Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

On Edge Crowd Traffic Counting System using Deep Learning on Jetson Nano for Smart Retail Environment

Muhammad Hafizuddin Mohd Razif[1], Ahmad Puad Ismail[1], Syahrul Afzal Che Abdullah[2], Mohd Affandi Shafie[1], Iza Sazanita Isa[1], Siti Noraini Sulaiman[1], Zainal Hisham Che Soh[1,*]

[1] Electrical Engineering Studies, College of Engineering, Universiti Teknologi MARA, Cawangan Pulau Pinang, 13500 Permatang Pauh, Pulau Pinang, Malaysia
[2] School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA Shah Alam, 40450 Shah Alam, Selangor, Malaysia

ARTICLE INFO

Article history:
Received 22 June 2023
Received in revised form 28 October 2023
Accepted 11 February 2024
Available online 26 March 2024

Keywords:
Crowd traffic counting system; Edge computing; Jetson nano; Computer vision; Smart retail; Deep learning; MobileNet SSD; YOLOv5

ABSTRACT

Most supermarket don't have a crowd traffic counting systems to track and counting customer entering their shop lot. In practise, these methods are useful for businesses in managing customer flow and optimize staffing and marketing efforts. Furthermore, the information can be used to estimate the popularity of the shop with relation to people entering the shop lot. The information also useful for the shop owner in determining the renting value. Therefore, the paper presents a system for tracking and counting people in a retail store using on the edge device, a Jetson Nano board. The comparison of the performance of two algorithms for people detecting of YOLOv5 and MobileNet-SSD are used in this work, YOLOv5 is a state-of-the-art object detection model that is known for its accuracy, but it is computationally intensive and may not be suitable for running on resource-constrained devices such as the Jetson Nano. MobileNet-SSD is a lightweight object detection model that is designed to run efficiently on mobile devices and embedded systems. Next is to track people using SORT, SORT is a real-time multi-object tracking algorithm that is based on the Kalman filter and the Hungarian algorithm. The results show that YOLOv5 was able to achieve the highest accuracy in detecting and counting people, but it was slower than the other two algorithms. MobileNet-SSD was the fastest algorithm, but it had lower accuracy compared to YOLOv5. In conclusion, the choice of algorithm will depend on the trade-off between accuracy and computational resources, and SORT is a good option for real-time people counting on resource-constrained devices.

1. Introduction

In recent years, technologies that count the number of people in a space and measure how crowded it is have gained increasing popularity for use in the management of shopping stores. These people counting and crowdedness measurement technologies can provide valuable insights into the behaviour of customers, which can then be used by store managers to make informed business decisions.

[*] *Corresponding author.*
E-mail address: zainal872@uitm.edu.my

https://doi.org/10.37934/araset.42.1.113

Counting the number of people entering and exiting a shop can provide valuable information for the shop owner to adjust pricing of the rent. By understanding the foot traffic in the shop, the owner can determine the demand for the location and adjust the rent accordingly. For example, if the shop is in a busy area with high foot traffic, the owner may be able to charge a higher rent. On the other hand, if the foot traffic is low, the owner may need to lower the rent to attract tenants. By using the technologies, which uses computer vision and AI to do people counting, could potentially be extended to include estimating the rent of shops. This could be done by using computer vision algorithms to detect and count the number of people in the images captured by the camera. The number of people could then be used as an additional feature in the machine learning model to make more accurate rent predictions.

Moreover, these technologies can provide store managers with real-time information about changing customer demand. In addition to staffing levels, these technologies can also be used to gain a deeper understanding of customer movement patterns within the store. For instance, if there is a sudden increase in foot traffic, the store manager can use the data to respond quickly. By analysing this data, store managers can identify areas of the store that are particularly popular and make changes to the layout or product placement to optimize the shopping experience. These systems make use of several different algorithms to properly track the number of individuals going into and out of a store as well as the patterns of movement those people participate in while they are inside the store. By combining people counting with other relevant features such as the size of the shop, location, and number of windows, the owner can have a more complete picture of the demand for the location and make more informed decisions about adjusting the rent.

Work in [1,2] compare human detection models on low-cost embedded platforms like the NVIDIA Jetson TX2 and Nano. Their work investigates the capabilities of the NVIDIA Jetson to execute models such as PedNet, multiped, SSD MobileNet V1, SSD MobileNet V2, and SSD inception V2. As processing speed is a key factor in embedded systems, Moreover, they also conducted comprehensive comparisons among those deep learning techniques to find the most efficient model. Findings of this study can be summarized as follows: integration and optimization of people detection algorithms in real applications into embedded platforms, end- to-end comparisons between existing people detection models in terms of accuracy and performance, and datasets that can be used to detect and track people in the building. From this study, an embedded platform called Jetson TX2 can execute computer vision applications quickly and efficiently (7.5W - 15W). As a result, it offers a way to create real-time AI object-detecting software. Jetson Nano is less powerful (5–10W) and less expensive than TX2, while having a smaller size and slower performance.

