Multi-Class Object Classification for Smart Retail Store Management Using "You Only Look Once" Model

X. Bai [1], S.Z. Jiang [2]
[1] Department of Artificial intelligence, Guangxi Polytechnic University, Nanning, 530226, China;
[2] Department of Intelligent Manufacturing, Guangxi Polytechnic University, Nanning, 530226, China

Abstract

Smart Retail Management relies heavily on multi-class item classification to automate inventory management and security. While deep learning techniques have shown promise, many methods struggle to achieve both high accuracy and real-time processing speeds. This paper evaluates the performance of the state-of-the-art You Only Look Once v8 network for the specific task of multi-class jewelry classification in a retail setting. To facilitate this, we developed a custom dataset of jewelry items. The You Only Look Once v8 architecture was trained on this dataset, and its performance was rigorously validated through extensive testing. Our results demonstrate that the standard You Only Look Once v8 model can deliver highly accurate and real-time classification, meeting the demanding requirements of smart retail management and establishing its viability for practical deployment.

Keywords: multi-class object classification, deep learning, smart retail management, YOLOv8, computer vision

Citation: Bai X, Jiang SZ. Multi-Class Object Classification for Smart Retail Store Management Using "You Only Look Once" Model. Computer Optics 2026; 50 (5): 1713. doi: 10.18287/COJ1713.

Introduction

With the rapid advancement of digital technologies, smart retail management (SRM) has emerged as a critical area of innovation in modern commerce [1, 2]. In such environments, automated systems capable of accurately identifying and classifying multiple types of products are essential for efficient inventory tracking [3], security monitoring [4], and customer service [5]. Multi-class object classification enables these systems to recognize diverse items [6], while object detection identifies both their categories and precise locations within an image or video frame [7, 8].

Recent advancements in computer vision and deep learning (DL) have significantly enhanced the accuracy of object detection models [9, 10]. Convolutional Neural Networks (CNNs) and their modern extensions have become central to this progress [11], offering the ability to automatically learn complex visual features and support real-time decision-making in retail contexts [12]. Despite these advancements, achieving both high accuracy and real-time performance remains challenging in practical deployments. Factors such as crowded store layouts, frequent object occlusions, reflections from glass surfaces, and variability in lighting conditions can still degrade recognition accuracy and processing efficiency.

Addressing these issues requires models that are not only fast but also resilient to real-world environmental factors. The YOLO family of architectures has become one of the most prominent solutions due to its end-to-end design and capability for real-time detection. The latest generation of this model introduces several enhancements in precision, feature extraction, and inference speed, making it an ideal foundation for specialized retail applications.

This study investigates the application of a modern YOLO-based model for multi-class jewelry classification in smart retail environments. A custom dataset of jewelry items was developed to train and evaluate the model under realistic conditions. Through systematic experimentation, the study assesses the model's ability to maintain both real-time processing speed and high classification accuracy. The findings demonstrate that this approach effectively meets the demands of automated retail management, providing a robust and practical solution for intelligent inventory and security systems.

The primary contribution of this work is a comprehensive evaluation of the YOLOv8 network for real-time, high-accuracy jewelry classification in smart retail management. While YOLOv8 is an existing architecture, its application and validation in this specific domain are novel, addressing a critical gap where current systems often fail to meet the dual requirements of accuracy and real-time performance. The main contributions of this work are as follows:

1. Development of a custom, annotated dataset tailored for jewelry item recognition in retail environments.
2. Rigorous performance analysis of the YOLOv8 model on this task, establishing strong benchmarks for precision, recall, and speed.
3. Practical demonstration that an off-the-shelf, state-of-the-art model can be effectively deployed to solve a challenging problem in smart retail, without the need for architectural modifications.

This document is organized as follows: Section 2 reviews the related literature, Section 3 presents the research methodology, Section 4 discusses the findings and results, and Section 5 concludes the study.

1. Review of Related Works

Recent YOLO models have dominated real-time retail detection tasks through improved accuracy–speed trade-offs. Ultralytics' YOLOv8 is a versatile anchor-free detector that supports multiple vision tasks; its family provides efficient

backbones and streamlined training APIs [13]. YOLOv9 further advances the series by introducing Programmable Gradient Information (PGI) and a new Generalized Efficient Layer Aggregation Network (GELAN) backbone to mitigate information loss during training [14]. These innovations enable YOLOv9 to achieve state-of-the-art COCO performance, such as 49.4 AP in 12 epochs (ResNet-50) and 63.2 AP with a Swin-L backbone after pretraining on Objects365 [14]. In practical retail benchmarks, fine-tuned YOLOv8 and YOLOv9 perform strongly. Skinderowicz fine-tuned the YOLOv8 and YOLOv9 models on the RPC checkout dataset, achieving an average accuracy of 95.16 % on the test split [15]. This demonstrates that pretraining and task-specific augmentation can effectively handle cluttered retail scenes. Similarly, a large-scale, multi-domain study found that among YOLO variants, YOLOv8 consistently outperformed on retail imagery [16]. These results underscore that both YOLOv8 and YOLOv9 variants are well-suited to multi-class retail product recognition, with YOLOv9's architectural advances boosting accuracy and YOLOv8's optimized pipeline yielding top domain performance.

