Molecular & Cellular Biomechanics 2025, 22(2), 688.
https://doi.org/10.62617/mcb688

Article

Enhancing supply chain resilience: The role of security practices and performance in mitigating disruptions in ghana's manufacturing sector

Fugang Guo[1,*], Mohd Azwardi Md Isa[2], Noor Azura Azman[2]

[1] Ghazalie Shafie Graduate School of Government, Universiti Utara Malaysia, Sintok 06010, Malaysia
[2] International business department, Universiti Utara Malaysia, Sintok 06010, Malaysia
*** Corresponding author:** Fugang Guo, Guofuganga@gmail.com

CITATION

Guo F, Isa MAM, Azman AN. Enhancing supply chain resilience: The role of security practices and performance in mitigating disruptions in ghana's manufacturing sector. Molecular & Cellular Biomechanics. 2025; 22(2): 688. https://doi.org/10.62617/mcb688

ARTICLE INFO

Received: 1 November 2024
Accepted: 13 November 2024
Available online: 23 January 2025

Abstract: At the current stage, the retail industry is undergoing unprecedented changes. From traditional physical stores to online shopping platforms, and then to the new retail model that integrates online and offline, customers' demand for shopping experience is constantly changing. To meet these demands, retailers need to constantly explore and apply new technologies to optimize the retail environment and enhance customer experience. Biomechanics is the study of the internal and external mechanical behavior of living organisms, which is concerned with the structure, function and motion laws of living organisms. Applying the knowledge of biomechanics to retail environment design can effectively improve customers' shopping experience. Based on this, this paper takes intelligent container as an example, gives a visual solution of detecting goods in intelligent container based on deep neural network, and proposes a twin-based pairwise image difference detection algorithm named DiffNet as the core algorithm of intelligent container solution, which aims to help enterprises deploy intelligent container flexibly, safely and at low cost. Enhance the customer's offline self-service shopping experience.

Keywords: biomechanics; retail environment; customers; shopping experience; digital technology

1. Introduction

With the popularization of the Internet and e-commerce, shopping and consumption are becoming more and more convenient. People can directly buy various commodities from all over the world on the e-commerce platform through the Internet, and then have them delivered to their homes by manual logistics [1]. However, after a period of rapid development, the e-commerce industry has gradually entered a bottleneck period, facing many challenges such as the weak growth of online traffic and the increase of logistics and distribution costs. The ceiling of the development of traditional e-commerce has appeared, and only through reform can there be a breakthrough. In the face of the challenges, domestic e-commerce enterprises put forward the concept of "new retail" [2].

In this context of development, smart container came into being, aiming to take consumers as the center, contact online e-commerce and offline stores at the same time, for consumers to achieve "scan code open the door—select goods—close the door automatic settlement" real-time convenient services [3]. However, because the smart container is a new business model, there are many solutions, but no solution can occupy the majority of the market share with absolute advantages, the smart container solutions on the market have gravity sensing technology, RFID technology or pure visual solutions.

From the current academic research, the intelligent container solution usually includes the customer level, the intelligent container level and the background business management level. The process at the customer level is to open the container door, close the door after selecting the goods, and pay according to the settlement results given by the intelligent container; the task at the intelligent container level is to obtain the customer's consumption information to provide information support for the subsequent tasks; the tasks at the backstage business management level include transaction management, commodity management, container management, data monitoring, and so on, which provide services for consumers and operators. According to the different methods of obtaining consumption information, the solutions can be categorized into two types: non-visual solutions and visual solutions. There are two types of non-visual solutions, namely gravity sensing technology and RFID radio frequency identification technology. Gravity sensing technology uses sensors to sense changes in weight on the shelves to detect the behavior of goods being picked up or put back. RFID radio frequency identification technology through radio frequency signals to achieve the target object for the required data information reading, the use of high-frequency microwave for short-range identification communication, its most important feature is non-contact identification. RFID advantage is that the technology has been quite mature, the product is easy to realize the scale of the landing, the market has been put into use today, such as Amazon, Jingdong intelligent container, the vast majority of the use of RFID. The vast majority of the use of RFID technology. Visual solutions are divided into dynamic solutions and static solutions. Compared to non-vision solutions, vision solutions have a higher technical threshold, while the actual cost is cheaper because of the marginal cost effect. Although vision algorithms based on convolutional neural networks can achieve a fairly high accuracy in the laboratory environment, but in terms of the actual landing and large-scale deployment of the product, there is no unified solution, whether it is a dynamic or static solution.