An AI Traffic Control System, in the study detected and tracked cars using a machine learning SSD algorithm instead of YOLO inside NVIDIA Jetson Nano [3]. This system proposes an advanced traffic control system consists of AI, machine learning, image processing and IoT. It represents a smart traffic light management prioritizing by vehicle density and a complete solution for whom breaks the traffic rules. For machine learning purposes, the SSD algorithm has been used instead of YOLO. Because the NVIDIA Jetson Nano is a small AI computer. It can only use four threads. If the YOLO algorithm is used and a server installed, the system will run very slowly. So, the efficiency of the system will be hampered. Because the NVIDIA Jetson Nano is a small AI computer. It can only use four threads.

As the base line to if the specific model can run in the targeted hardware a study by Haogang Feng *et al.,* investigate the inference workflow and performance of YOLO network [4], which is the most popular object detection model, in three different accelerator-based single-board computers (SBC), which are NVIDIA Jetson Nano, NVIDIA Jetson Xavier NX and Raspberry Pi 4B with Intel Neural Compute Stick2. Two versions of YOLO network across the above three SBCs to detect different video contents are benchmarked. By comparing the inference performance of the three SBCs, they discussed and evaluated the resource characteristics of the three SBCs to give empirical recommendations for reference when deploying AI applications on these three types of intelligence edges. Table 1 shown that the memory usage of GPU-based SBC, Jetson Nano, and Jetson Xavier NX is much bigger than the memory usage of ASIC-based SBC, Raspberry Pi with Intel Neural Compute Stick2 when using a same AI model, Jetson Nano always consumes the most energy for the model inference. In contrast, though Jetson Xavier NX has the highest average power (reaches 15.2W when running YOLOv3), it is extremely high performance enables it to complete inference work in a very short time, so Jetson Xavier NX always consumes the least energy for model inference. In terms of Raspberry Pi with Intel Neural Compute Stick 2, it always has the lowest average power, no matter which model is running for inference, but its FPS is higher than Jetson Nano, which makes it have better energy consumption performance than Jetson Nano.

Table 1
Benchmarking Jetson Nano, Jetson Xavier NX and Raspberry PI with Neural Stick 2 [4]

	Models	Accelerator-based SBCs	Mean Confidence (%)	FPS	CPU Usage%	Memory Usage (GB)	Power(W)	Time (s)
Video1	YOLOv3	RPI+NCS2	99.3	2.5	4.3	0.33	6.0	690
		Nano	99.7	1.7	26.5	1.21	7.9	967
		NX	99.7	6.1	22.5	1.51	15.2	256
	YOLOv3-tiny	RPI+NCS2	0	18.8	15.5	0.11	6.5	121
		Nano	57.9	6.8	28.8	1.00	7.2	236
		NX	57.9	41.1	30.5	1.33	13.5	46
Video2	YOLOv3	RPI+NCS2	85.8	2.5	9.8	0.41	6.2	496
		Nano	71.5	1.6	38.8	1.36	8.0	587
		NX	71.5	5.9	36.8	1.69	15.2	162
	YOLOv3-tiny	RPI+NCS2	61.5	19.0	34.8	0.18	6.8	162
		Nano	54.1	6.6	37.3	1.16	7.4	152
		NX	54.1	35.6	55.5	1.47	13.2	31

YOLOV5-based DeepSORT [5] pedestrian target tracking algorithm (YOLOV5-DeepSORT), which introduces the high-performing YOLOV5 algorithm into the DeepSORT algorithm, which detects the tracking video frame by frame, and then predicts the target position using Kalman filtering while matching the same target using the Hungarian algorithm. To verify the performance of the YOLOv5-DeepSORT algorithm, a comparison of it with the YOLOv3-DeepSORT algorithm experiments is carried out. YOLOv5-DeepSORT algorithm can detect, and track pedestrians better compared to the YOLOV3-DeepSORT algorithm.

For intelligent pig monitoring applications with low-cost embedded boards, such as the NVIDIA Jetson Nano, light-weight object detection and tracking algorithms are being used [6,7] . By reducing the computational cost in TinyYOLOv4 [8] and Deep- SORT, they can detect and track pigs in real time on an embedded board, without losing accuracy. The accuracies of light-weight object detection and tracking algorithms are not perfect, the counting accuracy of 99.44%, even when some pigs pass through the counting zone back and forth.

Another study by Zhongxing *et al.,* [9] Real-Time Enumeration of Metro Passenger Volume Using Anchor-Free Object Detection Network on Edge Devices. They record footage from the cameras and create CircleDet, an anchor-free object detection network, to identify passengers' heads. Instead of using a conventional bounding box to locate and bind the target, CircleDet predicts a circle. After that, they use a straightforward yet efficient circular IoU-based approach to locate and identify the passengers in the movies. CircleDet can operate at 7.8 frames per second (FPS) on an NVIDIA Jetson

Nano device and up to 111 FPS on an NVIDIA RTX 2080 graphics card. Their proprietary metro object detection (MOD) dataset demonstrates an enumeration accuracy of up.

