machines

MDPI

Article

An Integrated Visual Perception and Soft Robotic Grasping System for Adaptive Handling of Railway Maintenance Tools

Pan Fan [1,2], Meng Tian [1], Yuhang Du [1], Guodong Lang [3], Liang Li [2] and Yafeng Li [1,*]

[1] School of Computer, Baoji University of Arts and Sciences, Baoji 721016, China; fanpan@bjwlxy.edu.cn (P.F.); tianmeng@stu.bjwlxy.edu.cn (M.T.); duyuhang@stu.bjwlxy.edu.cn (Y.D.)
[2] Shaanxi Key Laboratory of Advanced Manufacturing and Evaluation of Robot Key Components, Baoji University of Arts and Sciences, Baoji 721016, China; liliang@bjwlxy.edu.cn
[3] State Key Laboratory of Robotics and System, Harbin Institute of Technology, Harbin 150001, China; 22b908022@stu.hit.edu.cn
* Correspondence: liyafeng@bjwlxy.edu.cn

Abstract

To address the challenges of severe background interference and unstable grasping of irregular maintenance tools in complex railway ballast environments, this paper proposes a robotic system that integrates enhanced visual perception with bio-inspired soft grasping. The core components of the system include a lightweight detection network (RA-YOLO), asymmetric "Fin Ray" soft fingers, and a visual servoing control framework. By embedding the CBAM attention mechanism and incorporating Mosaic data augmentation, RA-YOLO achieves robust feature extraction under complex backgrounds. The fingertip topology is optimized using the Yeoh constitutive model and finite element analysis, thereby improving stiffness under heavy loads and overall adaptability. Experimental results demonstrate that proposed RA-YOLO achieved a mAP@0.5 of 93.6% on the standard test set with an inference speed of 105 FPS. The visual-servo localization experiment an average Euclidean positioning error of 1.03 mm, with the maximum component-wise absolute error remaining below 2.5 mm. In system-level grasping experiments involving five categories of irregular tools, the integrated system achieved an overall grasping success rate of 91.8%, indicating its potential for automated tool recovery in unstructured railway maintenance environments.

Keywords: railway maintenance; object detection; soft gripper; visual servoing; YOLOv8

1. Introduction

With the rapid development of China's high-speed railways, the efficiency of operation and maintenance (O&M) and the safety of railway infrastructure have become critical for ensuring the stability of the transport network. By the end of 2025, the operating mileage of China's high-speed rail had exceeded 45,000 km, ranking first in the world and serving as critical infrastructure for both the national economy and public mobility [1]. China's high-speed railway lines operate in highly diverse and complex environments, extensively traversing special geographical regions such as alpine permafrost zones, hot-humid areas, and Gobi desert regions. Railway infrastructure is therefore subjected to coupled effects of alternating high-speed train loads, thermal stress, and environmental erosion over long periods, which significantly accelerates performance degradation and imposes unprecedentedly stringent requirements on maintenance precision, frequency, and efficiency [2].

Check for updates

Academic Editor: Zhuming Bi

Received: 8 April 2026
Revised: 22 May 2026
Accepted: 27 May 2026
Published: 1 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Traditional railway O&M operations rely heavily on manual work during nighttime maintenance windows for tool handover, foreign-object clearance, and equipment inspection/repair, resulting in long-standing bottlenecks that are difficult to overcome. Ballast zones are characterized by complex terrain, insufficient nighttime illumination, and frequent adverse weather, leading to high labor intensity, low efficiency, and significant safety risks such as hazards from passing trains on adjacent tracks and falls from height. In addition, substantial manpower is consumed in tool handover; poor mobility in ballast areas not only reduces transfer efficiency but also increases the risk of tools being dropped onto the track. These limitations have motivated increasing research interest in robotic inspection and maintenance systems for unstructured railway environments [3,4].

As the direct execution unit for tool handover tasks, the end-effector performance directly determines the success rate and stability of grasping operations [5,6]. Although conventional rigid grippers provide excellent load-bearing capacity, they have poor shape adaptability and high contact stiffness, making them unsuitable for the diverse and irregular maintenance tools encountered in railway O&M; their fault tolerance is particularly low in unstructured ballast environments [7]. In contrast, bio-inspired soft grippers leverage the compliance of hyperelastic materials and passive self-adaptation to achieve contour-conforming grasping of irregular objects without complex force-control systems, and have become a mainstream approach for grasping irregular objects in unstructured environments [8]. Among them, Fin Ray soft fingers inspired by fish-fin biomechanics can achieve enveloping grasps through force-induced bending, offering advantages such as simple structure, convenient fabrication, and high compliance [9,10]. Extensive studies have been reported on driving methods and topological design. However, most existing studies focus on light-load scenarios [11], while research on grasping 2 kg-class heavy maintenance tools in railway maintenance remains limited. The inherently insufficient lateral stiffness of Fin Ray structures can cause buckling instability during grasping [12,13], and existing improvements [14,15] still struggle to simultaneously satisfy compliant enveloping and rigid load-bearing requirements under heavy loads. Meanwhile, complex railway environments impose higher demands on the coordinated control and operation planning of robotic systems [16,17]. Traditional low-level code-centric development often suffers from long development cycles and difficult subsystem integration. However, there remains a gap in current research regarding the deep integration of visual perception, soft grasping, and robotic-arm motion control into a coherent, full-process automation pipeline [18].

The visual perception module is essential for target recognition and accurate localization in railway maintenance robots. Convolutional neural network-based object detection algorithms have been widely adopted in industrial inspection and railway scenarios. Among them, the YOLO family, with end-to-end architecture and excellent real-time performance, has become a mainstream solution in visual detection for robotic grasping applications. YOLOv8 [19] has been comprehensively optimized in accuracy, speed, and lightweight deployment [20–24], and has been validated in scenarios such as railway freight grasping and fastener-defect recognition [3]. However, when directly applied to ballast environments, high-frequency texture backgrounds tend to drown out small-scale tool features, and random tool poses combined with complex lighting easily lead to false detections and missed detections. In addition, many existing improvements boost detection performance at the cost of significantly increased model parameters, making it difficult to balance detection accuracy and lightweight deployment under severe background interference, failing to meet the strict real-time and high-precision requirements of embedded edge devices used in railway maintenance robots.

To address the challenges of tool detection difficulty, unstable grasping, and insufficient control precision in unstructured railway ballast environments, this paper proposes an

adaptive robotic operation system that integrates enhanced visual perception, bio-inspired soft grasping, and multi-module coordinated control. The main scientific contributions of this study are summarized as follows:

1. Biomechanical Topological Optimization: We designed an asymmetric Fin Ray soft finger and optimized its topological parameters based on the Yeoh hyperelastic constitutive model and finite element analysis [25]. This design improves the balance between lateral stiffness under 2 kg-class loads and compliant enveloping of irregular tools.
2. Vision-Guided Collaborative Control Framework: We established an intelligent control framework based on Simulink and ROS [18,26]. By utilizing a finite-state machine (FSM), the framework integrates visual perception, manipulator kinematics, and soft grasping into an automated operation pipeline.
3. Robust Perception in Complex Backgrounds: We proposed the RA-YOLO detection network, based on the lightweight YOLOv8n architecture [27,28]. By embedding the CBAM attention module and introducing a Mosaic augmentation strategy, the network achieved high-precision, real-time tool detection, with an mAP@0.5 of 93.6% at 105 FPS against severe interference from railway ballast environments.

2. Materials and Methods

2.1. Overall System Architecture

As shown in Figure 1, this study proposes a robotic system framework integrating visual perception and bio-inspired grasping, aimed at achieving automated sorting of railway maintenance tools. The hardware system consists of a host computer, an Intel RealSense D435i depth camera, an Aubo-i5 six-axis collaborative robotic arm, and the asymmetric Fin Ray bio-inspired soft gripper designed in this work [28,29].

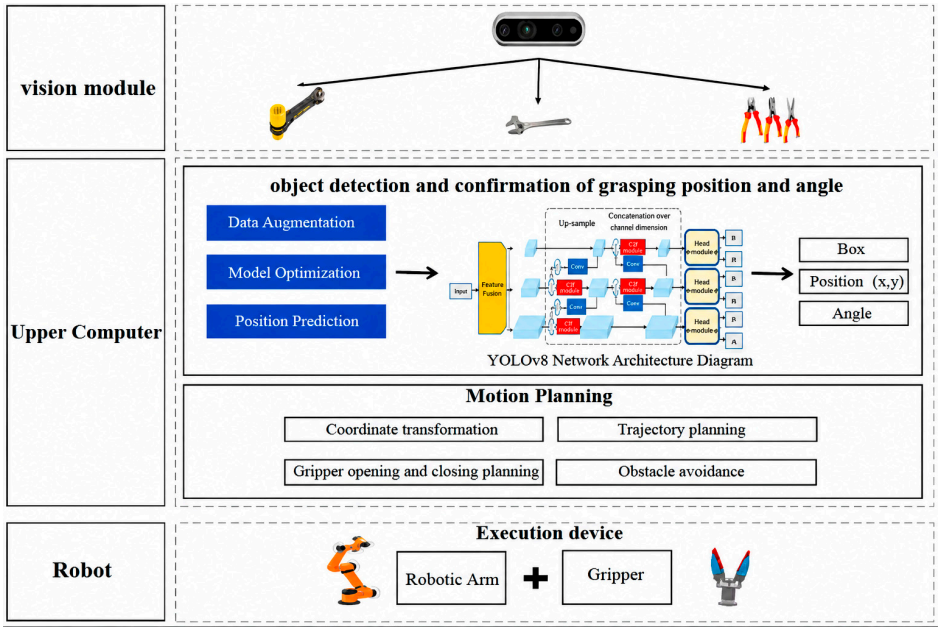


Figure 1. Schematic diagram of the robotic system.

The operational workflow of the system is divided into three main stages: perception, decision-making, and execution. In the perception stage, the depth camera acquires a real-time image stream, and the improved RA-YOLO network identifies target tools in complex ballast backgrounds and extracts their pixel coordinates. In the decision-making stage, the system uses a hand-eye calibration model to transform pixel coordinates into the robotic arm base frame, and invokes an inverse kinematics algorithm to compute the target

joint angles [28,29]. A quintic polynomial interpolation algorithm then generates a smooth motion trajectory [18,27]. In the execution stage, the robotic arm moves to the target pose along the planned path. A stepper motor then drives the asymmetric soft hand to close, allowing the fingers to conform to the target object through deformation and accomplish stable transfer operations.

2.2. Soft Gripper Design and Modeling

2.2.1. Asymmetric Bio-Inspired Design Concept

Given that the workpieces encountered at railway maintenance sites are geometrically irregular and made of hard materials, conventional rigid end-effectors frequently suffer from limited contact points, susceptibility to surface damage, and detachment under high dynamic loads. Soft fingers based on the fin-ray effect have attracted wide attention for their excellent compliance; however, the standard symmetric Fin Ray structure tends to undergo buckling instability due to insufficient lateral stiffness when grasping 2 kg-class heavy tools [9–13].

To reconcile the conflicting requirements of compliant enveloping and rigid load-bearing, and guided by the actual operational specifications of railway inspection tasks summarized in Table 1, this paper proposes an asymmetric Fin Ray topological structure [13,17].

Table 1. Design specifications of the soft gripper.

Attribute	Specific Requirements
Overall Weight	≤1 kg
Grasping Objects	Common tools for railway inspection, featuring irregular shapes and hard surfaces
Size Range	10–120 mm
Maximum Payload Weight	≤2 kg

The asymmetric Fin Ray structure is designed to enhance enveloping adaptability for irregular workpieces and heavy-load stability through the synergistic effect of multiple geometric and material features. As shown in Figure 2, its core features include an asymmetric rib distribution, an arc-shaped outer sidewall, and a composite anti-slip surface layer [11,12].

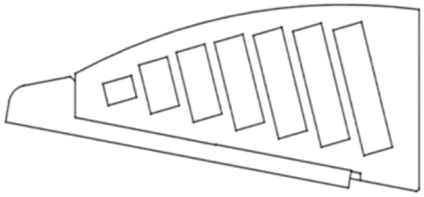


Figure 2. Structural schematic of the asymmetric Fin Ray soft finger.

Unlike the conventional symmetric layout, this structure arranges the inner ribs at an inclination along the grasping direction. This asymmetric topology breaks the mechanical equilibrium at the initial contact stage, inducing a directional inward-curling deformation of the finger pad, which more closely conforms to irregular workpiece contours and effectively increases the contact area [13].

The outer wall adopts an arc-shaped cross-section instead of the common flat wall. The arc cross-section provides a greater second moment of area during bending deformation [9,10], thereby significantly suppressing the tendency for outward buckling under load and enhancing the overall lateral stiffness and anti-instability capability of the structure [16,17].

To address the problem of slippage on smooth metal tool surfaces, a silicone layer is composited onto the finger-pad region. This surface layer increases the interfacial static friction coefficient, ensuring reliable gripping during heavy-load lifting and transfer operations.

The overall gripping mechanism adopts a lead-screw-linkage transmission scheme, as shown in Figure 3, which efficiently converts the rotational motion of the motor into parallel opening and closing motion of the fingers, achieving structural compactness and lightweight design while maintaining transmission efficiency.

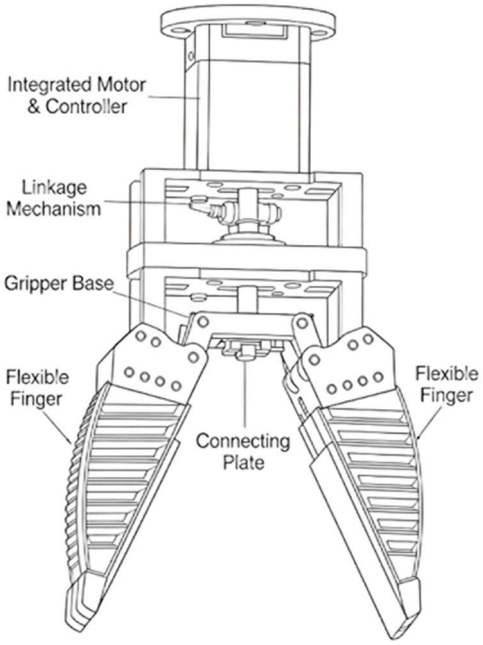


Figure 3. Overall structural design of the soft gripper.

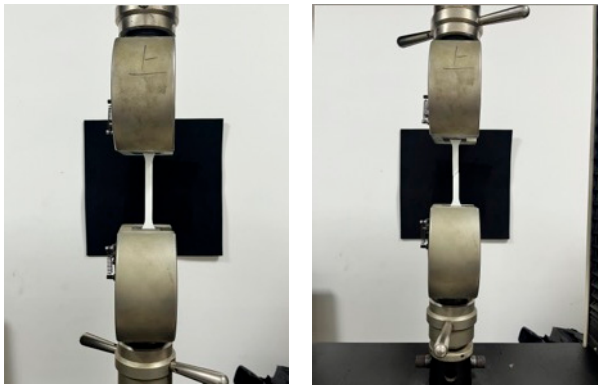
2.2.2. Material Characterization and Constitutive Modeling

The mechanical response of the soft fingers is primarily determined by the properties of the base material. BASF Elastollan® 1185A [30] thermoplastic polyurethane (TPU) was selected as the substrate due to its stable mechanical performance over a wide temperature range and excellent abrasion resistance, making it suitable for outdoor railway maintenance applications [9–12]. The key mechanical properties provided by the datasheet are summarized in Table 2.

Table 2. Mechanical properties of BASF Elastollan® 1185A.

Mechanical Property	Value	Unit	Test Standard
Tensile Strength	>45	MPa	ISO 527-1/-2
Elongation at Break	>500	%	ISO 527-1/-2
Shore A Hardness	87	Shore A	ISO 868
Tear Strength	70	kN/m	ISO 34-1
Abrasion Resistance	25	mm^3	ISO 4649

To determine the constitutive parameters of the soft fingers, uniaxial tensile tests were performed on specimens made from the same material. The tests were conducted on a universal testing machine. Three specimens were tested, and the mean engineering stress–strain curve was used to fit the constitutive model (Figure 4). Figure 4a shows the specimen mounted on the testing machine, while Figure 4b illustrates the specimen during elongation [31,32].