In recent years, deep learning has made a major breakthrough in various basic fields of image processing. The pure vision solution based on deep learning proposed for intelligent containers has high accuracy, fast speed and strong robustness, and the solution deployment is relatively convenient. Compared with traditional unmanned containers, the utilization rate of the space inside the container is higher. The vision solution based on deep learning is the future development trend of intelligent container.

2. Container commodity detection based on object detection algorithm

At the current stage, the common practice of intelligent container manufacturers using static vision solutions on the market. Although this solution can achieve a very good accuracy index, it also has the problems of heavy workload, high time cost and high labor cost for commodity detection model data training. This chapter proposes a pairwise image difference commodity detection algorithm based on deep neural network. The input of the algorithm is the commodity images in the container two moments before and after consumption. The output of the algorithm is the location of the different objects on the two images [4]. On the basis of the results obtained by our algorithm, a commodity type recognition model is used to identify the goods where the different objects are located. By comparing the two different commodities at the two moments, functions such as automatic settlement function and inventory margin can be completed. The process is shown in **Figure 1** below.

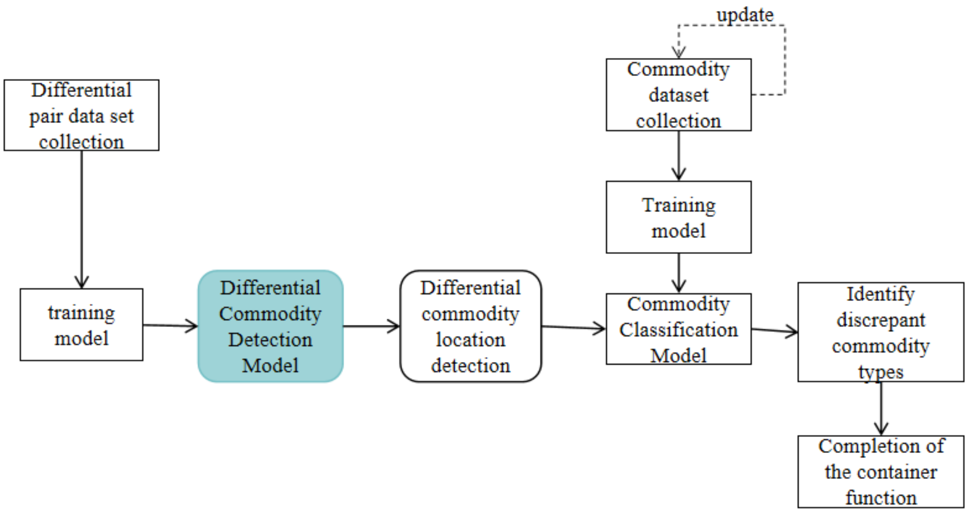

Figure 1. Intelligent container solution based on differential commodity detection.

(1) Differential detection model training. Collect and calibrate the difference pair data set, and train the difference detection model.

(2) Commodity classification model training. Collect and calibrate commodity identification data set, and train commodity classification model.

(3) Detection and identification of differentiated commodities. The container images of the two moments before and after consumer consumption are input into the difference detection network to obtain the location of the difference goods on the two charts, and then cut the difference goods down to enter the commodity classification model to identify the commodity type, analyze the difference goods at the two moments, and complete the automatic settlement, inventory margin management and other functions.

(4) Optimize and update. When there are new products on the shelves, update the product type data set, train the new product classification model, and complete the container product category update.

3. Pair image difference detection network

The proposed algorithm for paired image difference detection in this paper is named DiffNet and its specific structure is shown in **Figure 2**.

