



CUSTOMER BEHAVIOUR ANALYTICS USING CENTROID TRACKING ALGORITHM

ZUBIYA RAHMAN, AINAAN SIDDIQUA, MADIHA MUDDASIR, SYEDA SURRAYYA JABEEN

STUDENT

STUDENT

STUDENT

ASSISTANT PROFESSOR

COMPUTER SCIENCE ENGINEERING
ISL Engineering College, Hyderabad, India

Abstract: Understanding clients on-site, earlier than they take a look at out, has constantly been hard as control has been confined to choices primarily based totally on not on time and incomplete facts. Video analytics assist outlets to decorate the consumer purchasing enjoy with the aid of using lowering customer support time. Retail Video Analytics is the discovery, interpretation, and communicate of significant styles in facts, from the video content, and making use of one's styles toward powerful decision-making. Today retail video analytics has long gone past the conventional area of protection and loss prevention with the aid of using presenting outlets with insightful enterprise intelligence consisting of shop visitor's records and queue facts. Also, outlets use this fact to evaluate in-shop promotional techniques and degree the overall performance of aisle and product placements.

IndexTerms –Customer Analytics, MobileNet SSD, Centroid Tracking Algorithm

I. INTRODUCTION

Our venture is primarily based totally on the idea of robotically reading video streams to come across and decide temporal and spatial events. Video Analytics makes use of mathematical algorithms to monitor, examine and control massive volumes of video. The software program makes use of video photos given as entering and extracts beneficial records for the users. Our venture aims to decorate consumer enjoyment and offer beneficial insights for outlets. Our software program offers insights into predominant modules. The shop entry/go out evaluation module allows us to gain evaluation outcomes on the entry/go out factor of the shop. The 2d module is the Instore Crowd Analysis module, which in contrast to the Entry/Exit module specifically focuses to attract insights inside the shop or a selected benchmarked vicinity interior of a shop.

Advantages over the general existing systems:

- This version can go for walks over each and enter video streams and in addition to in real-time via a webcam (or a related IP camera), giving us a better FPS throughput price than R-CNNs.
- The software program person (or) the shop administrator may even allow mailer capability. By doing this, they may acquire an email alert if the quantity of humans found in the shop exceeds their unique threshold
- We have additionally supplied the capability of a log document being created at the stop of the day, which could be beneficial for footfall evaluation.
- The person also can set the timer as much as which a video runs after which stops robotically Video summarized text is displayed on the window.
- Output frames may be written to a video documentary on a neighbourhood disk (if the person specifies - - output argument throughout run-time). The output video may be recorded in any supportive layout like Mkv, mp4v, or avi.
- Automatic clean-up paintings are likewise carried out after the video time elapses.

Advantages of our approach:

- We used Single Shot Detector (SSD) with MobileNet structure - (1) SSD calls for simplest an unmarried shot to come across more than one gadget with inside the image, while local concept networks (RPN) are primarily based on strategies consisting of R-CNNs require shots, one for producing location proposals and one for detecting the item of every concept. Thus, SSD is a good deal quicker as compared to RPN-primarily based es. (2) Also, SSD achieves a better map than Faster R-CNN and YOLOv1.

- MobileNet is a DNN designed to run on resource-limited gadgets like IP cameras, smartphones, and raspberry pi, etc., supporting us to run over numerous enter video streams.
- Here we used a hybrid method that makes use of levels i.e.; item detection and item tracking, which runs simultaneously. The gain of this method is that came to follow fantastically correct item detection strategies without a good deal of a computational burden.
- While constructing item detection networks we generally use a current community structure just like the VGG or ResNet after which use it within the item detection pipeline. The trouble is that those community architectures may be very massive in size (2 hundred to 500 MB). Networks consisting of those are incorrect for resource-limited gadgets and subsequently, we used MobileNet here.

II. Data and Sources of Data

As the implementation of this software program is primarily based totally on a deep getting to know the methodology, thereby requiring us to teach our version over a neural community. We used MobileNet alongside Single Shot Detector as a pre-skilled neural community over a hybrid structure giving us the maximum correct outcomes.

To teach our version over a MobileNet over an SSD structure, we've carried out the usage of the Caffe framework.

