Computer Life
ISSN: 1819-4818 | Vol. 13, No. 1, 2025

HP-YOLO11n: A lightweight model for surface defect detection of liquor bottle caps based on improved YOLO11n

Jiting Han [1], Yingqian Zhang [2], Dazhi Yang [1, *], Haojun Liang [1], Feiyu Chen [1]

[1] College of Mechanical Engineering, Sichuan University of Science and Engineering, Zigong, China
[2] College of Civil Engineering, Sichuan University of Science and Engineering, Zigong, China
[*] **Corresponding author**

Abstract: In production, defect detection is crucial for ensuring product quality and consumer satisfaction. To address the issues of surface defect detection in liquor bottle caps and the large number of algorithm parameters, this study improves YOLO11n and proposes a more lightweight and higher-precision HP-YOLO11n algorithm. Firstly, We employ the improved HGNetv2 backbone network as our model backbone, which makes the model more lightweight while ensuring the accuracy of model detection. Secondly, added a P2 detection layer to YOLO11n, incorporating high-resolution feature maps and rich detailed information to enhance the model's overall recognition performance. Finally, we remove the P5 layer used for detecting large targets, which reduces the number of parameters and computational load while maintaining accuracy. The experimental results show that the HP-YOLO11n algorithm achieves a mean average precision mAP@0.5 of 87.06%, which is 1.52 percentage points higher than the original YOLOv11n algorithm, while reducing the number of parameters by 44.57%, making it more accurate and lightweight.

Keywords: Defect detection; YOLO11n; HGNetv2; P2; Bottle cap.

1. Introduction

Currently, the competition in the liquor industry has become increasingly intense [1,2], and the market environment has undergone significant changes. According to data from the National Bureau of Statistics, from 2016 to 2024, The Chinese liquor industry undergoes deep adjustment, with production continuously declining from 13.584 million kiloliters to 4.145 million kiloliters, marking a cumulative drop of 69.5% [3,4]. Figure 1 shows the bar chart of Chinese Baijiu Production Volume from 2016 to 2024.

Figure 1. 2016-2024 Chinese Baijiu Production Volume bar Chart

This reflects the trend of industry consoli -dation and also signifies increasing competi -tive pressure. Faced with these challenges, Liquor manufacturers focus not only on the taste of the wine, but also on the design and appearance of the packaging [5]. During the production of bottled liquor, due to factors such as the production environment, equipment, and manufacturing processes, various defects in cap sealing often occur, such as broken cap points, deformation, spins [6]. These defects will affect the appearance quality of the product, thereby reducing consumers' purchase intention [7]. Currently, defect detection mainly relies on manual inspection. Although traditional mach -ine learning methods have partially addressed the defect issue, their accuracy and speed still need improvement.

In recent years, deep learning-based object detection methods have become mainstream, which can be roughly divided into two categories: one is the two-stage detection algorithm based on candidate regions, and the other is the regression-based one-stage detection algorithm [8]. Two-stage algorithms such as the Fast R-CNN [9] series and Mask R-CNN [10] perform object classification and bounding box regression by extracting regions of interest from the input image.Although these algorithms demonstrate excellent detection performance, their numerous model parameters and high computational complexity result in slow detection speeds, making them less suitable for industrial applications.Relatively speaking, one-stage algorithms with faster detection speeds are more commonly used in industry, such as the YOLO (You Only Look Once) [11] series and RetinaNet [12]. However, detecting surface defects in bottled liquor production is quite challenging, as the types and sizes of defects vary significantly, especially with small defects being difficult to detect, which often leads tomissed or false inspections.Toxqui-Quitl et al. [13] proposed a defect detection method for PET bottles, which employs frequency filtering technology to correct the captured bottle images through adaptive gamma correction, aiming to achieve high-precision inspection of the bottle surface, wall, and bottom. However, this method often fails to effectively focus on specific defects and struggles to identify defects in complex backgrounds. In contrast, Zhou et al. [14] proposed a novel localization method called Entropy Rate Superpixel Circle Detection (ERSCD), which combines least squares circle detection, entropy rate superpixel (ERS), and an improved randomized circle detection to accurately obtain the region of interest at the bottle bottom. Although this approach is innovative, it requires extensive experimental validation to optimize the judgment criteria, thereby increasing its complexity. This limita -tion also restricts its capability to simultane -ously detect multiple types of defects. Liu et al. [15] proposed a defect detection method based on EfficientDet, but its model still suffers from slow detection speed and frequent missed detections.

