JSLHU
JOURNAL OF SCIENCE
OF LAC HONG UNIVERSITY

ISSN: 2525 - 2186
Tạp chí Khoa học Lạc Hồng, **2025**, 22, 033-039

ĐÁNH GIÁ TUÂN THỦ PLANOGRAM BẰNG HỌC MÁY VÀ THỊ GIÁC MÁY TÍNH TRONG MÔI TRƯỜNG BÁN LẺ

Phạm Thành Tài[1], Thái Hoàng Tân[1], Lê Hương Thanh[1], Nguyễn Thanh Thảo[1], Cao Minh Thành[1], Nguyễn Lê Vân Thanh[2*]

[1]Trường Đại học Sài Gòn, Thành phố Hồ Chí Minh, Việt Nam
[2]Trường Đại học Sư Phạm Kỹ Thuật, Thành phố Hồ Chí Minh, Việt Nam
* Tác giả liên hệ: *thanhnlv@hcmute.edu.vn*

THÔNG TIN BÀI BÁO

Ngày nhận: 14/4/2025
Ngày hoàn thiện: 19/5/2025
Ngày chấp nhận: 21/5/2025
Ngày đăng: 15/9/2025

TỪ KHÓA

Học máy;
Thị giác máy tính;
YOLO;
DBSCAN;
Hungarian Algorithm.

TÓM TẮT

Nghiên cứu này tập trung vào việc tìm ra phương pháp mới để đánh giá sự tuân thủ planogram tự động trong môi trường bán lẻ, với trọng tâm là thị trường Việt Nam. Để khắc phục những hạn chế của phương pháp kiểm tra thủ công—mất nhiều thời gian, dễ xảy ra sai sót và khó mở rộng—hệ thống được đề xuất tích hợp các tiến bộ gần đây trong lĩnh vực thị giác máy tính và học máy. Cụ thể, phương pháp này sử dụng các mô hình phát hiện đối tượng tiên tiến, YOLOv11 và YOLOv12, được huấn luyện trên các hình ảnh kệ hàng đã được chú thích thu thập từ các môi trường bán lẻ thực tế. Các sản phẩm phát hiện được sẽ được tổ chức không gian bằng thuật toán phân cụm DBSCAN, trong khi thuật toán Hungarian được sử dụng để so khớp các bố cục phát hiện được với các planogram đã được định sẵn và tính toán điểm tuân thủ. Kết quả thí nghiệm cho thấy độ chính xác cao trong việc phát hiện và đánh giá sự tuân thủ, ngay cả trong các điều kiện bán lẻ phức tạp. Sự kết hợp giữa các mô hình YOLO tiên tiến và các kỹ thuật lý luận không gian đã chứng tỏ hiệu quả trong việc xử lý các thử thách đặc thù của cảnh bán lẻ Việt Nam, như việc tổ chức kệ hàng không đồng nhất và bao bì đa dạng. Công trình này đóng góp một giải pháp có thể mở rộng, chính xác và thực tiễn để nâng cao hiệu quả thực thi bán lẻ và tối ưu hóa hoạt động.

APPLYING MACHINE LEARNING AND COMPUTER VISION FOR PLANOGRAM COMPLIANCE EVALUATION IN RETAIL ENVIRONMENTS

Pham Thanh Tai[1], Thai Hoang Tan[1], Le Huong Thanh[1], Nguyen Thanh Thao[1], Cao Minh Thanh[1], Nguyen Le Van Thanh[2*]

[1]Saigon University, Ho Chi Minh City, Vietnam
[2] Ho Chi Minh University of Technology and Education, Ho Chi Minh City, Vietnam
*Corresponding Author: *thanhnlv@hcmute.edu.vn*

ARTICLE INFO

Received: Apr 14th, 2025
Revised: May 19th, 2025
Accepted: May 21st, 2025
Published: Sep 15th, 2025

KEYWORDS

Machine Learning;
Computer Vision;
YOLO;
DBSCAN;
Hungarian Algorithm.