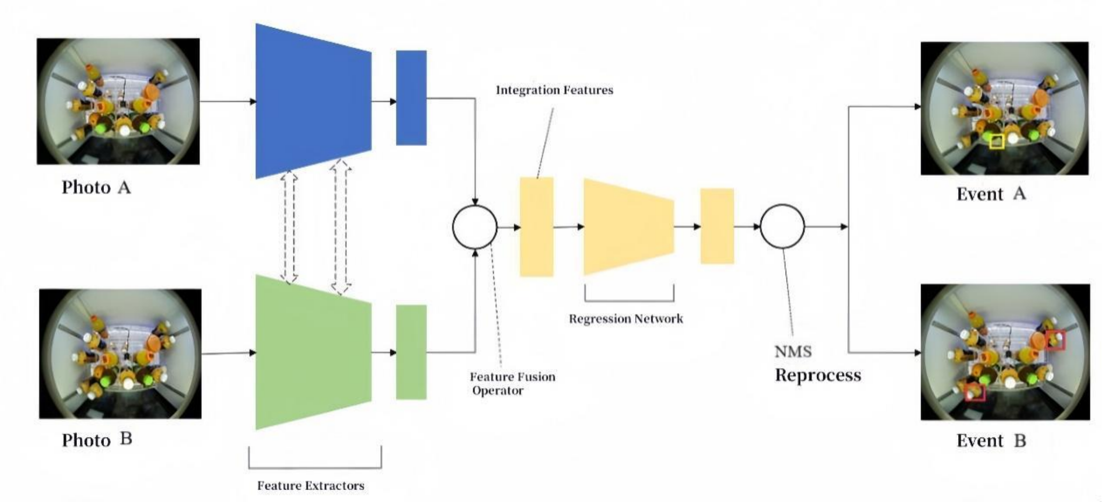

Figure 2. DiffNet structure diagram.

Since spatial features are very important for detecting the different positions of two images, we design the network as a full convolutional network to preserve the spatial features of the original image pairs. The two feature extractors of parameter sharing map the two input RGB images to the same feature space, and then use the feature fusion operator to extract the preliminary difference information of the two inputs according to certain rules; Finally, the difference features obtained by the fusion operator are input into a simple regression network, and the predicted bounding boxes and the images to which these bounding boxes belong are output. After the NMS (non-maximum suppression) post-processing, the final prediction results can be obtained.

3.1. Feature extractor

The input to the network is a pair of similar images, and the network uses two feature extractors with shared weights to extract features from the pair of input images separately. Since the parameters of the two feature extractors are shared, the two input images are mapped to the same feature space, which is extremely important for subsequently obtaining the difference in spatial location between the two images.

3.2. Feature fusion operator

Fusing the features extracted by the feature extractor to get the difference information is the key of this algorithm. There are many methods to get the difference of two input features, such as connecting two features into one feature by channel, subtracting by elements to get the difference features of two features, etc. After comparison experiments, we found that the fusion features obtained by subtracting by elements can get the best test results [5]. Subtracting by elements means to get a new feature map by calculating the difference between the corresponding elements of two feature maps, this method can capture the difference information between two features, which is good for change detection, anomaly detection and other tasks.

3.3. Regression network

After obtaining the fused features, the features are fed into the regression network, and after down sampling the feature map to a dimension matching the label size, the target's enclosing frame coordinates as well as its belonging pictures are quickly regressed [6]. In order to be able to better parse the dependency between the spatial differences of the fused features and the global features, the attention mechanism is considered to be introduced into the model, and two types of attention modules in spatial and channel dimensions are added to the regression network.

3.4. Network training

(1) Tag parsing:

DiffNet's detection scheme is closer to the single-step target detection method, calculating losses during training and doing backpropagation. During training, the labels of each pair of images are first parsed according to specific rules to obtain a label vector for the network to learn [7]. The input image is divided into $S \times S$ grids, each grid corresponds to a label represented by a vector of the form as follows:

$$[P(Obj), P(A|Obj), P(B|Obj), midx, midy, w, h]$$

In the formula, $P(Obj)$ indicates whether the center point of the label box falls into the grid, $P(A|Obj)$ and $P(B|Obj)$ indicate whether the center point of the label box falls on photo A or photo B (**Figure 2**), w and y represent the width and height of the label box.