People tracking is a crucial task in various applications such as video surveillance, traffic monitoring, and human-computer interaction. Tyagi *et al.,* [10] develop automated systems that can detect anomalies in the diverse and complex outdoor environments, especially at public places to ensure safety and to avoid crowd disasters such as human stampede, mob lynching, and riots using deep learning to analyse the crowded areas possibility of occurrence of suspicious and violence activities. Abdullah *et al.,* [11] inventing semantic segmentation on foreground extraction for pedestrians counting and tracking by the weighted averaging method that removes non-humans and non-pedestrians from the scene. Alamri *et al.,* [12] proposed system can detect and intelligently analyse the pattern of crowd activity to implement contingency plans, reducing accidents, ensuring public safety, and establishing a smart city. An intelligent crowd-monitoring system that can monitor the crowd movement makes it easier to maintain social distancing safety [13-15] in small and large public areas. The results reveal that the proposed model can identify crowd density status in real-time, implying that it can effectively assist crowd management tasks such as monitoring, guiding, and managing crowds to ensure safety. As for other application, Wang *et al.,* [16] proposed a Crowd Tracking and Counting in subway passage using an Improved FairMOT Method. De Sanctis *et al.,* [17] used a modular approach to detect people waiting in line.

In recent years, advances in computer vision and deep learning have led to significant improvements in the accuracy and efficiency of people tracking systems. Therefore, there a few papers proposed a system that uses image processing to track and count people as they walk through a certain area [18-25].

This paper presents an On-Edge Crowd Traffic Counting System (OECTCS) using Deep Learning on Jetson Nano for counting the detected person during entering or exiting a retail store. This research work only covers the parts on analysing the best model for the OECTCS on Jetson Nano. The Deep Learning Technique used in this research is YOLOv5 and Mobilenet-SSD. The remainder of the paper is organized as follows. Section 2 describes the methodology of OECTCS. Next, in section 3 discussed the experimental result of YOLOv5 and Mobilenet-SSD human detection accuracy on different hardware and human counting performances for both models. Finally, section 4 provides the concluding remarks and points out the ideas for future extension of this work.

2. Methodology

The proposed system using image processing techniques to accurately detect and count the number of people in a shop. The system will use a camera as a video source for detection and tracking but will not be able to identify individual faces or bodies. The goal of the project is to build a reliable system that can provide useful data to shop owners on the number of customers in their store, which can be used to optimize staffing levels, improve the shopping experience, and respond more quickly to changing customer demand.

The system will be developed specifically for usage by shop owners and will be designed to store data on a local storage device. The data collected by the system will only be accessible to the shop owner, and measures will be taken to ensure that the privacy of customers is respected. The system will be implemented using embedded boards and will be designed in a cost-effective manner. The goal is to create a system that can accurately and efficiently track the number of people in the hallway in real time.

2.1 Hardware

One popular platform for implementing people counting and crowdedness measurement applications for shopping store management is the NVIDIA Jetson Nano. This low-cost, low-power computing platform is well-suited for deploying machine learning models and computer vision applications in real-time at the edge. The NVIDIA Jetson Nano is a popular platform for edge computing due to its combination of affordability and high-performance computing capabilities. There are several advantages to using the Jetson Nano for people counting and crowdedness measurement applications in shopping stores. First, Nano's compact size and low power consumption make it easy to integrate into a store's existing infrastructure. It can be mounted discreetly on a wall or ceiling or incorporated into a smart camera or another device.

One of the primary benefits of the NVIDIA Jetson Nano is its inexpensive price, which makes it accessible to a variety of retail businesses. The platform is designed with robust hardware that enables it to handle demanding machine learning models and computer vision applications in real-time despite its low cost. There are several advantages to using the Jetson Nano for people counting and crowdedness measurement applications in shopping stores. First, Nano's compact size and low power consumption make it easy to integrate into a store's existing infrastructure. It can be mounted discreetly on a wall or ceiling or incorporated into a smart camera or another device.

2.2 Object Detection Techniques

Both Mobilenet-SSD and YOLOv5 are object recognition techniques that will be tested in the person counting system [4]. Mobilenet-SSD is a single-shot object detection model that uses the MobileNet architecture developed by Google. This model makes predictions for each object in the input image in a single pass, which makes it faster and more efficient compared to other models that use a sliding window or region proposal method. Mobilenet-SSD is designed to be fast and lightweight, so it can run on devices with limited resources, making it ideal for use in the person counting system.

On the other hand, YOLOv5 is an enhanced version of YOLOv3, a lightweight detection network presented by Utralytics in 2020. YOLOv5 consists of four components: Input, Backbone, Neck, and Prediction. The input component takes in the image, while the Backbone extracts features from the image. The Neck component processes the features, and the Prediction component outputs the object detections. YOLOv5 is a fast and efficient object detection model that is widely used in a variety of applications, including security systems and self-driving cars.