(**a**) (**b**)

Figure 4. Uniaxial (single-axis) tensile test setup of 3D-printed TPU specimens: (**a**) specimen mounted; (**b**) specimen during elongation.

During grasping, the soft fingers exhibit significant geometric and material nonlinearity. To accurately describe this large-deformation behavior, a third-order Yeoh hyperelastic model was adopted. Its strain energy density function is expressed as:

$$W = \sum_{i=1}^{3} C_{i0}\left(I_1 - 3\right)^i \quad (1)$$

where I_1 is the first strain invariant and C_{i0} are the material constants.

Based on the measured tensile stress–strain data within the 0–300% strain range, the Yeoh material constants were fitted using the least-squares method:

$$C10 = 2.054 \text{ MPa}, C20 = -0.0753 \text{ MPa}, C30 = 0.00231 \text{ MPa}$$

The coefficient of determination $R^2 = 0.9938$ indicates that the fitted Yeoh model accurately captures the large-deformation response of the TPU material. The resulting stress–strain curve and Yeoh model fit are shown in Figure 5, providing a reliable basis for subsequent finite element analyses [14].

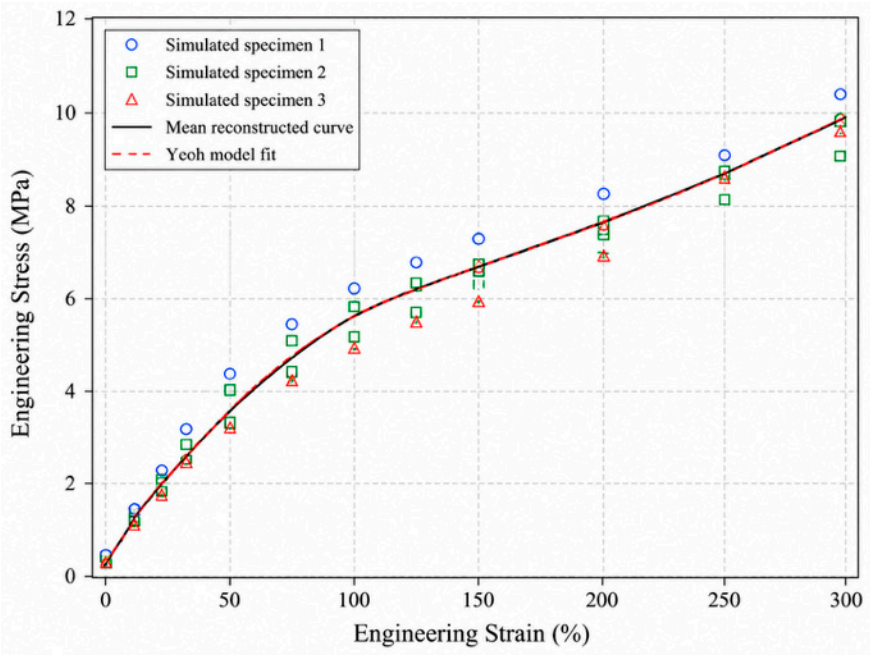


Figure 5. Engineering stress–strain curve of 3D-printed TPU specimens (mean ± SD) and fitted Yeoh model (0–300% strain).

2.2.3. Finite Element-Based Optimization and Orthogonal Analysis of Key Structural Parameters

To investigate the influence of geometric parameters on the mechanical response of the soft fingers, finite element-based parametric analyses were conducted using the previously fitted Yeoh hyperelastic model. Three structural parameters were examined: rib inclination angle (θ), rib spacing (d), and outer wall thickness (t). The study primarily focused on the stress concentration and deformation trends associated with different structural configurations [13,14].

The rib inclination angle (θ) mainly affects the internal force transmission path of the asymmetric Fin Ray finger. As shown in Figure 6, when θ = 0°, stress concentration appears at the rib roots because the load is primarily transferred through local bending. Increasing θ to 15° only slightly improves the stress distribution, indicating that a small inclination is insufficient to change the load path effectively. When θ = 30°, the inclined ribs form a more favorable truss-like transmission path, and the normalized stress index reaches its minimum value among the tested angles. Meanwhile, the inward-curling deformation becomes more pronounced, which is beneficial for conformal grasping of irregular tools. However, when θ increases to 45°, excessive rib inclination weakens radial support and increases shear-dominated deformation, causing the stress index to rise again.

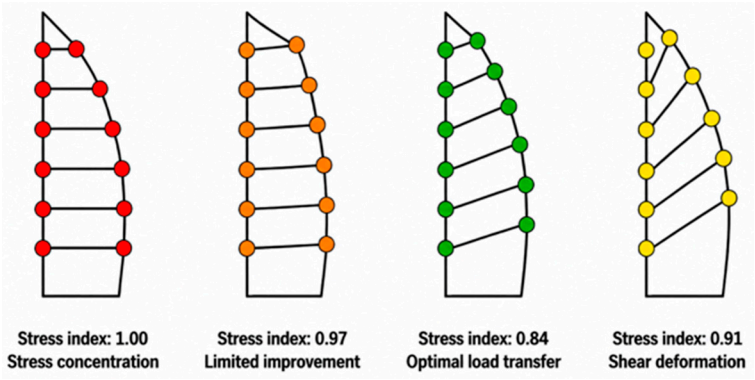


Figure 6. Comparison of stress distribution mechanisms for different rib inclination angles (θ). The colored dots indicate representative rib nodes used to illustrate stress-concentration regions and load-transfer paths under different rib inclination angles.

Accordingly, θ = 30° was adopted because it provided the most favorable compromise among stress reduction, load transfer, and inward deformation.

The rib spacing (d) directly affects the internal clearance during structural bending and the interference behavior between ribs. As shown in Figure 7, d = 4 mm, the ribs are too densely arranged, resulting in physical contact and interference under large-curvature bending and a relatively high peak-stress level. Increasing d to 8 mm provides sufficient internal deformation space, effectively reducing rib interference and lowering the peak-stress concentration. Although further increasing d to 12 mm can reduce local stress to some extent, the reduced rib number weakens the internal load-bearing path and lateral support.

A rib spacing of 8 mm was therefore adopted to reduce rib interference while maintaining sufficient structural support.

The outer wall thickness (t) directly determines the bending stiffness and deformation adaptability of the soft finger under load. As shown in Figure 8, when t = 2 mm, the structure exhibits local buckling due to insufficient radial support, accompanied by stress concentration. Increasing t reduces the maximum stress, but excessive wall thickness makes the finger behave more like a rigid beam and weakens its adaptive enveloping capability for irregular contours.

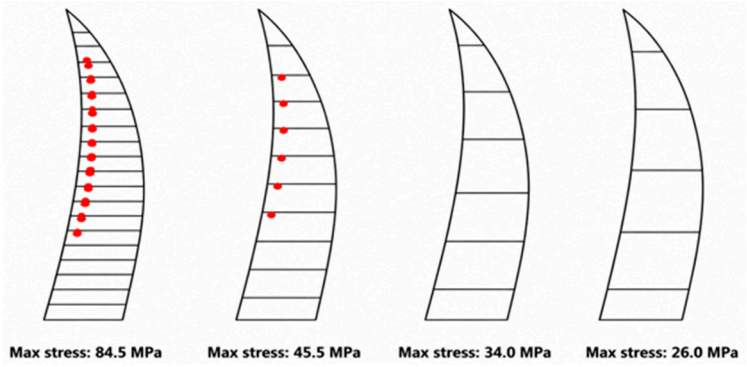


Figure 7. Internal structural interference analysis for different rib spacings (d). The red dots indicate representative regions where adjacent ribs tend to contact or interfere during large-curvature bending.

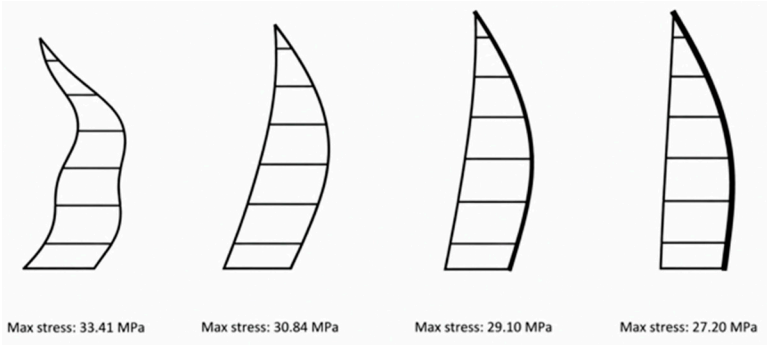


Figure 8. Effect of outer wall thickness (t) on stiffness and deformation mode.

Thus, t = 3 mm was selected to provide adequate bending stiffness without substantially reducing contour conformity [28].

To ensure that the material deformation performance meets the requirements of the railway environment, the influence of each structural parameter on the key mechanical response indicators was further quantified. Figure 9 summarizes the quantitative trends of the normalized stress index, maximum principal stress, and deformation-related indicators with respect to rib inclination, rib spacing, and outer wall thickness.

Based on the single-factor analysis, a three-factor, three-level orthogonal design was introduced to further evaluate the combined influence of multiple geometric variables. The rib inclination angle θ, rib spacing d, and outer wall thickness t were selected as the orthogonal design factors. To remain consistent with the preceding parametric analysis, the three levels were defined as 0°, 30°, and 45° for θ; 4 mm, 8 mm, and 12 mm for d; and 2 mm, 3 mm, and 6 mm for t, as listed in Table 3.

Table 3. Factors and levels of the L9(3^3) orthogonal design.

Factor	Symbol	Level 1	Level 2	Level 3
Rib inclination angle	θ	0°	30°	45°
Rib spacing	d	4 mm	8 mm	12 mm
Outer wall thickness	t	2 mm	3 mm	6 mm

For each parameter combination, the structural response was evaluated under the same loading condition. Three normalized response indices were introduced: stress safety score f_σ, adaptive deformation score f_δ, and lateral stability score f_S. The response scores were scaled to the range of 0–1 across the nine orthogonal combinations.

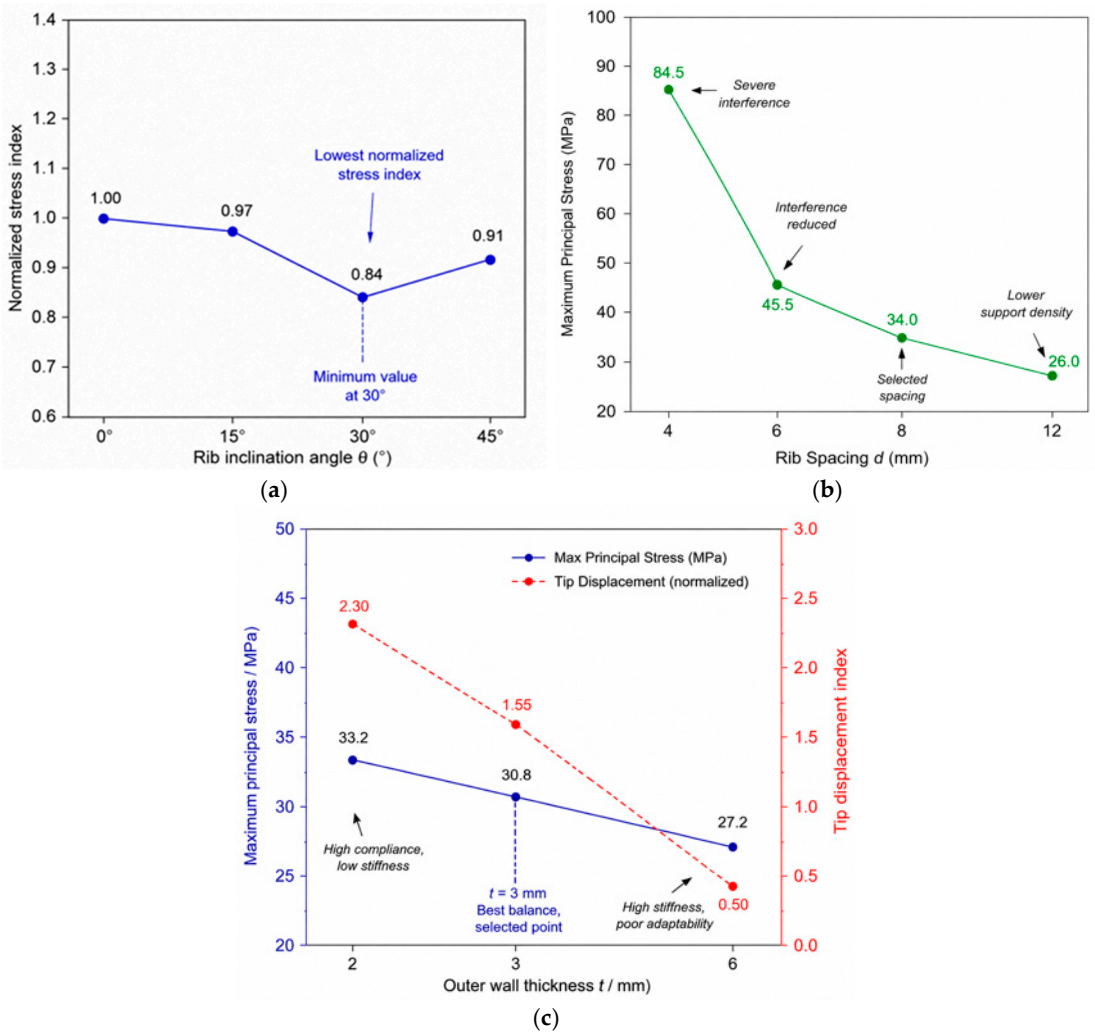


Figure 9. Quantitative analysis of single-factor structural parameter trends: (**a**) normalized stress index under different rib inclination angles; (**b**) influence of rib spacing; (**c**) influence of outer wall thickness.

For the stress index, inverse normalization was used so that lower stress corresponded to a higher f_σ, whereas direct normalization was used for adaptive deformation and lateral stability. A higher fσ indicates lower stress concentration and better structural safety, whereas higher f_δ and f_S indicate better contour-conforming capability and anti-instability performance, respectively.

To balance these competing requirements, a normalized comprehensive score was defined as:

$$F = 0.4f_\sigma + 0.3f_\delta + 0.3f_S$$

where f_σ, f_δ, and f_S denote the normalized sub-scores of stress safety, adaptive deformation, and lateral stability, respectively. Considering the repeated grasping requirement in railway maintenance scenarios, the stress safety index was assigned a relatively higher orthogonal design matrix and normalized response evaluation results are listed in Table 4.

As shown in Table 4, different parameter combinations lead to distinct trade-offs between compliant deformation and structural stability.

The level-average and range analysis of the comprehensive score is further summarized in Table 5.

Table 4. Orthogonal design matrix and normalized response evaluation of the soft finger structure.

Test No.	A:θ	B:d	C:t	Stress Safety Score f_σ	Adaptive Deformation Score f_δ	Lateral Stability Score f_S	Comprehensive Score (F)
1	0°	4 mm	2 mm	0.64	0.78	0.62	0.68
2	0°	8 mm	6 mm	0.70	0.62	0.78	0.70
3	0°	12 mm	3 mm	0.72	0.70	0.76	0.73
4	30°	4 mm	6 mm	0.76	0.65	0.84	0.75
5	30°	8 mm	3 mm	0.92	0.86	0.90	0.90
6	30°	12 mm	2 mm	0.68	0.88	0.58	0.71
7	45°	4 mm	3 mm	0.68	0.78	0.64	0.70
8	45°	8 mm	2 mm	0.62	0.84	0.56	0.67
9	45°	12 mm	6 mm	0.70	0.60	0.76	0.69

Table 5. Level-average and range analysis of the orthogonal design results.