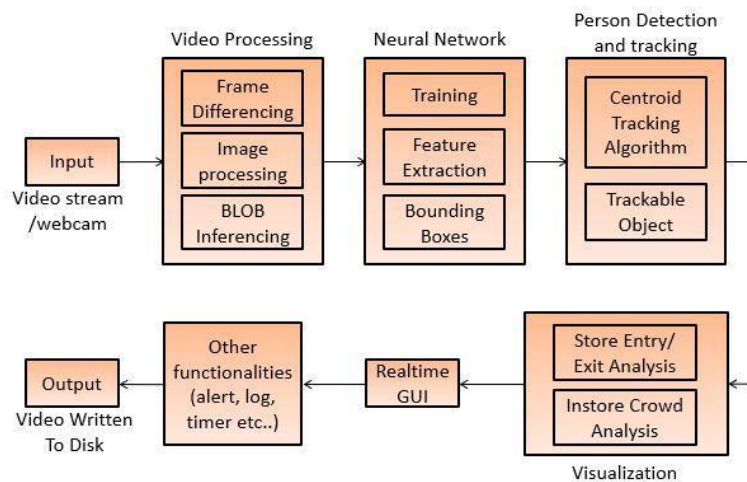


Figure 1- Architectural framework

III. Theoretical implementation of MobileNet SSD:

There are a few realistic boundaries even as deploying and going for walks on complicated and high-strength-ingesting neural networks in real-time applications. Since SSD is impartial to its base community, MobileNet changed into used because the base community of SSD to address this trouble. Rather than the usage of the usual convolution layers, it makes use of Depth's clever separable convolution layers. What makes this version stand out is that its structure lessens the computational price and really low computational strength is wanted to run or follow switch getting to know. Fig 2 offers the architectural illustration of the MobileNet-Primarily based on SSD structure.

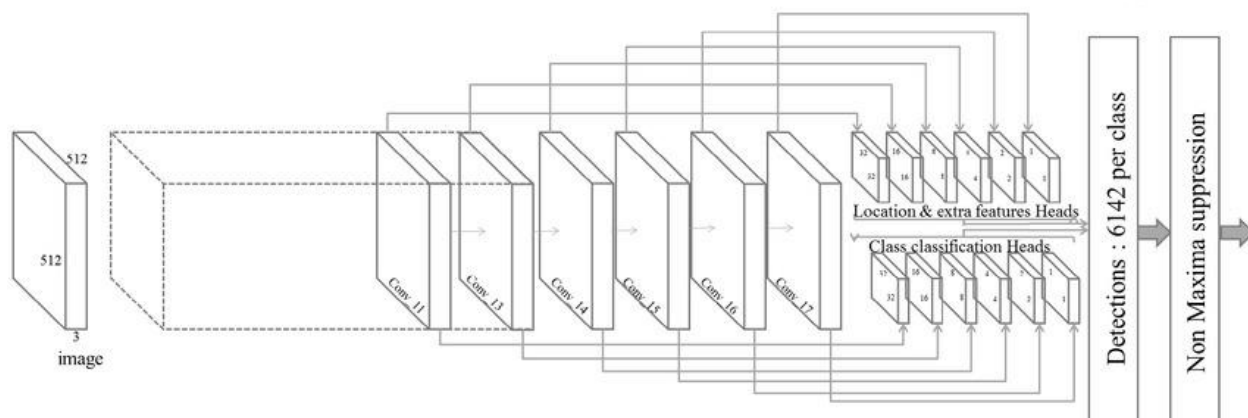


Figure 2- a pictorial representation of MobileNet based SSD architecture

When MobileNet is used alongside SSD, the previous few layers consisting of the FC, Maxpool, and Softmax are omitted. So, the outputs from the very last convolution layer within side the MobileNet are used, alongside convoluting it some extra instances to gain a stack of characteristic maps. These are then used as inputs for its detection heads. Its structure may be changed as consistent with required. Table1 beneath offers certainly considered one among its architectures in detail.