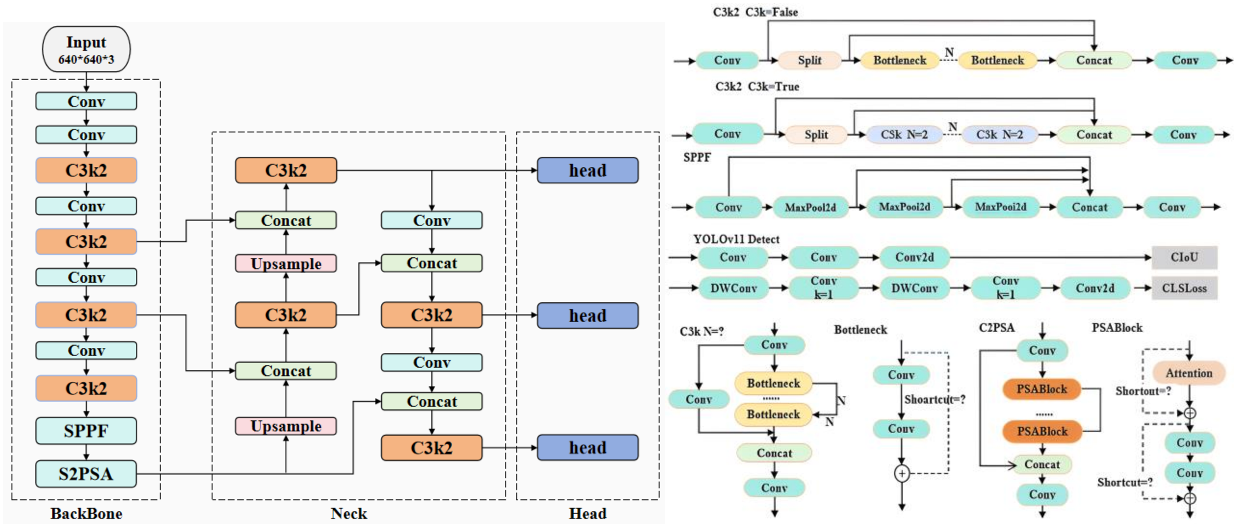
Therefore, these traditional methods have many limitations in practical applications, especially in terms of complexity, detection efficiency, and handling various types of defects. To address these challenges, This study proposes HP-YOLO11n, a lightweight model for surface defect detection in bottled liquor production. By introducing the P2 detection layer and removing the P5 detection layer, it achieves fast and high-precision detection of small target defect features on the surface of this industrial product.

2. METHODOLOGY

2.1. YOLO11n model

YOLO is a single-stage object detection algorithm that relies on convolutional neural networks for implementation [16]. Ultralytics introduced YOLO11 [17] in 2024, with its overall core architecture comprising three main components: the backbone network, the neck network, and the detection head. As a lightweight version within the YOLO11 series, YOLO11n significantly reduces the total number of parameters and computational load. The backbone network, which is used to extract image features, primarily consists of Conv modules, C3K2 modules, SPPF modules, and C2PSA modules. The neck network enables the fusion of feature maps from different network layers. The detection head is primarily responsible for performing the final object detection and classification tasks.