Transformer architectures have emerged as powerful end-to-end detectors. The original DETR model eliminated post-processing (NMS) via set-based prediction, and subsequent variants (DAB-DETR, DN-DETR) improved convergence. In particular, DINO demonstrated that a contrastive denoising training scheme and mixed queries greatly accelerate DETR training: DINO achieved 49.4 AP (12 epochs) and 51.3 AP (24 epochs) on COCO with ResNet-50, an improvement of +6.0 AP over previous DETR-like methods [17]. More recently, real-time transformer detectors have appeared. For example, Baidu's RT-DETR series targets low-latency detection: RT-DETR introduced a hybrid convolution/transformer encoder and IoU-aware query selection to match YOLO-like inference speeds [18]. RT-DETRv3 further adds hierarchical dense supervision during training to address sparse Hungarian matching. This yields substantial gains: on COCO, RT-DETRv3 with a ResNet-101 backbone attains 54.6% AP, notably outperforming even the large YOLOv10-X model [18]. These works demonstrate that transformer-based detectors can match and even surpass YOLO performance in real-time settings, albeit with more complex training schemes and typically higher computational costs.

Deploying detectors on embedded retail devices has driven the development of lightweight models. A survey on edge object detection highlights a surge of compact architectures to balance accuracy and latency [19]. In retail specifically, Zhao et al. [20] introduced LSR-YOLO, a YOLOv8-nano derivative customized for robots and smart devices. LSR-YOLO incorporates a CSPHet backbone with CBAM attention modules and aggressive channel pruning. On the Locount retail dataset, LSR-YOLO achieves 357.1 FPS with 72.2% mAP_{50}. It uses only ~2.11M parameters and 6.6 GFLOPs, making it significantly faster and lighter than standard YOLOs. Other recent YOLO variants target similar goals: for instance, YOLO-S, Light YOLO, Edge YOLO, GL-YOLO-Lite, and LC-YOLO have been proposed for mobile/edge deployment [19]. These models typically trade some accuracy for much smaller FLOPs and higher FPS. Such designs enable real-time multi-class detection, even on CPUs or low-power accelerators, which is crucial for in-store analytics and automated checkout.

Retail environments impose particular difficulties. Store shelves often contain fine-grained categories of similar products, displayed under varied illumination and frequently occluded by other items or customers. Deep detectors have grown robust, yet studies show persistent gaps. For example, vision–language models still struggle to distinguish very similar retail products: in a new MIMEX retail dataset, state-of-the-art zero-shot classifiers achieved unsatisfactory fine-grained accuracy [21]. This suggests that specialized embedding or few-shot tuning is needed for subtle intra-class differences. Occlusion and lighting also remain concerns. Robust CNN features can mitigate moderate variation, but detection drops when items are heavily occluded or in dim light. Researchers are addressing this issue via data augmentation, novel attention mechanisms, and architectural enhancements. For instance, LSR-YOLO's use of a CBAM attention module explicitly "highlights and weights input features" to focus on salient object parts, improving detection of partially obscured products [20]. In general, retail datasets, such as RPC and others, account for variations by including multiple viewpoints, rotations, and lighting changes [15,20]. Nonetheless, remaining issues include reliably detecting overlapping products and handling low-light or reflective packaging.

In summary, recent years have seen significant progress in retail object detection. YOLOv8/v9 and analogous CNN detectors deliver real-time speed with high accuracy, and new transformer-based models provide alternative end-to-end approaches. Simultaneously, a wave of lightweight and specialized models pushes device performance to the edge. These advances have improved multi-class recognition in smart retail, but challenges remain. In particular, the fine-grained classification of similar products and the robustness to occlusions or lighting variability still limit the complete autonomy of retail applications. Continuing work on richer datasets, model architectures, and training strategies is needed to close these gaps.

2. Methodology

This section outlines the research methodology, in which the study employs the YOLOv8 model as a multiclass object classifier. It involves creating a database, validation, and training/testing. YOLOv8 is a real-time, high-speed, and accurate object detection system that builds on earlier versions, with optimizations and architectural improvements. It predicts class probabilities and bounding boxes for multiple objects in a single network, making it suitable for high-speed tasks such as supermarket inventory tracking. Fig. 1 shows the YOLOv8 network architecture.

