International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 04 | Apr 2021 www.irjet.net p-ISSN: 2395-0072

AI EYE PROGNOSTICATOR

Karthik Pillai[1], Mohit Saini[2], Satyam Yadav[3]

Information Technology, VPPCOE&VA, Mumbai University, Mumbai, India.

Abstract - In this 21st century, we see CCTV cameras all around us whether it be our houses, schools, colleges, society, hospitals or on the roads. The raw data collected by these CCTV cameras are not used to the fullest in the current scenario. We go through the CCTV footage only after some mishap. We are just creating a huge amount of data everyday but not using that data for any further use. Aim of this project is to implement a system which is capable of extracting information from the CCTV footage by using human detection algorithms and human recognition algorithms to derive valuable insights from the footage. The system uses YOLOv3 human detection algorithm (CNN Based Algorithm) to detect humans. This data is further used by Machine Learning algorithms to make predictions regarding the human population density at a particular location, human count and human recognition.

Key Words: Human Detection, Human Recognition, CNN, YOLOv3, Machine Learning.

1. INTRODUCTION

In AI Eye Prognosticator we develop a system which uses computer vision & machine learning algorithms to detect humans & different objects. This data is then used to get useful insights from the humans and objects data. For different objects vehicle detection is also implemented. This data is further used to make predictions like space utilization, people density at a point of time, number of people incoming and outgoing and number of vehicles incoming and outgoing. The data stored is also used to analyse how the space is utilized in a supermarket, mall, shop, highway and road. By using different machine learning algorithms these data can be used to predict customer or human behaviour in crowded places, supermarkets, malls and shops.

2. RELATED WORK

The related work section is divided into three sections, each focusing on a specific topic within this manuscript. Section A starts by discussing literature on human detection in challenging situations, where section B continues on human detection in surveillance videos. Finally, section C focuses on the shopping behaviour analysis.

2.1 Human detection in a challenging situation

Reliable 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 therefore the image resolution must be high. This paper aims to develop an efficient method for estimating the amount of individuals and locate each individual during a low resolution image with complicated scenes. The contribution of this paper is threefold. First, post processing steps are performed on background subtraction results to estimate the amount of individuals during a complicated scene, which incorporates people that are moving only slightly. Second, an Expectation Maximization (EM)-based method has been developed to locate individuals during a coffee resolution scene. In this method, a replacement cluster model is employed to represent everyone within the scene. The method doesn't require a really accurate foreground contour. Third, the amount of individuals is employed as a priori for locating individuals supported feature points. Hence, the methods for estimating the amount of individuals and for locating individuals are connected. The developed methods are validated supported a 4-hour video, with the quantity of people within the scene ranging from 36 to 222. The best result for estimating the amount of individuals has a mean error of 10% over 51 test cases.

2.2 Human detection in surveillance videos

Detecting citizenry accurately during a visible television is crucial for diverse application areas including abnormal event detection, human gait characterization, congestion analysis, person identification, gender classification and fall detection for elderly people. The detection process starts with the detection of objects which are in motion. Background subtraction, optical flow and spatio-temporal filtering techniques can be used to perform object detection. Once detected, a moving object could be classified as an individual's being using shape-based, texture-based or motion-based features. A comprehensive review with comparisons on available techniques for detecting citizenry in surveillance videos is presented during this paper. The characteristics of few benchmark datasets also because the longer-term research directions on human detection have also been discussed.

2.3 Shopping behaviour analysis

In shopping behaviour analysis, by getting the count of individuals in shops or similar environments helps us to get valuable insights for operators of the store and provide us with key information about that store's layout (e.g. most frequently visited areas). Instead of using extra working staff, automated solutions are preferred more. These automated systems should be cost-effective, preferably on lightweight embedded hardware, add very challenging situations (e.g. handling occlusions) and preferably work real-time. This challenge was solved by implementing real-time TensorRT optimized YOLOv3-based pedestrian detector, on a Jetson TX2 hardware platform. By combining the detector with a sparse optical flow tracker, we assign a singular ID to every customer and tackle the matter of losing partially occluded customers. The detector-tracker based solution we used achieves a mean precision of 81.59% with a processing speed of 10 FPS.

