Journal of Applied Information Science
10 (1) 2022, 34-38
http://www.publishingindia.com/jais

A Web-Based Application for Monitoring Crowd Status in Stores

Salbin Sadik[1], Tushar Balakrishnan V.[2*], Vinayak M. R.[3] and Jain Stoble B.[4]
[1]Department of Computer Science & Engineering, Adi Shankara Institute of Engineering and Technology, Kerala, India. Email: shalbin.sadik@gmail.com
[2]Department of Computer Science & Engineering, Adi Shankara Institute of Engineering and Technology, Kerala, India. Email: tusharbala284@gmail.com
[3]Department of Computer Science & Engineering, Adi Shankara Institute of Engineering and Technology, Kerala, India. Email: sreevinayakam057@gmail.com
[4]Department of Computer Science & Engineering, Adi Shankara Institute of Engineering and Technology, Kerala, India. Email: jain.cs@adishankara.ac.in
*Corresponding Author

Abstract: Crowd counting and surveillance in public spaces and stores is a difficult task to complete. Because of the large population and variety of human activities, crowded scenes must be more frequent. The requirement for a mechanism to track crowd counting in certain periods, such as during the COVID outbreak, is critical. In this COVID scenario, large crowds without keeping social distance and without any measures can easily spread the COVID-19 virus. Store personnels may find it difficult to handle large crowds when a large group of people gathers at a store to make a purchase. This study proposes a method that can be used in stores to assist customers to plan their shopping trips properly, maintaining social distance. This software primarily aids users in locating shops in their area that they wish to visit, as well as offering information on the quietest and busiest times of the day to visit each shop. The user interface of the system is designed in such a way that the search function and GPS function aid you in identifying nearby retailers as well as monitoring the state of local establishments to determine if they are crowded or not.

Keywords: Counting, Crowd, Distance, GPS, Retailers, Shopping, Store, Surveillance, Visit.

I. Introduction

In conventional shopping, customers usually strive to get the greatest things, regardless of the crowds in the stores, and then pay for them. Customers may no longer be forced into crowded stores as a result of the COVID-19 pandemic. It is tough for them to shop in these actual establishments. These factors have an impact on consumer behaviour. While internet buying is convenient during these COVID times, it is not the ideal option for routine purchases. It is always in the best interests of the consumers if there is a method for them to go shopping while avoiding crowds and maintaining social distance.

In stores, we mostly use video surveillance from CCTV to analyse crowds, which is an important component of modern urban security. When combined with computational analytics, this functionality may be substantially increased, including facial recognition, motion detection, traffic and crowd monitoring, and more. Traditional manual monitoring of CCTV footage and human patrols to conduct crowd monitoring and tracking remain popular in the security and management sectors.

In industry and research, big data applications are taking up the majority of the space. Video streams from CCTV cameras play an equally essential part in big data as other sources such as social media data, sensor data, farm data, medical data, and data derived from space research. Surveillance videos contribute significantly to unstructured big data. CCTV cameras are installed in all areas where security is a priority. Motion information extraction and anomalous behaviour modelling are two challenges in crowd behaviour analysis utilising video surveillance. Crowd tracking is usually what the former entails. It's a method of estimating the crowd's speed, direction, and location in a video sequence. Anomaly occurrences can be detected using higher-level models of crowd behaviour. Individual tracking in a crowd is not the same as crowd movement tracking. When tracking individuals, data is computed at the individual level. When we talk about crowd tracking, though, we're talking about the mobility of a crowd as a collection of little pieces whose structure is always changing. han following the same character throughout the video. The use of crowd monitoring to create models of crowd behaviour and detect aberrant behaviours at the crowd level rather than at the individual level is one application of crowd monitoring. Crowd analysis is also used in crowd simulation, crowd management, disaster management, and planning, among other things. In order to distinguish individuals, cars, objects, and events, video surveillance employs computer software programmes that analyse the sounds and visuals from video surveillance cameras.

A non-rule based kind of A.I. for security dubbed "behavioural analytics" has been developed for video surveillance. This software is completely self-learning, requiring no programming input from the user or security contractor at the outset. Based on its own observations of patterns of various characteristics such as size, speed, reflectivity, colour, grouping, vertical or horizontal orientation, and so on, the A.I. learns what is usual behaviour for people, cars, equipment, and the environment in this sort of analytics.

II. Previous Work

This research necessitates the counting of people in a crowd. In most cases, sensors are used to determine the count, but with the rise in popularity of video surveillance systems, they are becoming more common. The number of people in stores is calculated using CCTV footage from those stores. Each individual in a crowd is recognized as an object, necessitating the use of a variety of object detection techniques. A variety of object detection algorithms have been used in the past in a number of studies. Object detection is the process of locating objects of a specific type within an image.