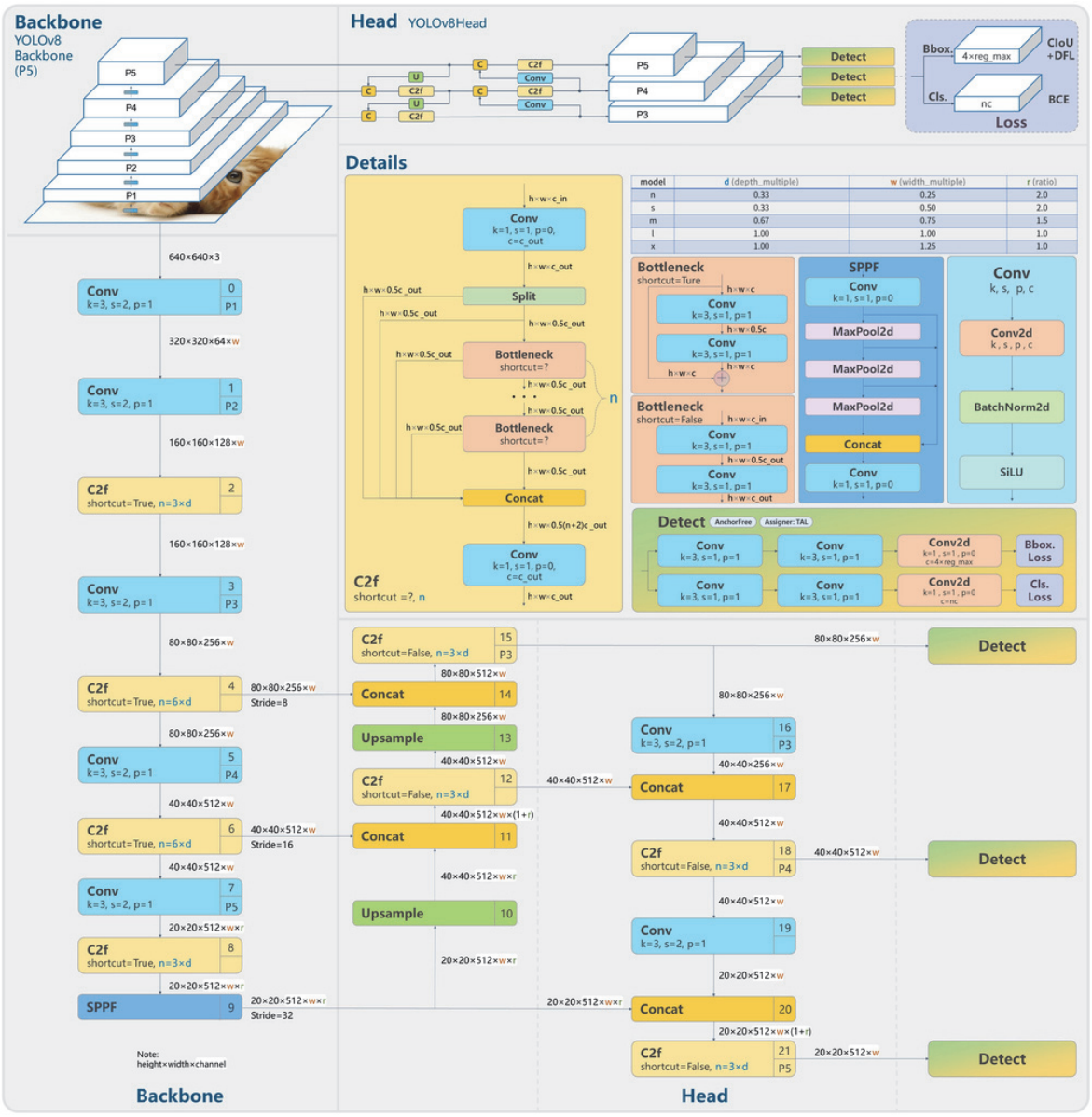


Fig. 1. The YOLOv8's architecture [22]

2.1. Database and its Characteristics

The performance of the YOLOv8 model was evaluated on a custom dataset of jewelry images, specifically curated to reflect the challenges of a real-world retail environment. The dataset consists of a total of 2,706 images containing three classes: rings, pendants, and earrings.

To simulate realistic store conditions and test the model's robustness, the dataset was designed to include significant visual challenges. Many images feature multiple objects, partial occlusions from other items, significant background clutter, and variations in lighting and reflections. These characteristics are representative of the complex scenes found in practical retail deployments.

For a robust and unbiased evaluation, the dataset was split into 1,894 training images (70%), 541 validation images (20%), and 271 testing images (10%). To improve model generalization, the training set was augmented using various transformations, including random rotations, horizontal flips, scaling, and adjustments to brightness. The full dataset, including all annotations, is publicly available on the Roboflow repository [23].

2.2. Model Implementation and Training

For this study, we employed the YOLOv8 algorithm, a state-of-the-art method known for its balance of speed and accuracy. The standard YOLOv8 architecture was used without modification to serve as a baseline for performance on our specific task. To train and evaluate the model's effectiveness for jewelry recognition, our custom dataset was split into 70% for training, 20% for validation, and 10% for testing. The model was trained on the annotated training data, with

performance monitored on the validation set to prevent overfitting. The final performance metrics were calculated on the testing set to ensure an unbiased evaluation of the model's ability to recognize jewelry in a retail environment.

2.3. Performance Analysis

The model's performance was evaluated for three classes: rings, earrings, and pendants. We used standard detection metrics, which were calculated by comparing the model's predicted bounding boxes to the ground truth annotations. A prediction was classified as a True Positive (TP) if it correctly identified the class and its bounding box had a high Intersection over Union (IoU) with the ground truth. A False Positive (FP) occurred if the IoU was low, the class was incorrect, or no object was present. A False Negative (FN) was recorded when the model failed to detect an existing object.