3. METHODOLOGY

Object detection could also be a field of Computer Vision and Image Processing that deals with detecting instances of various classes of objects (like a private, book, chair, car, bus, etc.) during a digitally captured Image or Video. This domain is split into many various sub-domains like Activity recognition, Face detection, Image annotation, etc. There are many important areas of applications for Object Detection like Object Tracking, Self-Driving cars, Video Surveillance, robots, etc.

3.1 Challenges

1. Variable Number of Objects
Object Detection is that the matter of locating and classifying a variable number of objects during a picture. The important thing is that the "variable" part the number of objects to be detected could vary from image to image. Therefore, the most problem associated with this is often that in Machine Learning models, we usually need to represent the data in fixed-sized vectors. Since the quantity of objects within the Image is unknown to us beforehand, we'd not know the proper number of outputs which we may require some post-processing which adds up the complexity.

2. Multiple Spatial Scales and Aspect Ratios
The Objects within the pictures are of multiple spatial scales and aspect ratios, there could even be some objects that cover most of the image and yet there will be some we'd want to hunt out but are as small as a dozen pixels (or a very small percentage of the Image). Even the same objects can have different Scales in several images. These varying dimensions of objects pose a problem in tracking them down. Some algorithms use the concept of sliding windows for the aim but it's extremely inefficient.

3. Modeling
Object Detection and Object Localization are two approaches which are required for object detection. Not only we might wish to classify the thing but we also want to locate it inside the Image. to affect these, most of the researchers use multi-task loss functions to penalize both misclassification errors and localization errors. because of this duality behavior of the loss function, repeatedly it lands up performing poorly in both.

4. Limited Data
The limited amount of annotated data currently available for object detection is another hurdle within the method. Object detection datasets typically contain annotated examples for about dozen to 100 classes while image classification datasets can include up to 100,000 classes. Gathering rock bottom truth labels in conjunction with the bounding boxes for each class remains a very tedious task to unravel.

5. Speed for Real-Time detection
Object detection algorithms need to be not only accurate in predicting the category of the thing in conjunction with its location, but it also must be incredibly fast in doing of those things to cope-up with the wants of the real-time demands of video processing. Usually, a video is shot at almost 24 fps and to make an algorithm which can achieve that frame rate is kind of a difficult task.

3.2 YOLO Algorithm

YOLO, also referred to as You Only Look Once is one among the foremost powerful real-time object detection algorithms. It's called that way because unlike previous object detector algorithms, like R-CNN or its upgrade Faster R-CNN, Single Shot MultiBox Detector (SSD), Retina-Net, it only needs the image (or video) to pass just one occasion through its network.

YOLO divides the given image into a grid of n*n cells (13*13):

Fig -1.1

Each of those cells is liable for predicting 5 bounding boxes. A bounding box is nothing but the rectangular box that encloses an object.

YOLO also outputs a confidence score that tells us how certain it's that the anticipated bounding box actually encloses some object. This score doesn't say anything about what quiet object is within the box, just if the form of the box is any good.

The predicted bounding boxes may look something just like the following (the higher the arrogance score, the fatter the box is drawn):

Fig -1.2

For each bounding box, the cell also predicts a category, it gives a probability distribution over all the possible classes. The version of YOLO we're using is trained on the COCO dataset. COCO stands for Common Objects in Context which may detect 80 different classes such as: bicycle, boat, car, cat, dog, person then on

COCO is a very large-scale image dataset. It includes annotations for image segmentation, object detection, image labelling and key points. The COCO team itself prepares these segments, labels, key points and lots of more that's why COCO is reliable to use and enables us to make robust models. COCO has several features:

1. Object segmentation
2. Recognition in context
3. Super pixel stuff segmentation
4. 330K images (>200K labeled)
5. 1.5 million object instances
6. 80 object categories
7. 91 stuff categories
8. 5 captions per image
9. 250,000 people with key points

The confidence score for the bounding box and therefore the class prediction is combined into one final score that tells us the probability that this bounding box contains a selected sort of object. for instance, the large fat yellow box on the left is 85% sure it contains the thing "dog":