In this study, Sindagi and Patel (2018) compared 27 crowd counting methodologies. Traditional approaches and convolutional neural network (CNN)-based approaches are the two sorts of ways they classify. Traditional methods recognize humans by extracting handmade elements from a picture, such as face characteristics, head and interest areas. Traditional techniques can be divided into three categories: detection, regression, and density estimation. While CNN-based systems measure the number of persons in an image by leveraging CNN's technological breakthroughs. The network property and training methodology may be used to further classify CNN-based approaches. Experiments (Sindagi & Patel, 2018) reveal that CNN-based techniques outperform other approaches in high-density crowds when object size and scene perspective are varied.

In their survey, Ryan *et al.* (2015) assessed 22 crowd counting methods and classified them as holistic, intermediate, or local. The holistic approaches use global picture attributes to describe each image, and then use a regression or classification model to map these attributes with a crowd size estimate. Local techniques, on the other hand, use local elements in an image to identify individuals and then collect them as a result. The intermediate ones collect information about nearby items into histogram bins, which are then displayed holistically.

Saleh *et al.* (2015) examined a number of crowd counting and density estimate methods. They divide the approaches under consideration into two categories: direct and indirect. The direct method (object-based target detection) counts people by detecting specific segments in a crowd and then adding them up.

Though the three classifications have distinct names, they are all identical in terms of crowd counting methods (i.e. how people are counted) and whether or not neural networks are used. Traditional approaches (Sindagi & Patel, 2018), for example, are similar to direct methods (Saleh *et al.*, 2015) and local approaches (Saleh *et al.*, 2015). (Ryan *et al.,* 2015). CNN-based methods are similar to indirect methods and holistic approaches in terms of their classification. In order to pick methods for counting the crowd size, we use the classification proposed by Saleh *et al.* (2015) in this study, namely direct and indirect approaches.

III. Methodology

The software we want to build analyses real-time CCTV footage to detect human objects, with the object count serving as a shop's threshold value, and a graphical user interface designed to make it simple for people to operate the software. This real-time CCTV footage must be run through some sort of object identification algorithm in order to acquire a count of persons. To do this, the object recognition algorithm must be both quick and efficient, as well as give accurate results.

The YOLO (You Only Look Once) Algorithm is used for object identification. It is a network that detects objects. Object detection requires discovering and recognizing certain objects in an image. YOLO processes the whole image in a single instance, and then outputs a vector of bounding boxes and class predictions using a neural network that mimics a typical CNN. The most significant benefit of adopting YOLO is its incredible speed; as it can analyse 45 frames per second. YOLO understands the idea of generic object representations as well.

YOLO's Mode of Operation

A single neural network projects several bounding boxes and class probabilities for those boxes at the same time. YOLO improves detection performance by training on the entire photo. During training and testing, YOLO sees the complete image, so it implicitly stores contextual information about classes as well as their appearance. To forecast each bounding box, the network takes characteristics from the entire image. It also predicts all bounding boxes for an image across all classes at the same time. The input image is divided into a SxS grid by the system. If the item's centre falls inside a grid cell, that grid cell is in charge of detecting the object. It mainly uses CNN to detect images.

A. Network Architecture

The CNN mainly uses different layers in our Yolo framework as shown below:

Convolutional Layer: The convolutional layer is the first. A filter (also known as a kernel) is applied to the input of the convolutional layer. A filter, which is also an array of numbers, projects itself on a region of the input image (the receptive field) and glides across all of the picture's areas. Astride is a single filter step or slide that varies depending on the CNN. Prior to being applied to the input, the filter is specifically switched. This filter's depth must match the depth of the input. By computing element-by-element multiplications, the result is obtained for each slide. These numbers are added together to form a single number that is entered into the appropriate field.

Pooling Layer: The feature map is then sent to the pooling layer or down sampling layer after it has been updated. The pooling layer, like the convolutional layer, reduces the number of parameters and computations required to process the data in the network, minimizing the spatial size of the representation. Only the most important features are saved, and the model's effective training process can continue. Aside from that, the reduced number of parameters prevents the model from overfitting. When your model is in the evaluation phase, it performs well on the training set but poorly on the test set, this is known as overfitting.

Fully Connected Layer: The model can now grasp the features thanks to the hidden layers, and we can begin the classification process. The pooling layer's output is flattened into a single long vector and sent through a fully connected layer, which is a feed-forward neural network (and backpropagation is applied to every iteration of training). Each value in this vector denotes the likelihood that a particular feature belongs to one of the classes. Thanks to the hidden layers, the model can now understand the features, and we can begin the classification process. The output of the pooling layer is flattened and transmitted through a fully connected layer, which is a feed-forward neural network (and backpropagation is applied to every iteration of training). The likelihood that a certain feature belongs to one of the classes is represented by each value in this vector.

