International Journal of Computer Science and Information Technology
ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 7, Number 2, Year 2025
DOI: https://doi.org/10.62051/ijcsit.v7n2.04
Journal homepage: https://wepub.org/index.php/IJCSIT/index

The Application of Computer Vision in Part Recognition

Genming Liu, Zhanpeng Li, Xiaobei Jin

College of Mechanical Engineering, Tianjin University of Technology and Education, Tianjin 300222, China

ABSTRACT

Industrial part recognition, localization, and sorting are essential components of intelligent manufacturing in the Industry 4.0 era. With the rapid development of deep learning, the YOLO series of detectors has become the mainstream solution for industrial vision due to their balance of real-time performance and accuracy. Recent studies have not only improved YOLO architectures—through lightweight design, attention mechanisms, and optimized loss functions—but also explored hybrid strategies that combine deep learning with classical geometric methods to enhance robustness in complex environments. In parallel, complete pipelines that integrate recognition, localization, and robotic grasping have demonstrated feasibility in real-world production scenarios. Experimental results report high accuracy (mAP exceeding 95%), inference speeds above 30 FPS, and localization precision within the sub-millimeter range. Nonetheless, challenges such as limited domain generalization, difficulty with reflective or occluded parts, and the lack of standardized 3D pose benchmarks remain unresolved. Future progress is expected to be driven by synthetic data and domain adaptation, multi-modal sensing, unified multi-task architectures, and industrial-grade deployment pipelines. This review provides a comprehensive synthesis of state-of-the-art approaches and highlights promising directions for advancing intelligent robotic manufacturing.

KEYWORDS

Machine vision; YOLO; Part recognition; Pose estimation

1. INTRODUCTION

Industrial part recognition, localization, and sorting are cornerstones of modern manufacturing and assembly processes. With the advent of Industry 4.0, intelligent production lines increasingly rely on automated inspection and robotic handling to replace manual labor. Traditional machine vision methods based on handcrafted features such as edge descriptors or shape analysis have achieved partial success but suffer from limited robustness under varying illumination, clutter, and reflective materials [1]. These limitations restrict scalability and hinder industrial deployment.

The emergence of deep learning has transformed this landscape. Among object detection frameworks, the YOLO (You Only Look Once) family has been particularly influential, offering a balance of accuracy and real-time performance. From YOLOv3 to YOLOv8, successive generations have introduced architectural refinements, feature fusion strategies, and optimized loss functions, yielding significant improvements in both precision and inference speed [2]. This evolution has made YOLO-based detectors the default choice for industrial part recognition, especially when real-time robotic manipulation is required. However, the YOLO series is not the sole solution applicable to all scenarios. In recent years, other object detection architectures such as Faster R-CNN, SSD, PointNet, as well as various model fusion and integration methods, have also demonstrated significant effectiveness in industrial part recognition.

Nevertheless, the manufacturing domain introduces unique challenges absent in conventional computer vision benchmarks. These include small and reflective parts, stacked or heavily occluded arrangements, and constraints imposed by lightweight embedded systems [3-5]. To overcome these, researchers have explored a spectrum of solutions: improvements to YOLO's backbone and head design, integration of attention mechanisms, lightweight convolutional operators, and hybrid approaches combining deep learning with geometry or stereo vision. In parallel, complete recognition–localization–sorting systems have been proposed, validated in robotic assembly lines and demonstrating feasibility under real-world conditions.

This review aims to systematically synthesize these developments. Specifically, we analyze recent studies and categorize them into three major research directions: YOLO-based improvements, hybrid and traditional methods, and integrated robotic systems. By comparing contributions, evaluating performance, and identifying persistent limitations, this work provides a comprehensive overview of state-of-the-art methods and highlights promising avenues for future research.