ABSTRACT

This paper presents a novel approach for automated planogram compliance assessment in retail environments, with a focus on the Vietnamese market. Addressing the limitations of manual inspection methods—which are time-consuming, error-prone, and difficult to scale—the proposed system integrates recent advances in computer vision and machine learning. Specifically, the method leverages state-of-the-art object detection models, YOLOv11 and YOLOv12, trained on annotated shelf images collected from real retail settings. Detected products are spatially organized using the DBSCAN clustering algorithm, while the Hungarian algorithm is employed to match detected layouts with predefined planograms and compute compliance scores. Experimental results demonstrate high detection accuracy and reliable compliance evaluation, even under complex retail conditions. The combination of advanced YOLO models with spatial reasoning techniques proves effective in handling challenges unique to the Vietnamese retail landscape, such as inconsistent shelf organization and varied packaging. This work contributes a scalable, accurate, and practical solution for enhancing retail execution and operational efficiency.

Doi: https://doi.org/10.61591/jslhu.22.714

Available online at: https://js.lhu.edu.vn/index.php/lachong

1. INTRODUCTION

As the retail industry in Vietnam continues to expand, optimizing shelf display and merchandise management has become a critical factor in enhancing customer experience and improving business efficiency. However, conventional planogram compliance inspection methods are still predominantly manual, relying heavily on visual observation, which is time-consuming and prone to human error.

In recent years, numerous studies have explored the application of computer vision and machine learning to automate planogram compliance assessment in retail settings. Laitala and Ruotsalainen (2023), for instance, proposed a computer vision-based system capable of automatically identifying product types and their respective locations on store shelves, which are then compared to a predefined planogram. Nevertheless, several challenges have been noted in the literature. These include reduced recognition accuracy under varying conditions such as lighting, camera angles, product occlusion, and image quality. Moreover, integrating deep learning techniques with traditional algorithms often introduces system complexity, making optimization and parameter tuning more difficult. Effective training of machine learning models also requires a diverse and representative dataset, which can be difficult to acquire and annotate.

This study adopts state-of-the-art image recognition models, particularly the YOLO (You Only Look Once) architecture, a prominent deep neural network in object detection, to detect and classify products on retail shelves. Additionally, the research incorporates advanced versions—YOLOv11 and YOLOv12—to improve detection accuracy. Tools such as Roboflow are employed to facilitate efficient data annotation and enhance training performance. The experimental dataset consists of real-world images collected from retail stores in Vietnam, where each product is meticulously labeled to build a high-quality training dataset. This approach lays the foundation for developing an automated, accurate, and efficient planogram compliance assessment system tailored to dynamic retail environments.

2. RESEARCH METHODOLOGY

This study adopts a machine learning and computer vision-based approach to automate the verification of retail shelf displays against predefined planograms. The methodology is structured into the following key stages:

Stages 1. Image Data Collection and Preprocessing:

Images of retail shelves were collected from various stores across Vietnam under diverse conditions, including different camera angles, lighting environments, and shelf arrangements. Detailed product annotations were conducted using the Roboflow platform. Image preprocessing involved resizing normalization, automatic orientation correction (Auto-Orient), and data augmentation techniques such as blurring and horizontal flipping to enhance model generalizability.

Stages 2. Development of Product Detection Models:

Two state-of-the-art object detection models—YOLOv11 and YOLOv12—were employed to detect and classify shelf products. These models were trained on the annotated dataset and evaluated using standard performance metrics, including mean Average Precision (mAP), Precision, and Recall, to assess detection accuracy and robustness.

Stages 3. Planogram Compliance Verification System:

A compliance verification system was developed to compare detected shelf layouts with the reference planogram. The DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm was utilized to group products into rows based on the spatial coordinates of their bounding boxes. The Hungarian Algorithm for Product Matching was then applied to match detected products with the corresponding planogram entries, considering their position, quantity, and category. To reduce redundancy in detections, the Non-Maximum Suppression (NMS) technique was implemented.

Stages 4. Model and System Performance Evaluation:

Model performance was assessed through quantitative evaluation on test datasets using the aforementioned metrics. The entire system was validated on real-world data to determine its applicability and effectiveness in practical retail scenarios.

