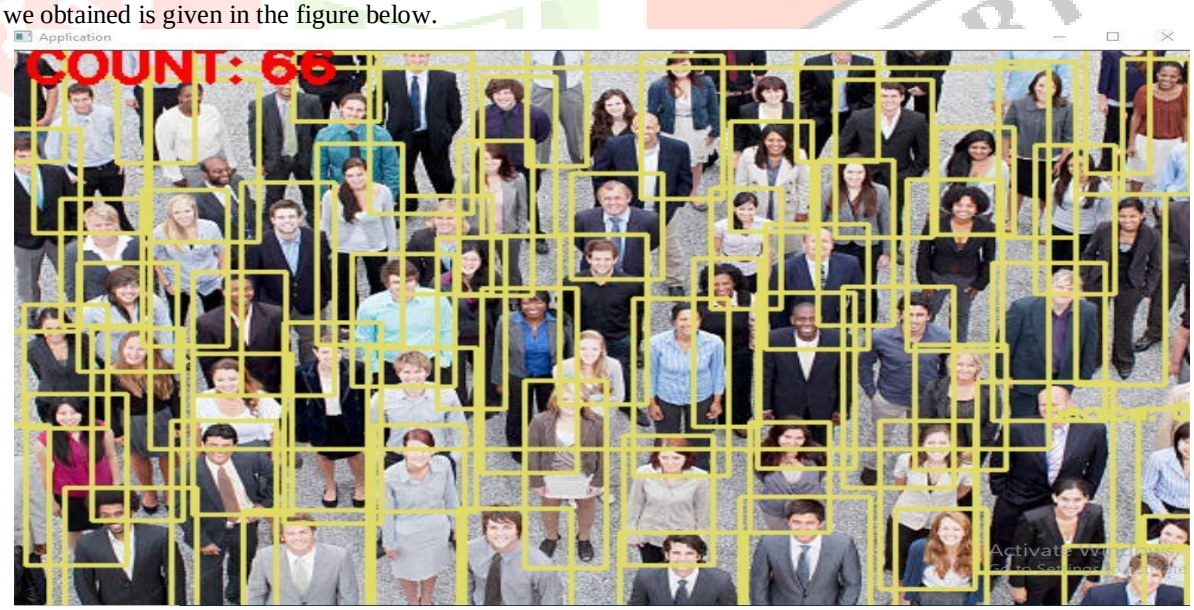

Figure 9: Output

The model performance depends upon various factors. Image quality of video quality is the first factor. It is seen that black & white images the detection is not well performing whereas in Digital color images we have decent performance. Next factor is the position of the camera. Camera should be placed in an appropriate position such that there are no other obstacles hindering the camera view. While implementing in public places Trees, Electric post etc. can cause this issue. The proposed system can do much more than a human detection. Counts detected by the system can be displayed with the help of a mobile application. In the last page users can see the organization details and the real-time count of people currently present at the organization. With this real time information obtained we can choose our own time which is more comfortable to visit organization such as Banks, Supermarkets and Public distribution systems. The Model still has some scope for improvements in terms of accuracy and latency. In Future, we can modify the existing system by attaching motion detection so that we can sense the moving object also. By making simple training heuristics and small architectural changes that can make YOLO v3 perform better. Further, the efficiency and performance of the

system can be increased. Cuda packages and GPU acceleration technique helps to enhance the accuracy and processing speed of the system.