TYPE/STRIDE	FILTER SHAPE	INPUT SIZE
Conv/s2	3 x 3 x 3 x 32	300 x 300 x 3
Conv dw/s1	3 x 3 x 32 dw	150 x 150 x 32
Conv/s1	1 x 1 x 32 x 64	150 x 150 x 32
Conv dw/s2	3 x 3 x 64 dw	150 x 150 x 64
Conv/s1	1 x 1 x 64 x 128	75 x 75 x 64
Conv dw/s1	3 x 3 x 128 dw	75 x 75 x 128
Conv/s1	1 x 1 x 128 x 128	75 x 75 x 128
Conv dw/s2	3 x 3 x 128 dw	75 x 75 x 128
Conv/s1	1 x 1 x 128 x 256	38 x 38 x 128
Conv dw/s1	3 x 3 x 256 dw	38 x 38 x 256
Conv/s1	1 x 1 x 256 x 512	38 x 38 x 256
Conv dw/s1	3 x 3 x 512 dw	38 x 38 x 512
Conv/s1	1 x 1 x 512 x 512	38 x 38 x 512
Conv dw/s1	3 x 3 x 512 dw	38 x 38 x 512
Conv/s1	1 x 1 x 512 x 512	38 x 38 x 512
Conv dw/s1	3 x 3 x 512 dw	38 x 38 x 512
Conv/s1	1 x 1 x 512 x 512	38 x 38 x 512
Conv dw/s1	3 x 3 x 512 dw	38 x 38 x 512
Conv/s1	1 x 1 x 512 x 512	38 x 38 x 512
Conv dw/s1	3 x 3 x 512 dw	38 x 38 x 512
Conv/s1	1 x 1 x 512 x 512	38 x 38 x 512
Conv dw/s1	3 x 3 x 512 dw	38 x 38 x 512
Conv/s1	1 x 1 x 512 x 512	38 x 38 x 512
Conv/s2	3 x 3 x 512 x 1024	38 x 38 x 512
Conv/s1	1 x 1 x 1024 x 1024	19 x 19 x 1024
Conv/s1	1 x 1 x 1024 x 256	19 x 19 x 1024
Conv/s2	3 x 3 x 256 x 512	19 x 19 x 256
Conv/s1	1 x 1 x 512 x 128	10 x 10 x 512
Conv/s2	3 x 3 x 128 x 256	10 x 10 x 128
Conv/s1	1 x 1 x 256 x 128	5 x 5 x 256
Conv/s2	3 x 3 x 128 x 256	5 x 5 x 128
Conv/s1	1 x 1 x 256 x 128	3 x 3 x 256
Conv/s1	3 x 3 x 128 x 256	3 x 3 x 128
Conv/s1	1 x 1 x 256 x 128	1 x 1 x 256
Conv/s1	3 x 3 x 128 x 256	1 x 1 x 128

3.1 Video Processing:

An input video stream can be considered as a set of multiple frames, where each frame is an individual image. Before training the input video over the neural network, we have pre-processed and optimized the individual frames to obtain better training/testing results.

3.2 BLOB Inferencing:

OpenCV's deep neural network (dnn) module contains two functions that can be used for preprocessing images and preparing them for classification via pre-trained deep learning models. OpenCV provides two functions to facilitate image preprocessing for deep learning classification:

`Cv2.dnn.blobFromImage` and `cv2.dnn.blobFromImages`

These two functions perform 1. Mean subtraction 2. Scaling 3. And optionally channel swapping

3.3 Detection and Tracking:

1. Person Detection

During the detection phase, we run the detector to, (1) Detect if new people have entered the view. (2) See if we can find people that were lost during the tracking phase.

For each detected person we create or update the person tracker with new bounding box coordinates. Person detection is computationally expensive, hence we only run this phase once every N frame.

2. Person Tacking:

When we are not in the detection phase, we are in the tracking phase. For each of our detected persons, we create a person tracker to track the person as it moves around the frame. We will continue tracking until we reach the Nth frame and then re-run our person detector. For this, we have created an algorithm called the "Centroid Tracking Algorithm."

- **Centroid Tracking Algorithm:**

Step 1 - Firstly we accepted bounding box coordinates from the mobilenetSSD object detector's output layer. Then we computed the centroid of the bounding box. Since these are the first initial set of bounding boxes presented to our algorithm, we assigned them with unique IDs.

Step 2 - For every successive frame in our video stream, we applied the above step (i.e.; step 1). But instead of assigning a unique ID to every detected person, first, we determined if we can associate the new people centroids with the old people centroids. To accomplish this process, we computed the Euclidian distance between each pair of existing people centroids and input people centroids.

Step 3 - The centroid tracking algorithm assumes that pairs of centroids with minimum Euclidean distance between them must be the same person ID. Using this we updated the (x, y) coordinates of existing people.

Step 4 - While performing the previous step, on many occasions, we incurred more input detections than existing people being tracked. Hence, we registered the new person with its corresponding new person ID.

Step 5 - If a person has been lost or has left the field of view, we simply deregistered the person.

We repeated steps 2 to 5 for every frame in the video.

