Bulletin of Electrical Engineering and Informatics
Vol. 14, No. 4, August 2025, pp. 3299~3308
ISSN: 2302-9285, DOI: 10.11591/eei.v14i4.9533

A method for brand image recognition for ordering payment in supermarket

Nhat Nam Huynh, Tran Tan Hai Bui, Quyen Thi Nguyen
Faculty of Information Technology, School of Technology, Van Lang University, Ho Chi Minh City, Vietnam

Article Info

Article history:

Received Nov 8, 2024
Revised Jun 1, 2025
Accepted Jul 5, 2025

Keywords:

F1-score
GroZi-120
GroZi-3.2K
Precision
Product brand recognition
YOLOv8

ABSTRACT

This paper presents a product brand recognition method based on the YOLOv8 algorithm. The performance evaluation of the proposed method is conducted on two datasets consisting of GroZi-120 and GroZi-3.2K. The results show that the proposed method can achieve high accuracy. The precision and F1-score on the GroZi-120 and GroZi-3.2K datasets reach of {74.77%, 80%} and {99.86%, 100%}, respectively. The comparison with previous studies shows that the precision and F1-score obtained by the YOLOv8 method outperform some previous studies. Additionally, the effectiveness of the proposed method is also evaluated on a dataset of 6,170 images for twelve real products collected from supermarkets for use in order payment. The results show that the proposed method can be applied in single-order payment as well as multiple simultaneous orders with high accuracy in product recognition ranging from 94% to 98%. Therefore, the proposed method can be applied in order quick payment at supermarkets.

Corresponding Author:

Quyen Thi Nguyen
Faculty of Information Technology, School of Technology, Van Lang University
69/68 Dang Thuy Tram Street, Binh Loi Trung Ward, Ho Chi Minh City, Vietnam
Email: quyen.nt@vlu.edu.vn

1. INTRODUCTION

As making payments for goods at supermarkets or convenience stores, employees often use barcode scanning as a payment method [1]. This payment method helps reduce labor costs and provides a better shopping experience. However, this method still poses some significant risks. Quick response (QR) code payments can only scan one product at a time, leading to long waiting times when there are many customers. Ayodeji *et al*. [2], have shown that long checkout times negatively impact customer satisfaction. Additionally, blurry barcodes that the scanner cannot recognize or incorrect scanning angles can also lead to payment errors. Based on these risks, many researchers have applied various algorithms to solve problems related to object detection and classification based on their images [1].

Product recognition and classification is one of the object detection problems [3]. This problem can be solved by conventional methods such as scale-invarian-feature-transform or speeded-up-robust-features and deep learning techniques [1], [4]. In which, deep learning models have become more efficient than conventional models in addressing complicated jobs [4]. The goal of product recognition and classification is to find and identify a specific object within an image or video. The general principle of product recognition for purpose of checkout system in supermarket is that a detector such as a camera is used to collect image data that includes multiple products. Then the individual product images are cropped. Next, multiple single-product images are extracted from the original image, which includes several products. Finally, each extracted image is fed into the classifier, transforming the product recognition process into an image classification task. This process typically involves three main stages: detection, feature extraction, and classification. Currently, there are some researches utilizing deep learning models to tackle object detection problems. Among these, you-only-look-once (YOLO) is recognized as a highly efficient method. There have been numerous YOLO versions developed, wherein YOLOv8 introduces novel features and optimizations, making it an ideal choice for various object detection tasks across a wide range of applications. For instance, a system for detecting defective packaged products in industrial quality control is presented in [5]. Additionally, a modification of the YOLO network has been applied to detect metal surface defects in real-time [6]. YOLOv8 has been successfully applied to detect and localize counterfeit objects in videos [7]. YOLO is presented to detect Camellia oleifera fruit maturity [8]. YOLO is used for wildlife detection to balance accuracy and speed [9]. The YOLO is successful applied to detect lightweight vehicle [10]. YOLO is also proposed for detecting of insulator defect in power system [11]. Furthermore, YOLO has aslo been successfull applied for many problems in different fields such as classification image of blood cell [12], recognition of human activity [13], detection of of rice grains [14], person detection in infrared images [15], safety monitoring [16], and fault location of the power system [17], [18]. In addition to the above fields, YOLO has also been successfully applied in the field of goods management in supermarkets or convenience stores. Selvam and Koilraj [19] have proposed a framework that combines retail product detection, product-text detection, and product-text recognition using YOLOv5. Do and Pham [20], the authors have proposed a method based on YOLOv3 to manage goods in supermarkets. These studies collectively highlight the potential of YOLO algorithms in enhancing the accuracy and efficiency of inventory management in retail settings.