For counting persons, we detected individual objects and keep the region of interest (RoI) for detection, exclusive non-RoI. The YOLOv5 and models are widely used for many object detection applications. In this study, YOLOv5n, which is a variant and a lightweight version of YOLOv5, is proposed to perform object detection on embedded boards. In addition, a multi-object tracking algorithm was performed based on the detection results to track individual objects.

2.3 Tracking and Counting Technique

For the tracking techniques, we will be using the sort tracking SORT (Simple Online and Realtime Tracking) is an object tracking algorithm developed by Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft. It is a real-time algorithm that can track multiple objects in a video stream. SORT uses a combination of object detection and data association techniques to track objects as they move through the frame. It is based on the Kalman filter, a mathematical model used to predict the future state of a system based on its past states. SORT is designed to be simple and efficient, making it well-suited for use in real-time applications.

SORT works in two steps. Detections are made in the first step utilising an object detection algorithm, such as YOLO or Mobilenet-SSD. The detections are then correlated with the entities they belong to in the second step using a data association approach. This is accomplished using a mix of tracking by detection and tracking by prediction, in which each object's position and velocity are calculated and utilised to anticipate its future location.

SORT then uses a Kalman filter to associate each detection with a specific object. The Kalman filter is an algorithm that estimates the state of a system over time, based on the past observations and a prediction of the future state. In the case of SORT, the Kalman filter is used to estimate the position and velocity of each object, and to associate the detections with the objects they correspond to.

In conclusion, SORT is a multiple object tracking algorithm that operates in two stages: object detection and data association. It uses a Kalman filter to associate detections with objects and to estimate the position and velocity of each object over time. SORT is designed to be fast and efficient, making it a popular choice for real-time object tracking applications.

3. Results

This section discusses the results obtained for performance evaluation of object identification of a person using the YOLOv5 and MobileNet-SSD. Firstly, compare the accuracy of YOLOv5 and MobileNet-SSD, use a commonly used benchmark dataset, such as (Common Objects in Context) COCO in this sub section. Next, evaluates (frames per second) FPS, for running the object detection models on the Jetson Nano or other hardware and measure the average FPS over several frames. The FPS will give an idea of how fast the models can process each frame and thus how suitable they are for real-time applications is discussed. Then, track the number of people using YOLOv5 and MobileNet-SSD, use the models to detect people in a video stream and count the number of detections in each frame. You can then compare the results to see which model performs better in terms of accuracy and speed. Finally, concluded that counting of people using MobileNet-SSD with SORT, you can use the SORT algorithm to track the detected people over several frames.

3.1 Accuracy YOLOv5 and MobileNet-SSD

In Table 2 terms of accuracy, YOLOv5 has been reported to perform very well on various object detection benchmarks, including the COCO dataset, which is a commonly used dataset for evaluating object detection models. On the COCO dataset, YOLOv5 has achieved an average precision (AP) of 44.6% at an IoU (Intersection over Union) threshold of 50% on the test-dev split. This is a relatively high accuracy compared to other object detection models.

Table 2
Accuracy for YOLOv5n

Metric	Value
mAP	44.6%
IoU	50%

Next in Table 3, MobileNet-SSD is generally considered to be a relatively fast but less accurate object detection model compared to some of the more complex models. On the COCO dataset, which is a commonly used benchmark for object detection, MobileNet-SSD has achieved an average precision (AP) of 21.3% at an IoU (Intersection over Union) threshold of 50% on the test-dev split

Table 3
Mobilenet-SSD

Metric	Value
mAP	21.3%
IoU	50%

3.2 Frame-Per-Second Comparison of YOLOv5n and Mobilenet-SSD

Table 4 shown that YOLOv5 can process images at 6 FPS (frames per second), approximately one frame every 167 milliseconds. This means that it may be suitable for applications that require a high degree of accuracy but may not be suitable for real-time applications that require faster processing speeds. In contrast, MobileNet-SSD can process images at 30 FPS, which is approximately one frame every 33 milliseconds. This makes it well-suited for real-time applications where fast processing speeds are required. However, MobileNet-SSD may not be as accurate as YOLOv5.

Table 4
YOLOv5n and Mobilenet-SSD

Model	FPS
YOLOv5	6
Mobilenet-SSD	30

3.3 YOLOv5 Performance on Different Hardware

The performance of YOLOv5 may vary depending on the specific hardware configuration and the complexity of the input data. For example, if YOLOv5 is running on a Jetson Nano at 6 FPS, it can process approximately one frame every 167 milliseconds. This may be sufficient for some applications but may not be suitable for real-time applications that require faster processing speeds.