Stages 5. System Testing and Optimization:

The system was tested in simulated retail environments to analyze operational performance. Based on test results, algorithmic parameters and model configurations were adjusted iteratively to optimize detection accuracy and system efficiency.

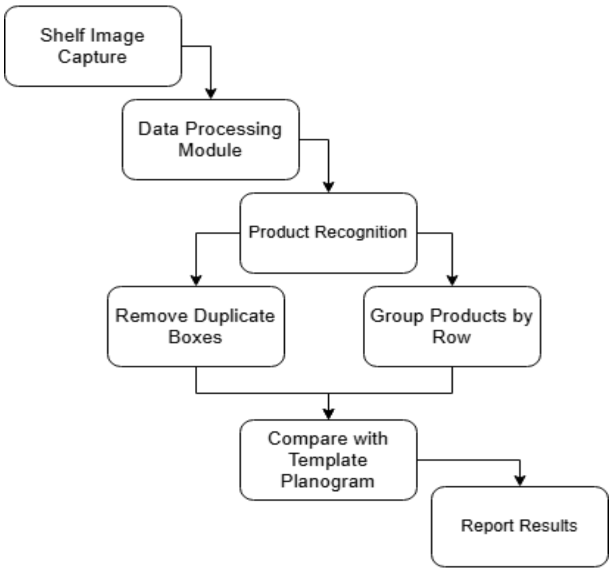

Figure 1. *Workflow of the Planogram Compliance Evaluation System.*

2.1 Image Acquisition Process

2.1.1 Image Data Collection

Data Collection Principles:

Diverse Camera Angles: Images were captured from various perspectives, including front-facing, angled, and top-down views, to simulate realistic observation scenarios in retail environments.

Varied Lighting and Environmental Conditions: The dataset includes images taken under multiple lighting settings (e.g., natural light, fluorescent lighting, and low-light conditions) and environmental backgrounds to improve model generalizability and robustness.

Comprehensive Product Representation: All target product categories were included in the dataset, with a balanced number of samples per class to mitigate data imbalance issues during training.

Minimization of Occlusion and Noise: Efforts were made to ensure that products were not excessively obstructed by other objects or affected by noise-inducing factors such as human movement or strong light reflections.

Sufficient Image Resolution**:** Images were captured at high enough resolution to ensure that even small products are clearly visible for accurate labeling and detection.

Collection Procedure:

Collection Sites: Images were taken directly from product displays in retail stores and supermarkets across various locations in Vietnam.

Dataset Size and Product Types: A total of 42,472 images were collected, covering 47 distinct product types.

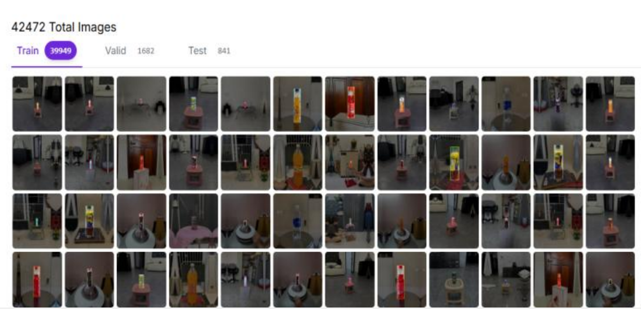

Figure 2. *Collected Image Dataset.*

Camera Angles:

Each product was photographed from three different distances: close-up, medium-range, and long-range. Approximately 100 images were captured for each angle per product to simulate various real-world shelf-viewing scenarios.

Product Code Reference Table:

A standardized reference table was created to document all product identifiers and labeling rules. Each product instance is classified into two view types: front-facing with visible branding/logo (True) and non-front-facing (False). This distinction enhances brand recognition accuracy when assessing shelf arrangements.

Dataset Splitting:

After collection, the dataset was divided into three subsets for model training and evaluation:

- Training Set: 85% (39,949 images)
- Validation Set: 10% (1,682 images)
- Test Set: 5% (841 images)

This data partitioning strategy was consistently applied for each object detection model used in the study.