YOLOv8 developed in 2023 is used to detect objects in images, classify images and discern objects from each other. YOLOv8 has many improvements over previous versions and delivers high speed and accuracy performance while maintaining a streamlined design that makes it suitable for various applications [21]. Therefore, this research utilizes the YOLOv8 algorithm for the detection and classification of products using brand images obtained from supermarkets. The algorithm's performance is assessed on established datasets like GroZi-120 and GroZi-3.2K [1]. Furthermore, a database comprising 6,170 images of twelve real found in convenience stores and supermarkets has been created to facilitate product identification and classification, ultimately improving the supermarket checkout process, lowering labor costs, and enhancing the customer shopping experience. The contribution of this paper is summarized below:

- Application of the YOLOv8 algorithm for product recognition.
- Evaluation results on two datasets consisting of GroZi-120 and GroZi-3.2K demonstrate that the proposed method has a high accuracy in product recognition compared to previous methods.
- Successful application of the YOLOv8-based product recognition method for product identification in convenience stores/supermarkets to support order payment. Experimental results show that the proposed method is capable of accurately recognizing products for both single and multiple order payments.

The structure of the paper is organized as follows: The introduction is presented in this section. The product recognition problem model using the YOLOv8 algorithm and the implementation steps are presented in sections 2 and 3. The results of applying the proposed method to the two datasets of GroZi-120 and GroZi-3.2K, as well as its application in order payment are presented in section 4. The conclusion is presented in the final section.

2. PRODUCT RECOGNITION MODEL USING THE YOLOV8 ALGORITHM

YOLOv8 has an architecture divided into three layers: backbone, neck, and head, as shown in Figure 1. The backbone network is utilized by the YOLOv8 model to extract features from input images. It converts the image into a set of features in the form of tensors, representing essential elements such as shapes, textures, and colours of objects in the image. YOLOv8 employs the cross-stage partial network architecture for its backbone, which helps reduce the computational cost while maintaining accuracy. The neck serves as the link between the backbone and the detection head. Its purpose is to merge the feature information from the three feature maps extracted by the backbone network, allowing for feature retention and the extraction of more detailed information. The neck combines features from different levels such as from shallow and deep layers to create a multi-scale representation. This helps the model effectively recognize objects of various sizes and scales. The neck network then generates new feature maps as output. Finally, the head of the network uses the processed features to predict bounding boxes, confidence scores indicating the presence of objects, and class probabilities for each detected object. This enables the model to not only identify the location but also classify objects in the image. Finally, non-maximum suppression is used to eliminate overlapping bounding boxes, keeping only those with the highest confidence score for each object. This process ensures that each object is detected only once with the highest accuracy, improving the reliability of the output.

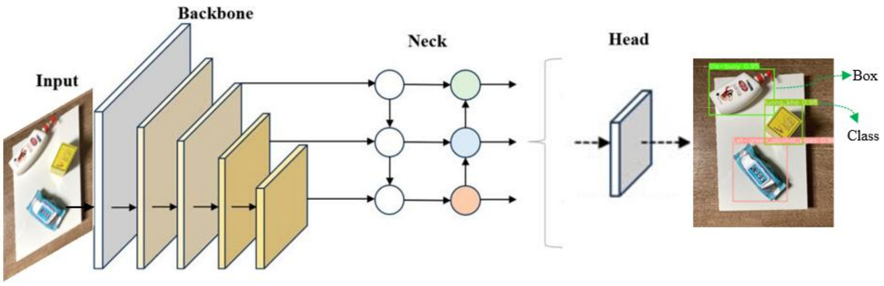


Figure 1. YOLOv8 model for product recognition

3. STEPS OF THE YOLO ALGORITHM FOR PRODUCT RECOGNITION

3.1. Step 1: data collection and labelling

A large amount of image data is collected from supermarkets, including products on shelves, in shopping carts, and in various other situations. Then, these images are labelled, indicating the specific location and type of product to create an accurate training dataset. Note that the size of all images used is adjusted to a standard size of 640×640.

3.2. Step 2: model training

YOLOv8 is trained on a labelled dataset with the goal of enabling the model to recognize product features from input images. The training process will optimize the model's ability to accurately and quickly detect and classify products. The data set is splitted into training, validation, and testing sets with the ratio of 80%, 10%, and 10%, respectively. The model's training and evaluation parameters are shown in Table 1.

Table 1. Parameters used for training and evaluating the YOLOv8 model in product recognition

Parameters	YOLOv8
Image size	640×640
Epochs	75
Plots	True
Patience	100

The model's training process is evaluated using the following metrics consisting of accuracy precision, recall, and F1-score. Precision indicates the proportion of true positive predictions among all positive predictions, while F1-score evaluates the model's overall performance. The values of these metrics are determined by (1)-(3):

$$Precision = \frac{TP}{TP+FP} \tag{1}$$

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

$$F1-score = 2 \times \frac{Precision \times Recall}{Precision+ Recall} \tag{3}$$

wherein, true positive (TP) and false positive (FP) represent the counts of correctly and incorrectly classified positive instances, respectively. True negative (TN) and false negative (FN) denote the counts of correctly and incorrectly classified negative instances, respectively. The structured flow of training and evaluation of the proposed approach for product recognition problem is shown in Figure 2.