Compared to YOLOv8[18], YOLO11n intr -oduces several improvements. Additionally, the backbone and neck incorporate the C3K2 module: C2f relies on convolutional layers and the Bottleneck structure, whereas C3K2 enhances efficiency through multi-scale convolutional kernels and a channel splitting strategy. The input features are divided into two parts: one part passes directly through standard convolution, while the other part is processed by the C3K or Bottleneck structure. They are then concatenated and fused via 1x1 convolu -tion. C3K supports dynamically adjusting the kernel size to expand the receptive field and optimize feature extraction in complex scena -rios.Users can choose to use either the C3K or C2f module by setting the hyperpara -meter c3k, a design that significantly enhances flexibility. In the detection head, the two conventional convolutional layers of the classification branch are replaced with depthwise separable convolu -tions, markedly reducing both the parameter count and computational load [19]. The overall frame -work of YOLO11n is illustrated in Figure 2.



(a) Network structure diagram of YOLO11n (b)Module structure diagram of YOLO11n

Figure 2. Overall framework diagram of YOLO11n

2.2. HP-YOLO11n model

Based on the YOLO11n,we propose a more lightweight and higher-precision deep learning algorithm called HP-YOLO, as shown in Figure 3. The model improves in three aspects: First, the improved HGNetv2 is selected as the model backbone, which not only reduces the number of parameters but also ensures the detection accuracy of the model; Secondly, a P2 detection layer was added, incorporating high-resolution feature maps and rich detailed information to enhance the model's recognition performance for small objects. Finally, the backbone P5 layer used for detecting large objects was removed, This reduces the number of model parameters while maintaining accuracy. The improved algorithm effectively reduces the number of parameters and enhances detection accuracy.

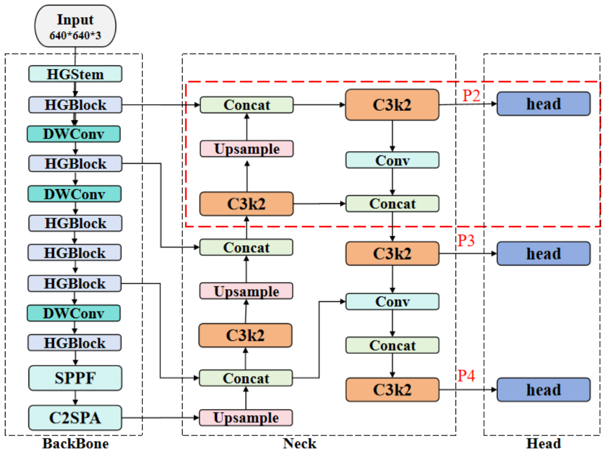


Figure 3. HP-YOLO11n network framework

2.2.1. Improved HGNetv2

RT-DETR is the first real-time vision transformer (ViT) model launched by Baidu, which outperforms the YOLO series in real-time detection [20]. It utilizes two backbone networks, one being HGNet and the other ResNet. this study attempts to replace the backbone of YOLOv11 with an improved HGNetV2. HGNetv2 structure diagram is illustrated in Figure 4.

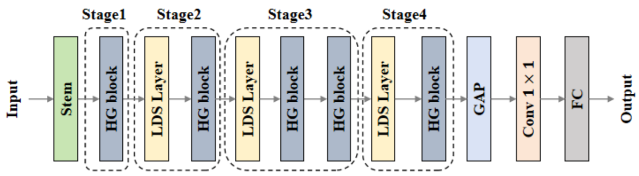


Figure 4. HGNetv2 structure diagram

HGNetV2 consists of multiple HG blocks, and its structure is shown in Figure 5.

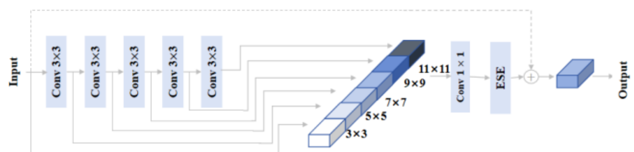


Figure 5. HG-Block structure diagram

It includes several convolutional layers with different filter sizes to capture diverse features. From the architecture of HGNetV2, it can be observed that this network contains a substantial number of convolutions. Therefore, we can optimize HGNetV2 by employing the more lightweight depthwise separable convolu -tion (DWConv) [21]. The improved HGNetV2 model is shown in Figure 6.