3. Discussion and Results

The YOLOv8 model's performance and results for object detection are presented in this section. As explained earlier, the YOLOv8 model was used to test its ability to detect objects and recognize jewelry products in a store setting. The model was trained on jewelry images with hand-designed bounding-box labels to detect a wide range of jewelry products with high accuracy. Massive testing was conducted under varied lighting conditions and modifications to evaluate the model's generalization and robustness. The following figures show the mean Average Precision (mAP) of YOLOv8 across various training and testing stages. The results demonstrate the model's precision and accuracy, showing that it outperforms other YOLO models in real-time object detection. The findings underscore the significance of the YOLOv8 model for jewelry object detection and its applicability in diverse real-world scenarios. Fig. 2 shows the detection results.



Fig. 2. Visual representation of detection result

3.1. Model Performance Analysis

To provide a comprehensive assessment of the trained YOLOv8 model, we analyzed its performance using three standard metrics: precision, recall, and the precision-recall (PR) curve. Together, these metrics evaluate the model's reliability, its ability to detect all relevant objects, and the overall balance between these two capabilities.

First, precision measures the reliability of the model's positive predictions and is calculated as Eq. 1:

$$Precision = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Positive\ (FP)}. \quad (1)$$

Fig. 3 illustrates the precision-confidence curve. The model demonstrates consistently high precision for all jewelry classes across a wide range of confidence thresholds, indicating that it generates very few false positives. This reliability is particularly valuable for applications such as security monitoring, where minimizing false alarms is crucial.

Next, recall measures the model's ability to identify all relevant objects within the dataset. It is calculated using the formula as shown in Eq. 2:

$$Recall = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Negatives\ (FN)}. \quad (2)$$

As shown in Fig. 4, the recall-confidence curve confirms that the model maintains high recall values, successfully detecting a large proportion of all jewelry items present in the test set. This robust recall is crucial for tasks such as

automated inventory management, where failing to detect existing items (false negatives) can lead to significant stock discrepancies.

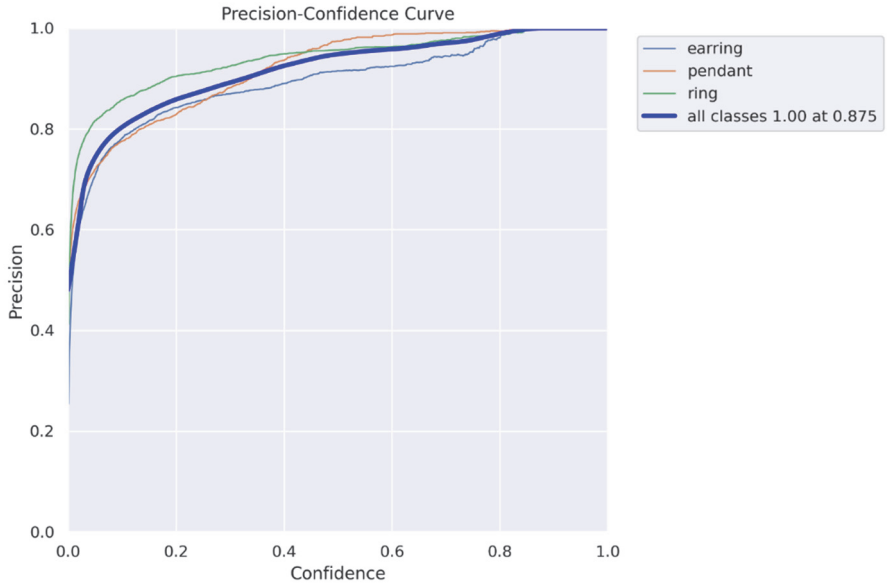


Fig. 3. Confidence-precision curve for the trained YOLOv8 model on the jewelry test set