2.1.2 Image Data Preprocessing

Image Annotation:

The image annotation process was conducted using Roboflow, an advanced web-based platform designed to assist in creating, managing, and analyzing datasets for computer vision tasks. Roboflow provides tools for data labeling, model training, and result visualization, streamlining the entire workflow.

The annotation process followed these steps:

Project Initialization and Class Definition: A new object detection project was created on Roboflow, with product categories defined based on a pre-compiled product code table. For instance, "MC-C1L5" represents "Mirinda 1.5L Orange Bottle," "MSK" denotes "Mirinda Soda Kem Bottle," and "STCD" stands for "Sting Red Bottle." This naming convention minimized labeling errors and ensured consistency across the dataset.

Dataset Upload: The collected image dataset was uploaded to the Roboflow platform. Each image could be previewed individually, allowing the team to verify image quality before annotation.

Manual Annotation: Using Roboflow's built-in bounding box tools, each product within an image was manually annotated and assigned to the correct class. Shortcut keys were utilized to accelerate the labeling process.

Label Review and Validation: Post-annotation, the dataset was reviewed by the team. Roboflow’s analytics features allowed the team to track the number of bounding boxes per class and identify potential imbalances or missing labels. Around 10–15% of annotated images were cross-checked by different team members to ensure label consistency and accuracy.

Data Export: The final annotated dataset was exported in YOLO (TXT) format, compatible with the object detection models used in this study. Roboflow automatically split the data into training (85%), validation (10%), and test (5%) sets according to predefined ratios.

Image Preprocessing:

To standardize the input and improve model performance, the images underwent several preprocessing steps:

Resize: All images were resized to a fixed dimension to ensure consistent input size for the model and avoid training instability caused by variable image resolutions.

Auto-Orientation: Images were automatically rotated to the correct orientation to eliminate recognition errors resulting from tilted or inverted inputs.

Data Augmentation:

To improve the model's robustness and generalization, the following augmentation techniques were applied:

Blur (up to 1.1 px): Simulates blurry images to improve the model's robustness under suboptimal imaging conditions.

Horizontal Flip: Enhances the model's ability to recognize products in mirrored or reversed viewpoints.

Dataset Splitting:

After preprocessing, the dataset was divided into three subsets for training and evaluation:

- Training set: 85% (39,949 images)
- Validation set: 10% (1,682 images)
- Test set: 5% (841 images)

2.2 Algorithms Used

2.2.1. Non-Maximum Suppression (NMS Algorithm)

Non-Maximum Suppression (NMS) for Redundant Bounding Box Elimination in Object Detection

Non-Maximum Suppression (NMS) is a post-processing technique widely used in object detection models to eliminate redundant bounding boxes and ensure that each detected object is represented by a single, most confident bounding box. The algorithm takes as input a set of predicted bounding boxes, their associated confidence scores, and two threshold parameters: the score threshold and the NMS threshold. Initially, all bounding boxes with confidence scores below the predefined score_threshold (set to 0.3 in this study) are discarded.

Figure 3. *Initial Detection Results Before Applying the NMS Algorithm*

The remaining boxes are then sorted in descending order based on their confidence scores. The algorithm iteratively selects the box with the highest confidence and computes the Intersection over Union (IoU) between this box and all other remaining boxes. Any box with an IoU exceeding the nms_threshold (set to 0.4) is considered redundant and is subsequently removed. This process continues until no further boxes meet the elimination criteria. The final output is an optimized set of bounding boxes, effectively reducing false positives by preventing multiple detections of the same object. In this implementation, the bounding boxes are represented in the format [x, y, width, height], accompanied by their respective confidence scores.

As illustrated in Figure 3, the object detection model generates multiple overlapping bounding boxes for the same object instance. This redundancy indicates that the model has identified the same object multiple times with varying confidence scores. Such overlapping detections can lead to ambiguities in downstream tasks and misinterpretations in evaluation metrics. Therefore, the application of Non-Maximum Suppression (NMS) is essential to refine the detection results by retaining only the most confident bounding box for each object.