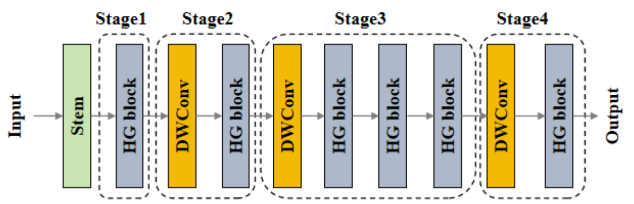


Figure 6. Improved HGNetv2 structure diagram

DWConv replaces the LDS Layer in HGNetv2, achieving further lightweight optimi -zation without modifying the architecture. Figure 7 illustrates the DWConv structure.

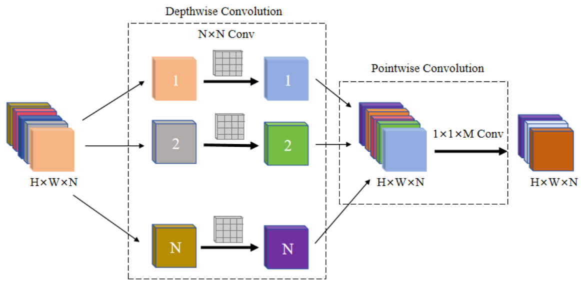


Figure 7. DWConv structure diagram

2.2.2. Neck introduces the P2 layer and removes the P5 layer

The traditional YOLOv11 architecture includes three output layers—P3, P4, P5 for object detection. We optimized the YOLOv11 network structure by adding a P2 detection layer and removing the P5 detection layer. The core objective of this adjustment is to enhance the detection capability for small targets while reducing the model's complexity and parameter count. Specifically, the P2 detection layer is positioned at a lower level of network, enabling it to better capture detailed features. By introducing the P2 detection layer, YOLO11 can more accurately identify small target defects such as bottle cap breakpoints and damaged edges, effectively reducing missed detections. On the other hand, the P5 detection layer is located at a higher level of the network and provides limited auxiliary support for small object detection. Therefore, after removing the P5 layer, YOLO11 can focus more on mid-and low-level feature extraction, improving the recognition efficiency of small objects and making the network more efficient and concise during processing [22]. As shown in the red box in Figure 1, the model now has three output layers—P2, P3, and P4.

3. Experiments and Results

3.1. Environment and Parameter Settings

The experimental environment is set up on the Windows 10 Professional operating system, with parameter settings: epochs at 300, batch size at 16, workers at 4, initial learning rate at 0.01, momentum at 0.937, weight decay at 0.0005, optimizer using SGD, and image size uniformly set to 640*640. The training hard -ware consisted of a 12th Gen Intel(R) Core(TM) i5-12490F@3.00GHz CPU and an NVIDIA GeForce RTX2070 GPU. The deep learning framework employed was PyTorch 2.0.1, with GPU acceleration via CUDA 11.7 + cuDNN 8.9.6, and the programming language used was Python 3.10.

3.2. Dataset Construction and Preprocessing

The training utilizes image data of liquor bottle caps from the Alibaba Cloud Tianchi Competition's liquor defect dataset [23], which includes 2,789 images of defective bottled liquor. Each image contains one or multiple defects. The defects are divided into six categories: cap break point, deformation, Coding anomaly, Coding normal, bad edge, spins. Figure 8 shows the six defect categories of bottle caps.

In order to facilitate the training and evaluation of the model, the dataset is divided into a training set, a test set, and a validation set in a ratio of 8:1:1. Data augmentation techni -ques such as flipping, brightness adjustment, and noise addition were employed to expand the training data and enhance its diversity. After applying these augmentation techniques, the training set was expanded to a total of 5,448 images. Figure 9 shows the quantity and proportion of each defect category for bottle caps.

