Queue Time Estimation in Checkout Counters Using Computer Vision and Deep Neural Network

Aseem Rajvanshi and Dhanya Pramod*
Symbiosis Centre for Information Technology, Symbiosis International (Deemed University), Pune, Maharashtra, India
Email: *dhanya@scit.edu

Abstract. In this busy world, everyone wants to make decisions that are more efficient and save their time and not cause/ incur them loss. Sometimes they even get skeptical about making the decision. A similar situation arises when we are required to decide which queue, we should join for billing of the items during the weekend rush in a posh supermarket. To crack this problem, this paper recommends using Computer Vision and Deep Learning Algorithms to estimate the time closely, a customer will take to reach the counter if they join a particular queue. The algorithm will estimate the number of items present in the trolley or the basket and the number of such trolleys in the queue. The next thing the program will measure is the employee's efficiency who is present in the checkout/billing counter in terms of scanning the items, accepting the payment, and handing over the bill. This will help the customers to make better decisions and save time; it will also make the supermarket look less crowded and use all the counters/resources effectively.

Keywords: Computer Vision, Deep Neural Network, Queue Time Estimation, Retail Stores.

1. Introduction
Digital transformation and technology adoption has been rapidly changing the way retail stores operate. Customer experience is one of the key focus areas of any business, and the same is the case with the retail sector, too [1]. Retail stores use computer vision techniques for product recognition and counting in stores which in turn helps to deal with out-of-stock and misplacement of products. It is also used for the classification of products on the shelf and product class recognition which enables the customer to pick up items without wasting much time. Order picking robots have also been experimented in stores to assist the customers. Many stores have also invested in technologies that analyze store traffic statistics and queue behavior which enabled informed decision-making through business intelligence and analytics [2].

Shopping is the most favorite hobby/ thing many people like to do. People shop for clothes, cosmetics, groceries, food, etc., but the irritating or frustrating part comes when they have to stand in a queue for checkout. Queues are boring and if you select the wrong one it can be very frustrating. When there is a customer in a big shopping center and the shopping center is crowded, one of the dilemmas that the customer's faces are, which queue to join for the checkout, as there are multiple checkout counters and all of them have long queues. As specified in the psychology of waiting:
“Principle 4: Uncertain Waits are Longer than Known, Finite Waits.”, we have experienced it ourselves as well [3].

![Figure 1: Check-out Counter Scenario](image)

This would be better understood with the help of an example of a common scenario. Suppose it’s the weekend, and the house is short on groceries and other house-hold items. The family plans to visit the famous supermarket because it offers quality products at a reasonable price, furthermore has decent offers [4]. After reaching there, the family realized that they were not the only ones with that plan and the supermarket was crowded more than usual because it’s the weekend. A common setup of a supermarket is displayed in Figure 1, where there is an entry on one side of the supermarket and exit on the other. After completing their shopping, they choose the checkout counters closest to the exit which are 7, 8, 9 and 10. As they start their shopping, they move from rack 1 to rack 5, exploring the required products. They somehow managed to get all the stuff that was required and are ready to head towards the checkout counter [5]. As there is already a crowd near the counters in the end, the counters in the middle and near the entry are less occupied. This leads to the underutilization of a few of the counters.

This paper would like to propose a model that would help the customers to efficiently select the queue they should join that would help them to reduce the waiting time. The model would give them a rough idea of the time they have to be in the queue so that they are mentally prepared for the same [6]. It would improve customer’s experience, but it would also be beneficial for the shopping center by utilizing all the checkout counters efficiently and reducing the average queue time. To overcome this problem of underutilization and increased wait time, the proposed solution plans to install a digital display over the counters, which would display the estimated time for the next joining customer. Because the digital display would be kept at the top of the counter, there would be no difficulty faced by the customer to see it. To estimate or predict the waiting time, artificial neural network and computer vision technology would be used [7].

2. Literature Review
A significant amount of work has been done in managing the queues and enhancing the customer’s experience in the retail sector. The team was able to successfully develop a system that would track the customers with the help of their trolleys. The proposed system used a binary pattern and color marker on the trolleys to identify and track them [8]. This also useful in identifying the buying pattern and provide value-added services to the customer inside the shopping malls. The system was very economical as it used the existing security cameras, and it could be easily scaled up or down by just adding or removing the cameras, without changing the software worked on a system to assist the shop owners in keeping track of the products in the rack. The system would alert the owner of the shop if
exposed product were misplaced from its original position or if some product was going out of stock [9]. The proposed model collaborated with Speed Up Robust Features (SURF) and color histograms to accurately identify its category. The model also used 3 model images instead of one for improved accuracy and precision.