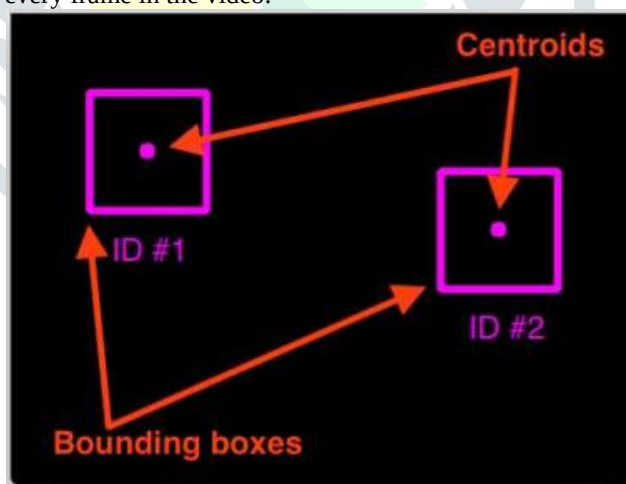
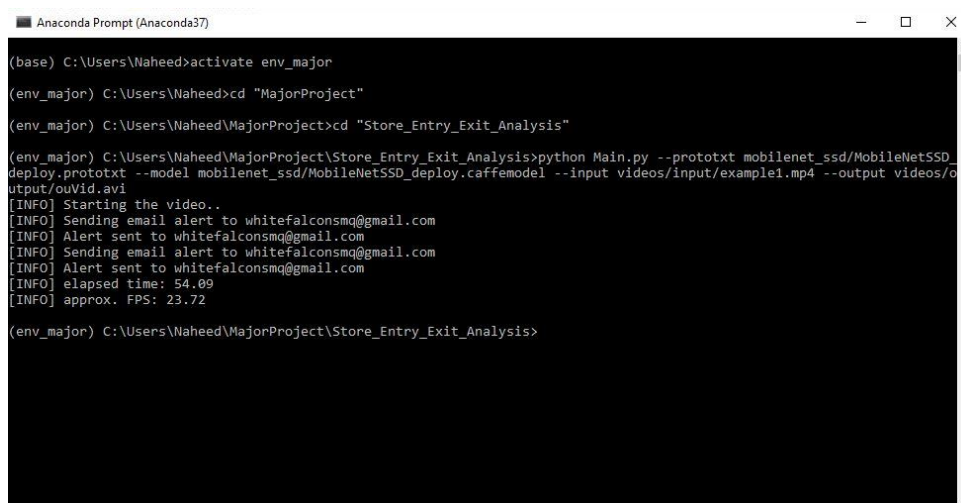


Figure 3- Centroid Tracking

IV. RESULTS

Specify the --input argument leading to the corresponding path of the pre-recorded input video (or) if you want to enable the webcam to specify URL=0 in the config file.



```

Anaconda Prompt (Anaconda37)

(base) C:\Users\Naheed>activate env_major
(env_major) C:\Users\Naheed>cd "MajorProject"
(env_major) C:\Users\Naheed\MajorProject>cd "Store_Entry_Exit_Analysis"
(env_major) C:\Users\Naheed\MajorProject\Store_Entry_Exit_Analysis>python Main.py --prototxt mobilenet_ssd/MobileNetSSD_deploy.prototxt --model mobilenet_ssd/MobileNetSSD_deploy.caffemodel --input videos/input/example1.mp4 --output videos/output/ouVid.avi
[INFO] Starting the video..
[INFO] Sending email alert to whitefalconsmq@gmail.com
[INFO] Alert sent to whitefalconsmq@gmail.com
[INFO] Sending email alert to whitefalconsmq@gmail.com
[INFO] Alert sent to whitefalconsmq@gmail.com
[INFO] elapsed time: 54.09
[INFO] approx. FPS: 23.72
(env_major) C:\Users\Naheed\MajorProject\Store_Entry_Exit_Analysis>
  
```

Figure 4- Anaconda run-time terminal.

4.1 Visualizing using GUI

With the help of a Graphical User Interface, we can easily display the output of our algorithm on the screen. The admin can monitor the results in real-time as well as gather useful insights from the text displayed on the video. The output can be permanently written on the local disk specified by the admin during run-time. In addition to the GUI useful functions like the Mailer, Log File, Timer and Scheduler add to the usefulness of our software.