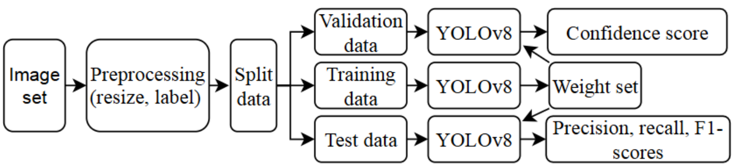


Figure 2. Training and evaluation process of YOLOv8 model for product recognition

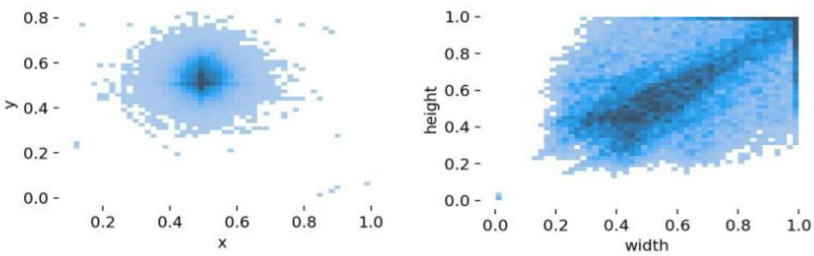
3.3. Step 3: object detection
In this stage, YOLOv8 rapidly identifies the location and size of products in an image using bounding boxes. The model is capable of processing high-resolution input images and detecting products even when they are occluded or in complex viewpoints.

3.4. Step 4: feature extraction
YOLOv8 employs feature layers to extract essential information from the detected objects. These features include shape, size, color, and other physical attributes, enabling the model to differentiate between similar products with subtle differences.

3.5. Step 5: product classification
Finally, YOLOv8 applies classification layers to determine the product category based on the extracted features. The model can classify products into specific groups such as food, household items, and beauty products, thereby accurately and efficiently completing the product recognition and classification process.

4. RESULTS AND DISCUSSION
4.1. Dataset
The GroZi-120 dataset consists of 120 product categories, 676 training images, and a total of 4,973 images for testing, with annotations for each product. These images were captured under good lighting conditions, and each image contains only a single product. In addition to images captured at various brightness levels, the dataset also includes 29 videos with a total duration of 30 minutes, covering all products in the training set. GroZi-3.2K is a dataset designed to evaluate zero-shot object detection performance, particularly for grocery products. It comprises 3,207 product categories with a total of 47,057 annotated images. The dataset covers a wide range of grocery items such as fruits, vegetables, dairy products, and packaged foods, providing a diverse set of objects for benchmarking object detection models.

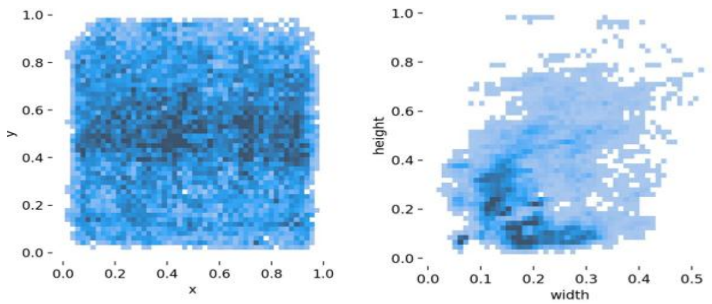
4.2. Numerical results
Figure 3 presents the distribution of bounding box positions and sizes for the GroZi-120 dataset. This visualization provides valuable insights into object appearance within the images. Figure 3(a) shows the density distribution of bounding boxes along the x-axes and y-axes, indicating the relative positions of objects. Most objects are centered in the image, with a high concentration in the middle. This information can be used to adjust the model's input size or configure convolutional layers to optimize object detection. Figure 3(b) analyzes the sizes of bounding boxes, revealing that objects are generally small, with only a few larger instances. The distribution of width and height shows that most bounding boxes have a width and height ranging from 0.2 to 0.8 of the total image size. This information is crucial for configuring appropriate convolutional layers to ensure the model can accurately detect and process objects of these sizes.



(a) (b)

Figure 3. Bounding box position and size of GroZi-120 dataset: (a) density distribution and (b) sizes