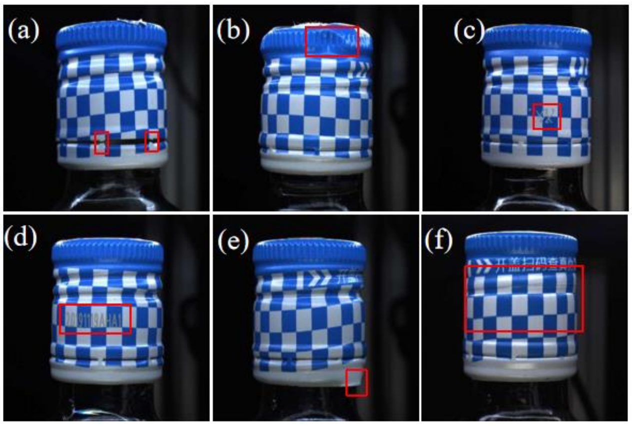


Figure 8. The six defect categories of bottle caps:(a) break point, (b) deformation, (c) Coding anomaly, (d)Coding normal, (e)bad edge, (f)spins

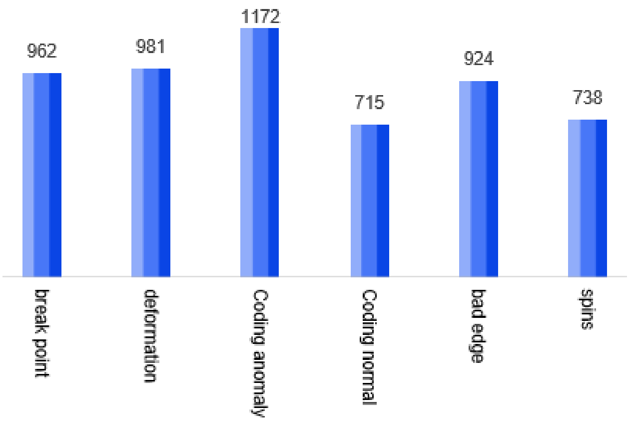


Figure 9. Number of each defect category;

3.3. Evaluation metrics

The evaluation metrics adopted in this study include precision(P), recall(R), mean average precision(mAP), model parameter cou -nt(Params),floating-point operations(FLOPs), and frames per second(FPS).

The calculation formula for precision is shown in equation (1):

$$P = \frac{TP}{TP + FP} \quad (1)$$

The calculation formula for recall is shown in equation (2):

$$R = \frac{TP}{TP + FN} \quad (2)$$

The mean Average Precision (mAP) is the average of the average precision across all classes, where AP represents the average precision for a single class, and its calculation formula is shown in equations (3) and (4).

$$mAP = \frac{1}{N}\sum_{i=1}^{n} AP_i \quad (3)$$

$$AP = \int_0^1 P(R)dR \quad (4)$$

3.4. Ablation experiment

To evaluate the contributions of the improved HGNetv2 and the detection layers P2P3P4 (P234) to the enhanced HP-YOLO11 model, we conducted ablation experiments on the dataset. The experimental results are summarized in Table 1.

According to the experimental results, replacing the YOLOv11 backbone network with the improved HGNetV2 network can enhance both detection accuracy and detection speed. Not only do the parameter count and FLOPs decrease by 19.4% and 14.3% respectively, but the mAP@0.5 also improves by 0.53 percentage points. The enhanced HGNetV2 effectively reduces model weight and computational redundancy. Adding a P2 detection layer and removing the P5 detection layer significantly improved the model's performance, with the mAP@0.5 value increasing by 1.12 percentage points. By introducing the P2 detection layer, the model can better capture the local details and positional features of small objects, thereby enhancing the recognition accuracy of small targets. The P5 layer used for detecting large objects is removed, which reduces the number of parameters and computational costs while maintaining accuracy. When HGNetv2 is combined with P234, compared to the baseline model, although the floating-point operations increase and the detection speed decreases slightly, both precision and recall rates show certain improvements without modifying the structure, with the mAP@0.5 value increasing by 1.52 percentage points and the number of parameters decreasing by 44.57%.Overall, the improved HP-YOLO11n model has achieved a significant enhancement in its overall performance for object detection tasks, This optimized combination not only improves the model's recognition accuracy but also significantly reduces the number of model parameters, providing strong support for the detection and deployment of surface defects in bottle caps in practical applications.