Explained the benefits of computer vision and deep learning in the retail sector and how it is beneficial for the shop owners to easily manage the stocks. It helped the shop owners to retain the customers and provide them with a better experience [10]. The paper also discussed the shortcomings of RFID Tags and weight sensors. The training of the neural network model does take some time but once it's done, it makes the task very smooth and reliable.

The judgement of various deep neural network (DNN) models was used to detect a person. The Region-based Convolutional Neural Network (RCNN) uses selective search that extracts 2000 regions, and feature matching is done on those. RCNN may have many advantages and benefits, but still, it's not fast enough for real-time application of any kind [11]. Region-based Fully Convolutional Network (R-FCN) creates a positive sensitive score map that identifies the object much faster than RCNN while maintaining accuracy. You Only Look Once (YOLO) algorithm is one of the best algorithms for real-life object detection. YOLO divides the input image into a square matrix and identifies the object was falling in the center of the box using a global context. In terms of speed and accuracy, Single Shot multi-box Detector (SSD) is object detectors. Feature maps extraction, also convolution filters application is the two main steps used to detect objects. Squeeze Det is the fastest object detector algorithm with high accuracy, which is used in self-driving cars. It uses a full convolution layer instead of feature extraction and it only contains a single advancing pass over the network, making it tremendously fast [12]. The paper, in the end, concludes that the algorithm which provides accurate results and takes reasonable time is YOLO v3-416. Hence, this model best considered for person detection using embedded platforms.

Mentioned in the paper, that stereo vision is computationally expensive as it is hard to find the corresponding pixels. Real-time suitability is retained while examining and improving the quality of correlation-based stereo [27]. To overcome the problem of correlation, three techniques/methods have been proposed. To decrease the errors at the border of an object and increase the correct matches, the Novel Multiple Window approach is used. To invalidate uncertain matches, General Error Filter's correlation function is put in use [13]. Finally, Object borders are further improved in a post-processing step of Border Correction Method. Every method had its own strong and week points, with the main weakness being computational cost. All things considered, the blend of techniques improved the nature of the ongoing relationship-based sound system fundamentally by diminishing the mistake to half by holding the computational expense.

Studied the scenario of a telecommunication shop and analyzed its queuing system. It was observed that an increasing number of customers in the queue and as time passes, it takes a toll on the quality of the service and causes discomfort to the customers in the queue [14]. The real-world dataset was collected from the store and various approaches for predicting queue time were analyzed and implemented using attendance time history. The aim was to regulate the quality of service provided to the customer and increase the satisfaction score. The model did achieve the aim; however, various other relevant aspects can be added to the same to enhance the model [15].

Various industries and the advantages of machine learning algorithms in predicting the queue time is explored. They collected a bank queue dataset available online and trained a fully connected neural network on it. They were able to predict the queue waiting time with a mean absolute error of 3.35 minutes [16]. They further worked with Queue For Me, which creates a virtual queue for the customers while allowing them the freedom to roam around. The parameters to calculate the queue time can be customized and altered depending on the type of scenarios and industries. In the end, they concluded that machine learning and neural network could act as a viable replacement for queueing theory.
3. Proposed Model in Brief
The proposed model is set to use 2 cameras per checkout counter and some computational power to calculate the time. There would be 2 cameras installed on the top of every checkout counter, through which live video would be captured [17]. In the first step, several trolleys or hand baskets present in the counter's queue are calculated. When the trolleys and baskets are identified in the queue, the next step includes calculating the quantity of the items present in the trolley or basket. Here a technique called stereo vision is used, in which both the cameras are used to make a 3-dimensional model of the trolley. The created 3-dimensional model would help us calculate the quantity of the products present in the trolley or the basket [18]. This is the first part of the system, where using computer vision, we estimate the number of products each trolley or basket has. In the next step, the efficiency of the employee at the checkout counter is measured and calculated. The efficiency of the employee can be measured by how many products/items he/she can scan in a minute. Another factor that impacts efficiency is how fast the employee can handle the transactions, either cash-based or card-based. After getting and calculating all the above-mentioned parameters, it is passed into the Artificial Neural Network to obtain the perfect weights or bias. The model would be ready to use and implement in the supermarket after being trained on real-life data [19].