On a full PC with an Intel Core i3-12100F processor and a GTX 1650 graphics card, YOLOv5 may be able to achieve a frame rate of 30 FPS as shown in Table 5, which is approximately one frame every 33 milliseconds. This may be sufficient for real-time applications and may allow YOLOv5 to reach its full potential in terms of accuracy and performance.

Table 5
YOLOv5 performance on different hardware

Hardware	FPS
Jetson Nano	6
Corei312100F/Gtx1650	30

3.4 Number of People That Can Track a Given Video Clips

To evaluate the effectiveness of the system under realistic conditions, a series of video clips that simulate a crowded hallway was used to test its performance. These clips allowed us to observe the system's ability to accurately track and count the number of people present in the hallway, even in situations where the environment is busy and crowded.

Fig. 1. Video Clip 1

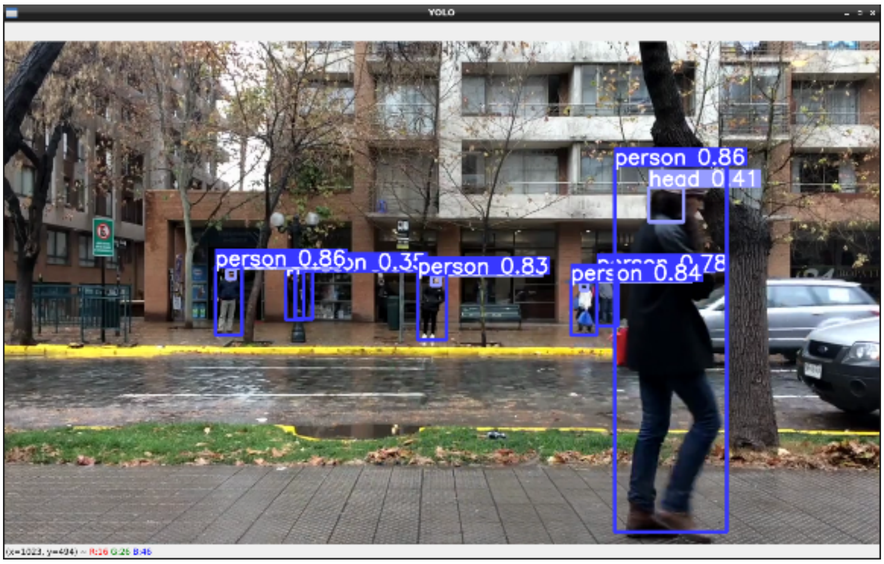

Fig. 2. Video Clip 2

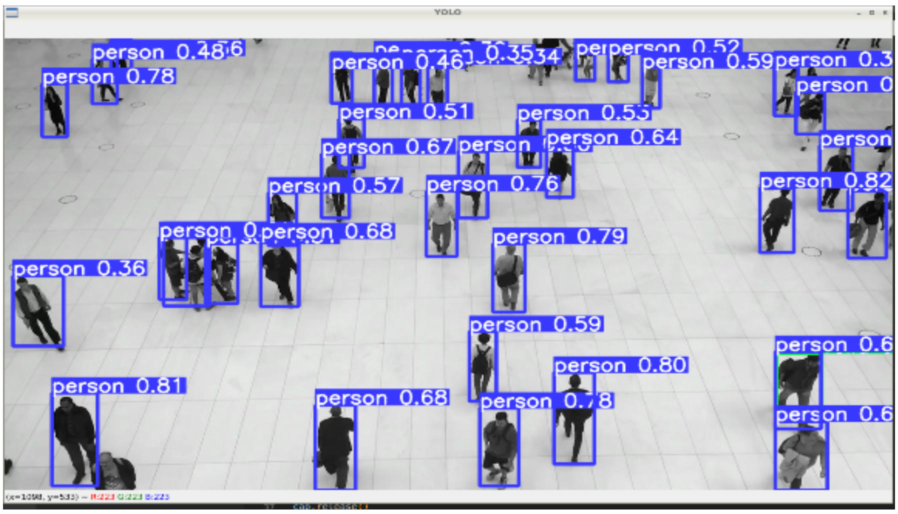

Fig. 3. Video Clip 3

From both the Table 6 and 7 the combination of YOLOv5 with SORT has been found to be more accurate than MobileNet-SSD for object tracking, but it comes with the cost of lower frames per second (FPS). YOLOv5 is a highly accurate object detection model, and when combined with SORT, it can further improve the accuracy of object tracking. However, YOLOv5 is a more complex model than MobileNet-SSD and may have a slower processing speed, resulting in lower FPS.

Table 6
Counting results in detail for each clip Mobilenet-SSD with SORT

	The number of people	Detected number of people	FPS
Clip 1	10	10	30
Clip 2	9	8	30
Clip 3	41	75	20

Table 7
Counting results in detail for each clip YOLOv5 with SORT

	The number of people	Detected number of people	FPS
Clip 1	10	10	6
Clip 2	9	9	6
Clip 3	41	33	5

On the other hand, MobileNet-SSD is a faster and more efficient model, but it may have lower accuracy compared to YOLOv5 with SORT. The choice between the two will depend on the specific requirements of the application and the relative importance of accuracy and speed. The increased number of people in each frame can affect the performance of an object detection and tracking system. As the number of people increases, the system may have to process more data and make more calculations to track each person, which can potentially decrease the frame rate and increase the processing time.