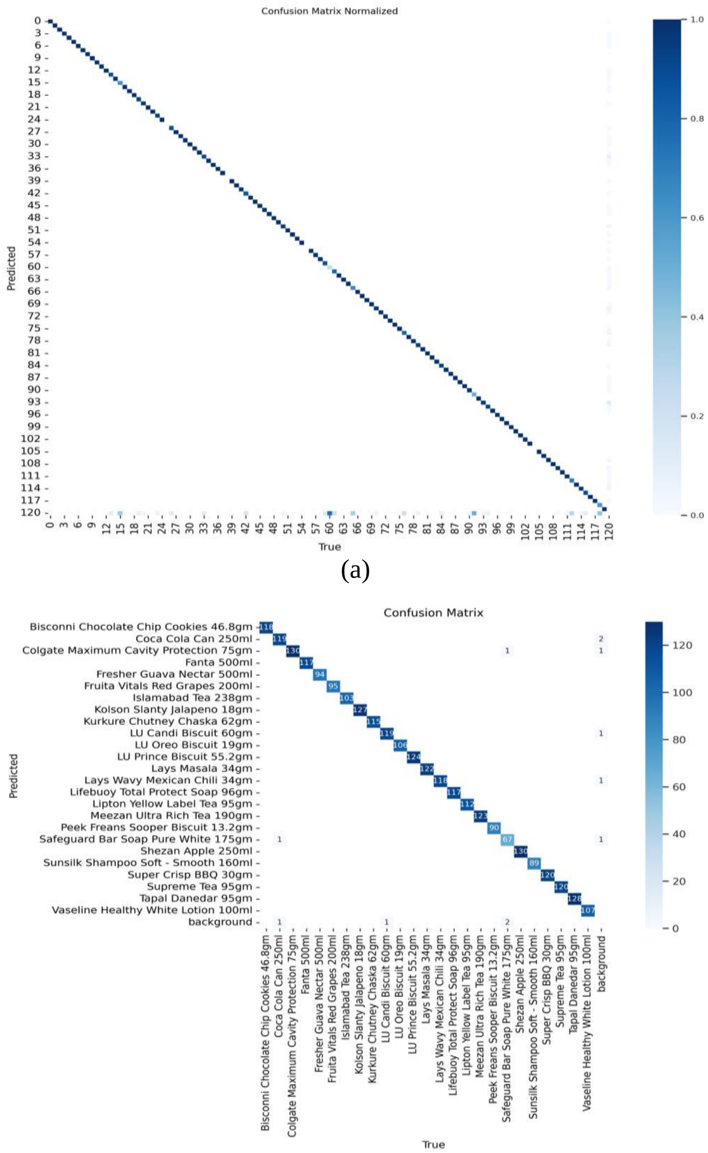
The distribution of bounding box positions and sizes for the GroZi-3.2K dataset is presented in Figure 4. Figure 4(a) shows that the centers of most bounding boxes are located near the center of the image, with x and y coordinates clustering around 0.5, further confirming that most objects in the dataset are positioned near the image center. The distribution of sizes in Figure 4(b) shows that most bounding boxes have a width and height ranging from 0.1 to 0.4 of the total image size. This finding can be leveraged to set object detection parameters for real-world applications where the central location of objects is crucial.



(a) (b)

Figure 4. Bounding box position and size of GroZi-3.2K datasets: (a) density distribution and (b) sizes

The confusion matrix of the model, as depicted in Figure 5 with GroZi-120 shown in Figure 5(a) and GroZi-3.2K in Figure 5(b), illustrates the model's effectiveness in accurately predicting the class of each object. Diagonal elements show correct predictions, while off-diagonal elements indicate incorrect ones. Each off-diagonal element reveals the confusion between two specific product classes.

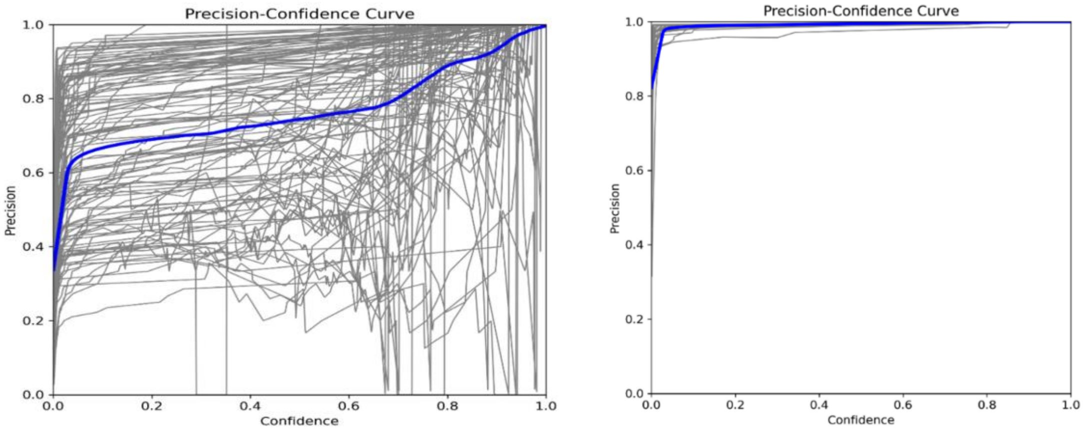


(a)

(b)