Factor	Level 1	Level 2	Level 3	Range (R)	Rank
A:θ	0.70	0.79	0.69	0.10	1
B:d	0.71	0.76	0.71	0.05	3
C:t	0.69	0.78	0.71	0.09	2

Based on the range analysis, the rib inclination angle θ and the outer wall thickness have the most significant impact on the overall performance, followed by the rib spacing. This indicates that:

1. The rib inclination primarily governs the internal force transmission path and the inward-curling deformation mode;
2. The outer wall thickness mainly controls the trade-off between stiffness and compliance;
3. The rib spacing affects internal interference and support density.

The main-effect plots of the orthogonal analysis are presented in Figure 10. The comprehensive score its highest average value at θ = 30°, d = 8 mm, and t = 3 mm. Therefore, the analysis indicates that A2B2C2 represents the optimal factor-level combination.

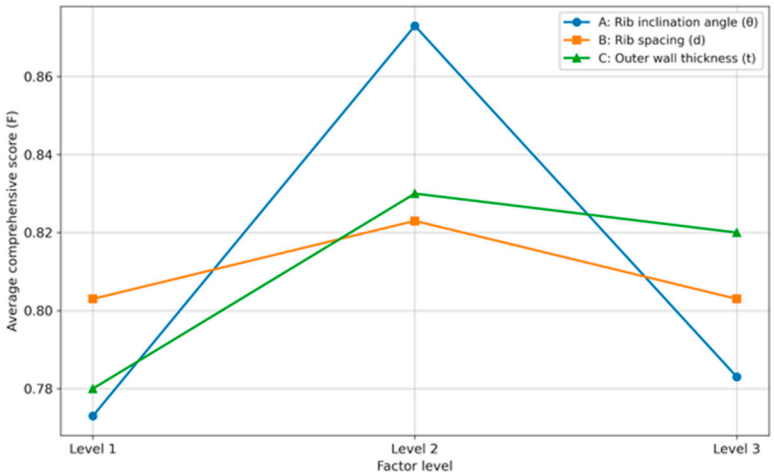


Figure 10. Main-effect plots of the orthogonal analysis for the comprehensive performance score.

Based on the single-factor finite element analysis and orthogonal evaluation [24], θ = 30°, d = 8 mm, and t = 3 mm were selected as the final structural parameters of the soft finger, as summarized in Table 6.

Table 6. Selected structural parameters of the asymmetric Fin Ray soft finger.

Parameter	Symbol	Selected Value
Rib angle	θ	30°
Rib spacing	d	8 mm
Outer wall thickness	t	3 mm

This parameter combination ensures that the working stress remains below the tensile strength of the TPU material while achieving a balanced compromise between stress safety, structural support, and geometric adaptability.

2.2.4. Kinematic and Mechanical Modeling of the Gripping Mechanism

The driving mechanism adopts a lead-screw-dual-crossbeam symmetric structure, as shown in Figure 11. The servo motor converts rotational motion into linear vertical displacement of the dual crossbeams via a lead-screw nut assembly, which in turn drives the fingers to open and close synchronously through a four-bar linkage mechanism.

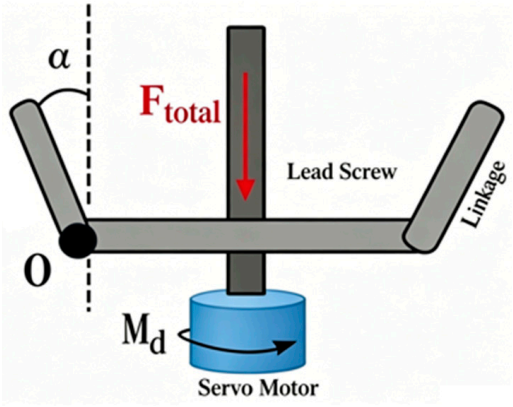


Figure 11. Schematic diagram of the lead-screw-dual-crossbeam driving mechanism.

Neglecting transmission friction losses, based on the principle of helical transmission, the motor output torque M_d is converted into the total axial thrust acting on the crossbeam F_{total}:

$$F_{total} = \frac{2\pi \cdot \eta \cdot M_d}{P} \quad (2)$$

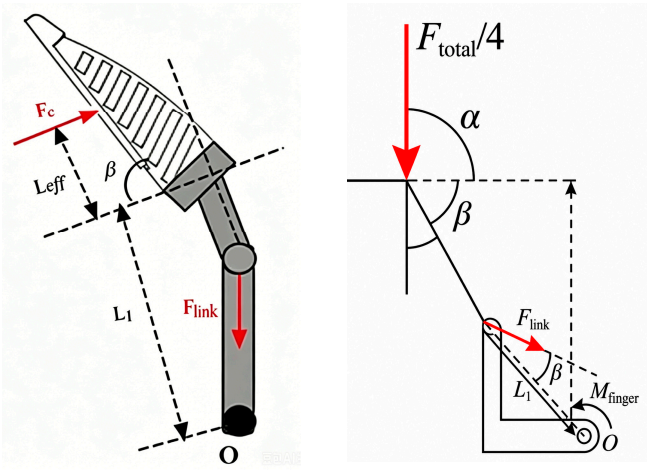
where P is the lead-screw pitch and η is the transmission efficiency. Assuming perfect structural symmetry, the total thrust is evenly distributed among four transmission chains, and the vertical force component acting on a single support link is $F_y = F_{total}/4$.

Taking a single-finger transmission chain as the free body for force analysis, as shown in Figure 12a, let the angle between the support link and the lead-screw axis be α; the thrust transmitted axially by the link to the finger connector is F_{link}:

$$F_{link} = \frac{F_y}{\cos \alpha} = \frac{\pi \cdot \eta \cdot M_d}{2P \cdot \cos \alpha} \quad (3)$$

This thrust acts on the kinematic hinge point of the finger connector, as shown in Figure 12b. Let the driving moment arm length be L_1, and the angle between the link thrust and the rotation radius be β. The driving torque M_{finger} obtained by the finger base about the rotation center O is:

$$M_{finger} = F_{link} \cdot L_1 \cdot \sin \beta \quad (4)$$



(**a**) (**b**)

Figure 12. Force analysis model of the rigid transmission mechanism: (**a**) geometric schematic of the link drive; (**b**) free-body force analysis of the single-finger transmission chain.

2.2.5. Rigid-Flexible Coupled Modeling Based on the Pseudo-Rigid-Body Model

Classical rigid-body dynamics cannot describe the large-deformation behavior of Fin Ray fingers during grasping. When the finger contacts a workpiece, the energy input from the servo motor is split into two parts: one part is converted into effective clamping work on the workpiece, and the other is stored as elastic strain energy in the TPU material. Neglecting the deformation energy dissipation would lead to a significantly overestimated clamping force.

To address this, a pseudo-rigid-body model is introduced to approximate the continuous deformation as an equivalent rigid-link system, as shown in Figure 13. The elastic restoring torque generated by finger bending can be expressed as M_{res}:

$$M_{res} = K_{eq} \cdot \Delta\phi \quad (5)$$

where K_{eq} is the equivalent torsional stiffness coefficient reflecting the elastic modulus of TPU and the cross-sectional second moment of area, and $\Delta\phi$ is the angular displacement of the finger. Based on energy conservation and moment equilibrium, the corrected moment balance equation is established:

$$M_{finger} = M_{res} + F_c \cdot L_{eff} \quad (6)$$

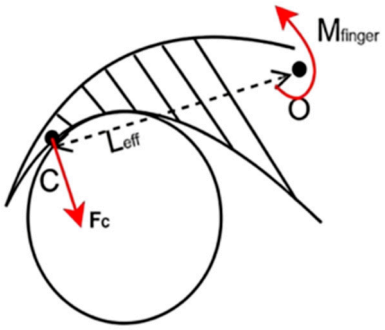


Figure 13. Rigid–flexible coupled model of the soft finger. The arrows indicate the contact force F_c, the equivalent elastic restoring moment M_{finger}, and the effective moment arm L_{eff} used in the pseudo-rigid-body analysis.

Combining Equations (3)–(6), the system-level contact force mapping model incorporating the deformation energy dissipation term is obtained:

$$F_c = \frac{1}{L_{eff}}\left(\frac{\pi\eta L_1 \sin\beta}{2P\cos\alpha}\cdot M_d - K_{eq}\cdot\Delta\phi\right) \quad (7)$$

This model reveals the physical essence of soft grasping: there exists a minimum actuation torque threshold. Only when the rigid driving term exceeds the flexible deformation dissipation term ($K_{eq}\cdot\Delta\phi$) can the gripper output a positive clamping force.

2.2.6. Grasping Stability and Anti-Slip Critical Conditions

In railway maintenance scenarios, the core constraint is to ensure that heavy maintenance tools do not slip during high-speed transfer. The force analysis is shown in Figure 14: the workpiece is subjected to the combined action of lifting forces and friction forces from four fingers.

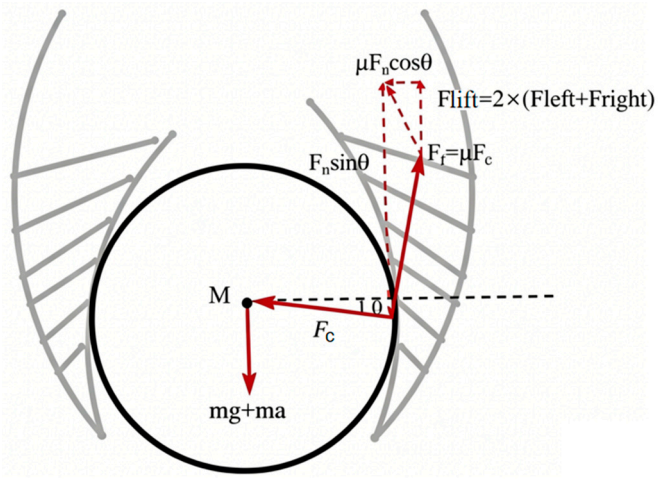


Figure 14. Stability analysis. The red arrow F_c represents the normal clamping force from a finger, the curved arrow M indicates the bending moment, and dashed lines denote the effective moment arms.

Let the workpiece mass be m and the maximum operational acceleration be a_{max}. The vertical lifting force component provided by a single finger consists of a normal clamping force component (F_c) and a tangential friction force component (F_f). To prevent the workpiece from slipping, the total lifting force of the system must overcome both gravity and inertial force; the stability criterion is:

$$4F_c(\sin\theta+\mu\cos\theta) \geq m(g+a_{max}) \quad (8)$$

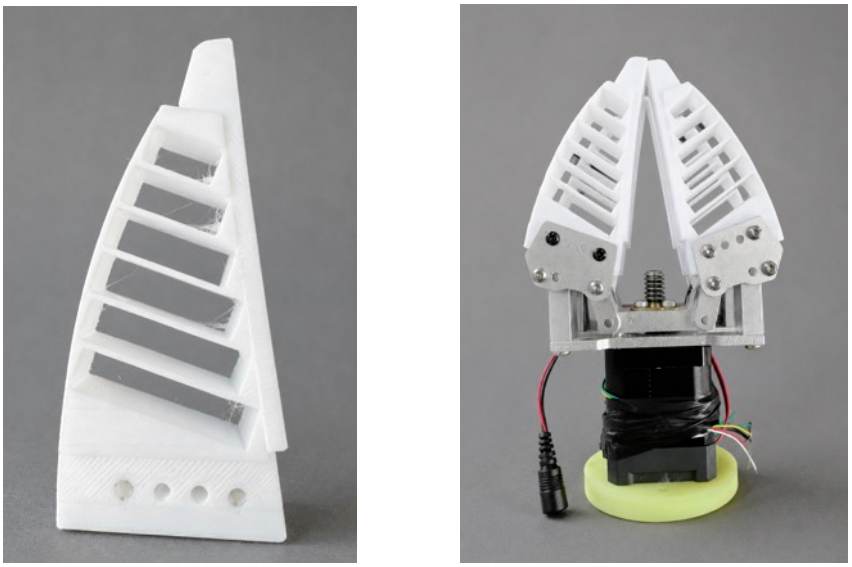
Substituting the contact force model (Equation (7)) into the stability criterion, the minimum safe driving torque required to grasp a workpiece of mass m can be derived M_{d_min}:

$$M_{d_min} \geq \frac{1}{K_{trans}}\left[\frac{m(g+a_{max})\cdot L_{eff}}{4(\sin\theta+\mu\cos\theta)} + K_{eq}\cdot\Delta\phi\right] \quad (9)$$

where $K_{trans} = \frac{\pi\eta L_1 \sin\beta}{2P\cos\alpha}$ is the torque amplification factor of the transmission mechanism. This inequality defines the lower bound for motor selection: the larger the enveloping angle θ, the greater the required deformation compensation term $K_{eq}\Delta\phi$, which requires the drive system to have sufficient torque reserve to provide adequate normal clamping force while ensuring large-deformation enveloping.

2.2.7. Fabrication and Integration of the Soft Gripper Prototype

To validate the theoretical model and simulation results, the asymmetric Fin Ray soft fingers were fabricated and the gripper was assembled based on the optimized topological parameters. The soft finger substrate was printed from BASF Elastollan® 1185A thermoplastic polyurethane (TPU) filament (BASF Polyurethanes GmbH, Lemförde, Germany) using fused deposition modeling (FDM) technology [13], as shown in Figure 15a. The infill density was set to 100% during printing to ensure the physical density of the finger pad under load and consistency with the constitutive model. To further enhance interfacial friction during grasping, a 2 mm-thick high-friction silicone coating was composited and cured onto the contact surface of the finger pad.



(**a**) (**b**)

Figure 15. Soft gripper prototype and integrated assembly: (**a**) a single 3D-printed soft finger; (**b**) overall assembly of the soft gripper system.

As shown in Figure 15b, the drive module employs an integrated stepper servo motor rigidly connected to the lead-screw-dual-crossbeam transmission mechanism via a flange. The 3D-printed flexible fingers are fixed to the aluminum alloy end links via high-strength pins, forming a complete rigid-flexible coupled end-effector. The total weight of the assembled soft gripper prototype is kept below 1 kg, satisfying the payload requirement of the Aubo-i5 collaborative robotic arm.

2.2.8. Contact Force Measurement and Preliminary Cyclic Durability Test

To measure the fingertip contact force of the fabricated gripper, a thin-film force sensor was placed between the soft fingertip and a rigid contact plate. The conditioned sensor signal was monitored using a dynamic signal analyzer. The hardware components used for contact-force acquisition are shown in Figure 16.

Before the contact force test, the thin-film force sensor was calibrated by applying a series of known weights vertically onto the rigid contact plate. The corresponding applied normal force was calculated as:

$$F = mg$$

where m is the applied mass and g is the gravitational acceleration. The calibration data are listed in Table 7.



(**a**) (**b**)

Figure 16. Hardware components for fingertip contact force measurement: (**a**) thin-film force sensor; (**b**) dynamic signal analyzer.

Table 7. Calibration data of the thin-film force sensor.

Test No.	Applied Mass (g)	Applied Force (N)	Conditioned Sensor Response (V)	Fitted Force (N)	Relative Error (%)
1	100	0.98	0.36	0.97	1.49
2	200	1.96	0.51	2.08	5.96
3	500	4.91	0.89	4.90	0.21
4	1000	9.81	1.56	9.86	0.51
5	1500	14.72	2.16	14.30	2.86
6	2000	19.62	2.91	19.85	1.18

As shown in Figure 17, the sensor response monotonically with the applied normal force. Within the tested force range, the relationship between the applied normal force and the sensor response was fitted as follows:

$$F = 7.41V - 1.70$$

where F is the applied normal force and V is the sensor response. The coefficient of determination reached $R^2 = 0.999$, indicating that the calibrated sensor provided sufficient linearity and measurement accuracy for subsequent fingertip contact force tests.