4. Implementation and Working

4.1 Computer Vision
To understand the proposed idea, one must understand the technology that is going to be a part of it and how it is going to be used. The discussion will begin with computer vision, as the name suggests, the ability to 'see' and 'interpret objects' is given to the machine, as the humans have. Using various techniques and algorithms, machines in some cases, are now able to process images and videos better than humans [20]. The technology has advanced so much that humans have now developed self-driving cars and machines that diagnose the disease accurately. Application of computer vision is in various sectors, some of them are:

✓ Automatic inspection in the manufacturing industry
✓ Assisting humans in identifying various biological species
✓ Controlling robots in the industry
✓ Visual surveillance
✓ Image modeling for medical image analysis
✓ Navigation for self-driving cars and autonomous mobile robots
✓ And many more.

The first aim of the proposed model is to identify the trolley or the basket that is there in the queue. For this purpose, YOLOv4 can be utilized as it is 10% faster than its predecessor YOLOv3, which was already very fast [21]. After identifying the trolleys and baskets, the number of trolleys and baskets is to be calculated. The program will identify the trolley or basket and track its movement in the queue. To calculate the quantity of the products present in the trolley, the two cameras installed on the top of the checkout counter will be used. To calculate the volume in the trolley/ basket, stereo vision technology would be used, which uses the dual-camera setup. Stereo Vision uses the concept of triangulation which uses the two known points to find the third point in the triangle. The two cameras are positioned at some distance from each other, which enables them to see the trolley from different angles, just as the human eye does. Analyzing the similarity between the two images, AI can create a 3-dimensional depth model of the trolley, which would help in determining the number of products present in the trolley/ basket [22]. This information would be further passed in the Deep Neural Network as a parameter to estimate the queue time.

4.2 Artificial Neural Network and Deep Learning
The next aspect used in the proposed framework is the Artificial Neural Network (ANN) and Deep Learning. In simple words, ANN attempts to better understand the Human Neural Network (HNN) by
implementing it on the machine and understanding it by working with it. ANN has an input layer from where it takes the input, one hidden layer, and one output layer. As there are neurons in HNN, there are nodes in ANN that take input from the previous node, apply weight or bias to it and pass it to another node [23]. The weights/ biases are calculated by training the model which is done by providing it with various inputs and their respective outputs. The weights/ biases are further improved by using special algorithms such as Gradient Descent and Back Propagation, which are coded in a way to bring the predicted output as close as possible to the actual output [24].

As the complexity increases, more hidden layers are added to the model and it is called Deep Neural Network. Training of this Deep Neural Network is often named Deep Learning. In the proposed system, Deep Learning forms the backbone of the model, where various parameters are passed, and the model calculates the queue time. It is also used in calculating the efficiency of the employee who is standing behind the checkout counter. The efficiency is calculated in 2 phases. The first phase calculates and evaluates the barcode scanning speed of the employee, for example, how many items they can scan in a unit time [25]. The second phase calculates and evaluates the time they take to proceed with payment, either in cash or using a card. As the employee gets better with practice and time, the weights/ biases get updated with time in the Neural Network's hidden layer. Once all these values are calculated they are used as input in the final model and the estimated queue time is calculated.

5. The Model

5.1 The Flowchart

Figure 2 represents the flowchart.
5.2 Advantage of using computer vision to calculate queue time

Conventional queuing theories have a mathematical formula to estimate the queue time by considering the inter-arrival time distribution, capacity of queue, and the number of servers. The queuing system can be designed with mixed queuing models with single-stage or two-stage queuing wherein weighing and cashier are different [26]. The queuing model could be a M/M/1 if the cashier and weighing are different. Fuzzy queuing models have also been experimented in counters servicing customers and found better performance compared to the conventional queuing models. Dong, Shah & Wong’s (DSW) algorithm was used for the fuzzy association function in this model.

Previous work in this field has been also done using various technologies such as predictive analytics, mobile sensing, RPROP Neural Network, etc. Using computer vision over previous other models and technology is advantageous as the proposed model calculates the queue time in real-time. Other models use past data and trends of the shop to give an estimate of the queue time which might not be accurate at every scenario, as trends change. This model takes care of real-time changes in the queue, such as a customer leaving the queue or a customer joining the queue. The model not only helps the customer in selecting the best queue but also the supermarkets to efficiently utilize all the counters. This is possible as the customer will always select the queue with less queue time which will prevent over-crowding and under-crowding of the queues. This will also reduce the time wastage of the employees and help and encourage them to perform better every day.