Figure 5. Confusion matrix of: (a) GroZi-120 and (b) GroZi-3.2K datasets

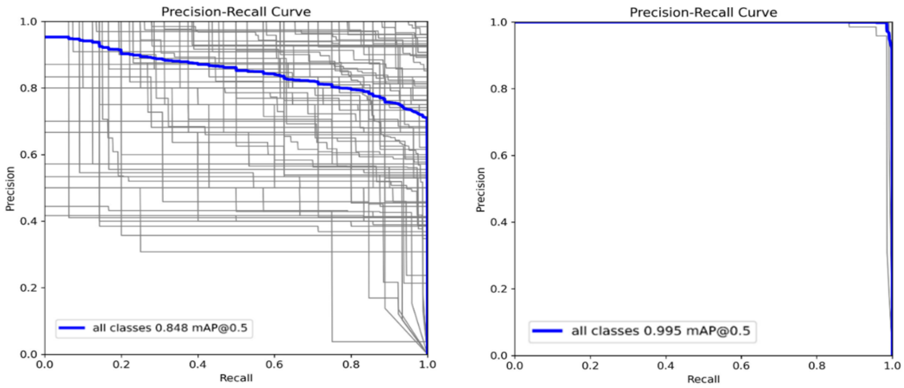
The model's classification performance on the two datasets is illustrated in Figure 6. The curve represents the relationship between precision and confidence in product prediction. Confidence refers to the threshold at which the model is certain that its prediction is correct. This value is typically calculated based on the probability assigned to each class by the model. Figure 6(a) shows that on the GroZi-120 dataset, the model achieves perfect precision at a confidence level of 1.00. Figure 6(b) reveals that for the GroZi-3.2K dataset, precision reaches nearly 1.0 when the confidence level approaches 0.925 and remains high as confidence increases further. This indicates that the model is highly accurate when it is highly confident in its predictions.



(a) (b)

Figure 6. Precision-confidence curve of: (a) GroZi-120 and (b) GroZi-3.2K datasets

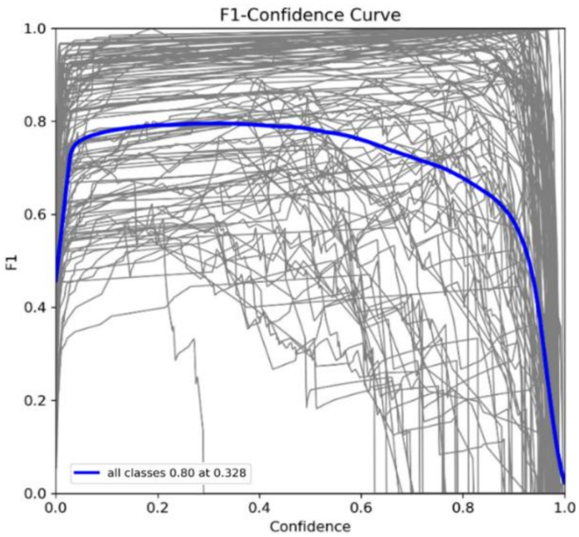
Figure 7 evaluates the model's performance by comparing precision and recall. Figures 7(a) and (b) provide an overview of the model's overall performance, with the blue curve representing the model's performance across all classes at high confidence thresholds. The results show that the model starts with high precision and gradually decreases as recall increases, which is typical for most classification models.



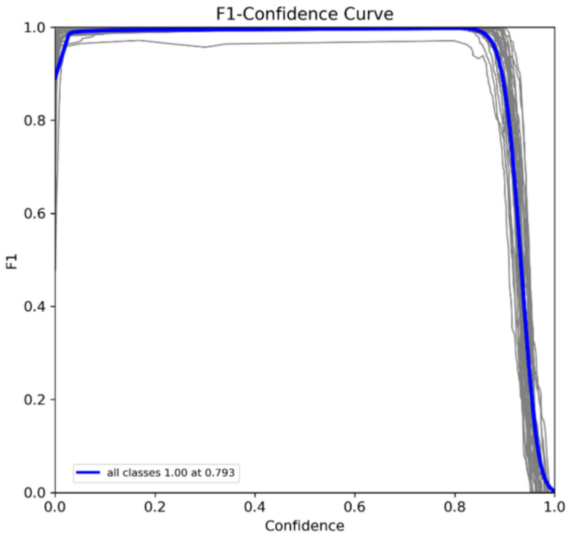
(a) (b)

Figure 7. Precision-recall curve of: (a) GroZi-120 and (b) GroZi-3.2K datasets

To evaluate the model's object detection performance, the F1-confidence curve is employed as shown in Figure 8. The confidence axis represents the confidence threshold ranging from 0 to 1, which determines whether a prediction is considered correct. A higher confidence threshold requires the model to be more certain about its prediction for it to be classified as correct. The F1-score axis represents the harmonic mean of precision and recall, offering a balanced assessment of the model's performance by considering both metrics. The primary curve marked in blue shows the model's F1-score at different confidence thresholds. The curve starts at a high F1-score at low confidence thresholds and gradually decreases as the confidence threshold increases, peaking at a certain point before dropping rapidly. This is typical when the model makes accurate predictions with high confidence, but the number of predictions decreases as the confidence threshold increases, leading to a decrease in precision or recall. Figure 8 shows that at a confidence threshold is approximately 0.328 in Figure 8(a), the model achieves an F1-score of 0.80, indicating a good balance between precision and recall at this confidence level. This could be the optimal threshold for applying this model in real-world scenarios where a balance between not missing objects and not making false predictions is required. At a confidence threshold of approximately 0.793 in Figure 8(b), the model achieves an F1-score of 1.00, indicating that the model is very strong when it is highly confident in its predictions, and it reaches its peak classification performance at a confidence threshold of approximately 0.793.