Figure 4. *Detection Results After Applying the NMS Algorithm*

As shown in Figure 4, after applying the Non-Maximum Suppression (NMS) algorithm, only a single bounding box is retained to represent each detected object. This result demonstrates the effectiveness of NMS in eliminating redundant overlapping boxes and improving the clarity and precision of the detection output. By preserving only the bounding box with the highest confidence score for each object, the post-processing step ensures more accurate and interpretable results.

2.2.2 DBSCAN Algorithm

The Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm is employed to automatically group products into rows on a retail shelf based on the spatial distance between their bounding boxes. This method does not require prior knowledge of the number of rows, making it suitable for flexible and unsupervised clustering in real-world retail environments. After applying Non-Maximum Suppression (NMS) to filter out redundant bounding boxes, DBSCAN is utilized using the vertical position (y-axis coordinates) and height of the remaining bounding boxes as input features.

In this experiment, the parameters for DBSCAN are set as follows:

- **ε (epsilon) = 30 pixels**: the maximum distance between two bounding boxes for them to be considered part of the same cluster.
- **MinPts = 1**: the minimum number of nearby bounding boxes required to form a valid cluster.

The algorithm iteratively scans all bounding boxes and assigns those with at least *MinPts* neighbors within a radius of ε to the same cluster. Bounding boxes that do not meet this criterion are labeled as noise, which may correspond to misplaced or misaligned products with respect to the intended planogram.

Compared to methods such as k-means, DBSCAN offers advantages in handling noise and does not assume a fixed number of clusters, which enhances its applicability to diverse shelf configurations. However, its performance is sensitive to the choice of ε and *MinPts*, requiring careful parameter tuning based on the characteristics of the specific shelf layout.

The resulting clusters provide a structural basis for evaluating planogram compliance by identifying rows of products, thereby contributing to more accurate and automated compliance assessment.

Figure 5. *Input Image for DBSCAN Clustering.*

Figure 5 presents the input to the DBSCAN clustering algorithm. The image shows a retail shelf containing various products, where each product has been localized and represented by a bounding box after applying Non-Maximum Suppression (NMS). These bounding boxes serve as the foundation for spatial clustering based on their vertical positioning and dimensions.

```
Hàng 1: RV RV-F RV RV RV RV-F RV-F
Hàng 2: RV-Z RV-Z-F RV-Z RV-Z-F RV-Z-F RV-Z-F RV-Z-F
RV-Z-F
Hàng 3: RV-CM-F RV-CM-F RV-CM-F RV-CM-F RV-CM-F
RV-CM RV-CM-F
```

Figure 6. *Output of DBSCAN: Product List Sorted by SKU Based on Shelf Rows.*

Figure 6 illustrates the result of applying the DBSCAN algorithm. Products are automatically clustered into horizontal rows according to their vertical alignment on the shelf. Within each row, the products are sorted by their respective product codes. This structured output enables efficient assessment of shelf arrangement and supports automated planogram compliance verification.

2.2.3 Hungarian Algorithm

The Hungarian algorithm, also known as the Kuhn–Munkres algorithm, is an efficient method for solving the assignment problem, where the goal is to find an optimal one-to-one matching that minimizes the total cost. In the context of planogram compliance checking, this algorithm is used to associate detected products from the real-world shelf image with the expected product layout defined in the reference planogram, ensuring accurate product placement and minimal deviation.

The algorithm operates on two input lists: the **Planogram** (representing the ideal arrangement of products) and the **Products** (representing the actual detected items on the shelf). To evaluate compliance, a **cost matrix** is constructed in which each element $Cost[i,j]$ represents the dissimilarity between the i-th product in the planogram and the j-th detected product. This dissimilarity may incorporate factors such as product mismatch, position deviation, or incorrect order

The total evaluation cost is defined as:

cost_matrix [i, j] = missing + extra + order_mismatch

Cost is computed as the sum of the number of missing products, extra products, and a fixed penalty for product order mismatch:

- Missing products refer to items that are present in the planogram but not detected on the shelf.
- Extra products are those that are detected on the shelf but do not appear in the planogram.
- Product order mismatch incurs a fixed penalty of 1 if the detected arrangement of products does not match the expected order defined in the planogram; otherwise, the penalty is 0.