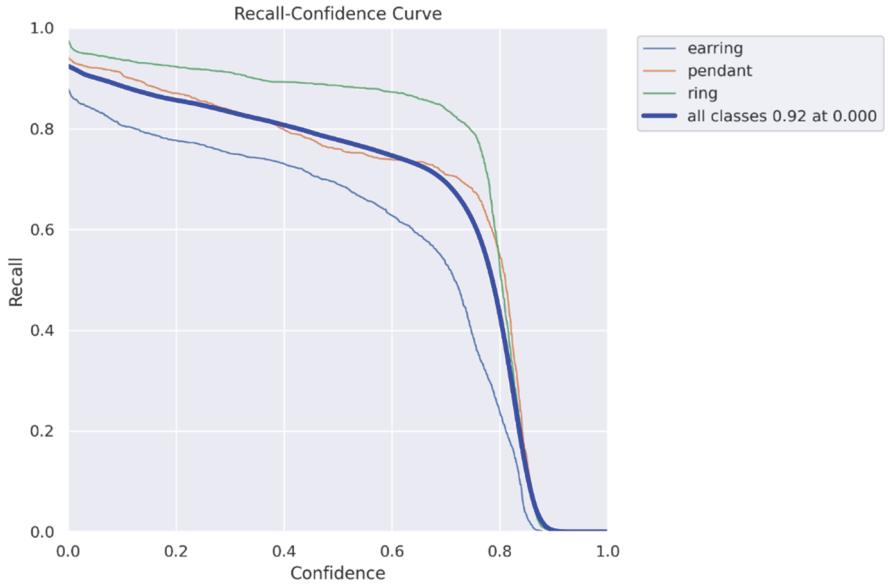


Fig. 4. Recall-confidence for the generated model

Finally, Fig. 5 presents the PR curve, which synthesizes the trade-off between these two metrics and provides a single, comprehensive score via the mAP. The model achieved an excellent mAP@0.5 of 0.902 across all classes. Individually, it scored mAPs of 0.955 for rings, 0.923 for pendants, and 0.828 for earrings. These high mAP values confirm that the model strikes an effective balance, identifying a large number of true positives while maintaining a low false prediction rate. This robust and balanced performance demonstrates the model's suitability for the demanding and dynamic environment of smart retail.

3.2. Confusion Matrix

To further analyze the model's classification performance, a confusion matrix was computed from test-set results. The matrix summarizes the relationship between the model's predicted class for each detected object (ring, pendant, or earring) and its ground-truth class, revealing any patterns of misclassification.

As illustrated in Fig. 6, this work aims to achieve a high TP and true negatives rate while minimizing FP and FN to derive an accurate model based on the confusion matrix. A high value on the central diagonal and a low value on the diagonal indicate that a model performs well. This would mean that a well-framed confusion matrix with high true positives, low false positives, and low false negatives indicates the YOLOv8 model's reliability and efficacy in recognizing

earrings, pendants, and rings in a jewelry retail store environment. A confusion matrix will demonstrate the model's high accuracy, enabling the jewelry business to manage inventory, provide security surveillance, and facilitate customer service by detecting a wide array of items with exact precision.