5.3 Performance evaluation measures

As the main task of the model is to predict and estimate the time that the customer will take to reach the counter, the model can evaluate itself by comparing the predicted time and the actual time taken to reach the counter. The model can be judged on various small parameters such as accuracy of identification of the trolley and keeping track of the same in the queue. The accuracy of the stereo vision algorithm used to calculate the number of products in the trolley will also be measured. To expand the accuracy of the model, training on real-time data is highly recommended.

5.4 Challenges in Implementation

Although theoretically, many things sound and look perfect, they might not be perfect at the time of implementation due to various uncertainties. As the behavior and movement of customers cannot be predicted, there are few challenges the proposed system may face.

- Customers Blocking the Camera's view of the trolley: As the customers are standing in a line, behind each other, and so are the trolleys, it will be a challenge to identify and calculate the number of products in the trolley.
- The calculation will be harder on the baskets: As baskets are smaller in size and usually hand-held by the customers, it is difficult to calculate the number of products in it.
- Shortage of cash or network error in the card machine: The employee on the counter might suddenly run out of change to return to the customer which could cause a delay in the line. Poor network reception on the card swiping machine could also cause the same kind of problem.
- Item missing barcode: This happens many times when the barcode tag gets dropped/removed or the packing gets wrinkled, which makes it difficult to scan. In this situation, the employee must type the code and make the entry that causes delay.
- Age of the Customer: It is observed that with the increase in age, the activeness of humans tends to decrease. Hence the younger customer might perform tasks faster, and older customers might take a little more time for the same task.

6. Conclusion

This paper proposed the technique to estimate the queue time with the help of Computer Vision which uses a dual camera setup and Deep Neural Network. As the queues are optimized, there is comparatively less waiting time and all the checkout counters are properly utilized. And as already
mentioned in the psychology of waiting: "Principle 4: Uncertain Waits are Longer than Known, Finite Waits", this helps the customer with a better queueing experience. Once this system is implemented effectively, various other factors that influence the queue time at the checkout counter can be added with time which might include the age of the customer. The younger customers are comparatively faster and active than the older customers and hence they take a little more time in the queue. Multiple cameras setup can be utilized for long queues to cover the customers standing behind. Other similar features can be added to the system to predict/ estimate the queue time more accurately.

Conflict of Interest: There is no conflict of interest among the authors

References
[1] Algburi, M. H., & Albayrak, S. (2018). Store products recognition and counting system using computer vision. Proceedings - 9th International Conference on Computational Intelligence and Communication Networks, CICN 2017, 2018-Janua, 221–224. https://doi.org/10.1109/CICN.2017.8319389.
[2] Baz, I., Yoruk, E., & Cetin, M. (2016). Context-aware hybrid classification system for fine-grained retail product recognition. 2016 IEEE 12th Image, Video, and Multidimensional Signal Processing Workshop, IVMSP 2016, 0–4. https://doi.org/10.1109/IVMSPW.2016.7528213.
[3] Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. May. http://arxiv.org/abs/2004.10934
[4] Bonk, E. (2016). The influence of activeness and independence on the quality of life of senior citizens. Polish Psychological Bulletin, 47(3), 338–345. https://doi.org/10.1515/ppb-2016-0040
[5] Borah, R., & Rajarajeswari, S. (2017). A study on application of machine learning and computer vision for retail projects. Asian Journal of Pharmaceutical and Clinical Research, 10, 476–480. https://doi.org/10.22159/ajpcr.2017.v10s1.20522
[6] Carvalho, A., & Belo, O. (2016). Predicting waiting time in customer queuing systems. 2016 IEEE International Conference on Knowledge Engineering and Applications, ICKEA 2016, 155–159. https://doi.org/10.1109/IICKEA.2016.7803010
[7] Connell, J., Fan, Q., Gabbur, P., Haas, N., Pankanti, S., & Trinh, H. (2013). Retail video analytics: an overview and survey. Video Surveillance and Transportation Imaging Applications, 8663, 86630X. https://doi.org/10.1117/12.2008899
[8] Cootes, T. F., & Taylor, C. J. (2001). Statistical models of appearance for medical image analysis and computer vision, in Medical Imaging. Image Processing -- Proceedings of SPIE, 4322, 238–248.
[9] Do, T. D., Duong, M. T., Dang, Q. V., & Le, M. H. (2018). Real-Time Self-Driving Car Navigation Using Deep Neural Network. Proceedings 2018 4th International Conference on Green Technology and Sustainable Development, GTSD 2018, 7–12. https://doi.org/10.1109/GTSD.2018.8595590
[10] Fiala, M. (2004). Vision guided control of multiple robots. Proceedings - 1st Canadian Conference on Computer and Robot Vision, 241–246. https://doi.org/10.1109/CCCRV.2004.1301450
[11] Hirschm, H. (2001). Comparison between Lemaitre and Gurson damage models in crac. 22. papers2://publication/uuid/27A7D934-36A3-460D-A17A-935AC9BA8C0B
[12] Kim, C. E., Dar Oghaz, M. M., Fajtl, J., Argyriou, V., & Remagnino, P. (2019). A comparison of embedded deep learning methods for person detection. VISIGRAPP 2019 - Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, 5, 459–465. https://doi.org/10.5220/0007386304590465
[13] Kruger, R. P., & Thompson, W. B. (1981). A Technical and Economic Assessment of Computer Vision for Industrial Inspection and Robotic Assembly. Proceedings of the IEEE, 69(12), 1524–1538. https://doi.org/10.1109/PROC.1981.12199
[14] Kyritsis, A. I., & Deriaz, M. (2019). A machine learning approach to waiting time prediction in queueing scenarios. *Proceedings - 2019 2nd International Conference on Artificial Intelligence for Industries, AI4I 2019*, 17–21. https://doi.org/10.1109/AI4I46381.2019.00013