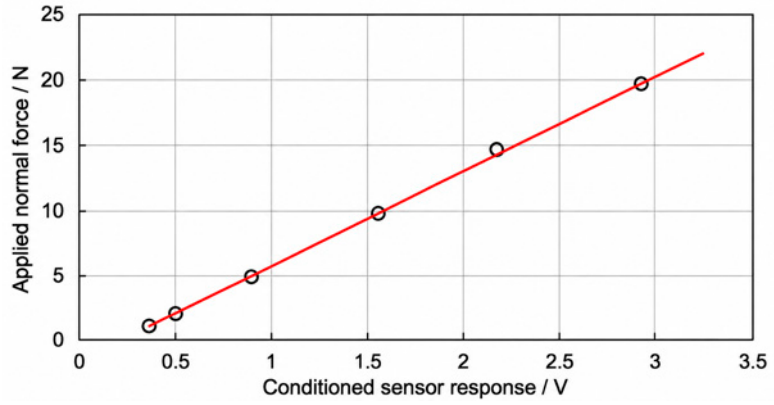


Figure 17. Calibration curve of the thin-film force sensor.

After calibration, the force-measurement system was used for the cyclic inward-curling test of the optimized soft finger with a rib inclination angle of 30° and an outer wall

thickness of 3 mm. The finger was driven to repeatedly curl inward and contact the rigid plate under the same actuation condition used for the fingertip contact-force measurement.

Each cycle consisted of inward curling, fingertip contact, force holding, and release. The test was performed for 3000 cycles, and the fingertip contact force was recorded every 500 cycles. After testing, the finger was inspected for cracks, permanent deformation, and local damage. The contact-force degradation ratio was calculated as:

$$\eta = \frac{F_0 - F_N}{F_0} \times 100\% \tag{10}$$

where F_0 and F_N are the initial contact force and the contact force after N cycles, respectively.

To preliminarily evaluate the durability of the assembled gripper under the rated load, a 2.0 kg payload was repeatedly grasped using a fixed robotic-arm trajectory. Each cycle consisted of gripper opening, payload grasping, lifting to 20 cm, holding for 5 s, lowering, and releasing. The test was repeated for 500 cycles. During and after the test, payload dropping, visible slippage, cracks, permanent deformation, surface wear, and local damage of the soft fingers were inspected.

2.2.9. Linear Fingertip Stiffness Identification

The linear fingertip stiffness of the asymmetric soft finger was computed from the measured force–displacement response using the following formula:

$$K_f = \frac{\Delta F}{\Delta \delta}$$

where K_f is the experimentally identified linear fingertip stiffness, ΔF is the change in measured normal contact force, and $\Delta\delta$ is the corresponding displacement increment. For each displacement, the fingertip was actuated incrementally, and the normal contact force was measured five times. The average value of these measurements was used to calculate the linear fingertip stiffness.

2.2.10. Anti-Slip Stability Evaluation Method

To evaluate the load-holding capability of the soft gripper during lifting, the available holding force was estimated based on the measured normal contact force and the interfacial friction coefficient. Considering that the vertical support provided by finger geometry varies with the actual contact posture, a friction-dominated holding model was adopted for experimental verification. The available holding force of the four-finger grasp was calculated as:

$$F_{available} = 4\mu F_n$$

where μ is the interfacial friction coefficient between the finger pad and the tool surface, and F_n is the measured normal contact force generated by a single finger.

The required holding force under lifting conditions was defined as:

$$F_{required} = m(g + a_{max})$$

where m is the mass of the grasped object and a_{max} is the maximum operational acceleration during transfer. The friction coefficient μ was estimated from preliminary sliding tests between the finger pad and the tool surface, and a_{max} was set to 1.0 m/s^2 according to the controlled transfer conditions.

Stable grasping was considered to be satisfied when $F_{available}$ exceeded $F_{required}$. This verification provides an experimental basis for evaluating whether the measured fingertip normal force can support stable tool lifting under the prescribed transfer condition.

2.3. Visual Perception System

Railway inspection operations are typically conducted in outdoor unstructured environments, facing the dual challenges of strong background noise interference and multi-scale feature suppression compared with standard industrial assembly lines. High-frequency ballast textures can easily cause false activations in convolutional neural networks, while reflections and shadows from metal tools under natural lighting further increase the difficulty of feature extraction.

To address these issues, a robust attention-enhanced network, RA-YOLO, is proposed to improve RGB-based tool detection under complex ballast backgrounds. In addition, a hand-eye localization model mapping from 2D pixel space and depth information to the robotic arm base frame is established to support subsequent grasp pose estimation.

2.3.1. RA-YOLO Network Architecture

The overall architecture of RA-YOLO is shown in Figure 18. Compared with the original YOLOv8n [33], the key improvement in this work is introduced at the junction between the Backbone and the Neck, aiming to enhance the ability of the feature extraction network to focus on critical target regions while suppressing the propagation of background noise.

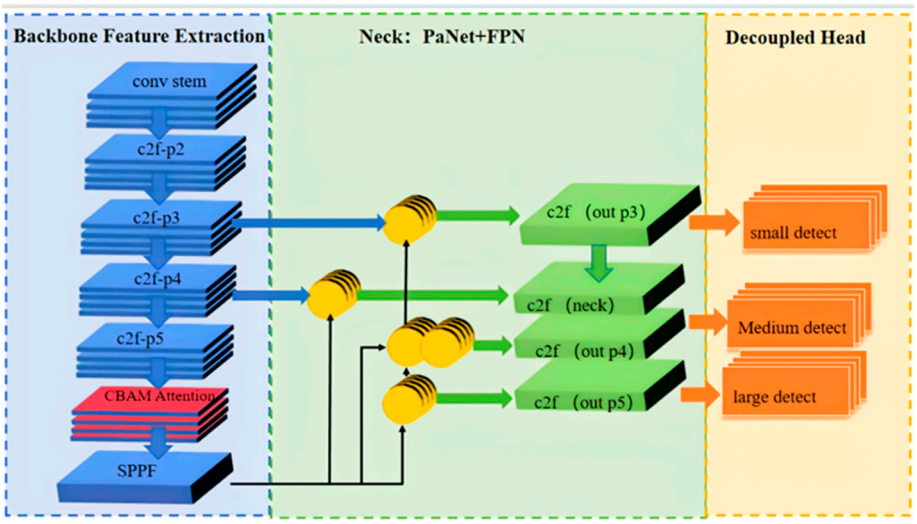


Figure 18. Overall architecture of RA-YOLO.

2.3.2. Embedding of the CBAM Dual-Attention Mechanism

To address the problem that metal workpieces have weak texture features against complex gravel backgrounds and that convolution operations struggle to capture global contextual information effectively, the convolutional block attention module (CBAM [34]) embedded between the C2f-P5 output layer of the backbone network and the SPPF module. This module adopts a channel-first, spatial-second serial mechanism to adaptively recalibrate feature maps, thereby enhancing the model perception of critical regions, as shown in Figure 19.

Specifically, CBAM first applies the channel attention sub-module, which aggregates spatial information via global average pooling and global max pooling, and learns channel-wise weights through a multi-layer perceptron to strengthen feature channels sensitive to metallic materials and tool structures while suppressing channels that respond to background gravel. Building on this, the spatial attention sub-module pools the recalibrated features along the channel dimension to generate a 2D spatial saliency mask, guiding the network to focus on the main body of the workpiece and thereby filtering out peripheral ballast noise in the spatial dimension.

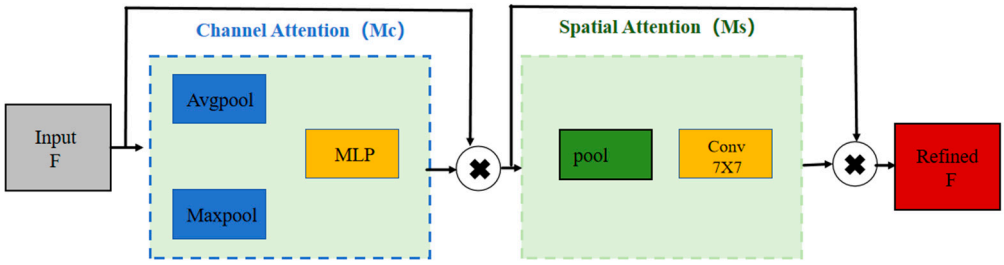
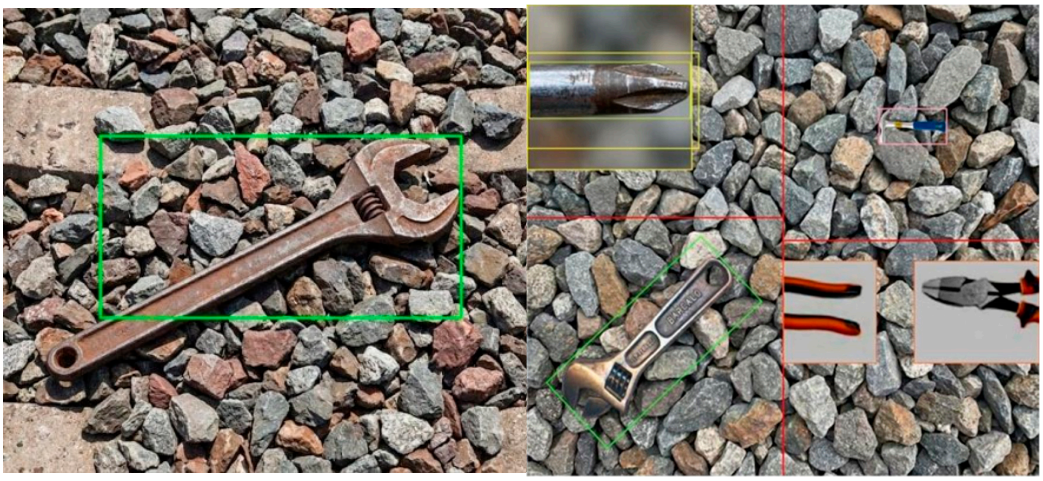


Figure 19. Internal mechanism of the CBAM module.

This attention mechanism mimics the focusing characteristics of the human visual system, significantly improving feature robustness under complex illumination, occlusion, and background interference with negligible additional computational overhead.

2.3.3. Mosaic Data Augmentation Strategy

To address the problems of single viewpoint, lack of scale variation, and insufficient occlusion diversity in publicly available railway maintenance tool datasets, and to improve the model generalization capability in real complex scenarios, the Mosaic data augmentation strategy is introduced [35] during training. In each training iteration, four original images are randomly selected, individually subjected to random scaling and cropping, and then stitched into a single composite training sample, as shown in Figure 20. This operation not only significantly increases the background complexity and diversity of training samples by fusing the background information of multiple images—which forces the network to better distinguish foreground tools from ballast backgrounds—but also effectively increases the pixel proportion of small-scale targets in the training set through random scaling, thereby enhancing the model robustness to multi-scale targets and partial occlusion.



(**a**) (**b**)

Figure 20. Examples of augmented training samples: (**a**) original single sample; (**b**) training sample after Mosaic stitching augmentation. Colored boxes denote annotated tool targets.

2.3.4. Dataset Construction and Partitioning

The scale and diversity of the dataset directly determine the convergence speed and generalization capability of the detection network. A dedicated railway maintenance tool dataset containing 3640 high-resolution samples was constructed in this study. To simulate real outdoor working conditions, three typical lighting conditions were covered during image acquisition: strong direct sunlight (1210 images), overcast/cloudy conditions (1220 images), and shadow occlusion conditions (1210 images). This multi-dimensional environmental distribution ensures stable feature recognition of the RA-YOLO model under

different weather conditions. The dataset was randomly partitioned in an 8:1:1 ratio. The training set 2912 images, providing sufficient parameter space for model weight optimization. The validation set and test set each contained 364 images, used for hyperparameter tuning and objective performance evaluation, respectively. The sample distribution across the five categories of maintenance tools is shown in Figure 21.

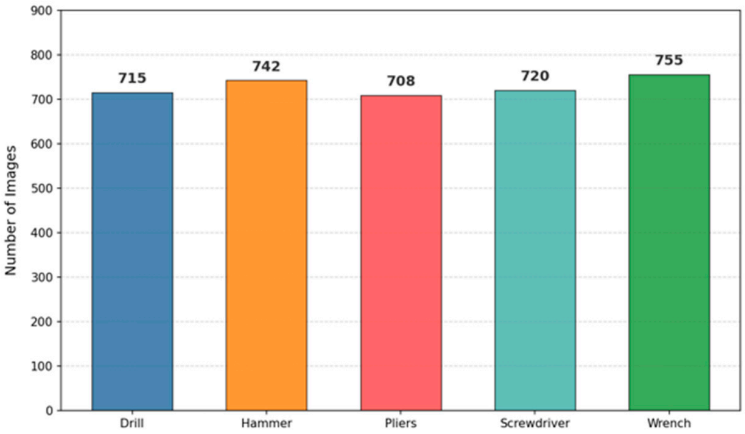


Figure 21. Sample count distribution of each category in the self-built railway tool dataset.

The detailed subset partitioning statistics are presented in Table 8. The total number of samples for each tool category maintained above 700, and this balanced distribution effectively class bias during training. In addition, image rotation and scaling augmentation were applied during training, further increasing the effective sample size and enhancing the system robustness to pose variations of irregular tools.

Table 8. Detailed scale and subset partitioning statistics of the railway tool dataset.

Tool Category	Training Set	Validation Set	Test Set	Total Samples
Drill	572	71	72	715
Hammer	594	74	74	742
Pliers	566	71	71	708
Screwdriver	576	72	72	720
Wrench	604	76	75	755
Total	2912	364	364	3640

To further evaluate the environmental generalization capability of RA-YOLO, an additional environmental robustness test set was constructed and used only for testing, without participating in model training or validation. The supplementary test set included images captured under different ballast background and moisture conditions, including dry ballast, wet ballast, and ballast with different particle color/texture. The trained RA-YOLO model was directly evaluated on this supplementary test set to assess its robustness to representative ballast moisture and texture variations, rather than exhaustive extreme-weather conditions.

2.4. Collaborative Control and Operation Planning Based on Simulink and ROS

Railway ballast environments exhibit pronounced unstructured characteristics, and traditional low-level code-centric development approaches often suffer from long development cycles and difficult subsystem integration. To cope with complex and variable field conditions, an intelligent collaborative control system based on Simulink and ROS constructed in this study [18,25,34,35]. As shown in Figure 22, the system integrates a visual recognition module, a soft gripper control module, and a robotic arm motion con-

trol module, realizing a coherent full-process automated operation pipeline from target approach and compliant grasping to tool collection.

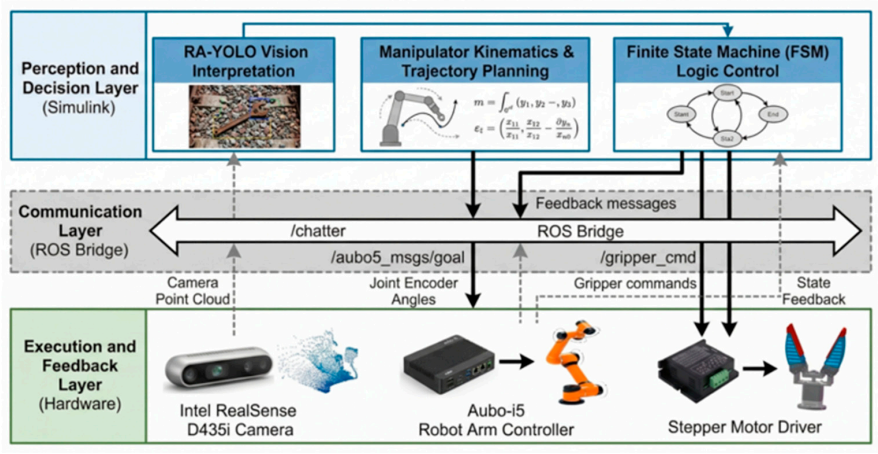


Figure 22. Overall collaborative control architecture of the system. Arrows denote the flow of control commands, sensor data, and feedback signals between the visual recognition, soft gripper, and robotic arm modules.

2.4.1. Hardware Collaboration and Low-Level Communication Mechanism

The stable operation of the system relies heavily on close collaboration and efficient communication among the underlying hardware components. In terms of hardware architecture, the system uses an industrial PC as the core control unit, combined with a driver board, a USB communication module, and a power supply module to form the complete low-level support infrastructure. The power supply module provides the necessary electrical support for the entire system; the industrial PC issues high-speed commands to and monitors the status of the Aubo-i5 robotic arm and the soft gripper via the driver board; and the depth camera works in close coordination with the actuators to ensure precise task execution, as shown in Figure 23.