2. LITERATURE REVIEW

2.1. YOLO-Based Improvements

In the field of object detection, the YOLO series, as a typical single-stage detection algorithm, has been widely applied in industrial visual tasks due to its efficiency and real-time performance. With the continuous increase in application demands, different versions of YOLO are also constantly evolving and optimizing. YOLOv3, YOLOv5, and YOLOv8 have been progressively adapted for industrial part recognition tasks. Each generation introduced enhancements in feature extraction, anchor design, attention mechanisms, and loss functions. And achieve a better result.

2.1.1. YOLOv3

In 2018,Joseph Redmon proposed YOLOv3 [6]. YOLOv3 utilized Darknet-53 as the feature extraction network and incorporated residual structures and deeper network layers. It adopted multi-scale prediction (based on the FPN concept), enabling it to detect small, medium, and large objects simultaneously, thereby enhancing the robustness of the detection.

Many scholars have made improvements to the YOLOv3 object detection algorithm from various aspects such as model structure optimization, preprocessing methods, feature enhancement, anchor box re-clustering, and lightweight networks, aiming to enhance the accuracy and detection speed of industrial part recognition. Among them, Feng Dantong [7] accelerated model training through transfer learning and achieved an mAP of 99.8% for 10 types of parts;Teng Chao [8] improved recognition accuracy to over 94.3% by combining feature point extraction and image enhancement with YOLOv3; Zhang Jing [9]. improved YOLOv3 by introducing Mish activation function, K-means clustering, and residual network,achieving robust and efficient part recognition in intelligent assembly. Based on YOLOv3 and combined with geometric feature extraction.

The JVadim P [10] combined the YOLOv3 and VGG19 neural network model for the identification of aerospace components. The VGG19 network was trained using the STL model to enhance accuracy. The recognition accuracy of this model on actual component photos reached 97%. Author Jongcheol p [11] utilized the YOLO algorithm to conduct real-time recognition of components. The developed system is capable of automatically tracking the progress of each process, thereby enhancing the visual management and data accuracy of the production process. Experimental results show that the accuracy rate of component recognition for this system has reached 90%. The application scenarios, recognition accuracy and detection speed of the part detection methods based on YOLOv3 in different studies are shown in Table 1.

Table 1. Improve YOLOv3

Model	Author	Application	Accuracy	Speed
YOLOV3	Teng Chao	Smart defect detection of parts	Accuracy Rate >94.3%	Detection time ⩽ 2.1 ms
	Feng Dantong	Multi-target assembly detection	mAP=99.8%	Detection cycle < 1 second
	Zhang Jing	Part detection in intelligent assembly systems	mAP=92.38%	Time=22.62ms
	JVadim P	Components of the rotor of a small gas turbine engine	Accuracy Rate =97%	-
	Jongcheol p	Optimization of the assembly process for ship engines	Accuracy Rate =90%	-

2.1.2. YOLOv5

In 2020, Ultralytics released YOLOv5 [12], which was implemented based on the PyTorch framework. It comes in various scales such as n/s/m/l/x, balancing speed and accuracy. It incorporates optimizations such as Mosaic data augmentation, automatic anchor box clustering, cosine annealing learning rate, model pruning and quantization. The greatest advantage is that it is extremely convenient for training and deployment, and supports exporting in formats like ONNX/TensorRT.

The scholars have focused on the improvements and optimizations of the YOLOv5 series of models in the field of object recognition. Liu Xiangde [13] proposed to integrate a lightweight backbone network, a Transformer module, and a noise purification mechanism to effectively enhance the accuracy of small target recognition and the speed of inference; Zhang Haoyang [14] optimized the detection ability for small-sized parts by adding small target detection layers and introducing the SE attention mechanism; Li Haoyu [15] combined the InternImage backbone network, the TSCode decoupling head, and the Slim-Neck to achieve a balance between high accuracy and lightweight in unstructured environments; Zheng Liangliang [16] addressed multi-scale and occlusion scenarios by adopting CBAM, Transformer encoder, and an improved NMS algorithm, significantly improving the robustness of part recognition in complex scenes. The experimental results show that all the improved methods have achieved significant improvements in terms of accuracy, speed and robustness. They have strong industrial application value and can support intelligent assembly, sorting and visual inspection tasks in complex scenarios. The application scenarios, recognition accuracy and detection speed of the part detection methods based on YOLOv5 in different studies are shown in Table 2.