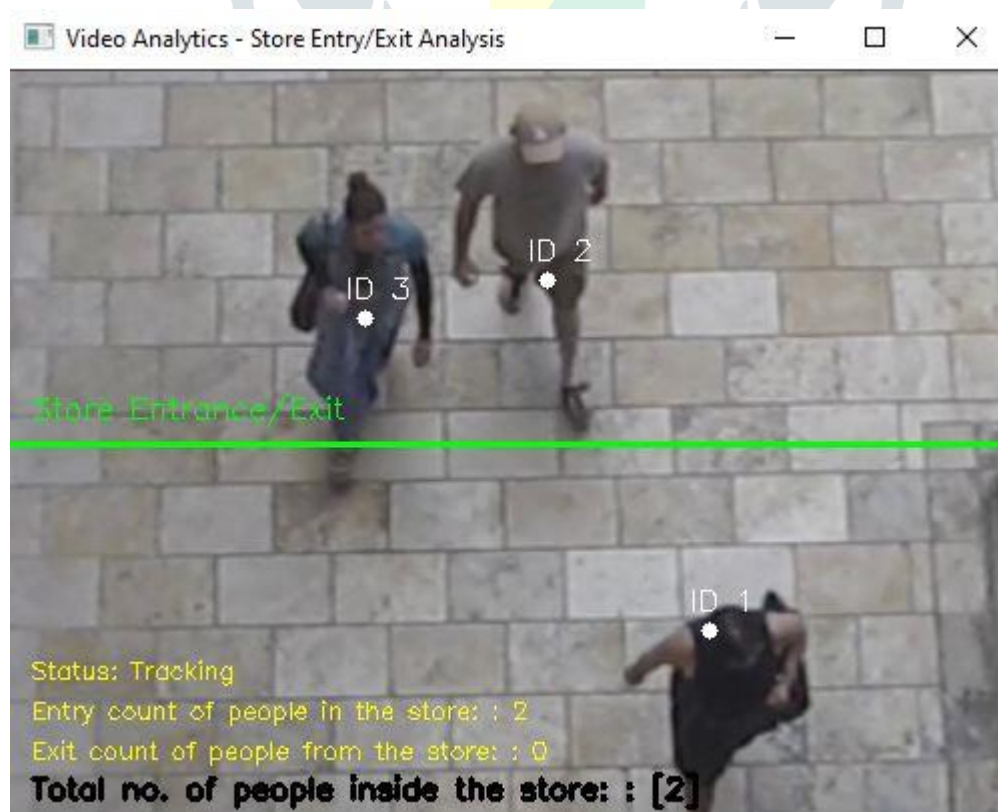


Figure 5- Store Entry/Exit Analysis



Figure 6- Instore crowd analysis

V Conclusion and Future Scope:

Video analytics help retailers enhance the customer shopping experience by reducing customer service time. We developed software on video analytics that will provide valuable insights and feedback to retailers on various aspects. We intend to provide a very cost-effective, simple, and automatic solution to smart retailers who are looking for ways to improve store management. We understand the challenges posed to retailers who practice traditional store management. Video content analytics allows retailers to detect traffic, and effectively respond and react. Driving the customer conversion rate is an important goal for a retailer. Therefore, ensuring efficient in-store management is important. Our software helps retailers to decide how he/she can enhance the customer shopping experience.

Our model is extensible. In the future, our algorithm can be upgraded with the availability of better technology. With the support of software developers, the model can be deployed completely as a market-ready product. Our model can be easily integrated with advanced surveillance systems. Also, a real-time admin dashboard can be integrated to monitor it effectively.

VI Acknowledgment

This work was done by Ms.Zubiya Rahman, Ms.Ainaan Siddiqua, and Ms.Madiha Muddasir ; the final year B.E. students under the guidance of Syeda Surrayya Jabeen (Assistant Professor) in the Department of Computer Science and Engineering, ISL Engineering College, Hyderabad, as a part of their final year dissertation. The authors express their gratitude to the CSE department and the Institute, for their efforts in developing the research environment, where this study was conducted. We also thank Google Colab for providing access to the computational resources used to conduct this study.

References:

- <https://www.einfochips.com/blog/video-analytics-in-retail-bringing-the-wow-factor-to-customer-experience/>
- <https://ieeexplore.ieee.org/document/4425348>
- <https://www.pathpartnertech.com/10-amazing-ai-based-video-analytics-use-cases-in-retail>
- <https://towardsdatascience.com/ssd-single-shot-detector-for-object-detection-using-multibox-1818603644ca>
- <https://medium.com/@cindy.trinh.sridykhan/a-tour-of-video-object-tracking-part-i-presentation-8a8aa9da9394>
- <https://tryolabs.com/resources/video-analytics-guide/>