(2) Preset anchor frame:

This paper introduces the preset anchor frame strategy. It is much simpler to predict the center coordinates and the offset of the length and width of the sample according to the preset anchor frame than the direct regression coordinates, which can simplify the regression problem and make the network easier to train [8]. For each prediction box, the network predicts the following: the probability that there is a target object in the prediction box, the probability that the object in the prediction box will fall on the two plots respectively, and the position of the prediction box. Specifically, for each grid we presuppose K anchor boxes, and for each anchor box we predict 1 object score, 2 position scores and 4 offsets relative to the anchor box shape (x,y,w,h). This results in 7K filters being applied around each grid in the feature plot, and a feature plot of size $m \times n$ yields a total of 7Kmn prediction boxes. The way the anchor boxes are matched during training is shown in **Figure 3** below, where the model models the task of detecting differences as a regression problem [9]. First, the image is divided into $S \times S$ grids, each of which predicts K bounding boxes. During the training period, only one anchor box is assigned to each true value, and the loss generated by other prior boxes only affects the object loss, and does not affect the coordinate loss and class loss.

Figure 3. Differences annotate the data set with examples.

(3) Loss function:

DiffNet loss function is divided into two parts: a priori frame loss and regression loss. Regression loss includes object loss, coordinate loss and class loss [10]. A priori loss only plays a role in the early T iterations of training, which is to help the network quickly learn the shape of the preset anchor frame in the early stage and make the network converge faster. The a priori frame loss function is defined as:

$$L_{prior} = \sum_{i=0}^{S\times S}\sum_{j=0}^{K} 1_{t<0}\left((x_{ij}-0.5)^2+(y_{ij}-0.5)^2\right)$$

In the formula, t is the total number of current training samples, and $1_{t<T0}$ indicates that the number of current training samples is less than the preset number $T._{0}$ Only when this condition is met, the prior frame loss is calculated. The regression function is defined as follows:

$$L_{noobj} = \sum_{i=0}^{S\times S}\sum_{j=0}^{K} 1_{ij}^{IOU<Thresh}\left(P_{ij}(Obj)-P_i(\hat{O}bj)\right)^2$$

$$L_{obj} = \sum_{i=0}^{S\times S}\sum_{j=0}^{K} 1_{ij}^{truth}\left(P_{ij}(Obj)-P_i(\hat{O}bj)\right)^2$$

In the formula L_{noobj}, whether the intersection ratio between the prediction box obtained from the JTH anchor point corresponding to the i-th cell in the feature map and the marked box falling into the cell is less than the preset threshold value *Thresh, the Thresh* used in the experiment = 0.6, if less, it is considered that there is no target object in the prediction box, and L_{noobj} is calculated for the prediction box. In L_{obj}, it indicates whether the JTH anchor box corresponding to the *i*-th cell matches the annotation box falling into the cell. If it matches, the loss value is calculated for the prediction box.

4. Experimental analysis

4.1. Data set setup

Because the algorithm proposed in this paper is relatively novel, there is no public data set for experiment. Therefore, the data set used in this paper is collected and labeled by itself. The task requirement is to detect the difference between two images, so we collect pairs of images taken in the container, and each pair of images is a simulation of the situation in the container two moments before and after consumer consumption [11]. Each image pair is composed of two images of picture A and picture B, which are taken by the camera in the container at the same Angle and at different times with A short time interval, so as to ensure that picture A and picture B are basically in the same light and background, but the commodities in the images are more or less different. The setting of this shooting condition is to simulate the situation of consumers in the automatic container for a consumption, generally speaking, the time of a consumption will not be too long, and the changes of goods brought about by consumption will not be particularly large.