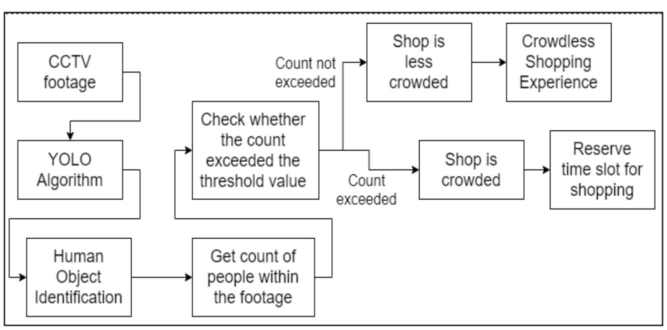

Fig. 1: Working of BusyFree Application

IV. Results

The YOLO algorithm is used to process images or videos captured by retail CCTV cameras for human object identification. It enables the model to examine the entire image at test time, allowing it to make predictions based on the picture's whole context.

It begins by dividing the picture into a 13x13 cell grid. Depending on the amount of the input, the size of these 169 cells varies. The cell size was 32 x 32 for a 416 x 416 input size that we utilized in our studies. After that, each cell is in charge of estimating a certain number of boxes in the picture. The network also predicts the confidence that a bounding box truly encloses an object and the likelihood that the enclosed object belongs to a certain class for each bounding box. The majority of these bounding boxes are removed because they have a low confidence score or because they enclose the same item as another bounding box with a high confidence score. The object is recognized as human if the confidence score is greater than 0.5.

Fig. 2(a): In the Picture above, Objects with a Confidence Score more than 50% are Identified as Humans

In Fig. 2(a), In this case, the confidence score is critical for identifying human objects inside a visual so that the system can predict if the count has exceeded the threshold number, which is the most important component in determining whether the shop is crowded or not.

Fig. (2b): Not Crowded

In Fig. 2(b), the business does not appear to be busy since it has not reached the threshold value, allowing consumers to enjoy a crowd-free shopping experience.

Fig. 2(c): Crowded

In Fig. 2(c), since the threshold amount has been exceeded, the shop looks to be packed.

Crowded shopping may not lessen the chance of the pandemic spreading. Hence, the user may reserve a time window so that they may shop without having to deal with crowds.

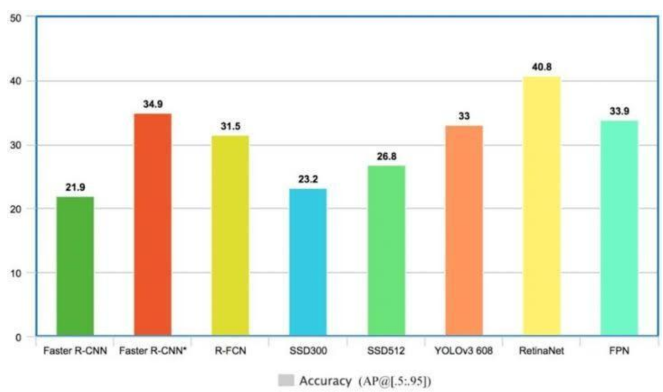

Fig. 2(d): Accuracy Comparison of Different Object Detection Algorithms

In Fig. 2(d), the accuracy of numerous algorithms other than YOLO is compared.

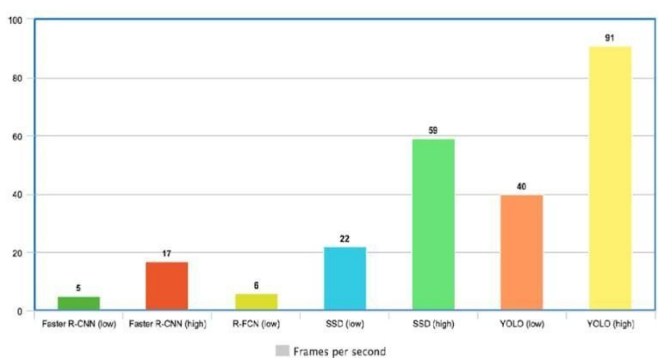

Fig. 2(e): On the Basis of Processing Speed Frames Per Second, YOLO was Compared to Various Different Object Detection Methods