[15] Luo, Q., Cao, Y., Liu, J., & Benslimane, A. (2019). Localization and Navigation in Autonomous Driving: Threats and Countermeasures. *IEEE Wireless Communications*, 26(4), 38–45. https://doi.org/10.1109/MWC.2019.1800533

[16] Mann, L. (1970). The social psychology of waiting lines. *American Scientist*, 58(4), 390–398.

[17] Murphy, C. (2018). Predicting Queueing Times Using Queueing Theory & Predictive Analytics (Doctoral dissertation, Trinity College).

[18] Rai, H. G. N., Jonna, K., & Krishna, P. R. (2011). Video analytics solution for tracking customer locations in retail shopping malls. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 773–776. https://doi.org/10.1145/2020408.2020533.

[19] Ramanathan, V., Huang, J., Abu-El-Haija, S., Gorban, A., Murphy, K., & Fei-Fei, L. (2016). Detecting Events and Key Actors in Multi-person Videos. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-Decem, 3043–3053. https://doi.org/10.1109/CVPR.2016.332.

[20] Satya Hermanto, R. P., Suharjito, Diana, & Nugroho, A. (2018). Waiting-Time Estimation in Bank Customer Queues using RPROP Neural Networks. *Procedia Computer Science*, 135, 35–42. https://doi.org/10.1016/j.procs.2018.08.147.

[21] Wäldchen, J., & Mäder, P. (2018). Plant Species Identification Using Computer Vision Techniques: A Systematic Literature Review. In *Archives of Computational Methods in Engineering* (Vol. 25, Issue 2). Springer Netherlands. https://doi.org/10.1007/s11831-016-9206-z.

[22] Wang, J., Wang, Y., Zhang, D., Wang, L., Chen, C., Lee, J. W., & He, Y. (2017). Real-time and generic queue time estimation based on mobile crowdsensing. *Frontiers of Computer Science*, 11(1), 49–60. https://doi.org/10.1007/s11704-016-5553-z.

[23] Zaki, N. H. M., Saliman, A. N., Abdullah, N. A., Hussain, N. S. A. A., & Amit, N. (2019). Comparison of queuing performance using queuing theory model and fuzzy queuing model at check-in counter in airport. *Mathematics and Statistics*, 7(4), 17–23. https://doi.org/10.13189/ms.2019.070703.

[24] Zhang, H. B., Yuan, K., & Liu, J. D. (2003). A fast and robust vision system for autonomous mobile robots. *Rissp 2003*, 2003-Octob(October), 60–65. https://doi.org/10.1109/RISSP.2003.1285549.

[25] Ziegler, M., Rudorfer, M., Kroischke, X., Krone, S., & Krüger, J. (2019). Point Pair Feature Matching: Evaluating Methods to Detect Simple Shapes BT - Computer Vision Systems. In *Computer Vision Systems. ICVS 2019*.

[26] Luo, R., & Shi, Y. (2020). Analysis and optimization of supermarket operation mode based on queueing theory: Queuing and pricing of personalized service. *ACM International Conference Proceeding Series*, 221–224. https://doi.org/10.1145/3380625.3380635.

[27] Rajavelu, T. (2015). Motion estimation search algorithm using new cross hexagon-diamond search pattern. *Int. J. MC Square Sci. Res*, 7(1), 104-111.