Table 1. Model performance comparison

Algorithm	P	R	Params/10^6	mAP@0.5	FLOPs/10^9	FPS
YOLO11n	85.31	82.71	2.58	85.54	6.3	**250.42**
YOLO11n+HGNetv2	86.72	82.59	2.08	86.08	**5.4**	248.66
YOLO11n+P234	**87.52**	82.86	1.94	86.67	9.8	149.82
HP-YOLO11n	87.02	**83.67**	**1.43**	**87.06**	8.7	177.43

3.5. Comparative experiment

Under identical conditions, the improved HP-YOLO11n model was compared with YOLO series models including YOLOv5n, YOLOv5s, YOLOv8n, YOLOv8s, and YOLO 11s algorithms. The experimental results are shown in Table 2.

According to the results in Table 2, Although the YOLOv5n, YOLOv8n, and YOLOv11n models have fewer FLOPs and higher FPS, they come with larger parameter counts and lower detection accuracy. While YOLOv5s, YOLOv8s, and YOLOv11s achieve high detection accuracy, they still underperform our model by 0.26% to 0.41%, and their parameter counts are also higher—all without modifying the architecture. Compared to the baseline model YOLO11n, it achieves a 2.31% improvement in precision, a 0.96%

increase in recall, and a 1.52% enhancement in mAP@0.5, while reducing the number of parameters by 44.57%. Figure 10 shows the mAP@0.5 curves of HP-YOLO11n and YOLO11.

Table 2. Comparison of HP-YOLO11n with other YOLO detection models

Algorithm	P	R	Params/10^6	mAP@0.5	FLOPs/10^9	FPS
YOLOv5n	85.46	83.92	2.18	85.52	**5.8**	229.75
YOLOv5s	84.47	**86.0**	7.82	86.65	18.7	162.3
YOLOv8n	86.40	82.36	2.69	86.17	6.8	**251.99**
YOLOv8s	85.78	85.39	9.83	86.80	23.4	141.41
YOLO11n	85.31	82.71	2.58	85.54	6.3	250.42
YOLO11s	86.33	84.99	9.42	86.70	21.3	144.38
HP-YOLO11n	**87.02**	83.67	**1.43**	**87.06**	8.7	177.43

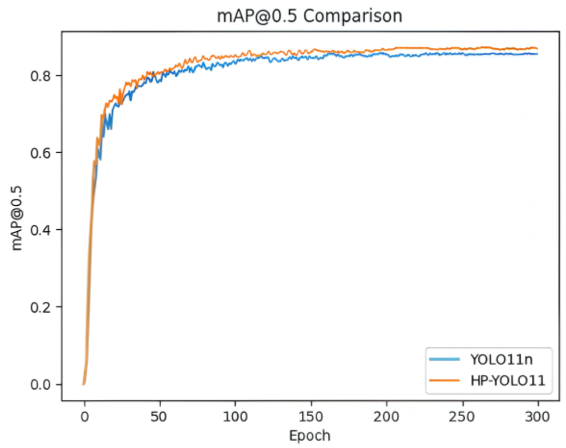


Figure 10. The mAP@0.5 curves comparing HP-YOLO11n with YOLO11

4. Conclusion

To address the issues of surface defect detection in liquor bottle caps and the large number of algorithm parameters, this paper proposes the HP-YOLO11n model to solve these problems. The model has two main improvements: First, we employ the improved HGNetv2 backbone network as our model's backbone, achieving a lightweight design while maintaining detection accuracy. Second, the capability to extract small target features is enhanced by adding a P2 detection layer and removing the P5 detection layer, thereby reducing unnecessary computations. Experi -mental results demonstrate that HP-YOLO11n outperforms other models in the YOLO series on the liquor bottle cap defect dataset. Compared to YOLO11n, the precision and recall rates increased by 2.31% and 0.96%, respectively, mAP@0.5 improved by 1.52%, while the number of parameters decreased by 44.57%."In the future, we will continue to optimize the network structure. In actual deployment, we will evaluate the practicality of the model in real-world environments."