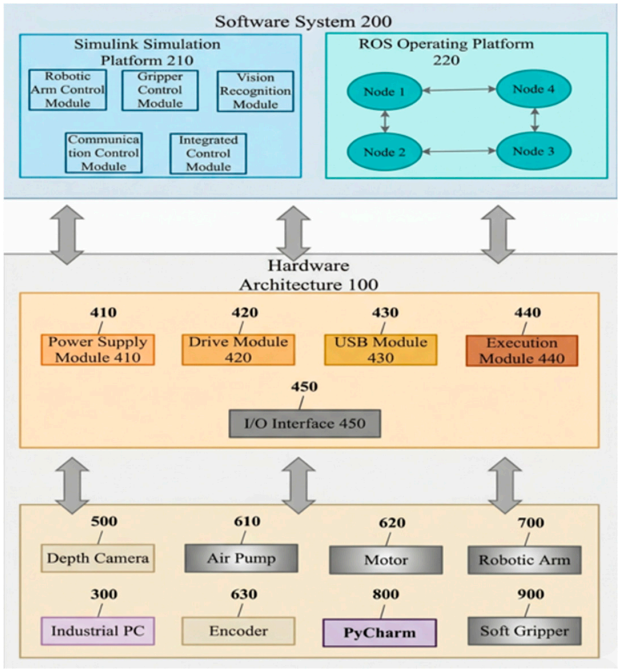


Figure 23. Low-level hardware architecture and communication topology of the robotic collaborative control system. Arrows indicate communication and feedback links among modules.

2.4.2. Operation Logic and Exception Handling Based on a Finite-State Machine

For the control of the end-effector, the opening and closing of the asymmetric soft gripper strictly depend on the high/low logic level states of the I/O ports of the gripper controller. To this end, a dedicated message format established in the ROS environment to implement the gripper control module, as shown in Figure 24.

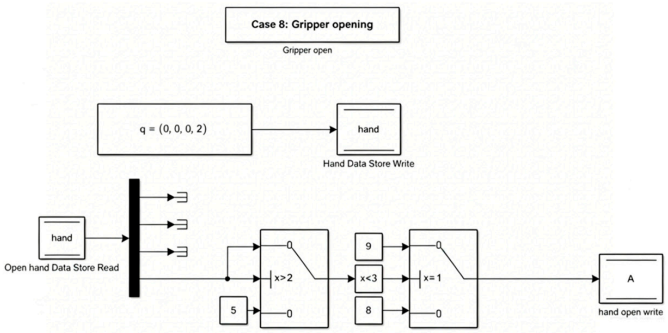


Figure 24. Gripper state control model based on the geometry_msgs/Quaternion message format.

The system uses the geometry_msgs/Quaternion message format to precisely control the state of the soft gripper. Within the Simulink control model, the system receives the specified quaternion values through dedicated module nodes. Control signals are exchanged at high frequency between the low-level terminal and the Simulink system via a designated ROS topic, converting logical commands into physical voltage levels to precisely drive the stepper motor to open or close the soft fingers, ensuring stable physical enveloping of irregular tools.

2.4.3. FSM-Based Operation Logic and Full-Process Automation

A finite-state machine (FSM) constructed in this study to realize full-process automated control of adaptive grasping of railway maintenance tools, as shown in Figure 25. The system operation logic strictly follows the closed-loop perception decision execution framework. In the S0 initialization stage, the robotic arm resets and drives the gripper to perform an open-close self-check to confirm that the hardware feedback loop is functioning normally. The system then enters the visual perception stage, where the D435i depth camera acquires real-time RGB-D images and the RA-YOLO detection network is invoked to identify target tools in complex ballast backgrounds.

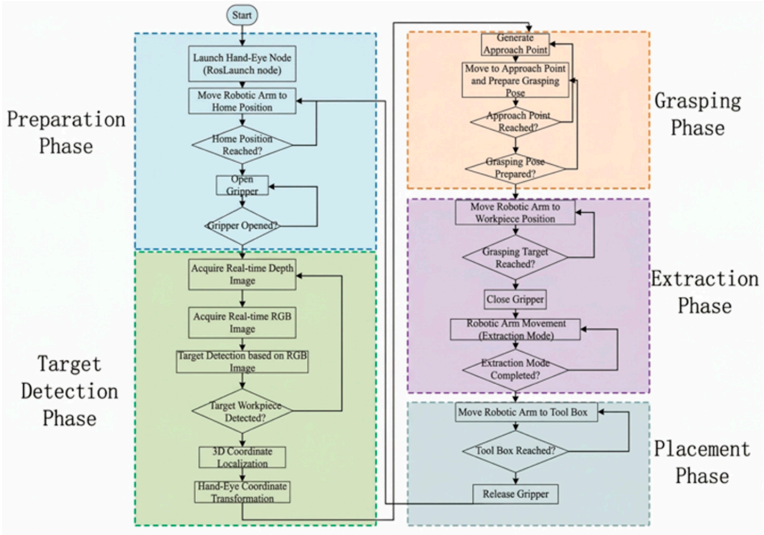


Figure 25. FSM-based control flowchart for maintenance tool sorting.

In the decision-making and execution stage, the end-effector is guided from a safe approach point to reach the target position with high precision. Finally, the stepper motor drives the asymmetric soft fingers to close, leveraging the passive deformation characteristics of the TPU material to achieve compliant enveloping of heterogeneous workpieces. Combined with the lateral extraction motion of the robotic arm, the tool is safely extracted from the ballast background and transferred to the collection area, completing a single operation cycle.

2.4.4. Model-Based Design Integration and Automatic Code Generation

The entire system achieves deep closed-loop integration of the visual recognition system, the soft gripper grasping system, and the robotic arm motion control system through a model-based design (MBD) pathway. This Simulink-ROS co-development approach not only achieves efficient decoupling of perception, decision-making, and execution at the logical level, but also fundamentally resolves the problems of difficult subsystem integration and tedious low-level code debugging that are inherent in traditional grasping robots, significantly shortening the development cycle. This collaborative control strategy ensures that complex sorting logic can be faithfully reproduced and stably executed with high precision in actual railway ballast scenarios.

3. Results

3.1. Visual Detection Experimental Results

3.1.1. Experimental Setup

To validate the effectiveness of RA-YOLO, experiments were conducted based on the PyTorch 2.0.1 framework. The hardware platform was configured with an Intel Core i9 processor and an NVIDIA RTX 3090 GPU. Training parameters were set as follows: SGD optimizer, initial learning rate of 0.01, momentum of 0.937, input size of 640 × 640, and 120 training epochs.

3.1.2. Convergence Analysis

The evolution curves of mean average precision (mAP@0.5) during training for RA-YOLO and the baseline model YOLOv8n are shown in Figure 26.

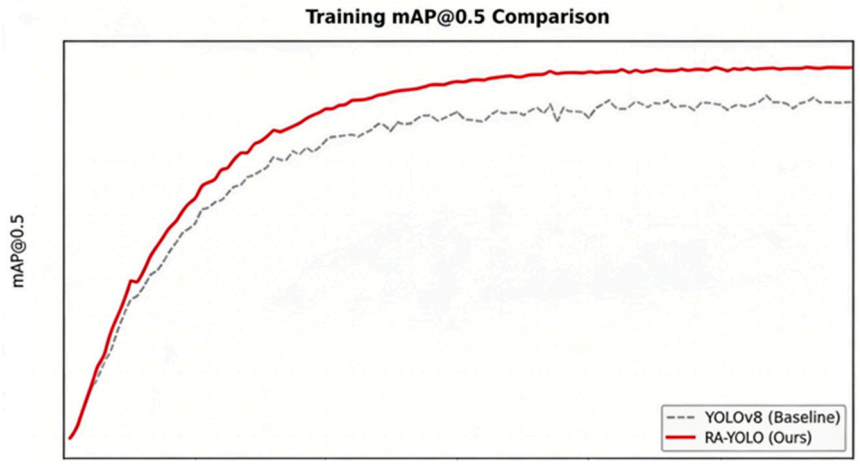


Figure 26. Comparison of mAP@0.5 training curves between the improved model and the baseline model.

In the early training phase, both models exhibited rapid accuracy improvement. As training progresses, the accuracy growth of YOLOv8n gradually and noticeable fluctuations, reflecting instability in feature extraction under complex ballast texture backgrounds. In contrast, RA-YOLO, which incorporates the CBAM attention mechanism, a

smoother convergence trend throughout the mid-to-late training stages, and ultimately higher stable accuracy.

This result indicates that the attention mechanism effectively suppresses interference from irrelevant backgrounds by enhancing the specificity of feature selection, thereby improving the model feature robustness and convergence efficiency in complex environments.

To further quantify the detection robustness of the model for maintenance tools with different geometric shapes, the precisionrecall (PR) curves of RA-YOLO on the test set analyzed, as shown in Figure 27.

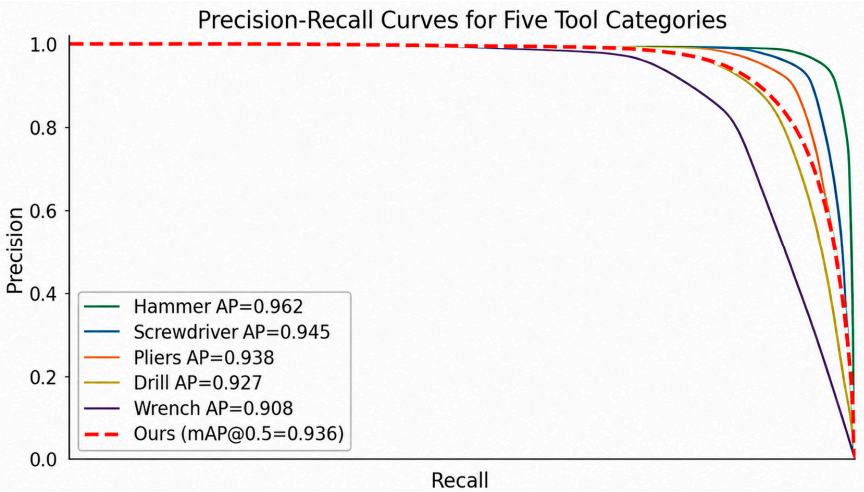


Figure 27. PR curves for the five tool categories.

3.1.3. Ablation Study and Attention-Module Selection

To further verify the individual contributions of the Mosaic data augmentation strategy and the CBAM attention mechanism to the railway inspection tool detection task and validate the rationality of the RA-YOLO improvement scheme, an ablation study was designed for quantitative analysis.

As shown in Table 9, introducing Mosaic augmentation mAP@0.5 by 2.4%, effectively enhancing the model robustness to complex backgrounds. Embedding CBAM slightly FPS to 105 but a 3.1% improvement in mAP@0.5 and the model's ability to focus on tool-related regions under ballast background interference. Ultimately, Group D, which integrates both strategies, the highest mAP@0.5 of 93.6%, representing a substantial 8.2 percentage-point improvement over the baseline network. This fully validates the positive synergistic effect of data diversity enhancement and feature recalibration on performance improvement.

Table 9. Ablation study results for different improvement strategies.

Group	Mosaic	CBAM	mAP@0.5 (%)	mAP@0.5:0.95 (%)	FPS
A	×	×	85.4	64.2	122
B	√	×	87.8	66.1	122
C	×	√	88.5	66.7	105
D (Ours)	√	√	93.6	68.7	105

Note: "×" denotes that the corresponding improvement strategy was not applied, whereas "√" denotes that the strategy was applied.

To further evaluate the rationale for selecting CBAM in RA-YOLO, several lightweight attention mechanisms, including SE [36], ECA [37], coordinate attention (CA [38]), and CBAM, were compared under identical training and inference settings. All attention modules were inserted between the C2f-P5 output layer and the SPPF module, while

Mosaic data augmentation was consistently applied across all models. Table 10 presents the comparison results of the different lightweight attention mechanisms.

Table 10. Comparison of different lightweight attention mechanisms.

Method	Attention Module	Params/M	GFLOPs	mAP@0.5 (%)	mAP@0.5:0.95 (%)
YOLOv8n	None	3.20	8.7	87.8	66.1
YOLOv8n + SE	SE	3.28	8.8	89.6	66.8
YOLOv8n + ECA	ECA	3.22	8.7	90.3	67.2
YOLOv8n + CA	Coordinate Attention	3.34	9.0	91.1	67.7
YOLOv8n + CBAM	CBAM	3.40	9.2	93.6	68.7

As shown in Table 10 and Figure 28, all lightweight attention mechanisms improved detection accuracy. Among them, CBAM achieved the highest detection accuracy, with mAP@0.5 and mAP@0.5:0.95 reaching 93.6% and 68.7%, respectively. Although the inference speed decreased slightly, the model still maintained real-time detection capability. CBAM provided the highest mAP@0.5 and mAP@0.5:0.95 among the tested modules while retaining real-time inference; it was therefore adopted in RA-YOLO.

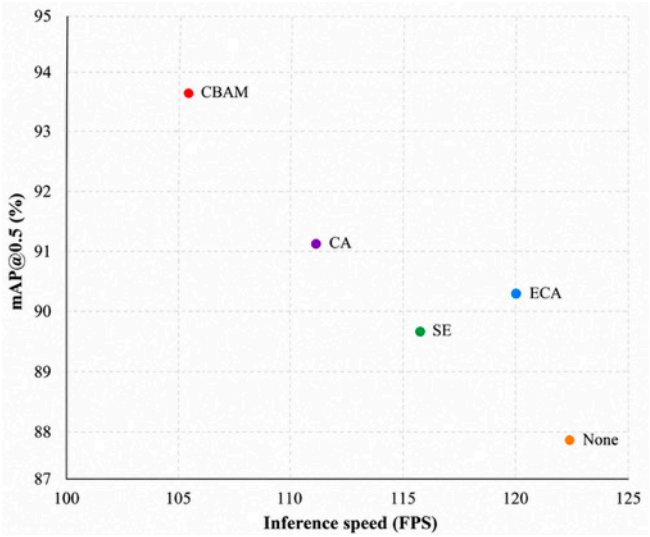


Figure 28. Accuracy–efficiency trade-off of different lightweight attention mechanisms.

3.1.4. Algorithm Performance Comparison and Multi-Platform Deployment Validation

To verify the advancement of the algorithm, RA-YOLO was compared with existing railway object detection algorithms (MACE-Net, SE-YOLOv5, Railway-YOLOv8s) [4–6], with results shown in Table 11. Meanwhile, to intuitively demonstrate the balance between detection accuracy and lightweight characteristics of each model, an algorithm performance balance scatter plot is presented in Figure 29.

Table 11. Algorithm performance comparison results.

Method	mAP@0.5 (%)	FPS	Params/M
MACE-Net	83.8	98	4.5
SE-YOLOv5	86.5	92	7.5
Railway-YOLOv8s	87.0	85	11.1
YOLOv8n	85.4	122	3.2
Ours	93.6	105	3.4

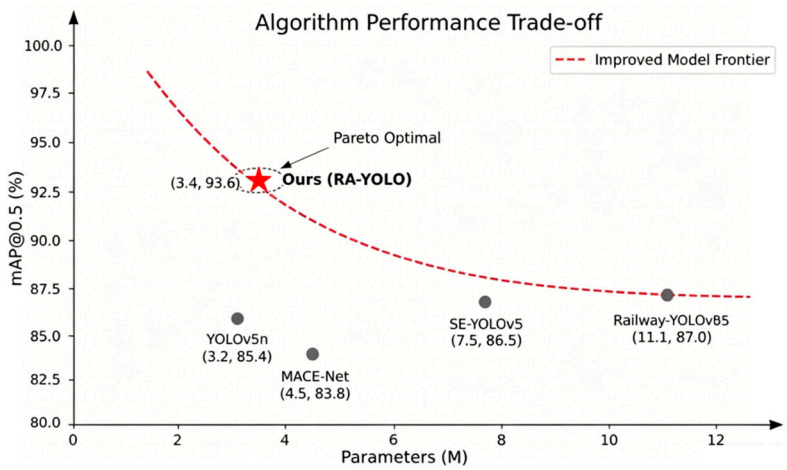


Figure 29. Algorithm performance balance scatter plot.