Table 2. Improve YOLOv5

Model	Author	Application	Accuracy	Speed
YOLOV5	Liu Xiangde	Industrial parts	AP50=86.7%	-
	Zhang Haoyang	Small-sized parts	mAP=85.95%	Single image time=0.592s
	Li Haoyu	Assembling parts in an unstructured environment	mAP=96.42%	Frame=191FPS
	Zheng Lianglian	Multi-scale parts and occluded parts	mAP=96.2%	Frame=50FPS

2.1.3. YOLOv8

In 2023, Ultralytics released YOLOv8 [17]. YOLOv8 adopts an Anchor-free architecture, avoiding the complexity of anchor box matching and making the detection process more efficient. It replaces the C3 module of YOLOv5 with an improved C2f module to enhance feature expression capabilities and optimize the loss function and post-processing strategies. YOLOv8 unifies the interfaces for tasks such as object detection, instance segmentation, and keypoint detection, representing the development direction of the YOLO series towards multi-tasking and generalization.

Subsequently, improvements were made based on YOLOv8, targeting different industrial part recognition scenarios. Liu Chengyi's [18] research focused on enhancing the accuracy of part recognition in workshops and achieving lightweighting, proposing the YOLOv8-REM algorithm, which strikes a balance between accuracy and speed. Wang Jun [19]. addressed the issue of low recognition rate for small target parts by introducing the CBAM attention mechanism and an improved bounding box loss function, significantly improving the accuracy of small target recognition. Cheng Yao [20]. studied instance segmentation for complex stacked parts, proposing to add SPD-Conv, Focal Modulation, and additional segmentation layers on the basis of YOLOv8s, achieving real-time high-precision recognition and segmentation of stacked parts. The application scenarios, recognition accuracy and detection speed of the part detection methods based on YOLOv8 in different studies are shown in Table 3.

Overall, compared with YOLOv3 and YOLOv5, YOLOv8 exhibits a more efficient anchor-free design and stronger generalization ability. This makes it highly suitable for complex industrial environments such as stacked or small part recognition.

Table 3. Improve YOLOv8

Model	Author	Application	Accuracy	Speed
YOLOV8	Liu Chengyi	Machining part	mAP@0.5=89.1%	Frame=334FPS
	Wang Jun	Small parts	mAP=82.5%	21ms per sheet
	Cheng Yao	Stacked parts	mAP=95.5%	Frame=76.6FPS

2.2. Other Popular Method

Hybrid approaches integrate classical vision with deep learning to enhance robustness. Examples include Wan Songfeng [21] proposed a part recognition method based on the MPL classifier. By extracting the roundness features of 3C product components to train the multi-layer perceptron network model, the automatic classification and position detection of components were successfully achieved. Guo Beitao [22] proposed an efficient method based on binocular vision for the recognition and positioning of stacked parts. The study improved the classification fusion technology to enhance the recognition accuracy, and adopted the improved random Hough transform to extract the edge information of the parts, thereby achieving precise positioning of the stacked parts. In addition, there are other deep learning models used for part recognition, such as Fast RCNN and SSD. Scholar Wang Yi [23] proposed an improved method for part recognition in Faster RCNN. In the original network model, the ResNet101 network was substituted for the VGG16 network, and two new anchor points were added. The traditional non-maximum suppression was modified to Soft-NMS. Finally, in the training model stage, a multi-scale training strategy was adopted to reduce the false detection rate and improve the model accuracy. Scholar Yu Yongwei [24] proposed a part recognition method based on the Inception-SSD algorithm. The Inception network structure was introduced into the additional layers of the SSD network, and batch normalization module (BN) and residual structure connections were used to improve the detection accuracy without affecting the detection speed.