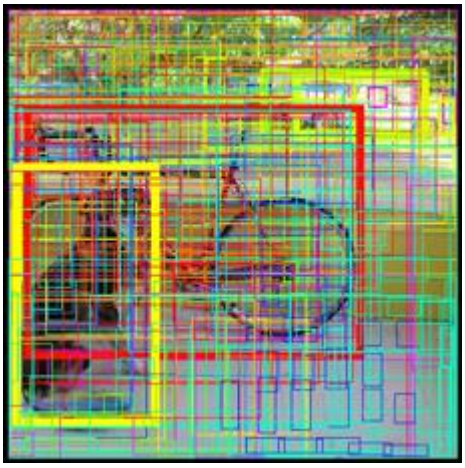

Fig -1.3

Since there are 13×13 = 169 grid cells and every cell predicts 5 bounding boxes, we find yourself with 845 bounding boxes in total. It seems that the majority of those boxes will have very low confidence scores, so we only keep the boxes whose final score is 30% or more (you can change this threshold counting on how accurate you would like the detector to be).

The final prediction is then:

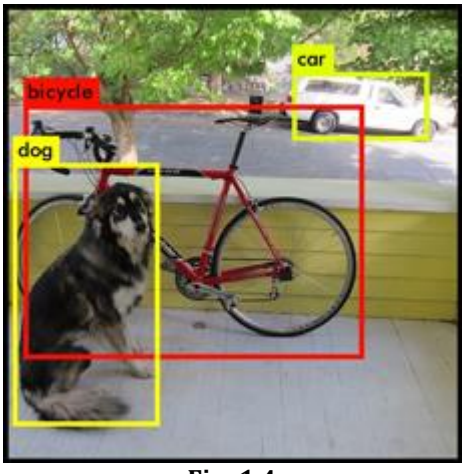

Fig -1.4

From the total 845 bounding boxes, only these three boxes are kept because they are giving us the simplest results. But note that albeit there have been 845 separate predictions, they were all made at an equivalent time — the neural network just ran once. This is what makes the YOLO algorithm so fast and powerful.

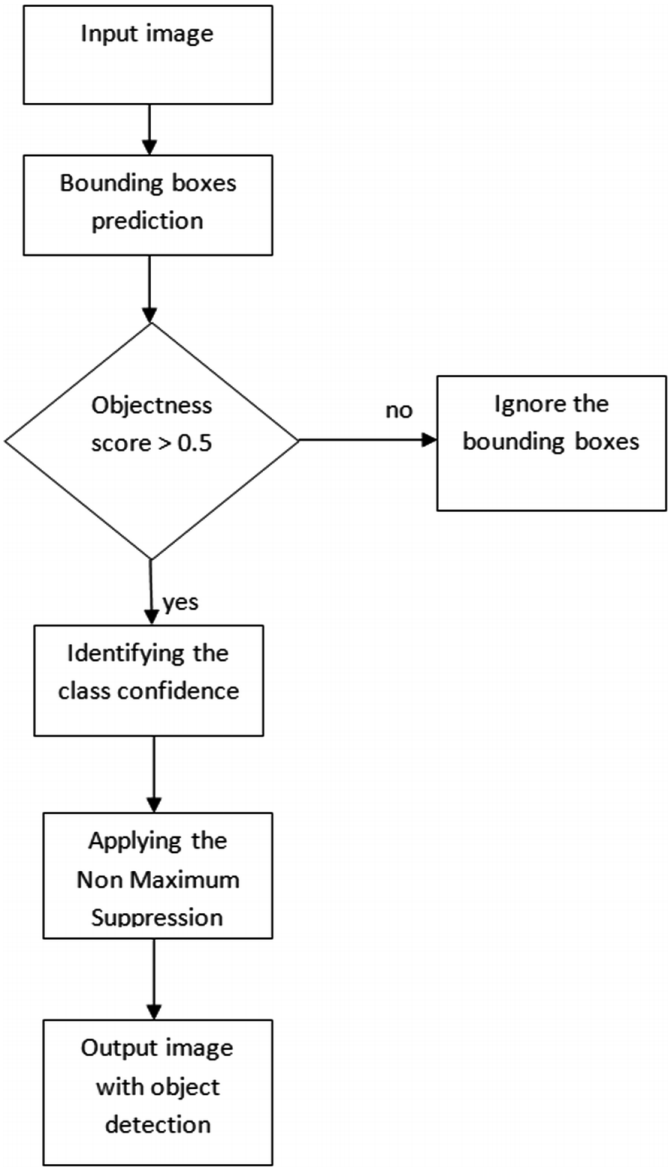

Fig -2: Flow chart of YOLO Algorithm

4. APPLICATIONS

The proposed system can be used:

1. To get insights like a population density at a shop inside the shopping mall.
2. To get a number of people entering into and exiting the mall.
3. To get data like which product is least visible to consumers at a retail shop.
4. To make optimum use of space left over at a retail store or at malls.

5. CONCLUSION

In this paper, we present a system which is capable of extracting valuable information from the existing data which is collected by the CCTV cameras around us. AI Eye Prognosticator can provide a solution for mapping the flow of customers in real-time based on human detection and human recognition, implemented on a powerful and efficient platform. This system is useful in understanding the human density in a particular space. It is useful in deriving important business insights in an off-line store. This system can do the task of managing the space effectively & efficiently. Thus, AI Eye Prognosticator can understand the people movement pattern in a place & make optimized utilization of space.

ACKNOWLEDGMENTS

The author acknowledges support by 'Ms. Supriya Chaudhary', mentor from Department of Information Technology Engineering, VPPCOE&VA, University of Mumbai, India, for providing this opportunity to work on real-life based research. Much gratitude is expressed to the open internet sources that have enhanced the knowledge, motivation and learnings.

REFERENCES

1. Ya-Li Hou, Student Member, IEEE, and Grantham K. H. Pang, Senior Member, IEEE, "People Counting and Human Detection in a Challenging Situation", IEEE Transactions on systems, man and cybernetics part A systems and humans, vol. 41, no. 1, January 2011.
2. Manoranjan Paul, Shah M E Haque and Subrata Chakraborty, "Human detection in surveillance videos and its applications - a review", Paul et al. EURASIP Journal on Advances in Signal Processing 2013, 2013:176 http://asp.eurasipjournals.com/content/2013/1/176.
3. Xiaohang Xu, Dongming Zhang, and Hong Zheng, "Crowd Density Estimation of Scenic Spots Based on Multifeature Ensemble Learning", Hindawi, Journal of Electrical and Computer Engineering Volume 2017, Article ID 2580860, 12 pages.
4. Joanna M. Burger, Frances E.C. Stewart, John P. Volpe, Jason T. Fisher, A Cole Burton, "Estimating density for species conservation: Comparing camera trap spatial count models to genetic spatial capture-recapture models", Global Ecology and Conservation, July 2018.
5. G. Sreenu and M. A. Saleem Durai, "Intelligent video surveillance: a review through deep learning techniques for crowd analysis", Sreenu and Saleem Dural Big Data (2019) 6:48 https://doi.org/10.1186/s40537-019-0212-5.
6. R Schrijvers S. Puttemans, T. Callemein', and T. Goedemé, "Real-time Embedded Person Detection and Tracking for Shopping Behaviour Analysis", EAVISE, KU Leuven, Jan De Nayerlaan 5, 2860 Sint-Katelijne-Waver, Belgium PXL Smart-ICT,

Hogeschool PXL, Elfde Liniestraat 24, 35000 Hasselt, Belgium.

7. Volker Eiselein, Ivo Keller, Hajer Fradi, Thomas Sikora, "Enhancing human detection using crowd density measures and an adaptive correction filter", Conference Paper August 2013, DOI: 10.1109/AVSS.2013.6636610.
8. Mayur D. Chaudhari, Archana S. Ghotkar, "A Study on Crowd Detection and Density Analysis for Safety Control", JSCE International Journal of Computer Sciences and Engineering, April,2018.
9. V. Keerthi Kiran, Priyadarsan Parida and Sonali Dash, "Vehicle Detection and Classification: A Review", Chapter - January 2021.
10. Sriashika Addala, Dept of CSE, LPU Punjab, "Research paper on vehicle detection and recognition", Experiment Findings May 2020.