According to Table 11 and Figure 29, RA-YOLO outstanding detection accuracy, with mAP@0.5 reaching 93.6%, representing improvements of 2.4% and 5.1% over SE-YOLOv5 and MACE-Net, respectively [4]. Even compared with Railway-YOLOv8s, which approximately three times the number of parameters, RA-YOLO a 1.9% accuracy advantage. In terms of model efficiency, RA-YOLO only 3.4 M parameters and an inference speed of 105 FPS, exhibiting excellent lightweight characteristics.

To further validate the deployment capability of the RA-YOLO algorithm on actual inspection robots, the inference performance of the model was compared between server-side and mobile terminal platforms. The target deployment hardware was an NVIDIA Jetson AGX Orin edge computing platform (NVIDIA Corporation, Santa Clara, CA, USA). With TensorRT acceleration enabled and an input resolution of 640 × 640, RA-YOLO achieved an average inference speed of 42.0 FPS on this terminal. Although this represents a decrease compared with the 105 FPS achieved on a laboratory server equipped with an NVIDIA GeForce RTX 3090 GPU (NVIDIA Corporation, Santa Clara, CA, USA), the end-to-end processing latency remained consistently below 25 ms. Performance comparison details across different hardware platforms are shown in Figure 30. This result demonstrates that the improved algorithm can fully meet the real-time visual perception performance requirements of railway maintenance field operations.

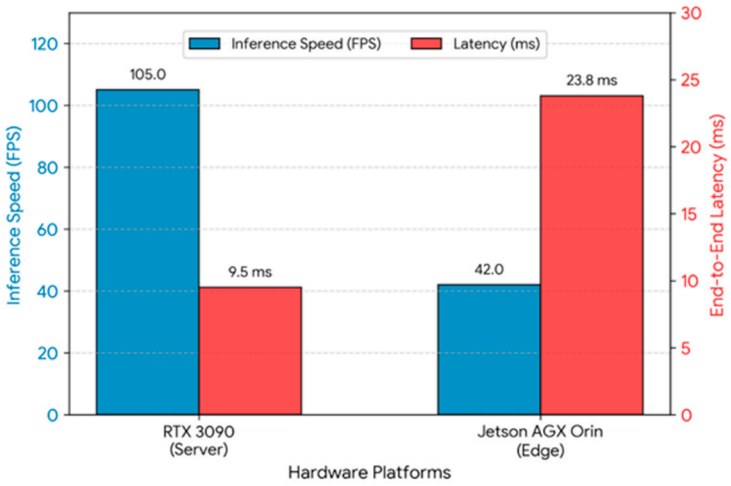
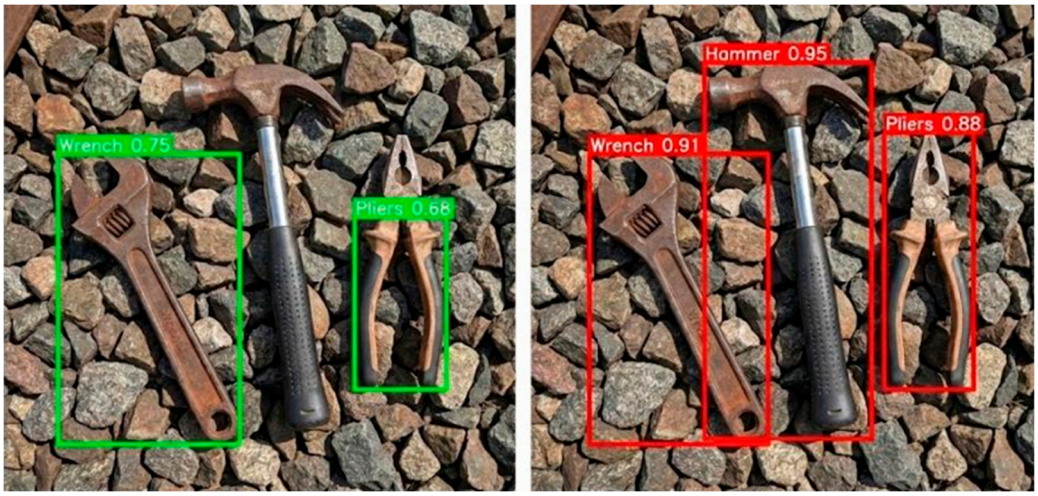


Figure 30. Performance comparison of RA-YOLO inference speed across different hardware platforms.

3.1.5. Detection Performance Under Complex Backgrounds and Environmental Variations

To intuitively verify the robustness of the improved model in actual railway environments, Figure 31 presents a comparison of detection results between the two models under typical ballast backgrounds.



(**a**) (**b**)

Figure 31. Detection performance comparison under typical railway ballast backgrounds: (**a**) original YOLOv8n results marked in green; (**b**) improved RA-YOLO results marked in red. The numbers denote confidence scores.

As shown in Figure 31a, in environments with complex background textures, the original YOLOv8n model is susceptible to gravel noise interference, resulting in missed detections. This occurs because in deep feature maps, small target features are easily overwhelmed by background clutter. In contrast, as shown in Figure 31b, the improved RA-YOLO model successfully all target tools, and the predicted bounding boxes fit the actual object contours more closely. By comparing confidence scores, it observed that the improved model generally higher confidence for easily confused targets. This visualization result confirms that the spatial attention mechanism in the CBAM module can effectively suppress activation values in background regions, enabling the network to focus on the tool body itself, thereby resolving the missed detection problem under complex backgrounds.

Based on the supplementary environmental robustness test set described in Section 2.3.4, the trained RA-YOLO model was evaluated under different ballast background and moisture conditions. The results are shown in Table 12.

Table 12. Detection performance of RA-YOLO under different ballast moisture and texture conditions.

Test Condition	Number of Images	mAP@0.5 (%)	mAP@0.5:0.95 (%)	Precision (%)	Recall (%)
Dry ballast condition	100	93.6	79.2	94.1	92.8
Wet ballast	80	92.5	78.1	93.0	91.5
Different ballast texture/color	80	91.8	77.5	92.2	90.8
Wet + different ballast	60	90.5	76.0	91.0	90.2
Mean	—	92.1	77.7	92.6	91.3

Table 12 summarizes the detection performance of RA-YOLO under varying ballast background and moisture conditions. The model achieved a lowest mAP@0.5 of 90.5% under the combined wet and different ballast condition, compared with 93.6% under the dry ballast condition. Across the supplementary test conditions, RA-YOLO maintained

a mAP@0.5 above 90.5%, while precision and recall remained above 90%. These results indicate that the model retained stable detection performance under the tested ballast-texture and moisture variations.

3.2. Localization Accuracy Experiments

3.2.1. Calibration Experiment

Based on the checkerboard calibration method proposed by Zhang [39], 15 sets of images at different poses were collected using a 10×7 checkerboard calibration board, as shown in Figure 32, to establish the hand-eye coordinate system mapping relationship.

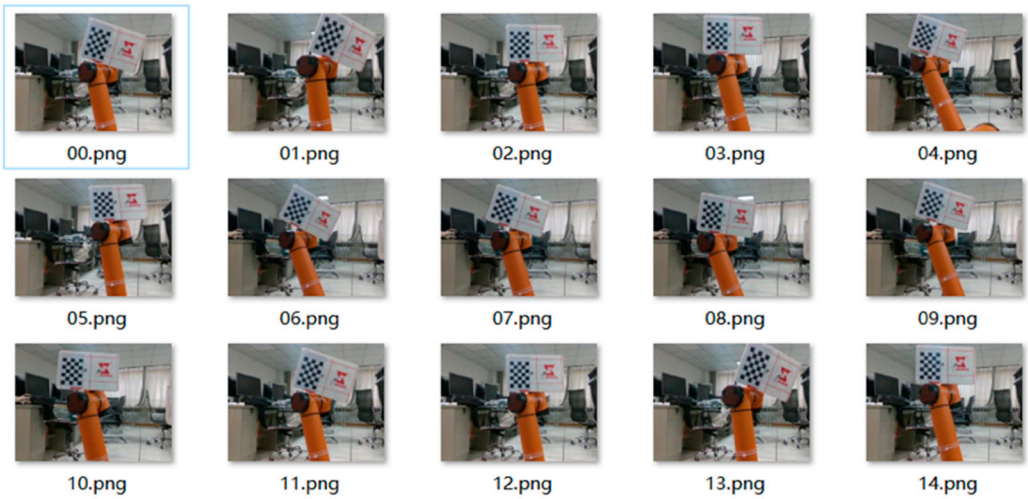


Figure 32. Multi-pose image acquisition during hand-eye calibration.

The solved camera intrinsic matrix and distortion coefficients are as follows:

$$\mathrm{K} = \begin{bmatrix} 637.504 & 0 & 318.102 \\ 0 & 637.963 & 214.713 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\mathrm{D} = [0.075,\ 1.728, -0.027, -0.001, -11.443]$$

By further combining the end-effector pose data of the robotic arm, the transformation matrix from the camera optical center to the robotic arm base frame is computed as follows:

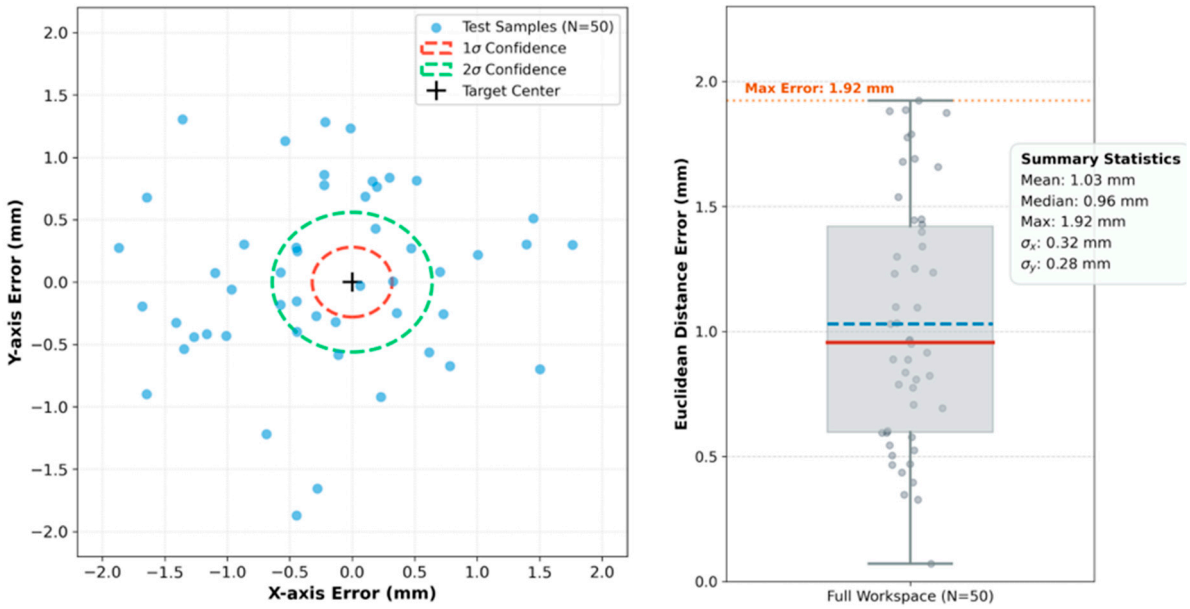
$$\begin{bmatrix} -0.970 & 0.038 & 0.239 & 303.616 \\ -0.242 & -0.084 & -0.967 & 360.975 \\ 0.016 & -0.996 & 0.090 & 373.104 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

This calibration result establishes the spatial geometric relationship between the vision system and the execution mechanism, serving as the prerequisite for achieving precise grasping.

3.2.2. Localization Accuracy Validation

To further verify the geometric localization accuracy of the vision-guided hand-eye calibration model, 50 random test points were selected within the workspace. By comparing the target positions estimated by the vision system with the physical coordinates touched by the robotic arm end-effector probe, the localization deviation distribution within the calibrated workspace was obtained.

The spatial error distribution shown in Figure 33a central clustering characteristics in the X-Y plane, with axial standard deviations of 0.32 mm and 0.28 mm, respectively. The compact distribution of data points and the confidence ellipse indicate good geometric consistency of the hand-eye calibration model within the tested workspace.



(**a**) (**b**)

Figure 33. Multi-dimensional statistical analysis of absolute localization error: (**a**) spatial error distribution on the X-Y plane; (**b**) Boxplot of localization errors; red line = median, blue dashed line = mean.

As shown in Figure 33b, the localization errors concentrated around an average Euclidean distance of 1.03 mm. The component-wise maximum absolute error below 2.5 mm, as further quantified in Table 13. These results verify the millimeter-level localization accuracy of the calibrated vision-guided hand-eye localization model under quasi-static validation conditions and provide a reliable basis for subsequent grasping experiments.

Table 13. Quantified absolute error data between vision-computed coordinates and actual robotic arm coordinates.

Sample	Vision Coords (x, y, z)	Actual Coords (x, y, z)	Absolute Error (Δx, Δy, Δz)
1	(221.5, −85.2, 10.5)	(220.1, −86.5, 9.8)	(1.4, 1.3, 0.7)
2	(356.8, 107.3, 11.2)	(355.2, 106.1, 10.4)	(1.6, 1.2, 0.8)
3	(425.6, 121.5, 10.8)	(423.8, 119.8, 9.9)	(1.8, 1.7, 0.9)
4	(510.2, −55.4, 11.0)	(508.1, −53.9, 10.1)	(2.1, 1.5, 0.9)
5	(573.6, 132.1, 10.6)	(571.2, 130.5, 9.5)	(2.4, 1.6, 1.1)

Further quantified error data are presented in Table 13.

Experimental data show that the average absolute errors in the X and Y directions controlled within 2.0 mm, the maximum absolute error 2.4 mm, and the Z-direction error approximately 1.0 mm. These errors are within the passive compliance range of the soft gripper, supporting subsequent system-level grasping tests.

Because localization was evaluated under quasi-static conditions, the reported errors mainly characterize the geometric accuracy of the handeye calibration model. Dynamic effects, including image acquisition latency, robotic-arm vibration, depth fluctuation, and

illumination-induced point-cloud sparsity, evaluated indirectly through the system-level grasping tests in Section 3.4.

3.3. Mechanical Validation of the Soft Gripper

To characterize the force-output behavior and anti-slip performance of the fabricated soft gripper, component-level experiments were conducted on the fabricated fingers. This section presents the equivalent contact-force calibration, force-output repeatability, cyclic inward-curling stability, and anti-slip performance for typical railway maintenance tools.

3.3.1. Force-Output Characterization of the Soft Finger

The force-output characteristics of the fabricated soft finger were evaluated using the calibrated force sensor at different actuation displacements. As the actuation displacement increased from 1 mm to 6 mm, the measured normal contact force increased from 2.74 N to 15.47 N, as summarized in Table 14.

Table 14. Experimental identification of linear fingertip stiffness and contact force validation.