Table 4. Hybrid and traditional approaches for part recognition

Method	Author	Application	Accuracy	Speed
MLP	Wan Songfeng	3C parts	Accuracy Rate=99.8%	Classification Yime <20 ms
DDQN	Guo Beitao	Stacked parts	Accuracy Rate=97.6%	-
FAST-RCNN	Wang Yi	assembly parts	Precesion=96.1%.Recall=93.2%	Recognition Time=0.65/s
SSD	Yu Yongwei	parts	Normal ap =99.3%; Complex ap = 97.8%;	Frame=41FPS

2.3. Integrated Recognition–Localization–Sorting Systems

Integrated recognition–localization–sorting systems represent the practical application stage of machine vision research, where detection algorithms are embedded into complete robotic workflows. Compared with single-module studies that focus only on recognition accuracy, these systems must handle the full pipeline, including object detection, pose estimation, coordinate transformation, and robotic grasp planning. In industrial environments, this means not only recognizing parts, but also locating them in 3D space and executing reliable grasping under uncertainties such as conveyor movement, lighting variation, or mechanical tolerance.

Several recent studies have demonstrated the feasibility of such end-to-end systems. For example, Gao Maoyuan [25] designed a robotic grasping system for stacked parts, achieving a grasping success rate of 94.6% by combining YOLO-based detection with pose estimation. Yin Yuxiang [26] proposed a YOLOv4-tiny + CBAM model for unordered part sorting, which was validated on a robotic platform and achieved high efficiency in practical scenarios. Sun Jiahui [27] developed a YOLOv5s-CEGS-based system integrated with depth cameras, reducing localization error to less than 0.8 mm, which is sufficient for precise robotic assembly tasks.

Figure 1 presents a typical recognition–localization–grasping pipeline, showing how detection modules are integrated with pose estimation and robotic execution.