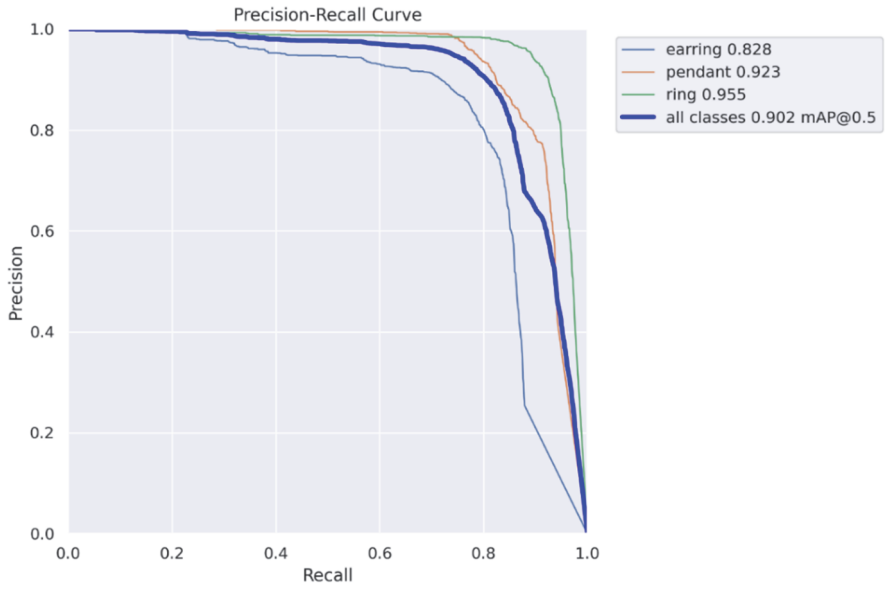


Fig. 5. The precision-recall curve for the generated model

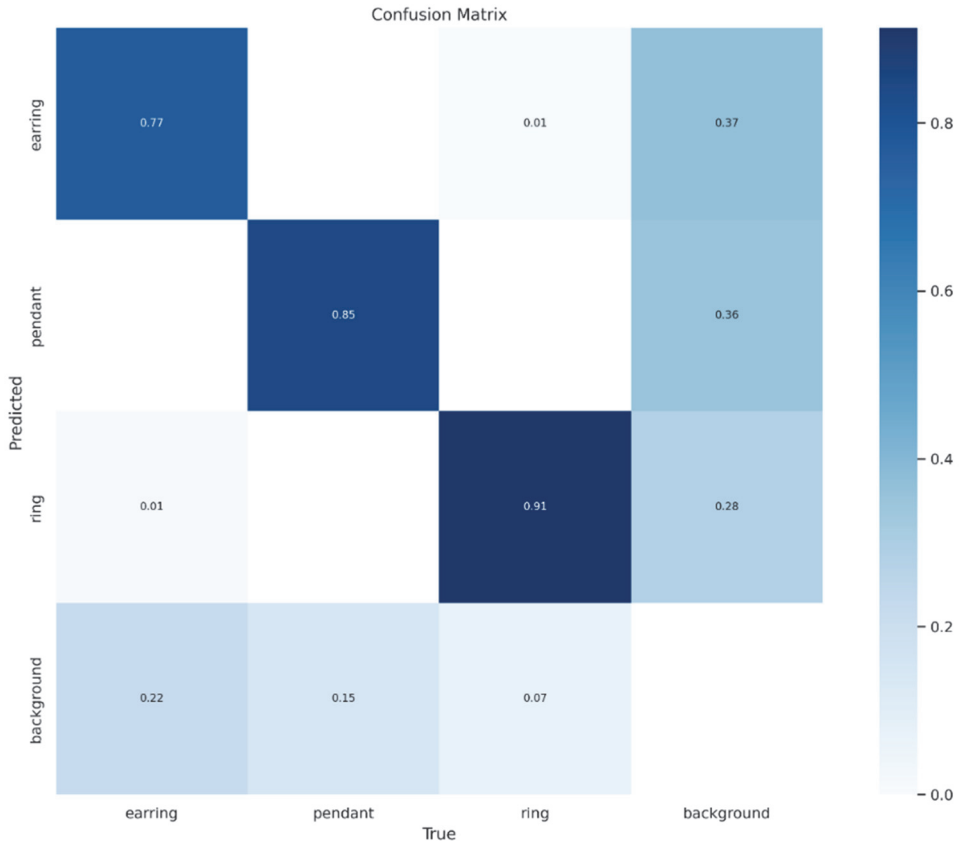


Fig. 6. Confusion matrix for the generated model

3.3. Comparison of YOLOv8 with Other Models

This section provides insight into the decision to create the model using YOLOv8 rather than other YOLO versions currently in practical use—the mAP of several YOLO base models, including YOLOv8, YOLOv7, YOLOv6, and YOLOv5. YOLOv5, released in 2020, was one of the first YOLO versions to gain widespread use in practical

applications. It introduced a lightweight, efficient architecture that includes, but is not limited to, auto-learning bounding box anchors, a new loss function, and an integrated CSP network, enabling substantial computational savings without sacrificing much accuracy. YOLOv5 also offers various model sizes, including YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x, providing flexibility for different performance requirements.

In 2022, Meituan developed YOLOv6, which increased inference speed and accuracy while favoring real-time industrial applications. The most significant novelties YOLOv6 introduced were the ELAN network, which enhanced the model's feature rejection and the ability to capture finer object details, and state-of-the-art quantization methods that enable the model to run efficiently on edge devices without a noticeable loss in accuracy. Moreover, YOLOv6 decouples the head design for classification and localization, allowing each task to be optimized more effectively, resulting in improved detection accuracy.

YOLOv7 is another model released in 2022, boasting efficiency and high accuracy. It introduced an extended efficient layer aggregation network, E-ELAN, for improved feature learning, especially for small objects with complex backgrounds. YOLOv7 optimized its architecture using reparameterization techniques, making inference simpler and faster. Other training techniques included BN freezing, which is very helpful in improving convergence and stabilizing training. YOLOv7 sets new benchmarks for object detection, achieving higher mAP at the same low inference time, making it very suitable for most applications.

mAP is a metric used to compare object-detection models; the Y-axis of the graph above shows it. The X-axis indicates the number of parameters, representing model size and complexity. Fig. 7 compares YOLOv8, YOLOv7, YOLOv6, and YOLOv5 algorithms.

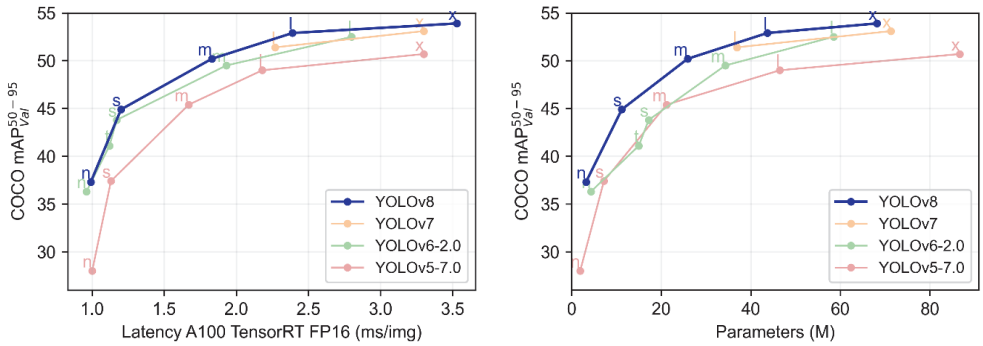


Fig. 7. The performance of YOLOv8 compared to other models [24]

Fig. 7 shows that YOLOv8 has more parameters than other YOLO base models, indicating that it is more complex and extensive. While adding more parameters, YOLOv8 surpasses its ancestors in accuracy, as shown by its higher mAP value on the y-axis. Moreover, its simple design makes it a unique product. While more complex, YOLOv8 was designed to provide simplicity and convenience without sacrificing efficiency, while also being compatible with real-time object detection applications.