After filtering out some invalid data, a total of 1896 difference pairs were obtained, of which 1638 image pairs were divided into the training set and 258 image pairs were divided into the test set. The tagging box of the image pairs is the location of the differences in each pair of images, and the tagging information is saved as an xml file. The tagging information records the surrounding box of the goods with differences in the same spatial position on the two images, and records the position and size in the form of coordinates of the upper left and lower right corner of the surrounding box [12]. The algorithm in this chapter only detects the different positions, but does not identify specific categories. Therefore, all the surrounding boxes are only considered as one class. If an image does not have an item that is different from another image, there is no corresponding annotation file for that image.

4.2. Evaluation criteria

In this paper, mAP indexes similar to common object detection are used to measure the performance of the model [13]. However, since the semantic information of categories in the algorithm in this paper is spatial, the average accuracy rate (AP) of each category in the mAP in this paper does not refer to the accuracy of various types specific to the object type, but to the picture A or B to which the object belongs.

4.3. Parameter settings

In order to better understand the algorithm in this paper and choose a better structure and training strategy, we have done some comparative experiments to observe the effects of components in the model and various Settings on the experimental results [14]. In all experiments, the other parameters of the training remained the same except that the specific control conditions were changed, and the image input size was fixed at 512 × 512. The experimental results are shown in **Table 1**, where the mAP index was obtained on a test set containing 258 pairs of different images, including 62 pairs of simple scene data and 196 pairs of complex scene data. The speed refers to the average reasoning time of the model for a single input image. The test environment is Ubuntu14.0 system and an NVIDIA GTX TITAN Xp graphics card with 12GB of video memory.

Table 1. DiffNet structure comparison experiment.

	Anchor Frame	Subtracting/joining	Random transform	Color enhancement	Attention module	mAP	Speed (ms/pair)
Experiment 1	×	Subtract	✓	-	-	92.30	7.71
Experiment 2	✓	Subtract	✓	-	-	94.20	7.16
Experiment 3	✓	Subtract	×	-	-	79.92	7.08
Experiment 4	✓	Connecting	✓	-	-	92.35	7.01
Experiment 5	✓	Subtract	✓	-	DA	92.40	8.58
Experiment 6	✓	Subtract	✓	-	CBAM	92.88	7.77
Experiment 7	✓	Subtract	✓	✓	-	95.56	6.98

Experiments 1 and 2 are compared to show that preset anchor frames have improved the accuracy of the model. Experiments 2 and 3 are compared to show that the data enhancement strategy of randomly replacing the order of the image pairs and their corresponding labels during the training process is very effective for training datasets with extremely heterogeneous label distributions, and the mAP is improved by close to 15% using this strategy. Experiments 2 and 4 are compared to illustrate that fusion of two input feature maps with elemental subtraction results in fused features that are more suitable for subsequent learning than fusion obtained by direct merging by channel [15]. Experiments 5 and 6 are compared to the inclusion of an attention module after the fusion features are obtained by elemental subtraction in the hope of enhancing the learning of the local importance of the difference features and constructing some kind of description of the difference environment. Experiment 7 is based on Experiment 2, adding more data enhancement strategies in terms of image color, and the experiment shows that more effective data enhancement has an improved effect on the accuracy of the model.

The algorithm in this paper has an inference time of only about 7 milliseconds for a single pair of inputs, achieving real-time detection in the true sense of the word, which is a favorable advantage for vision solutions for smart containers.

4.4. Result analysis

According to the comparative experiment in the previous section, the detection results of the model in experiment 7, which has the highest mAP index, are further analyzed. Considering that there are more complicated scenes in the sample scene than the simple container scene, the test set is divided into simple and difficult parts. In the simple test set, the commodity shape is more regular, while in the difficult test set, there are more types of commodities, most of which are irregularly shaped and placed in a more random way. There are 62 pairs of data in the simple test set, and the test accuracy rate is 100%, as shown in **Figure 4** below. The red rectangle in the figure is the mark box, and the green rectangle is the test result.