The **Hungarian algorithm** is then applied to the cost matrix to determine the optimal matching between planogram items and detected products, minimizing the overall deviation. The output is a list of matched pairs representing the best alignment between the expected and actual shelf layouts.

If certain products cannot be assigned reasonably (i.e., their deviation exceeds acceptable thresholds), this may indicate planogram violations, such as missing items or misplaced products. The output is a list of optimal product pairings used to assess compliance.

The algorithm is also employed to optimally match entire product rows, with the cost matrix penalizing missing products, surplus items, and incorrect ordering. This ensures that the overall assignment cost is minimized across all possible row pairings. However, the cost function used in the Hungarian algorithm is manually designed and may not fully capture subjective human judgment in all scenarios.

After applying the Hungarian algorithm, the system obtains the best matches between the products in each shelf row and their corresponding entries in the reference planogram. Based on this output, compliance is evaluated using three key metrics

Missing items: products present in the planogram but not detected on the shelf,

Extra items: products found on the shelf but not defined in the planogram,

Ordering errors: products placed in the wrong sequence compared to the standard layout.

These metrics are used to compute a deviation score for each shelf row and to derive an overall compliance score for the entire shelf display.

3. RESULTS AND DISCUSSION

The training process of the YOLOv11 and YOLOv12 models on a dataset comprising 841 real-world shelf images demonstrated high product detection performance. Detailed results are summarized in Table 1 below:

Table 1. *Performance Comparison Between YOLOv11 and YOLOv12 on the Dataset*

	mAP (%)	Precision (%)	Recall (%)	Number of Training Epochs (epoch)
YOLOv11	97.8%	94.5%	95.5%	140
YOLOv12	97.6%	93.7%	94.7%	120

From the results presented in Table 1, it is observed that YOLOv11 achieves a slightly higher mean Average Precision (mAP) compared to YOLOv12. However, YOLOv12 exhibits faster convergence during training and demonstrates superior performance in detecting small-sized products in real-world scenarios. Therefore, YOLOv12 was selected for integration into the planogram compliance checking system.

The proposed system is composed of three core components: the YOLOv12 model for product detection, the DBSCAN algorithm for grouping products into shelf rows, and the Hungarian algorithm for matching detected product positions with the reference planogram. Evaluation of the system was conducted using a test dataset independent from the training and validation datasets. The results indicate the following:

- The YOLOv12 model achieved a mean Average Precision (mAP) of 97.6% and a Precision of 93.7% on the validation set (Table 1).
- The overall system accuracy in detecting planogram non-compliance errors—including incorrect product type, position, or facing—was 91.6%, which is slightly lower than the object detection model's mAP. This discrepancy is primarily due to the **syst**em's strong dependence on the performance of the YOLOv12 model, which provides the input for subsequent stages.

In addition, several factors can degrade system performance, including poor image quality, blurred or tilted photos, and irregular product spacing. These factors negatively affect the accuracy of bounding box detection and product row grouping. Accurate product grouping plays a critical role, as it serves as the input to the Hungarian algorithm. If DBSCAN fails to correctly cluster products by rows, the Hungarian algorithm may be unable to perform reliable matching, thereby reducing the system's accuracy in identifying display errors.

In terms of processing speed, the system achieves an average inference time of 1.8 seconds per image, which is practical for real-world retail environments that involve large-scale data processing.

Despite promising performance, several limitations remain. Product detection accuracy decreases when products are partially occluded. Detecting incorrect facings requires further investigation, potentially through the integration of Optical Character Recognition (OCR) techniques to enhance label recognition accuracy. Distinguishing between visually similar product variants also remains a significant challenge. Since the performance of machine learning models is highly dependent on the diversity and quality of training data, expanding the dataset to include variations in viewing angles and lighting conditions is essential. The current study covers 47 product types, and future work should aim to scale up to accommodate a wider range of retail categories.

Future development directions include enhancing training datasets with more diverse samples, integrating OCR for label verification, refining the spatial matching algorithm, enabling real-time alert systems, and extending the application to various retail domains.