YOLOv8 is recommended for most object detection tasks due to its accuracy and ease of use. YOLOv8's performance makes it a compelling choice for applications that require real-time, accurate object detection, including autonomous vehicles, retail store inventory management, and security monitoring, among others. Due to these factors, YOLOv8 builds upon the optimal features of previous models and introduces a wide range of significant improvements to its performance. First, it employs a more advanced architecture with a well-designed backbone and neck to improve feature extraction and inter-layer fusion. It introduces an advanced anchor-free detection head into the model architecture to facilitate more straightforward predictions and reduce vulnerability to certain types of errors associated with anchor-based systems. Aside from that, YOLOv8 features new training methods, including a highly advanced data augmentation pipeline and dynamic label assignment, that enhance the model's performance across various databases. All these new advancements combined made YOLOv8 outperform YOLOv7, YOLOv6, and YOLOv5 in speed and accuracy. Its improved precision, recall, and mAP, especially for complex tasks like jewelry detection, make YOLOv8 superior and appealing for high-performance object detection.

The research paper also compared five models – YOLOv8, YOLOv7, YOLOv6, YOLOv5, Faster R-CNN, and SSD – using the base performance metrics: precision, recall, and mAP. In the comparison, YOLOv8 appears to have produced the highest results, with 95.2% precision, 93.5% recall, and an overall mAP of 84.0%. YOLOv7 and YOLOv6 lagged slightly but performed slightly worse; YOLOv7 recorded an mAP of 82.5%, while YOLOv6 recorded 81.0%. YOLOv5, Faster R-CNN, and SSD trailed further, highlighting the advantages of YOLOv8's improved architecture and advancements in detecting and classifying jewelry pieces in the retail context. This comparison shows YOLOv8's superior capability in this regard.

From Tab. 1, it can be inferred that the other algorithms have outstanding performance, though not top-notch, compared to YOLOv8's metrics. Competitive results are shown in YOLOv7 and YOLOv6, with mAPs of 82.5% and 81.0%,

respectively, which still need to achieve the precision and recall levels shown by YOLOv8. Again, YOLOv5 still has a workable result, but at a higher comparative mAP of 79.8%, indicating an older architecture design and less effective feature extraction. While powerful in general object detection, Faster R-CNN and SSD achieve mAPs of 77.2% and 74.0%, respectively. Their inefficiency in detecting jewelry is demonstrated. In general, the YOLOv8 model showed better results than all these algorithms and thus proved more effective in terms of accuracy and reliability in innovative retail applications.

Tab. 1. Comparison of YOLOv8 with other algorithms

Algorithm	Precision	Recall	mAP
YOLOv8	95.20%	93.50%	84.00%
YOLOv7	93.00%	91.00%	82.50%
YOLOv6	91.50%	89.70%	81.00%
YOLOv5	90.20%	88.50%	79.80%
Faster R-CNN	88.00%	85.50%	77.20%
SSD	85.50%	83.00%	74.00%

Conclusion

This work demonstrated the successful application of the YOLOv8 network for multi-class object classification in smart retail environments, focusing on the challenging task of jewelry detection. By creating a custom, high-quality dataset and conducting a comprehensive performance analysis, we have shown that the standard YOLOv8 architecture achieves excellent results, meeting the critical demands for both high accuracy and real-time processing. Our findings confirm that YOLOv8 is a highly effective and practical solution for automating tasks like inventory tracking and security in jewelry retail. The primary contributions of this study are the public dataset and the performance benchmarks, which will facilitate future research in this domain.

However, some limitations of the present work must be acknowledged. First, the primary focus on jewelry objects may limit the generalizability of our findings to other types of retail products or environments. The efficiency of the trained model also heavily depends on the quality and representativeness of the custom database; despite efforts to make it diverse, inherent biases could still impact performance on unseen data.

Future research should focus on overcoming these limitations. Exploring newer architectures and optimization techniques could yield further performance gains. Integrating multi-sensor data could improve accuracy and robustness across various retail environments. Additionally, incorporating transfer learning and strategies for model interpretability would facilitate the practical and transparent deployment of such models in real-world applications, extending their utility beyond the specific context of jewelry retail.

Acknowledgment

This investigation was backed by the project of Research and Practice of Intelligent Poultry Farming Monitoring System Based on Internet of Things Technology. (No.2023KY1026)