3.5 People Counting using Mobilenet-SSD with SORT

The brown and green zones represent the entrance and exit areas as shown in Figure 4, respectively. If a person is detected in the entrance area and then walking to the exit area, the counting of a person is complete.

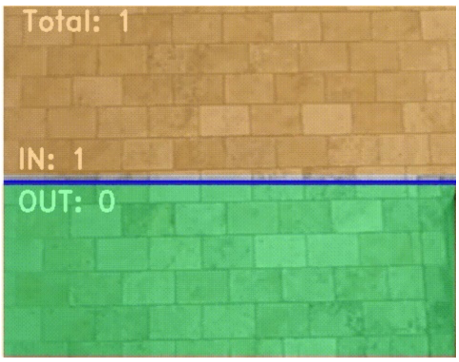

Fig. 4. Area of detection

In this case, the person was detected in the F-1 frame as shown in Figure 5, and the position of the person was in the entrance area.

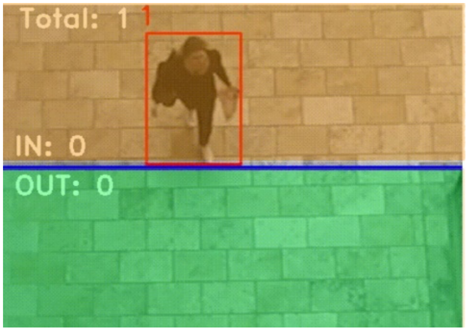

Fig. 5. F-1 frame depicts a person detected from the entrance area

The start status and end status are ‘0’. The person moves to the exit area in the F-2 frame as shown in Figure 6, the start and end status becomes ‘1’ increasing the (counting result) CR value by 1 as indicate in Table 8.

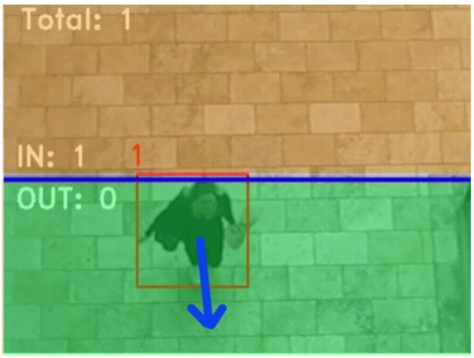

Fig. 6. F-2 frame depicts a person detected from the entrance exit

Table 8
Moving from entrance to exit

Frame	F-1	F-2
Start status	0	0
End status	0	1
CR	-	+1

The person was detected in the F-1 frame and the position is in the exit area as shown in Figure 7. The start and end states are ‘1’. In this situation the end status value is the same as the start value, the CR value does not change.

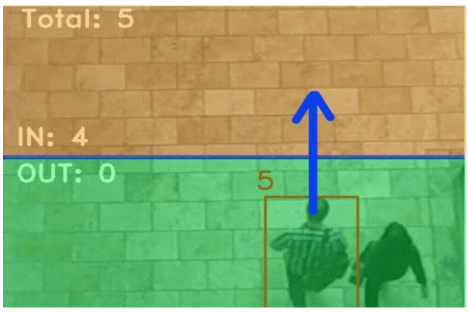

Fig. 7. F-1 frame depicts a person detected from the exit area

The person was detected in the F-2 frame and the position as shown in Figure 8 is in the entrance area.

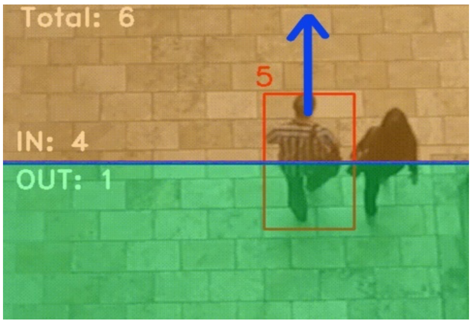

Fig. 8. F-2 frame depicts a person detected from the exit area

However, in this frame, the CR value decreases by 1 owing to changes between the end status ('0') and the start status ('1') on Table 9 below.