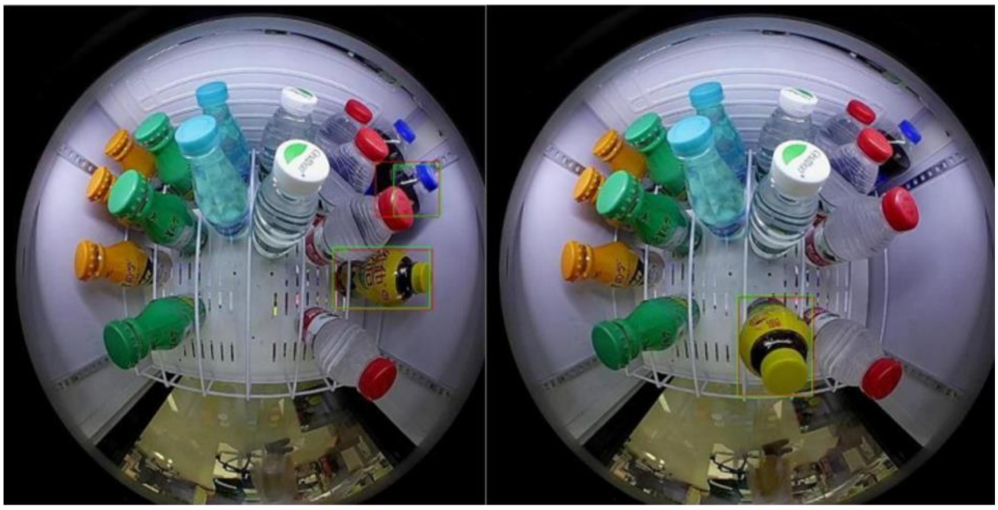

Figure 4. Visual result of simple test set.

The difficult test set has 196 pairs of data, the correct rate is 91.33%, and 17 pairs have detected problems. The reason for this is mainly because the goods are placed too densely, and the model has the probability of treating two items as one when detecting them, for which the problem needs to be continuously enriched with label parsing so that the model can recognize more goods in more complex environments. The visualization results of the difficult sample test are shown in **Figure 5** below. It can be seen that the model can also well detect the location of all the different items when the goods are densely placed and taken away a lot.

On this basis, 10 customers are invited to experience the smart container selling service optimized by this paper's algorithm, and the survey results show that all 10 customers are satisfied with this smart container, which also reflects the effect of this paper's retail environment optimization based on the DiffNet algorithm.

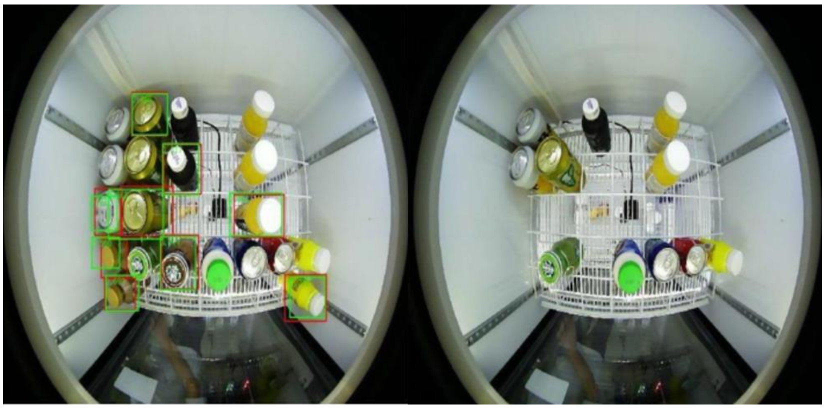

Figure 5. Visualized results of the difficulty test set.

5. Conclusion

The research topic of this paper is retail environment design based on biomechanics. A solution of smart container is presented. A twin-network pairwise image difference detection algorithm named DiffNet is proposed as the core algorithm of smart container solution. The algorithm can directly detect the consumer before and after the purchase of two containers in the scene image of the difference in the location of the goods, and then use the target recognition model to get the category of the different goods, you can complete the automatic settlement, inventory and other functions. Since DiffNet is a network that detects different locations, the marking work only marks the surrounding box information of the objects with differences in a pair of pictures. There is no need to accurately mark the surrounding box of all goods in the picture, which can save a lot of marking costs and updating costs, and has strong practical significance in the solution of intelligent containers.