(a)



(b)

Figure 8. F1-confidence curve of: (a) GroZi-120 and (b) GroZi-3.2K datasets

A comparison of our proposed model with other state-of-the-art methods is presented in Table 2. The results demonstrate that YOLOv8 achieves the highest accuracy of 74.77% on the GroZi-120 dataset. For the GroZi-120 dataset, YOLOv8 outperforms the methods proposed in [22]–[24] by 24.97%, 25.72%, 29.57%, and 29.07%, respectively. Similarly, on the GroZi-3.2K dataset, YOLOv8 achieves the highest accuracy of 99.86%, significantly outperforming the other models. The accuracy achieved by YOLOv8 is 47.70%, 7.67%, 26.76%, and 76.37% higher than the methods proposed in [22]–[24], respectively.

Table 2. Comparison of precision results between the proposed method and other methods for the GroZi-120 and GroZi-3.2K datasets

Method	GroZi-120 dataset Precision (%)	GroZi-3.2K dataset Precision (%)
YOLOv8	74.77	99.86
Karlinsky *et al.* [22]	49.80	52.16
Geng *et al.* [23]	49.05	92.19
Deep neural network [24]	45.20	73.10
Bag of visual Words [24]	45.70	-

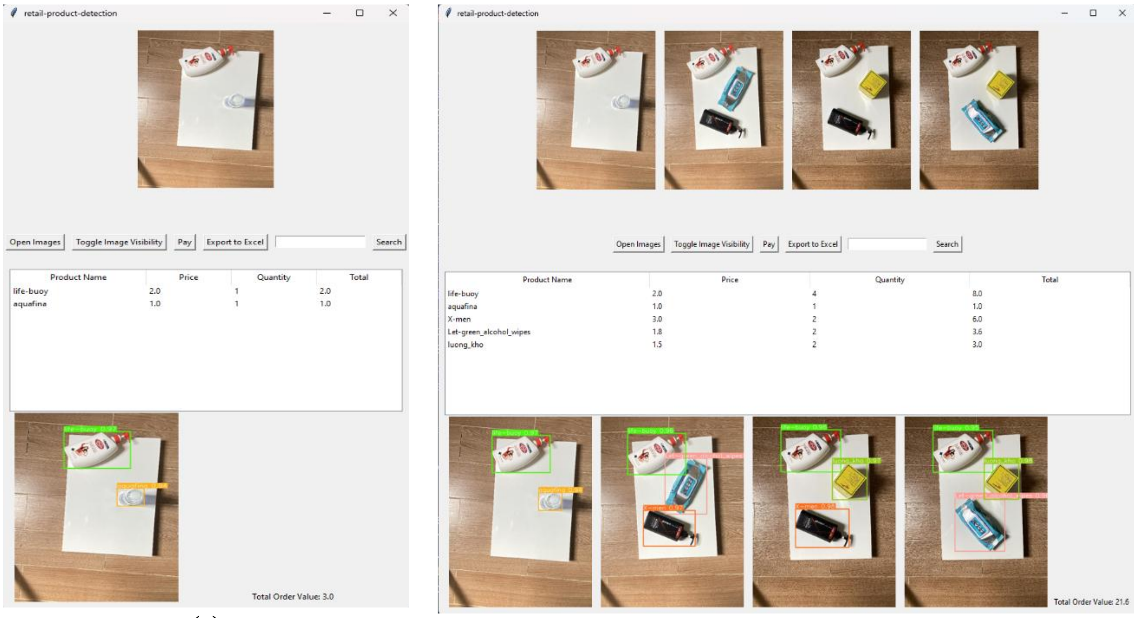
Additionally, Table 3 presents a comparison of F1-scores between YOLOv8 and other methods on the GroZi-120 and GroZi-3.2K datasets. The results demonstrate that YOLOv8 significantly outperforms the method in [25]. While YOLOv8 achieves F1-scores of 80% on GroZi-120 and 100% on GroZi-3.2K, the method in [25] only reaches 44.40% and 80.2%, respectively. These findings highlight the superior performance of YOLOv8 for product recognition problem.

Table 3. Comparison of F1-score between the proposed method and other methods for the GroZi-120 and GroZi-3.2K datasets

Method	GroZi-120 dataset F1-score (%)	GroZi-3.2K dataset F1-score (%)
YOLOv8	80	100
Santra *et al.* [25]	44.40	80.2

4.3. Application of the model for product recognition in supermarkets

In this section, the proposed method is applied to recognize 12 types of products in supermarkets and convenience stores. The dataset for these 12 products consists of 6,000 images for training, 120 images for validation, and 50 images for testing. This image dataset was collected from the internet and captured in various convenience stores and supermarkets. To ensure diversity and richness, the products were captured from single angles and under various lighting conditions. The user interface is shown in Figure 9 wherein the single order result is shown in Figure 9(a). When an image is selected from "Open image", a table appears displaying statistics such as product name, price, quantity, and total cost. Additionally, the interface allows for the visualization of images with labelled products and their corresponding recognition accuracy. For example, the recognition accuracy for the product "life-buoy" is as high as 0.97 and 0.94 for the product "aquafina". The recognition results of multiple orders for simultaneous payment are presented in Figure 9(b). These results demonstrate that the proposed method can accurately identify multiple images simultaneously and generate data based on the detected information. Specifically, products in these orders were recognized with high accuracy, ranging from 0.94 to 0.98. This result highlights the potential of applying brand image recognition methods to payment processes in stores or supermarkets.