Table 9
Moving from exit to entrance

Frame	F-1	F-2
Start status	1	1
End status	1	0
CR	-	-1

The last scenario is when the same person moves back and forth between the entrance area and the exit area as shown in Figure 9. First, the person is detected in the T-1 frame, and because it is in the entrance area, the start status and end status values are stored as '0'. In the T-2 frame, the CR value increases by 1 because the end status value is '1

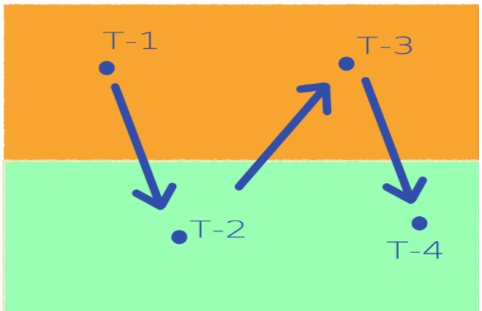

Fig. 9. Subject moving back and forth

. The end status value of the T-3 frame is the same as the start status value, and it does not affect the CR value. Finally, because the end statuses of the T- 4 frames are '1', the CR value increases by 1.

Table 10
Moving back and forth

	T-1	T-2	T-3	T-4
Start status	0	0	0	0
End status	0	1	0	1
CR	-	+1	-	+1

4. Conclusions

In this research, the paper presented a solution of an On Edge Crowd Traffic Counting System using deep learning model for counting the number of people entering and exiting a shop lot. Furthermore, this research can possibly provide the retail industry with an effective solution to the estimate the popularity of certain shop lot based on number of customers coming to shop. As mentioned above, a machine learning model for people entering and exiting a shop lot successfully being created. First, the YOLOv5 model is suitable and a better choice for system with better object detection accuracy. However, if you need to run the object detection model on a device with limited computational resources, such as a Jetson Nano, then MobileNet-SSD might be the better choice because it can run faster on these devices although with less accuracy. The SORT object tracker can reduce the computational expense of feature extraction without compromising tracking accuracy. Even when people are repeatedly entering and leaving the counting zone, this straightforward identification and tracking approaches, together with an exact counting algorithm, may give results with high precision and real-time speed. In conclusion, the human tracking and counting system operate with real-time speed and acceptable counting precision on an embedded NVIDIA Jetson Nano board.

Acknowledgement
The author would like to express sincere gratitude to Universiti Teknologi MARA, Cawangan Pulau Pinang for permitting us to use their facility to complete this research work. Special thanks for those who helped directly and indirectly.

References
[1] Rahmaniar, Wahyu, and Ari Hernawan. "Real-time human detection using deep learning on embedded platforms: A review." *Journal of Robotics and Control (JRC)* 2, no. 6 (2021): 462-468. https://doi.org/10.18196/26123
[2] Lopez-Montiel, Miguel, Ulises Orozco-Rosas, Moisés Sánchez-Adame, Kenia Picos, and Oscar Humberto Montiel Ross. "Evaluation method of deep learning-based embedded systems for traffic sign detection." *IEEE Access* 9 (2021): 101217-101238. https://doi.org/10.1109/ACCESS.2021.3097969
[3] Uddin, Md Imran, Md Shahriar Alamgir, Md Mahabubur Rahman, Muhammed Shahnewaz Bhuiyan, and Md Asif Moral. "Ai traffic control system based on deepstream and iot using nvidia jetson nano." In *2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, pp. 115-119. IEEE, 2021. https://doi.org/10.1109/ICREST51555.2021.9331256
[4] Feng, Haogang, Gaoze Mu, Shida Zhong, Peichang Zhang, and Tao Yuan. "Benchmark analysis of yolo performance on edge intelligence devices." *Cryptography* 6, no. 2 (2022): 16. https://doi.org/10.3390/cryptography6020016
[5] Gai, Yuqiao, Weiyang He, and Zilong Zhou. "Pedestrian target tracking based on DeepSORT with YOLOv5." In *2021 2nd International Conference on Computer Engineering and Intelligent Control (ICCEIC)*, pp. 1-5. IEEE, 2021. https://doi.org/10.1109/ICCEIC54227.2021.00008
[6] Kim, Jonggwan, Yooil Suh, Junhee Lee, Heechan Chae, Hanse Ahn, Yongwha Chung, and Daihee Park. "EmbeddedPigCount: Pig counting with video object detection and tracking on an embedded board." *Sensors* 22, no. 7 (2022): 2689. https://doi.org/10.3390/s22072689
[7] Jensen, Dan Børge, and Lene Juul Pedersen. "Automatic counting and positioning of slaughter pigs within the pen using a convolutional neural network and video images." *Computers and Electronics in Agriculture* 188 (2021): 106296. https://doi.org/10.1016/j.compag.2021.106296
[8] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection." *arXiv preprint arXiv:2004.10934* (2020).
[9] Zheng, Zhongxing, Weiming Liu, Heng Wang, Guici Fan, and Yuan Dai. "Real-time enumeration of metro passenger volume using anchor-free object detection network on edge devices." *IEEE Access* 9 (2021): 21593-21603. https://doi.org/10.1109/ACCESS.2021.3054938
[10] Tyagi, Bhawana, Swati Nigam, and Rajiv Singh. "A review of deep learning techniques for crowd behavior analysis." *Archives of Computational Methods in Engineering* 29, no. 7 (2022): 5427-5455. https://doi.org/10.1007/s11831-022-09772-1