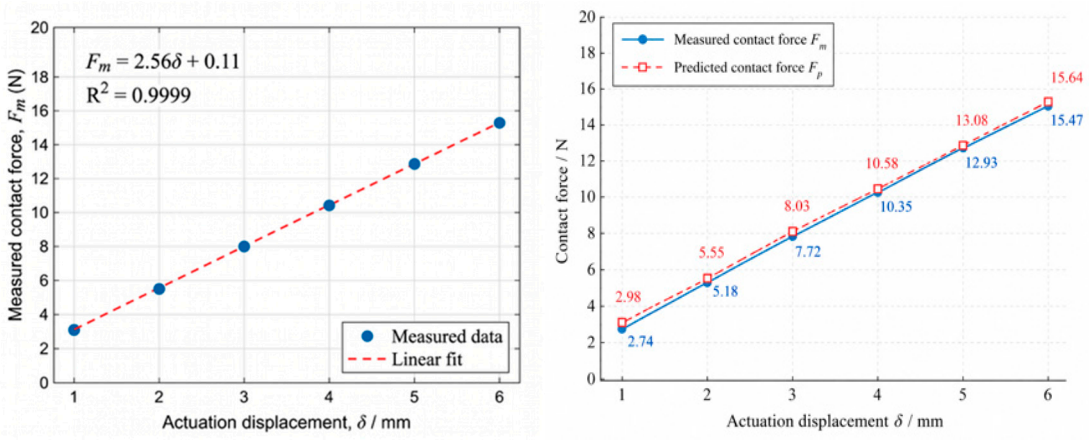
Test No.	δ (mm)	F_m (N)	F_p (N)	$\lvert F_p - F_m \rvert$ (N)	Relative Error (%)
1	1	2.74	2.98	0.24	8.76
2	2	5.18	5.55	0.37	7.14
3	3	7.72	8.03	0.31	4.02
4	4	10.35	10.58	0.23	2.22
5	5	12.93	13.08	0.15	1.16
6	6	15.47	15.64	0.17	1.10

As shown in Figure 34a, the measured contact force increased approximately linearly with the actuation displacement. The force–displacement relationship was fitted as:

$$F_m = 2.56\delta + 0.11$$

where F_m is the measured normal contact force and δ is the actuation displacement. Therefore, the linear fingertip stiffness of the asymmetric soft finger was identified as:

$$K_f = 2.56 \text{N/mm}$$



(**a**) (**b**)

Figure 34. Force-output characterization of the soft finger: (**a**) forcedisplacement fitting for Kf identification; (**b**) comparison of measured and predicted contact forces.

By using the experimentally identified linear fingertip stiffness to calibrate the force–displacement response of the soft finger, the predicted contact forces were compared with the measured results, as shown in Figure 34b. The predicted forces exhibited a mean relative error of 4.07%, and the relative error gradually as the actuation displacement. These results indicate that the model can reasonably describe the fingertip contact force output.

Five repeated measurements were conducted at each actuation displacement to evaluate force-output repeatability. The measured contact forces showed small fluctuations across repeated trials, indicating stable force output under repeated actuation.

3.3.2. Preliminary Cyclic Durability Evaluation

To further evaluate the cyclic stability of the optimized soft finger, the cyclic inward-curling test was performed for 3000 cycles. The fingertip contact force was recorded every 500 cycles, and the contact-force degradation ratio was calculated according to Equation (10). The results are summarized in Table 15.

Table 15. Contact-force variation during the cyclic inward-curling test.

Cycle Number	Contact Force (N)	Degradation Ratio (%)	Visual Condition
0	15.47 ± 0.08	0.00	No damage
500	15.39 ± 0.09	0.52	No damage
1000	15.32 ± 0.09	0.97	No damage
1500	15.26 ± 0.10	1.36	No damage
2000	15.19 ± 0.10	1.81	No damage
2500	15.11 ± 0.11	2.33	No damage
3000	15.04 ± 0.12	2.78	No visible damage

As shown in Table 15, the fingertip contact force showed only a slight decrease during the cyclic test. The initial contact force was 15.47 N, and the contact force after 3000 cycles was 15.04 N, corresponding to a degradation ratio of 2.78%. After the test, no visible cracks, permanent deformation, or local structural damage were observed. This test provides preliminary evidence of short-term rated-load durability, whereas long-term fatigue performance under high-frequency field operation requires further investigation.

A rated-load cyclic grasping test was further conducted using a 2.0 kg payload. The gripper completed 500 grasping cycles without payload dropping or visible slippage. After the test, no visible cracks, permanent deformation, surface wear, or local damage were observed, as summarized in Table 16.

Table 16. Results of the rated-payload cyclic grasping test.

Payload Mass (kg)	Cycles	Test Condition	Result
2.0	500	Lifted to 20 cm and held for 5 s in each cycle	No payload dropping, visible slippage, cracks, permanent deformation, or local structural damage

3.3.3. Anti-Slip Stability Verification

Based on the measured contact force output, the anti-slip performance was further evaluated using five types of railway maintenance tools: an electric drill, hammer, pliers, screwdriver, and wrench. The available friction force generated by the four-finger grasp was compared with the required holding force under the specified transfer acceleration. The results are summarized in Table 17.

Table 17. Anti-slip performance of the soft gripper for typical railway maintenance tools.

Test Object	Mass m (kg)	F_n (N)	μ	$F_{available}$ (N)	$F_{required}$	Margin (N)	Predicted State
Electric drill	1.80	15.47	0.46	28.46	19.46	9.00	Stable
Hammer	1.25	13.62	0.50	27.24	13.51	13.73	Stable
Pliers	0.62	10.35	0.48	19.87	6.70	13.17	Stable
Screwdriver	0.35	8.05	0.47	15.13	3.78	11.35	Stable
Wrench	0.50	9.22	0.44	16.23	5.41	10.83	Stable

As shown in Figure 35, the available friction force for all tested tools the required holding force, indicating that the gripper satisfies the conservative friction-only anti-slip verification criterion. Among the tools, the electric drill the smallest stability margin due to its larger mass. Nevertheless, its available friction force 28.46 N, which higher than the required holding force of 19.46 N, leaving a safety margin of 9.00 N.

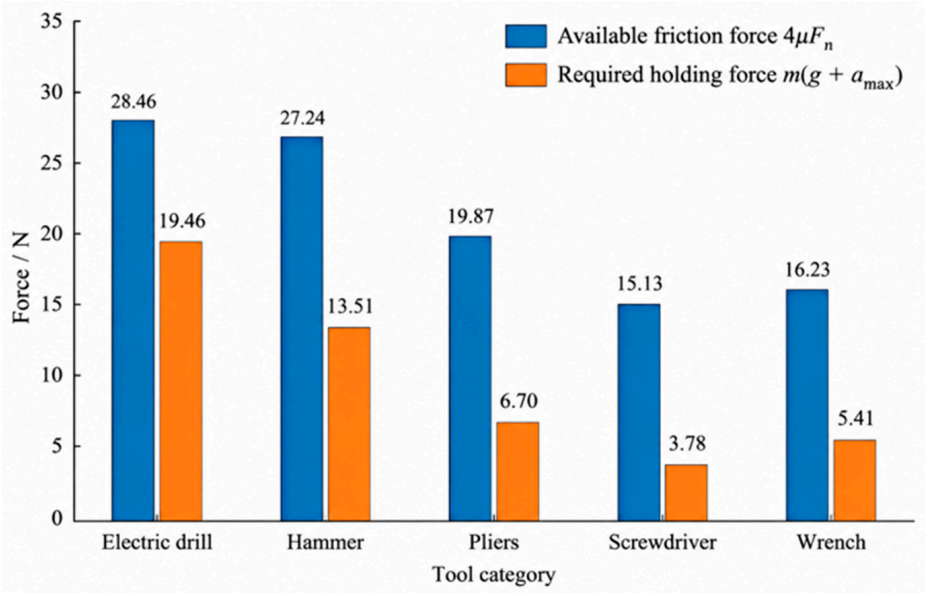


Figure 35. Comparison of available friction force and required holding force for different tools.

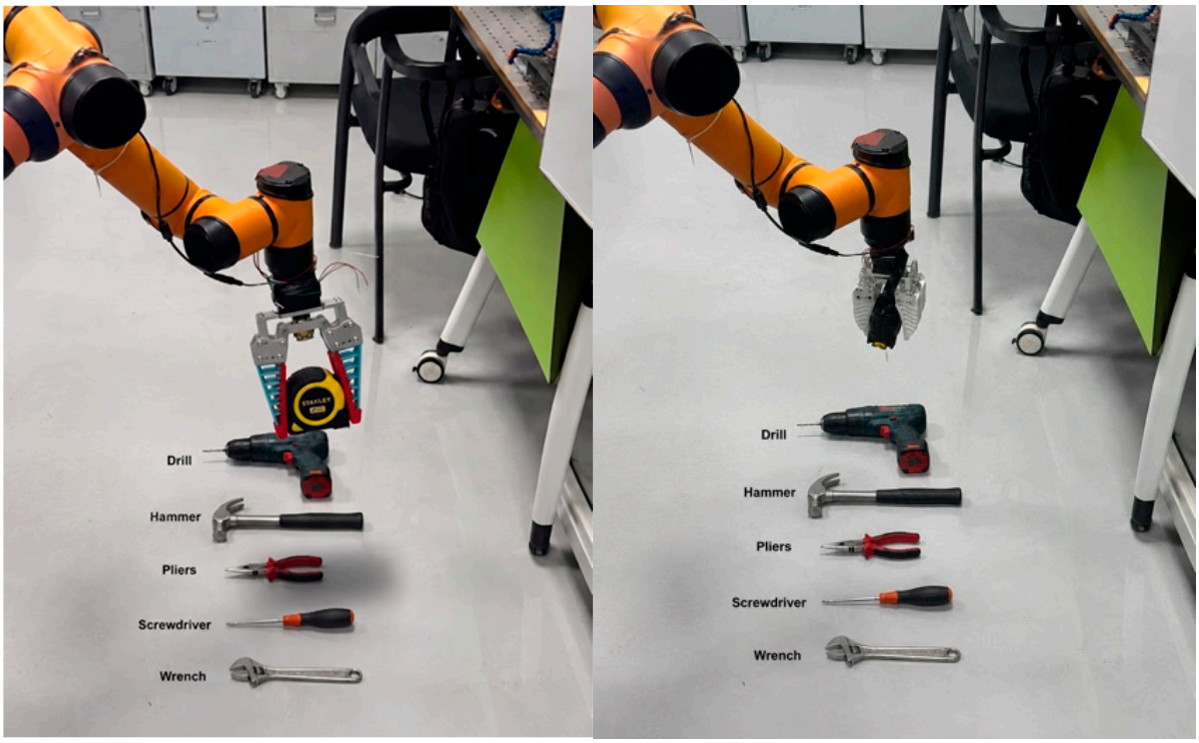
Therefore, the gripper met the conservative friction-only stability criterion for all tested tools.

3.4. Comprehensive Grasping Performance Evaluation

3.4.1. Controlled Laboratory Baseline Gripper Comparison

To quantify the contribution of the asymmetric Fin Ray topology, a controlled laboratory comparison was conducted between the proposed asymmetric Fin Ray gripper and a conventional symmetric Fin Ray baseline, as shown in Figure 36. The purpose of this experiment was to isolate the effect of the finger topology on grasping performance within the integrated robotic system. Therefore, the finger structure was the primary controlled variable in this comparison.

Five typical railway maintenance tools, including an electric drill, hammer, pliers, screwdriver, and wrench, were selected as target objects. Repeated grasping trials were conducted for each gripper across the five tool categories, and the success rates were calculated for each tool category, as shown in Table 18.



(**a**) (**b**)

Figure 36. Baseline comparison between the proposed asymmetric Fin Ray gripper and the symmetric Fin Ray baseline: (**a**) symmetric Fin Ray baseline; (**b**) proposed asymmetric Fin Ray gripper.

Table 18. Controlled laboratory comparison of grasping success rates between the proposed asymmetric Fin Ray gripper and the symmetric Fin Ray baseline.

Tool	Proposed Asymmetric Fin Ray (%)	Symmetric Fin Ray Baseline (%)	Improvement (pp)
Drill	92.2	79.7	12.5
Hammer	94.8	84.5	10.3
Pliers	91.9	80.6	11.3
Screwdriver	93.0	84.2	8.8
Wrench	88.1	76.3	11.9
Average	92.0	81.0	11.0

As shown in Table 18, the asymmetric gripper achieved higher grasping success rates than the symmetric gripper for all five tool categories. The overall success rate reached 92.0%, corresponding to an overall improvement of 11.0 percentage points. This result indicates that the asymmetric finger topology made a quantifiable contribution to grasping performance.

The improvements across the five tool categories ranged from 8.8 to 12.5 percentage points, as shown in Figure 37. This result indicates that the asymmetric topology provided consistent performance benefits for tools with different geometries and mass distributions. Compared with the symmetric Fin Ray baseline, the asymmetric structure promoted inward deformation and enhanced grasping stability for irregular maintenance tools.

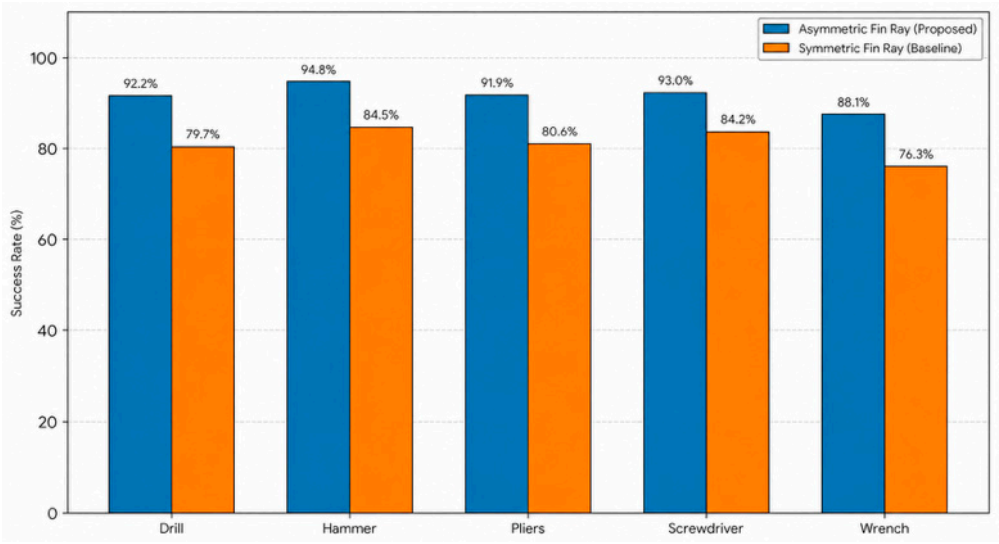


Figure 37. Overall grasping success rates for different tool categories.

3.4.2. Experimental Setup and Quantitative Results

To comprehensively evaluate the operational robustness of the system in real railway maintenance scenarios, the robotic system was deployed in an outdoor railway ballast environment for field testing. As shown in Figure 38, five typical railway maintenance tools—electric drill, hammer, pliers, screwdriver, and wrench—were selected as target objects. These tools exhibit significant differences in weight distribution, geometric features, and surface reflectivity under natural lighting, enabling thorough validation of the system's comprehensive performance in unstructured real-world scenarios.

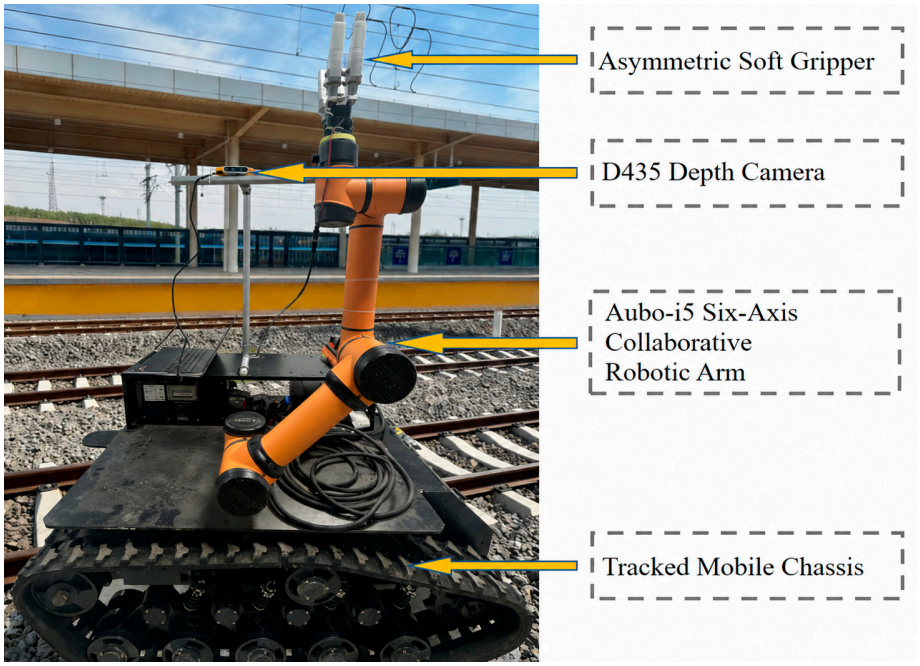


Figure 38. Hardware integration and field deployment.