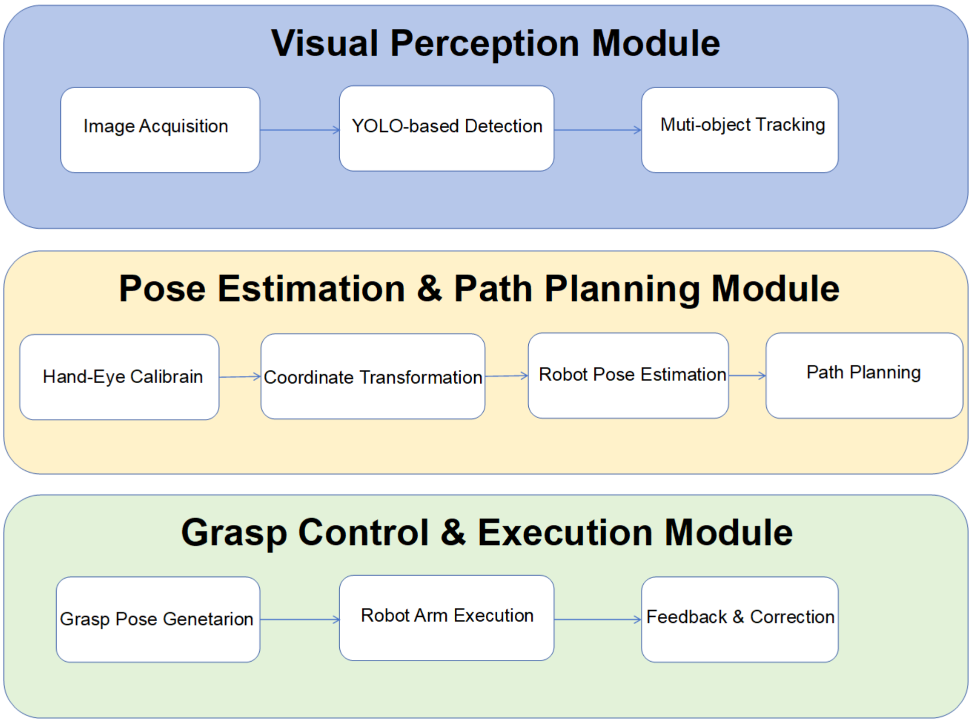

Figure 1. Recognition–Localization–Grasping Pipeline

3. CHALLENGES AND OPEN ISSUES

3.1. Domain Generalization

A major limitation of current part recognition models is their restricted generalization across domains. Models trained on specific datasets often struggle when exposed to unseen part types, new surface materials, or different lighting conditions. This is due to the inherent bias of training data, which is typically collected under controlled laboratory settings. For instance, YOLO-based detectors achieve high accuracy in curated datasets but may fail when deployed on factory floors where dust, reflections, or motion blur occur. The lack of large-scale, diverse, and standardized datasets for industrial parts further exacerbates this issue.

3.2. Pose Estimation and Benchmarks

Accurate pose estimation is critical for robotic manipulation, yet current methods vary widely in their evaluation protocols. Some works rely on stereo cameras, while others use monocular cues or depth sensors, making direct comparison difficult. Furthermore, the absence of standardized 3D benchmarks for part localization means that reported performance metrics are not directly comparable across studies. Without a unified evaluation framework, it is difficult to assess true progress or identify the most promising methods.

3.3. Lightweight Deployment Constraints

Real-time deployment often requires recognition models to run on embedded or mobile hardware, which imposes strict constraints on computation and memory. Lightweight YOLO variants (e.g., YOLOv5s-CEGS, YOLOv8-REM) demonstrate progress, but reductions in parameter count are still associated with trade-offs in accuracy. Balancing high precision with minimal computational cost remains an open problem, especially in scenarios where latency and power efficiency are critical, such as mobile inspection robots.

3.4. Robotic System Integration

Even when recognition accuracy is high, transferring detection results to robotic execution is non-trivial. Errors can arise from hand–eye calibration, coordinate transformations, or robot dynamics. Few studies address fault tolerance or adaptive control when recognition fails. Without robust integration strategies, recognition algorithms cannot fully meet industrial requirements for reliability and safety. This gap underscores the need for research on end-to-end pipelines that integrate recognition, localization, and grasp planning.

4. FUTURE DIRECTIONS

4.1. Synthetic Data and Domain Adaptation

To overcome the lack of diverse datasets, synthetic data generation and domain adaptation techniques hold significant promise. Simulation platforms can generate large volumes of annotated part images under varied lighting, noise, and occlusion conditions. Domain adaptation, including adversarial training and self-supervised fine-tuning, can then bridge the gap between synthetic and real-world data. This approach could drastically improve cross-domain robustness.

4.2. Multi-Modal and Advanced Imaging

Future recognition systems will increasingly integrate multi-modal sensing, such as polarization, hyperspectral, or structured-light cameras, alongside RGB inputs. These modalities can mitigate challenges associated with reflection and occlusion by providing complementary information. For example, polarization can reduce glare, while structured light can improve depth estimation for stacked parts. Combining modalities with deep learning has the potential to significantly increase robustness.

4.3. Unified Multi-Task Architectures

Current pipelines often treat detection, segmentation, and pose estimation as separate tasks, leading to error accumulation. A promising direction is the development of unified multi-task architectures that jointly predict part categories, segmentation masks, 6D poses, and grasp quality. Transformer-based backbones and decoupled heads are well-suited for this integration. Such models could reduce complexity and improve consistency across tasks, streamlining deployment in robotic systems.

4.4. Uncertainty Estimation and Robust Robotics

For safety-critical industrial applications, models must not only provide predictions but also quantify uncertainty. Incorporating Bayesian deep learning or Monte Carlo dropout could enable detectors to output confidence intervals for both detection and pose. This information would allow robots to make informed decisions, such as deferring to a fallback strategy when confidence is low. Coupling uncertainty estimation with adaptive robotic control will be crucial for achieving industrial-grade reliability.