Although the pairwise image difference detection algorithm DiffNet proposed in this paper can achieve completely correct differential commodity positioning in a relatively simple container scene, there will still be some misdetection and missed detection for complex container scenes with irregular placement and large difference in commodity shape, the next step is based on the analysis and summary of the wrong samples. Targeted data supplement for the wrong samples is conducive to the further improvement of the model.

Author contributions: Conceptualization, FG and MAMI; methodology, FG; software, MAMI; validation, FG, MAMI and NAA; formal analysis, NAA; investigation, FG; resources, FG; data curation, MAMI; writing—original draft preparation, NAA; writing—review and editing, FG; visualization, FG; supervision, MAMI; project administration, NAA; funding acquisition, FG. All authors have read and agreed to the published version of the manuscript.

Ethical approval: Not applicable.

Informed consent: Not applicable.

Conflict of interest: The authors declare no conflict of interest.

References

1. Liu L, Zhou B, Zou Z, et al. A smart unstaffed retail shop based on artificial intelligence and IoT. in: Proceedings of 2018 IEEE 23rd International workshop on computer aided modeling and design of communication links and networks(CAMAD). IEEE, 2018, 1–4.
2. Koehn D, Lessmann S, Schaal M. Predicting online shopping behaviour from clickstream data using deep learning[J]. Expert Systems with Applications, 2020, 150: 113342.
3. Kliestik T, Zvarikova K, Lăzăroiu G. Data-driven machine learning and neural network algorithms in the retailing environment: Consumer engagement, experience, and purchase behaviors[J]. Economics, Management and Financial Markets, 2022, 17(1): 57-69.
4. Shekokar N, Kasat A, Jain S, et al. Shop and Go: An innovative approach towards shopping using Deep Learning and Computer Vision[C]//2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT). IEEE, 2020: 1201-1206.
5. Cruz E, Orts-Escolano S, Gomez-Donoso F, et al. An augmented reality application for improving shopping experience in large retail stores[J]. Virtual Reality, 2019, 23: 281-291.
6. Hinton G, Thrun S. Using AI and Algorithms to Improve Online Shopping: Enhancing Customer Experience and Analyzing Market Trends[J]. AlgoVista: Journal of AI & Computer Science, 2024, 2(2).
7. Yi S, Liu X. Machine learning based customer sentiment analysis for recommending shoppers, shops based on customers' review[J]. Complex & Intelligent Systems, 2020, 6(3): 621-634.

8. Lee J, Won J, Lee D, et al. Customer shopping experience in a South Korea's Government-run home shopping channel for small and medium enterprises based on critical incident technique and unsupervised machine learning analysis[J]. Telematics and Informatics, 2022, 68: 101777.
9. Hawkins M. Metaverse live shopping analytics: retail data measurement tools, computer vision and deep learning algorithms, and decision intelligence and modeling[J]. Journal of Self-Governance and Management Economics, 2022, 10(2): 22-36.
10. Hussain M N, Harsha C, Shaik S, et al. Usage of deep learning techniques for personalized recognition systems in online shopping[C]//2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC). IEEE, 2023: 1739-1746.
11. Redmon J, Farhadi A. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767, 2018, 1.
12. Xinxiang Zhang. Ongoing trust and tourism O2O platform continuance: a two-trustee involved model with moderating variable[J]. SAGE Open, 2020:1–18.
13. P Duartea, S C Silvab, M B Ferreira. How convenient is it? delivering online shopping convenience to enhance customer satisfaction and encourage e-WOM[J]. Journal ofRetailing and Consumer Services, 2018,44:161-169.
14. Soni V. Deep Learning and Computer Vision-Based Retail Analytics for Customer Interaction and Response Monitoring[J]. Eigenpub Review of Science and Technology, 2021, 5(1): 1-15.
15. Sarwar M A, Daraghmi Y A, Liu K W, et al. Smart shopping carts based on mobile computing and deep learning cloud services[C]//2020 IEEE Wireless Communications and Networking Conference (WCNC). IEEE, 2020: 1-6.