References

[1] Hu W, Chen Y, Wu Z, et al. Study on the High Quality Development of Liquor Industry[J]. World Scientific Research Journal, 2023, 9(6): 109-121.

[2] Shen Y. A Study of the Operation Strategies of Kweichow Moutai Co., Ltd[J]. Academic Journal of Business & Management, 2023, 5(11): 54-58.

[3] Wang X, Lan X, Liu L. ESW-YOLO: A lightweight YOLO model for defect detection in bottled liquor[J]. 2024.

[4] China Liquor Industry Association. (2024). Anal ysis of output trend of Chinese liquor industry in 2023-2024 [R]. Beijing: CADA

[5] Liang, S.Qin, L, Zhang, M, Chu, Y, Teng, L, He, L.Win big with small:the influence of organic food packaging size on purchase intention. Foods 11(16), 2494 (2022)

[6] Chen B, Li C, Yuan P, et al. Research on defect detection of bottle cap interior based on low-angle and large divergence angle vision system[J]. Plos one, 2024, 19(5): e0303744.

[7] Abbey J D, Kleber R, Souza G C, et al. The role of perceived quality risk in pricing remanufactured products[J]. Production and Operations Management, 2017, 26(1): 100-115.

[8] Sharma V K, Mir R N. A comprehensive and systematic look up into deep learning based object detection techniques: A review[J]. Computer Science Review, 2020, 38: 100301.

[9] HU Z, WANG C. A small object detection algorithm of remote sensing image based onim proved Faster R-CNN[J]. Computer Engineering & Science, 2024, 46(06): 1063.

[10] ZHANG R, NING Q, LEI Y, et al. Garbage detection based on Mask R-CNN[J]. Computer Engineering & Science, 2022, 44(11): 2003.

[11] Nazir A, Wani M A. You only look once-object detection models: a review[C]//2023 10th International conference on computing for sustainable global development (INDIACom). IEEE, 2023: 1088-1095.

[12] Li Y, Ren F. Light-weight retinanet for object detec tion[J]. arXiv preprint arXiv:1905.10011, 2019.

[13] Toxqui-Quitl, C, Cardenas-Franco, J., Padilla -Vivanco, A., ValdiviezoNavarro, J: Bottle inspector based on machine vision. In: Image Processing: Machine Vision Applications VI, vol. 8661, pp. 299– 308 (2013). SPIE

[14] Zhou, X, Wang, Y, Zhu, Q, Mao, J, Xiao, C, Lu, X, Zhang, H: A surface defect detection framework for glass bottle bottom using visual attention model and wavelet transform. IEEE Transactions on Industrial Informatics 16(4), 2189–2201 (2019)

[15] Sheng,Z, Wang, G. Fast method of detecting packaging bottle defects based on eca-efficientdet. Journal of Sensors 2022(1), 9518910 (2022)

[16] Vijayakumar A, Vairavasundaram S. Yolo-based object detection models: A review and its applications[J]. Multimedia Tools and Applications, 2024, 83(35): 83535-83574.Khanam R, Hussain M. Yolov11: An overview of the key architectural enhancements[J]. arXiv preprint arXiv:2410.17725, 2024.

[17] Swathi Y, Challa M. YOLOv8: advancements and innovations in object detection[C]//International Conference on Smart Computing and Communi -cation. Singapore: Springer Nature Singapore, 2024: 1-13.

[18] Khanam R, Hussain M. Yolov11: An overview of the key architectural enhancements[J]. arXiv preprint arXiv:2410.17725, 2024.

[19] Zhao Y, Lv W, Xu S, et al. Detrs beat yolos on real-time object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2024: 16965-16974.

[20] Zhang P, Lo E, Lu B. High performance depthwise and pointwise convolutions on mobile devices [C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 34(04): 6795-6802.

[21] Wang Z, Su Y, Kang F, et al. Pc-yolo11s: a lightweight and effective feature extraction method for small target image detection[J]. Sensors, 2025, 25(2): 348.