4.5. Standardization and Industrial Deployment

The long-term success of part recognition systems depends on standardized benchmarks, evaluation protocols, and deployment pipelines. Establishing widely accepted datasets for 3D pose estimation would enable fair comparison across methods. At the same time, integrating recognition into production-ready MLOps frameworks will be essential for continuous monitoring, retraining, and optimization. Lightweight model distillation, ONNX/TensorRT acceleration, and edge deployment strategies will facilitate real-time performance in actual factories.

5. CONCLUSION

This paper has reviewed recent advances in machine vision-based part recognition and sorting for industrial automation. The YOLO family of detectors remains the dominant approach, offering continuous improvements in accuracy, speed, and robustness through architectural innovations and training strategies. Hybrid methods further enhance recognition by combining deep learning with classical geometric reasoning, while complete pipelines integrating recognition, localization, and robotic grasping demonstrate strong feasibility in industrial settings.

Despite these advances, current methods still face challenges such as domain generalization, reflective and small objects, heavy occlusion, lightweight deployment, and reliable robotic integration. Addressing these issues will require not only algorithmic improvements but also comprehensive solutions that combine synthetic and real data, multi-modal sensing, uncertainty estimation, and standardized evaluation frameworks. Looking forward, the convergence of multi-task architectures and industrial-grade deployment pipelines is expected to accelerate the practical application of intelligent vision systems, thereby promoting greater automation and efficiency in smart manufacturing.

REFERENCES

[1] Ma J., Liu X., Yang Y., et al. Application of computer vision technology in parts recognition [J]. China Equipment Engineering, 2020, (21):244-246.

[2] Mi Z., Lian Z. A review of YOLO methods for general object detection [J]. Computer Engineering and Applications, 2024, 60(21):38-54.ZX. Liu, Lightweight YOLOv5 for Part Recognition, 2024.

[3] Li R., Fu J., Zhai F., et al. Recognition and Pose Estimation Method for Stacked Sheet Metal Parts [J]. Applied Sciences, 2023, 13(7).

[4] Wang L. Research on industrial parts recognition method based on deep learning [D]. Lanzhou University of Technology, 2023. DOI:10.27206/d.cnki.ggsgu.2023.000415.

[5] Wang Q. Research on industrial stacked parts target recognition based on machine vision [D]. Anhui University of Science and Technology, 2024. DOI:10.26918/d.cnki.ghngc.2024.000790.J. Wang, Small-Target Part Recognition with Improved YOLOv8, 2024.

[6] Redmon, Joseph and Ali Farhadi.“YOLOv3:An Incremental Improvement.” ArXiv abs/1804.02767(2018).

[7] Feng D., Fang Z., Deng L. Real-time multi-object detection of assembly parts based on YOLOv3 [J]. Internal Combustion Engine & Parts, 2022, (12):91-93. DOI:10.19475/j.cnki.issn1674-957x.2022.12.039.

[8] Teng C., Chen M., Cai W., et al. Intelligent recognition method of mechanical parts based on YOLOv3 algorithm [J]. Information & Computer (Theory Edition), 2023, 35(21):53-55.

[9] Zhang J., Liu F., Wang R. Research on industrial parts recognition algorithm based on YOLOv3 in intelligent assembly [J]. Journal of Optoelectronics·Laser, 2020, 31(10):1054-1061. DOI:10.16136/j.joel.2020.10.0264.B. Guo, Stereo Vision-Based Recognition and Localization of Stacked Parts, 2024.

[10] Vadim P., Ekaterina P. Solving the part identification problem using their STL models [J]. Journal of Physics: Conference Series, 2022, 2373(7).M. Gao, Stacked Part Recognition and Robotic Grasping System, 2025.

[11] Jongcheol P., Wonho J., Namhyun Y. An empirical study on process management system using YOLO-based parts recognition [J]. Journal of Physics: Conference Series, 2021, 1888(1).