(a) (b)

Figure 9. The interface displays the information of: (a) selected shopping cart and (b) multiple selected shopping carts

5. CONCLUSION

In this work, the YOLOv8 algorithm was successfully applied for brand recognition in images. Experiments conducted on two datasets demonstrated the high accuracy of the proposed method compared to some previous approaches. Specifically, precision and F1-score on the GroZi-120 and GroZi-3.2K datasets reached {74.77%, 80%} and {99.86%, 100%}, respectively. Compared to previous studies, the Precision and F1-score achieved by the YOLOv8 method outperformed several existing approaches. Furthermore, when applied to a dataset of 12 products in supermarkets and convenience stores, the proposed method demonstrated the ability to process single or multiple orders with high accuracy, ranging from 94% to 98% for product recognition. However, the number of products when applying the proposed method to identify real products in supermarkets is still limited. Therefore, for future research, real-world datasets from supermarkets and retail stores can be collected to further evaluate the effectiveness of the YOLOv8 method. Additionally, integrating real-time image reading devices to complete the product recognition and payment system in stores or supermarkets can be further developed.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Nhat Nam Huynh					✓	✓	✓	✓	✓	✓				
Tran Tan Hai Bui			✓			✓	✓	✓		✓				
Quyen Thi Nguyen	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	

C : **C**onceptualization	I : **I**nvestigation	Vi : **Vi**sualization
M : **M**ethodology	R : **R**esources	Su : **Su**pervision
So : **So**ftware	D : **D**ata Curation	P : **P**roject administration
Va : **Va**lidation	O : Writing - **O**riginal Draft	Fu : **Fu**nding acquisition
Fo : **Fo**rmal analysis	E : Writing - Review & **E**diting	

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Derived data supporting the findings of this study are available from the corresponding author Q.T.N on request.

REFERENCES

[1] Y. Wei, S. Tran, S. Xu, B. Kang, and M. Springer, “Deep Learning for Retail Product Recognition: Challenges and Techniques,” *Computational Intelligence and Neuroscience*, vol. 2020, 2020, doi: 10.1155/2020/8875910.

[2] Y. Ayodeji, H. Rjoub, and H. Özgit, “Achieving sustainable customer loyalty in airports: The role of waiting time satisfaction and self-service technologies,” *Technology in Society*, vol. 72, 2023, doi: 10.1016/j.techsoc.2022.102106.

[3] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, “Deep learning for computer vision: A brief review,” *Computational Intelligence and Neuroscience*, vol. 2018, no. 1, p. 7068349, 2018, doi: 10.1155/2020/8875910.

[4] G. Al-Rasheedi and R. Ullah Khan, “Facemask detection and classification using you only look once version 7,” *IAES International Journal of Artificial Intelligence*, vol. 13, no. 3, pp. 3330–3338, 2024, doi: 10.11591/ijai.v13.i3.pp3330-3338.

[5] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, “A Review of Yolo algorithm developments,” *Procedia Computer Science*, vol. 199, pp. 1066–1073, 2022, doi: 10.1016/j.procs.2022.01.135.

[6] D.-L. Pham and T.-W. Chang, “A YOLO-based real-time packaging defect detection system,” *Procedia Computer Science*, vol. 217, pp. 886–894, 2023, doi: 10.1016/j.procs.2022.12.285.

[7] A. Kashyap, “A novel method for real-time object-based copy-move tampering localization in videos using fine-tuned YOLO V8,” *Forensic Science International: Digital Investigation*, vol. 48, p. 301663, 2024, doi: 10.1016/j.fsidi.2023.301663.

[8] X. Zhu, F. Chen, Y. Zheng, C. Chen, and X. Peng, “Detection of Camellia oleifera fruit maturity in orchards based on modified lightweight YOLO,” *Computers and Electronics in Agriculture*, vol. 226, p. 109471, 2024, doi: 10.1016/j.compag.2024.109471.

[9] L. Chen, G. Li, S. Zhang, W. Mao, and M. Zhang, “YOLO-SAG: An improved wildlife object detection algorithm based on YOLOv8n,” *Ecological Informatics*, vol. 83, p. 102791, 2024, doi: 10.1016/j.ecoinf.2024.102791.

[10] H. Pan, S. Guan, and X. Zhao, “LVD-YOLO: An efficient lightweight vehicle detection model for intelligent transportation systems,” *Image and Vision Computing*, vol. 151, p. 105276, 2024, doi: 10.1016/j.imavis.2024.105276.