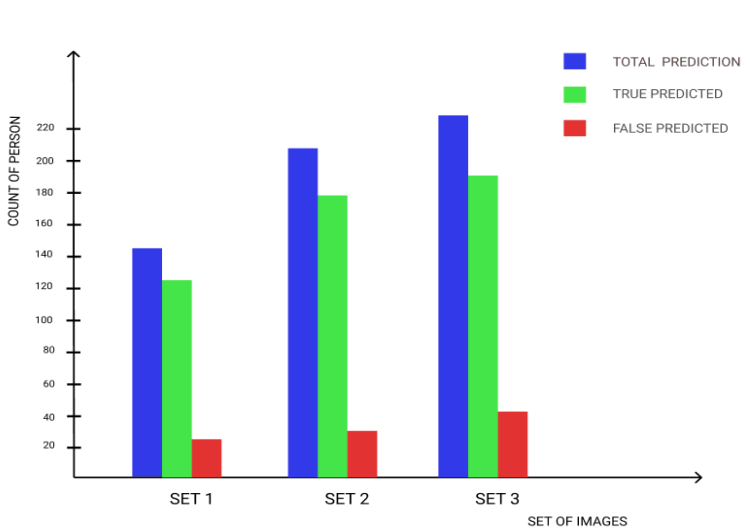

Figure10: Graph

Figure 11: Detailed Screen

Figure clearly shows the graphical analysis of the performance of the model. Accuracy of the model is 85.714 \%. For that we are taken three sets of images that are completely unique in those three sets. Each contains 30 images of distinct count of persons. Here set 1, set 2 and set 3 contain 146, 210 and 225 respectively. Once these are passed as an input to the model, it plainly finds 126, 178 and 194 persons accurately in each set of images. Apart from that the model also has false predictions.
False predictions are 20, 32 and 31 respectively.

REFERENCES

[1] D. Fehr, R. Sivalingam, V. Morellas, N. Papanikolopoulos, O. Lotfallah, and Y. Park, "Counting people in groups," 6th IEEE Int. Conf. Adv. Video Signal Based Surveillance, AVSS 2009, pp. 152–157, 2009, doi: 10.1109/AVSS.2009.55.

[2] H. Fernando, I. Perera, and C. De Silva, "Real-time Human Detection and Tracking in Infrared Video Feed," MERCon 2019 - Proceedings, 5th Int. Multidiscip. Moratuwa Eng. Res. Conf., pp. 111–116, 2019, DOI: 10.1109/MERCon.2019.8818862.

[3] Y. L. Hou and G. K. H. Pang, "People counting and human detection in a challenging situation," IEEE Trans. Syst. Man, Cybern. Part ASystems Humans, vol. 41, no. 1, pp. 24–33, 2011, DOI: 10.1109/TSMCA.2010.2064299.

[4] D. Konings and F. Alam, "LifeCount: A Device-free CSI-based Human Counting Solution for Emergency Building Evacuations," 2020 IEEE Sensors Appl. Symp. SAS 2020 - Proc., pp. 1–5, 2020, DOI: 10.1109/SAS48726.2020.9220032.

[5] P. Lestari and H. P. Schade, "Human Detection from RGB Depth Image using Active contour and Grow-cut Segmentation," 2019 Int. Conf. Comput. Control. Informatics its Appl. Emerg. Trends Big Data Artif. Intell. IC3INA 2019, pp. 70–75, 2019, DOI: 10.1109/IC3INA48034.2019.8949571.

[6] B. Li, J. Zhang, Z. Zhang, and Y. Xu, "A people counting method based on head detection and tracking," Proc. 2014 Int. Conf. Smart Comput. SMARTCOMP 2014, pp. 136–141, 2014, doi: 10.1109/SMARTCOMP.2014.7043851.

[7] S. Saxena and D. Songara, "Design of people counting system using MATLAB," 2017 10th Int. Conf. Contemp. Comput. IC3 2017, vol. 2018-January, no. August, pp. 1–3, 2018, DOI: 10.1109/IC3.2017.8284344.

[8] A. Shehzed, A. Jalal, and K. Kim, "Multi-Person Tracking in Smart Surveillance System for Crowd Counting and Normal/Abnormal Events Detection," 2019 Int. Conf. Appl. Eng. Math. ICAEM 2019 - Proc., pp. 163–168, 2019, doi: 10.1109/ICAEM.2019.8853756.

[9] L. Teng, F. Xue, and Q. Bai, "Rotation Invariance-Based HOG for Multi-posture Human Detection," 2019 5th Int. Conf. Control. Autom. Robot. ICCAR 2019, pp. 305–310, 2019, DOI: 10.1109/ICCAR.2019.8813709.

[10] F. Wang, F. Zhang, C. Wu, B. Wang, and K. J. Ray Liu, "Passive People Counting using Commodity WiFi," IEEE World Forum Internet Things, WF-IoT 2020 - Symp. Proc., pp. 1–6, 2020, DOI: 10.1109/WF-IoT48130.2020.9221456.