[12] Nelson J., Solawetz J. YOLOv5 is here: State-of-the-art object detection at 140 fps [EB/OL]. (2020-06-10). https://blog.roboflow.com/yolov5-is-here.D. Tong, Multi-target Real-time Detection of Assembly Parts with YOLOv3, 2022.

[13] Liu X., Ma H. Lightweight algorithm for parts recognition based on YOLOv5 [J]. Modular Machine Tool & Automatic Manufacturing Technique, 2024, (05):100-104+107. DOI:10.13462/j.cnki.mmtamt.2024.05.021.F. Dantong, Multi-target Real-time Detection of Assembly Parts with YOLOv3, 2022.

[14] Zhang H., He S., Meng D., et al. Research on improved YOLOv5 algorithm in mechanical parts recognition [J]. Computer & Digital Engineering, 2025, 53(04):1194-1200.C. Guanqu, Multi-object Part Recognition with Improved SSD, 2019.

[15] Li H., Chao Y., Wang C. Visual detection algorithm for assembly parts based on improved YOLOv5n model [J]. Modular Machine Tool & Automatic Manufacturing Technique, 2025, (01):10-14. DOI:10.13462/j.cnki.mmtamt.2025.01.003.

[16] Zheng L., Guo Y., Pu J., et al. Multi-scale parts recognition in complex assembly scenes based on improved YOLOv5s [J]. Modular Machine Tool & Automatic Manufacturing Technique, 2023, (07):1-5+9. DOI:10.13462/j.cnki.mmtamt.2023.07.001.

[17] Varghese R., Sambath M. YOLOv8: A novel object detection algorithm with enhanced performance and robustness [C]. 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS). Chennai: IEEE, 2024: 10533619.

[18] Liu C. Research on machining parts recognition method based on YOLOv8 [D]. Nanchang University, 2024. DOI:10.27232/d.cnki.gnchu.2024.001749.

[19] Wang J., Ma F., Fu C. Small target parts recognition method based on improved YOLOv8 [J]. Journal of Qingdao University (Engineering & Technology Edition), 2024, 39(02):39-46. DOI:10.13306/j.1006-9798.2024.02.006.

[20] Cheng Y., Zou G., Zhang J., et al. Improved YOLOv8 instance segmentation algorithm for stacked parts recognition [J]. Modular Machine Tool & Automatic Manufacturing Technique, 2025,(06):165-171. DOI:10.13462/j.cnki.mmtamt.2025.06.028.

[21] Wan S. Parts recognition based on MPL classifier [J]. Industrial Control Computer, 2022, 35(11):126-127+130.

[22] Guo B., Liu H., Zhang L. Research on recognition and localization of stacked parts based on binocular vision [J]. Modular Machine Tool & Automatic Manufacturing Technique, 2024, (08):43-47+53. DOI:10.13462/j.cnki.mmtamt.2024.08.009.

[23] Wang Y., Ma Z., Dong G. Research on parts recognition method based on improved Faster R-CNN [J]. Applied Optics, 2022, 43(1).

[24] Yu Y., Han X., Du L. Parts recognition based on Inception-SSD algorithm [J]. Optics and Precision Engineering, 2020, 28(08):1799-1809.

[25] Gao M., Wang T., Wang H., et al. Research on visual recognition and grasping system of parts in stacking scenes [J/OL]. Machine Tool & Hydraulics, 1-8 [2025-09-08]. Available at: https://link.cnki.net/urlid/44.1259.th.20250829.1343.002.

[26] Yin Y. Research on intelligent sorting method of unordered parts based on machine learning [D]. Yancheng Institute of Technology, 2023. DOI:10.44381/d.cnki.gycit.2023.000176.

[27] Sun J. Research on parts recognition and localization method based on machine vision and system design [D]. Shenyang University of Technology, 2024. DOI:10.27322/d.cnki.gsgyu.2024.000930.