[11] Abdullah, Faisal, and Ahmad Jalal. "Semantic segmentation based crowd tracking and anomaly detection via neuro-fuzzy classifier in smart surveillance system." *Arabian Journal for Science and Engineering* 48, no. 2 (2023): 2173-2190. https://doi.org/10.1007/s13369-022-07092-x
[12] Alamri, Abdullah. "Cloud of things in crowd engineering: a tile-map-based method for intelligent monitoring of outdoor crowd density." *Sensors* 22, no. 9 (2022): 3328. https://doi.org/10.3390/s22093328
[13] Valencia, Immanuel Jose C., Elmer P. Dadios, Alexis M. Fillone, John Carlo V. Puno, Renann G. Baldovino, and Robert Kerwin C. Billones. "Vision-based crowd counting and social distancing monitoring using Tiny-YOLOv4 and DeepSORT." In *2021 IEEE International Smart Cities Conference (ISC2)*, pp. 1-7. IEEE, 2021. https://doi.org/10.1109/ISC253183.2021.9562868
[14] Pandiyan, P., Rajasekaran Thangaraj, M. Subramanian, R. Rahul, M. Nishanth, and Indupriya Palanisamy. "Real-time monitoring of social distancing with person marking and tracking system using YOLO V3 model." *International Journal of Sensor Networks* 38, no. 3 (2022): 154-165. https://doi.org/10.1504/IJSNET.2022.121700
[15] Kanwal, Nadia. "Deep Learning based Customer Count/Flow Monitoring System for Social Distancing." (2021).
[16] Wang, Guankun, Yongze Yang, Xiaopin Zhong, and Yang Yang. "An Improved FairMOT Method for Crowd Tracking and Counting in Subway Passages." In *International Conference on Intelligent Transportation Engineering*, pp. 130-139. Singapore: Springer Nature Singapore, 2021. https://doi.org/10.1007/978-981-19-2259-6_11
[17] De Sanctis, Mauro, Simone Di Domenico, Davide Fioravanti, Eduardo Ballesteros Abellán, Tommaso Rossi, and Ernestina Cianca. "Rf-based device-free counting of people waiting in line: A modular approach." *IEEE Transactions on Vehicular Technology* 71, no. 10 (2022): 10471-10484. https://doi.org/10.1109/TVT.2022.3182548
[18] Alharbey, Riad, Ameen Banjar, Yahia Said, Mohamed Atri, Abdulrahman Alshdadi, and Mohamed Abid. "Human Faces Detection and Tracking for Crowd Management in Hajj and Umrah." *Computers, Materials & Continua* 71, no. 3 (2022). https://doi.org/10.32604/cmc.2022.024272
[19] Kumar, Shailender, Pranav Sharma, and Nitin Pal. "Object tracking and counting in a zone using YOLOv4, DeepSORT and TensorFlow." In *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, pp. 1017-1022. IEEE, 2021. https://doi.org/10.1109/ICAIS50930.2021.9395971
[20] Pervaiz, Mahwish, Ahmad Jalal, and Kibum Kim. "Hybrid algorithm for multi people counting and tracking for smart surveillance." In *2021 International Bhurban conference on applied sciences and technologies (IBCAST)*, pp. 530-535. IEEE, 2021. https://doi.org/10.1109/IBCAST51254.2021.9393171
[21] Sivachandiran, S., K. Jagan Mohan, and G. Mohammed Nazer. "Deep Learning driven automated person detection and tracking model on surveillance videos." *Measurement: Sensors* 24 (2022): 100422. https://doi.org/10.1016/j.measen.2022.100422
[22] Pervaiz, Mahwish, Yazeed Yasin Ghadi, Munkhjargal Gochoo, Ahmad Jalal, Shaharyar Kamal, and Dong-Seong Kim. "A smart surveillance system for people counting and tracking using particle flow and modified SOM." *Sustainability* 13, no. 10 (2021): 5367. https://doi.org/10.3390/su13105367
[23] Abdullah, Faisal, Yazeed Yasin Ghadi, Munkhjargal Gochoo, Ahmad Jalal, and Kibum Kim. "Multi-person tracking and crowd behavior detection via particles gradient motion descriptor and improved entropy classifier." *Entropy* 23, no. 5 (2021): 628. https://doi.org/10.3390/e23050628
[24] Haritaoglu, Ismail, and Myron Flickner. "Detection and tracking of shopping groups in stores." In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, pp. I-I. IEEE, 2001.
[25] Minh, Hieu Tran, Linh Mai, and Thanh Vo Minh. "Performance evaluation of deep learning models on embedded platform for edge ai-based real time traffic tracking and detecting applications." In *2021 15th International Conference on Advanced Computing and Applications (ACOMP)*, pp. 128-135. IEEE, 2021. https://doi.org/10.1109/ACOMP53746.2021.00024