The experimental results confirm that the proposed planogram compliance checking system is a feasible and effective solution for automating product display validation. The adoption of modern machine learning algorithms such as YOLOv12, DBSCAN, and the

Hungarian algorithm significantly improves accuracy, reduces manual effort, and provides a practical direction for the deployment of AI-powered solutions in the retail sector.

4. CONCLUSION

This study successfully developed a planogram compliance checking system by integrating the YOLOv12 model for product detection, the DBSCAN algorithm for clustering product rows on shelves, and the Hungarian algorithm for spatial matching with the reference planogram. The system was evaluated on a dataset of 841 real-world retail images, achieving a planogram deviation detection accuracy of 91.6% with an average processing time of 1.8 seconds per image, demonstrating high potential for practical deployment in retail environments.

Experimental results indicate that the proposed system is not only feasible but also effective in automating the product display compliance verification process. The integration of advanced machine learning algorithms significantly enhances detection accuracy while reducing manual labor, thus laying the groundwork for AI-driven solutions in the retail sector.

Future work may involve extending the system to support video data for continuous monitoring of product display changes, incorporating Transformer-based architectures to improve detection performance in complex environments, and integrating OCR technologies to directly recognize product names from packaging, thereby enhancing recognition accuracy and compliance evaluation.

5. REFERENCES

[1] Planogram Compliance Control via Object Detection, Sequence Alignment, and Focused Iterative Search

[2] Laitala, J.; Ruotsalainen, L. Computer Vision Based Planogram Compliance Evaluation. Appl. Sci. **2023**, 13, 10145

[3] N. V. Nguyen, T. M. Nguyen, and V. H. Le, "Deep learning for planogram compliance verification in retail environments," in Proc. *2022 International Conference on Advanced Technologies for Communications (ATC)*, Hanoi, Vietnam, **2022**, pp. 89–94.

[4] Giulia Pazzaglia, Marco Mameli, Emanuele Frontoni, Primo Zingaretti, Rocco Pietrini, Davide Manco, Valerio Placidi (**2021**)

[5] Mehwish Saqlain, Saddaf Rubab, Malik M. Khan, N. Ali, Shahzeb Ali (**2022**). Hybrid Approach for Shelf Monitoring and Planogram Compliance (Hyb-SMPC) in Retails Using Deep Learning and Computer Vision. Mathematical Problems in Engineering Volume 2022, Article ID 4916818, 18 pages

[6] Julius Laitala (**2021**). Computer vision based planogram compliance evaluation. Master's thesis, Master's Programme in Computer Science

[7] Z. Li, Y. Chen, X. Wang, and Q. Zhao, "An Efficient Retail Product Recognition System with YOLO-based Architecture," J. *Visual Communication and Image Representation*, vol. 85, p. 103578, **2022**.

[8] D. Zhang, J. Zhang, and Q. Guo, "Retail shelf monitoring using deep learning and planogram matching," in Proc. *IEEE International Conference on Image Processing (ICIP)*, **2021**, pp. 345–349.

[9] M. R. Amer, E. S. Al-Masri, and M. I. Khalil, "Computer Vision for Retail Planogram Compliance," *Procedia Computer Science*, vol. 65, pp. 786–793, **2015**.

[10] Z. Chen, Y. Hu, L. Xu, Y. Wang, and L. Jin, "Retail product recognition in shelf images using sparse features and deep learning," *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pp. 2082–2090, **2019**.

[11] R. Giri, V. Mittal, and M. Srivastava, "Shelf Monitoring Using Object Detection and Semantic Segmentation," *2021 6th International Conference on Signal Processing*, Computing and Control (ISPCC).

[12] Y. Yang, K. Sheng, Y. Zhang, and X. He, "Planogram Compliance Checking with Deep Learning and Planar Object Tracking," arXiv preprint arXiv:1907.00573, **2019**.

[13] M. V. Baraban, S. V. Poslad, and D. C. Greenhill, "A Deep Learning Framework for Planogram Compliance Checking," Electronics, vol. 9, no. 2, **2020**.