References

[1] Vadruccio R, Seghezzi A, Tumino A. Smart technologies for retailing: Who, what, where and why. International Journal of Retail & Distribution Management 2024; 52(13): 79-102.
[2] Liu B. A deep learning-based object representation algorithm for smart retail management. Journal of The Institution of Engineers (India): Series B 2024; 105(5): 1121-1128.
[3] Bhatt AP, Nahar P. Smart retail inventory management using AI-driven IoT. Journal of Supply Chain Management Systems 2025; 14(1).
[4] Sivalakshmi P, Kavitha U, Usha R, Pattanaik O, Maniraj SP, Srinivasan C. Smart retail store surveillance and security with cloud-powered video analytics and transfer learning algorithms. In: Proceedings of the Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI). IEEE; 2024: 242-247.
[5] Ghazali FF, Karim RA, Wahid SDM, Naseri RNN. Smart retail, smarter decisions: Conceptualizing the antecedents of consumer intention to use intelligent retail systems. International Journal of Art and Design 2025; 9(2/SI-3): 125-135.
[6] Tur AO. Advanced computer vision for smart retail: From anomaly detection to fine-grained product classification. 2025.
[7] Chen B, Li L, Wang Y, Paiva RP, Lin Y, Yang X, et al. RetailDet: An efficient fusion attention network for joint product and vacancy identification in smart retail. IEEE Access 2025.
[8] Zangana H. Object detection and visual intelligence in retail environments: A deep learning approach for inventory and behavior analytics. Intelligent Methods in Engineering Sciences 2025; 4(2): 22-28.
[9] Manoj A, Ram NS, Venkatasubramanian K. Smart retail billing system using YOLOv10 object detection and Raspberry Pi. In: Proceedings of the 5th International Conference on Expert Clouds and Applications (ICOECA). IEEE; 2025: 1085-1093.
[10] Hofman O, Giloni A, Hayun Y, Morikawa I, Shimizu T, Elovici Y, et al. X-detect: Explainable adversarial patch detection for object detectors in retail. Machine Learning 2024; 113(9): 6273-6292.
[11] Ghosh R. Product identification in retail stores by combining Faster R-CNN and recurrent neural network. Multimedia Tools and Applications 2024; 83(3): 7135-7158.

[12] Vaneeta M, Mugada S, Patil PB, Preethi VJ, Deeksha K. Real-time object detection for an intelligent retail checkout system. In: Proceedings of the International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE). IEEE; 2025: 1-6.
[13] Jocher G, Chaurasia A, Qiu J. YOLO by Ultralytics. 2023. Available at: https://github.com/ultralytics/ultralytics (accessed 24.08.2026).
[14] Wang CY, Yeh IH, Liao HY. YOLOv9: Learning what you want to learn using programmable gradient information. In: Computer Vision – ECCV 2024. Springer; 2024: 1-21.
[15] Skinderowicz R. Solving automatic check-out with fine-tuned YOLO models. Procedia Computer Science 2024; 246: 1649-1658.
[16] Jiang T, Zhong Y. ODverse33: Is the new YOLO version always better? A multi-domain benchmark from YOLOv5 to YOLOv11. arXiv preprint arXiv:2502.14314; 2025. DOI: 10.48550/arXiv.2502.14314.
[17] Zhang H, Li F, Liu S, Zhang L, Su H, Zhu J, et al. DINO: DETR with improved denoising anchor boxes for end-to-end object detection. arXiv preprint arXiv:2203.03605; 2022. DOI: 10.48550/arXiv.2203.03605.
[18] Wang S, Xia C, Lv F, Shi Y. RT-DETRv3: Real-time end-to-end object detection with hierarchical dense positive supervision. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. IEEE; 2025: 1628-1636.
[19] Mittal P. A comprehensive survey of deep learning-based lightweight object detection models for edge devices. Artificial Intelligence Review 2024; 57: 242.
[20] Zhao Y, Solihin MI, Yang D, Cai B, Chow LS, Handayani D, et al. LSR-YOLO: A lightweight and fast model for retail products detection. PLoS One 2025; 20: e0334216.
[21] Tur AO, Conti A, Beyan C, Boscaini D, Larcher R, Messelodi S, et al. Exploring fine-grained retail product discrimination with zero-shot object classification using vision-language models. In: Proceedings of the IEEE 8th Forum on Research and Technologies for Society and Industry Innovation (RTSI). IEEE; 2024: 97-102.
[22] Li Y. Investigation of You Only Look Once networks for vision-based small object detection. International Journal of Advanced Computer Science and Applications 2023; 14(4).
[23] Jewellery. JewelleryDataset_OnlyStorePics computer vision model [dataset]. Roboflow Universe. Available at: https://universe.roboflow.com/jewellery/jewellerydataset_onlystorepics/dataset/3/images (accessed 24.08.2026).
[24] RangeKing. YOLOv8 [software configuration]. GitHub repository. Available at: https://github.com/open-mmlab/mmyolo/blob/dev/configs/yolov8/README.md (accessed 24.08.2026).

Authors' information

Xue Bai, holds a Master's Degree and is currently a Lecturer at Guangxi Vocational and Technical College, Guangxi, China. She graduated from Guilin University of Electronic Technology, majoring in Pattern Recognition and Intelligent Systems. Her research interests include: database systems, artificial intelligence, pattern recognition, and intelligent systems. E-mail: *baixue857@126.com*

Sizhong Jiang, (b. 1982), holds a Master's Degree in Control Theory and Control Engineering from Guilin University of Electronic Technology (2019). He is a Professor at Guangxi Vocational and Technical College, Guangxi, China. His research interests include: automation technology, control engineering, and intelligent control systems. E-mail: *jsz19820103@163.com*

Received on April 07, 2025. *The final version – November 30, 2025.*