V. Conclusion

It's difficult to track and monitor crowds in public spaces and commercial establishments. As a result of the vast population and diversity of human activities, crowded conditions will grow more common. Using our web-based application (BusyFree App), we can identify the crowd pattern at a certain store, allowing us to avoid a crowded shopping experience. The proposed system's major function is to determine the crowd status in nearby shops that the user wants to visit in order to determine whether the shop is crowded or not at a given time. Users can also reserve a spot in the shop if it is not too busy. The system primarily functions by capturing real-time CCTV footage in the store. The CCTV data is fed into our software, which uses image processing and image recognition to calculate the number of people in the store. By entering a threshold value based on the shop's size, the administrator determines whether the shop is crowded or not. The user interface of the app is designed so that the user can see the crowd status of nearby stores, and if the store he wants to visit is crowded, he can enable a notification option that will send him a message when it is not crowded. Real-time footage from the shop's CCTV camera is primarily used to determine the crowd count. The threshold count derived from CCTV footage is used to determine whether or not the shop is crowded. People entering the store are detected using the OpenCV algorithm.

The OpenCV algorithm that was used was YOLO. It is a regression-based approach that predicts classes and bounding boxes for the entire picture in a single run of the algorithm, rather than picking relevant areas of an image. Object detection is the task of locating and classifying specific objects in an image. With a face detection accuracy of 95%, we take an image as input, run it through a neural network which is similar to a regular CNN, as an output, you'll get a vector of bounding boxes and class predictions.

In the event of another pandemic outbreak, this method can be used at that time as well. People are likely to practise social distancing to prevent the disease from spreading. As a result, using this method would be beneficial for a secure shopping experience. It is easy to use the application. This web-based application is also compatible with mobile phones. On mobile devices, the app is responsive. This app contains a chatbot function that assists the user in resolving any questions they may have about the app. The software supports both user and shop authentication. The app returns stores close to the user's current location using the search function. A database keeps track of the supermarket's characteristics as well as the crowd situation. The user may reserve a spot in the store and see how crowded it is. Shop owners may use the app to update the state of their company and gather information on user slot bookings, allowing for two-way communication between customers and shop owners.

References

[1] I. Culjak, D. Abram, T. Pribanic, H. Dzapo, and M. Cifrek, ‘‘A brief introduction to OpenCV,’’ in *Proc. 35th Int. Conv. MIPRO*, 2012, pp. 1725-1730.

[2] T. Ko, ‘‘A survey on behavior analysis in video surveillance for homeland security applications,’’ in *Proc. 37th IEEE Appl. Imag. Pattern Recognit. Workshop*, Oct. 2008, pp. 1-8.

[3] A. Hampapur, L. Brown, J. Connell, A. Ekin, N. Haas, M. Lu, H. Merkl, and S. Pankanti, ‘‘Smart video surveillance: Exploring the concept of multiscale spatiotemporal tracking,’’ *IEEE Signal Process. Mag*., vol. 22, no. 2, pp. 38-51, Mar. 2005.

[4] E. Grevelink, ‘‘A closer look at object detection, recognition and tracking,’’ Intel, Santa Clara, CA, USA, Tech. Rep., 2017. [Online]. Available: https://software.intel.com/en-us/articles/a-closerlook-at-object-detection-recognition-and-tracking

[5] Z. Li, K. Ishibashi, K. Takai, and K. Yada, “Shop area visit ratio, stay time, and sales outcomes: In-depth analysis based on RFID data,” *2015 2nd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE),* 2015, pp. 1-7, doi: 10.1109/APWCCSE.2015.7476231.

[6] H. Phillips, and R. Bradshaw, “How customers actually shop: customer interaction with the point of sale,” *Market Research Society Journal*, vol. 35, no. 1, pp. 1-10, 1993, doi: 10.1177/147078539303500104.

[7] N. Farooqi, “Intelligent safety management system for crowds using sensors,” *2017 12th International Conference for Internet Technology and Secured Transactions (ICITST),* 2017, pp. 144-147, doi: 10.23919/ICITST.2017.8356365.

[8] S. Rajarajeshwari, S. Archana, S. R. Boda, R. B. Sreelakshmi, K. Venkateswaran, and N. Sridhar, “A smart image processing system for hall management including social distancing - “SoDisCop”,” *2020 3rd International Conference on Smart Systems and Inventive Technology (ICSSIT),* 2020, pp. 1213-1219, doi: 10.1109/ICSSIT48917.2020.9214262.

[9] K. H. Cheong et al., “Practical automated video analytics for crowd monitoring and counting,” in *IEEE Access*, vol. 7, pp. 183252-183261, 2019, doi: 10.1109/ACCESS.2019.2958255.

[10] V. B. Subburaman, A. Descamps, and C. Carincotte, “Counting people in the crowd using a generic head detector,” *2012 IEEE 9th International Conference on Advanced Video and Signal-Based Surveillance*, 2012, pp. 470-475, doi: 10.1109/AVSS.2012.87.