[11] F. Fahim and M. S. Hasan, “Enhancing the reliability of power grids: A YOLO based approach for insulator defect detection,” *e-Prime-Advances in Electrical Engineering, Electronics and Energy*, vol. 9, p. 100663, 2024, doi: 10.1016/j.prime.2024.100663.

[12] B. Wu, S. Feng, S. Jiang, S. Luo, X. Zhao, and J. Zhao, “EB-YOLO:An efficient and lightweight blood cell detector based on the YOLO algorithm,” *Computers in Biology and Medicine*, vol. 192, 2025, doi: 10.1016/j.compbiomed.2025.110288.

[13] S. Rajakumar and R. B. Azad, “A novel YOLOv8 architecture for human activity recognition of occluded pedestrians,” *International Journal of Electrical and Computer Engineering*, vol. 14, no. 5, pp. 5244–5252, 2024, doi: 10.11591/ijece.v14i5.pp5244-5252.

[14] C. Liu *et al.*, “Grain-YOLO: An improved lightweight YOLO v8 and its android deployment for rice grains detection,” *Computers and Electronics in Agriculture*, vol. 237, 2025, doi: 10.1016/j.compag.2025.110757.

[15] M. Li and N. Yan, “IPD-YOLO: Person detection in infrared images from UAV perspective based on improved YOLO11,” *Digital Signal Processing: A Review Journal*, vol. 168, p. 105469, 2026, doi: 10.1016/j.dsp.2025.105469.

[16] J. Wu, "Design and Exploration of Real Time Monitoring System for Power Grid Operation Safety Based on YOLO Object Detection Algorithm," *Procedia Computer Science*, vol. 228, pp. 889–897, 2023, doi: 10.1016/j.procs.2023.11.117.

[17] S. F. Stefenon, L. O. Seman, A. C. R. Klaar, R. G. Ovejero, and V. R. Q. Leithardt, "Hypertuned-YOLO for interpretable distribution power grid fault location based on EigenCAM," *Ain Shams Engineering Journal*, vol. 15, no. 6, p. 102722, 2024, doi: 10.1016/j.asej.2024.102722.

[18] H. Li *et al.*, "A lightweight RKM-YOLO algorithm for transmission line fault inspection," *Energy Reports*, vol. 14, pp. 1509–1522, 2025, doi: 10.1016/j.egyr.2025.06.052.

[19] P. Selvam and J. A. S. Koilraj, "A deep learning framework for grocery product detection and recognition," *Food Analytical Methods*, vol. 15, no. 12, pp. 3498–3522, 2022, doi: 10.1007/s12161-022-02384-2.

[20] H.-T. Do and V.-C. Pham, "Deep Learning Based Goods Management in Supermarkets," *Journal of Advances in Information Technology*, vol. 12, no. 2, pp. 164–168, 2021, doi: 10.12720/jait.12.2.164-168.

[21] M. Sohan, T. S. Ram, and C. V. R. Reddy, "A review on yolov8 and its advancements," in *International Conference on Data Intelligence and Cognitive Informatics*, Springer, 2024, pp. 529–545, doi: 10.1007/978-981-99-7962-2_39.

[22] L. Karlinsky, J. Shtok, Y. Tzur, and A. Tzadok, "Fine-grained recognition of thousands of object categories with single-example training," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4113–4122, doi: 10.1109/CVPR.2017.109.

[23] W. Geng *et al.*, "Fine-grained grocery product recognition by one-shot learning," in *Proceedings of the 26th ACM international conference on Multimedia*, 2018, pp. 1706–1714, doi: 10.1145/3240508.3240522.

[24] A. Franco, D. Maltoni, and S. Papi, "Grocery product detection and recognition," *Expert Systems with Applications*, vol. 81, pp. 163–176, 2017, doi: 10.1016/j.eswa.2017.02.050.

[25] B. Santra, A. K. Shaw, and D. P. Mukherjee, "Part-based annotation-free fine-grained classification of images of retail products," *Pattern Recognition*, vol. 121, p. 108257, 2022, doi: 10.1016/j.patcog.2021.108257.

BIOGRAPHIES OF AUTHORS

Nhat Nam Huynh is an excellent student at the Faculty of Information Technology, School of Technology, Van Lang University. He did his graduation thesis with the topic of YOLO. In addition, he has experience in developing recognition applications. He can be contacted by email: nam.207ct40449@vanlanguni.vn.

Tran Tan Hai Bui is an excellent student at the Faculty of Information Technology, School of Technology, Van Lang University. He did his graduation thesis with the topic of YOLO. In addition, he has experience in developing recognition applications. He can be contacted by email: hai.207ct40227@vanlanguni.vn.

Quyen Thi Nguyen received M.S. degree in Information Technology from Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam, in 2015. Currently, she is a lecturer at Faculty of Information Technology, School of Technology, Van Lang University, Vietnam. Her interests include optimization algorithms, artificial intelligence, and data mining. She can be contacted at email: quyen.nt@vlu.edu.vn.