The experiment designed two typical working conditions, as shown in Table 19, to separately verify the system adaptive enveloping capability for single geometric shapes and the detection robustness of the RA-YOLO algorithm under complex backgrounds and partial occlusion.

Table 19. Tool grasping experimental scenario settings and evaluation focus.

Scenario	Type	Environment	Evaluation Focus
A	Isolated	Tools placed independently without mutual contact	Adaptive enveloping capability for diverse geometries
B	Cluttered	Randomly stacked tools with partial occlusions	Detection robustness of RA-YOLO in complex backgrounds

Repeated grasping tests were conducted for each tool under each experimental scenario. The success criterion was defined as follows:

The robotic arm successfully grasped the tool from the ballast surface, lifted it to a height of 20 cm, maintained it for 5 s without visible slippage, and finally transferred it to the collection area. The experimental results are summarized in Table 20.

Table 20. Grasping success rate statistics for different tools in each experimental scenario.

Tool Type	Scenario A	Scenario B	Overall
Drill	94.5%	87.9%	91.2%
Hammer	96.5%	92.5%	94.7%
Pliers	95.2%	88.6%	91.9%
Screwdriver	95.8%	89.2%	92.5%
Wrench	93.8%	83.6%	88.7%
Average	95.2%	88.4%	91.8%

It should be noted that the current field experiments mainly evaluated clutter interference and outdoor ballast backgrounds under representative operating conditions. Extreme weather conditions were not fully included in the present system-level validation.

As shown in Table 20, the system an overall grasping success rate of 91.8%. Benefiting from the enveloping advantage of the soft fingers, the system a 95.2% success rate in unoccluded Scenario A, while being limited to 88.4% in Scenario B with severe occlusion. This data distribution is logically consistent with the visual AP evaluation results, authentically reflecting the grasping characteristics in unstructured environments.

3.4.3. Task Efficiency Distribution and Consistency Analysis

To further evaluate the operational stability of the system under different loads, task completion time (TCT), a continuous variable, was introduced for statistical analysis. Figure 39 presents boxplots of TCT distributions for the five tool categories under different scenarios.

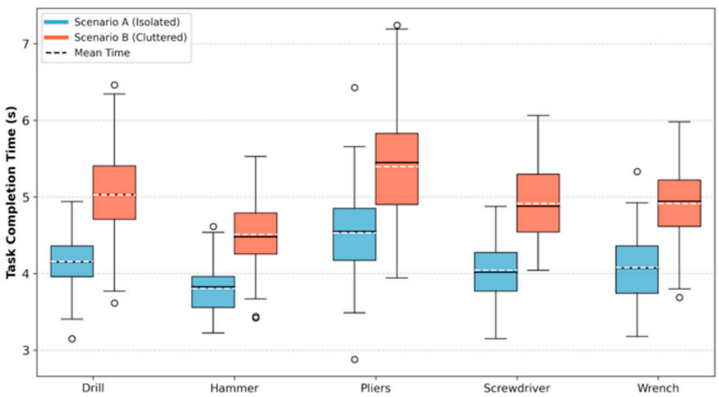


Figure 39. Comparison of task completion time (TCT) distributions for each tool category under different experimental scenarios.

Statistical results show that the average time consumption in Scenario A concentrated at 4.12 s, while in Scenario B, due to increased obstacle avoidance planning and environmental complexity, the average time consumption to 4.94 s. The compactness of the boxplots indicates that even in complex stacking scenarios, the variance in task time consumption for different geometric tool shapes is controlled within 0.7 s. This high degree of temporal consistency indirectly demonstrates the determinism of the perception algorithm when handling irregular objects, compensating for the insufficiency of the single success rate metric in describing the system dynamic performance.

3.4.4. Failure Mode Analysis and System Limitations

Although a high operational success rate was achieved, quantitatively decomposing the remaining 8.2% of failure cases is crucial for identifying system boundaries and guiding future optimization. As shown in Figure 40, the failures categorized into four types: perception interference, coordinate drift, dynamic slippage, and other errors.

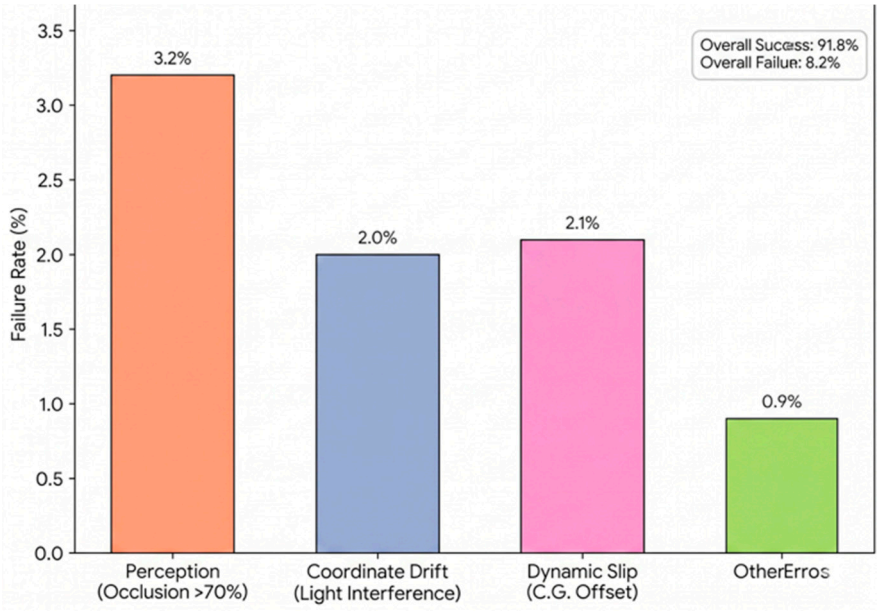


Figure 40. Statistical analysis of grasping failure causes.

Among the tested tools, the wrench exhibited the lowest grasping success rate, particularly under the cluttered condition of Scenario B (83.6%, Table 20). This reduction can be attributed to the combined influence of mechanical and perceptual factors. Mechanically, the thin and flat geometry of the wrench limits the effective contact area available to the asymmetric Fin Ray fingers, making stable multipoint enveloping more difficult than for bulkier tools such as drills or hammers. Perceptually, the highly reflective metallic surface is more susceptible to specular reflection under strong outdoor illumination, which may degrade the local quality of depth observations and lead to sparse depth point distributions around the grasp region. As a consequence, although the RGB-based detector can still provide a reliable target hypothesis in most cases, the associated 3D localization may exhibit residual pose deviation, thereby increasing the probability of misaligned contact during grasp execution.

Analysis shows that perception missed detections mainly occur in extreme stacking situations with occlusion rates exceeding 70%. In addition, depth point cloud voids caused by strong environmental lighting may increase local pose uncertainty in cluttered failure cases. Through this frequency analysis of failure causes, it is demonstrated that the system can robustly handle the vast majority of routine railway working conditions.

In addition, the failure-mode analysis that coordinate drift and dynamic slippage were observed at rates of 2.0% and 2.1%, respectively. These results suggest that dynamic motion introduced measurable but limited uncertainty under the current low-speed outdoor transfer conditions.

4. Discussion

Experimental results demonstrate that the integrated system an overall grasping success rate of 91.8%. This strongly validates the effectiveness of the perception-execution collaborative framework proposed in this work. Among these results, the system grasping success rate for heavy-load irregular tools 90%. This is primarily attributed to the mechanical advantages of the asymmetric Fin Ray structure. Compared with traditional symmetric structures, the asymmetric design actively induces an inward curling effect when subjected to force. This effect significantly increases the contact area between the finger pad and the workpiece, forming effective physical enveloping. Meanwhile, the arc-shaped outer sidewall increases the cross-sectional second moment of area of the finger, effectively suppressing lateral instability phenomena under heavy loads and ensuring the stability of rigid-flexible coupled grasping.

Under cluttered stacking scenarios, the average grasping success rate a 6.8 percentage-point decrease, mainly due to increased occlusion, contact uncertainty, and localization difficulty. Nevertheless, the RA-YOLO algorithm helped maintain reliable target detection under background interference. Railway ballast backgrounds possess extremely strong texture interference, and traditional detection models are prone to false activations. The CBAM module introduced in this work adaptively recalibrates feature weights, strengthening the response of target regions while suppressing background noise. Meanwhile, the Mosaic augmentation strategy enhances the model ability to distinguish overlapping targets. Furthermore, the system an overall grasping success rate of 91.8% across five tool categories, indicating that the integrated perception, hand-eye localization, trajectory planning, and compliant grasping modules can work cooperatively under representative outdoor ballast conditions.

The system proposed in this work demonstrates significant engineering application value in unstructured railway environments. Compared with existing railway visual recognition research, RA-YOLO the best balance between detection accuracy and inference speed, maintaining an average accuracy of 93.6% while achieving an inference speed of 105 FPS. This lightweight characteristic is highly conducive to deployment on mobile robot terminals. In terms of the end-effector, the soft gripper designed in this work stronger adaptive capability than traditional rigid grippers. It can passively accommodate workpieces of various geometric sizes, significantly reducing the complexity of low-level trajectory planning. This design not only improves the operational compliance of the system but also avoids secondary mechanical damage to workpiece surfaces.

Although this system excellently in multiple experimental scenarios, certain limitations remain under extreme working conditions. This limitation mainly affects the depth-sensing and 3D localization stage rather than the RGB-based target detection stage. Therefore, the proposed RA-YOLO module should be understood as improving detection robustness under complex backgrounds, while illumination-induced uncertainty in depth measurements may still remain under severe glare or highly reflective metallic surfaces. Although supplementary robustness tests were conducted under wet ballast and different ballast texture/color conditions, the present study not fully cover all harsh railway maintenance conditions. In particular, heavy rain, snow, mud contamination, severe dust, and extremely low illumination may further affect visual perception and grasping stability. Strong outdoor direct sunlight may interfere with the imaging quality of the depth camera,

thereby increasing random errors in hand-eye localization. In addition, although the cyclic inward-curling test confirmed stable short-term force output of the optimized soft finger, long-term fatigue wear of the TPU material under high-frequency field operations still requires further investigation. Future research will focus on integrating flexible pressure sensors into the soft finger pads. By introducing tactile feedback mechanisms, the system can achieve more precise closed-loop grasping force control. Meanwhile, multi-robot collaboration schemes are planned to address larger-size and heavier-load railway maintenance task requirements.

Moreover, the localization accuracy experiment in Section 3.2 mainly the quasi-static geometric accuracy of the hand-eye calibration model. In actual grasping-and-transfer operation, the effective positioning accuracy may be affected by dynamic factors such as robotic-arm vibration, image acquisition delay, transient depth fluctuation, and illumination-induced point-cloud sparsity. The system-level experiments indicate that the combined influence of these factors did not lead to severe performance degradation under the current low-speed operating conditions, as reflected by the overall grasping success rate of 91.8%. Nevertheless, coordinate drift and dynamic slippage were still observed in a small proportion of test attempts, suggesting that dynamic motion has a measurable but limited influence on the current system. Future work will further improve dynamic localization robustness through online visual compensation, visual-inertial fusion, and tactile-feedback-based closed-loop grasping control.

5. Conclusions

In this study, a railway inspection tool automated sorting system integrating improved visual perception and bio-inspired soft grasping was designed and developed. The system primarily consists of Fin Ray soft fingers with asymmetric rib topology, an attention mechanism-based RA-YOLO visual detection network, and a six-axis robotic arm collaborative control module. To address the challenges of inaccurate perception and unstable grasping in unstructured ballast backgrounds, the finger structural parameters were optimized based on the Yeoh hyperelastic constitutive model and finite element analysis (FEA), significantly enhancing lateral stiffness under heavy-load conditions while maintaining compliant enveloping, thereby resolving the instability problem of traditional soft grippers. In terms of visual perception, by embedding the CBAM module into YOLOv8n and combining it with Mosaic data augmentation, high-precision real-time detection of irregular metal tools under complex backgrounds was achieved, with mAP@0.5 reaching 93.6%. System integration experiments validated the effectiveness of the proposed design. Under quasi-static vision-guided hand-eye localization validation, the hand-eye calibration model achieved an average Euclidean positioning error of 1.03 mm, with the maximum component-wise absolute error remaining within 2.5 mm. Under representative low-speed outdoor ballast conditions, the integrated system achieved an overall grasping success rate of 91.8% across five typical tool categories, demonstrating good load adaptability, perception-execution coordination, and practical feasibility for unstructured railway maintenance environments. The system can effectively cope with unstructured environments, meeting the high-reliability requirements for intelligent O&M equipment under the Industry 4.0 context [40]. Furthermore, this research provides important theoretical and technical support for railway intelligent O&M equipment operations in unstructured environments.

Author Contributions: Overall scheme design, Y.L.; Final draft review, P.F.; Writing, M.T.; Create a chart, Y.D.; Literature search, G.L.; Data analysis, L.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Research Program of the Shaanxi Provincial Department of Education (No. 23JP004).

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Luo, W.; Zhu, B.; Wang, X.; He, J. Unmanned intelligent construction in railway infrastructure: A review. *Intell. Transp. Infrastruct.* **2026**, *5*, liag002. [CrossRef]
2. Liu, H.; Rahman, M.; Rahimi, M.; Starr, A.; Durazo-Cardenas, I.; Ruiz-Carcel, C.; Ompusunggu, A.; Hall, A.; Anderson, R. An autonomous rail-road amphibious robotic system for railway maintenance using sensor fusion and mobile manipulator. *Comput. Electr. Eng.* **2023**, *110*, 108874. [CrossRef]
3. Xiong, H.; Wu, L.; Lin, Y. A Robotic Grasping Framework for Railway Cargo Based on an Improved YOLOv8 Model. In Proceedings of the 2025 8th International Conference on Robotic Systems and Applications (ICRSA) 2025, Wuhan, China, 19–21 September 2025; pp. 191–198. [CrossRef]
4. Chen, X.; Tian, Y.; Li, M.; Lv, B.; Zhang, S.; Qu, Z.; Wu, J.; Cheng, S. Automatic detection of foreign object intrusion along railway tracks based on MACE-Net. *PLoS ONE* **2024**, *19*, e029303. [CrossRef]
5. Ning, S.; Ding, F.; Chen, B. Research on the Method of Foreign Object Detection for Railway Tracks Based on Deep Learning. *Sensors* **2024**, *24*, 4483. [CrossRef]
6. Verma, S.S.; Behera, C.K. Real-Time Railway Obstacle Detection in Variable Weather Conditions: A Novel Framework for Enhanced Safety Using YOLOv8. In *Machine Intelligence, Tools, and Applications; Learning and Analytics in Intelligent Systems*; Springer: Cham, Switzerland, 2024; Volume 40, pp. 387–398. [CrossRef]
7. Mohammad, A.; Sun, E.; Zhou, J.; Yang, G.; Zhang, G.; Munoa, J.; Barrios, A. Robotic end-effectors for manufacturing: Recent developments and future research challenges. *Int. J. Mach. Tools Manuf.* **2026**, *216*, 104367. [CrossRef]
8. Wang, Y.; Wang, Y.; Mushtaq, R.T.; Wei, Q. Advancements in Soft Robotics: A Comprehensive Review on Actuation Methods, Materials, and Applications. *Polymers* **2024**, *16*, 1087. [CrossRef]
9. Chancharoen, R.; Chaiprabha, K.; Chungsangsatiporn, W.; Piankitrungreang, P.; Saetia, S.; Viravan, T.; Phanomchoeng, G. Electro-Actuated Customizable Stacked Fin Ray Gripper for Adaptive Object Handling. *Actuators* **2026**, *15*, 52. [CrossRef]
10. Yamamoto, K.; Ishibashi, K.; Ishikawa, H.; Azami, O. Design methodology of hydraulically-driven soft robotic gripper for a large and heavy object. *Adv. Robot.* **2026**, *40*, 157–169. [CrossRef]
11. Al-Hadithi, B.M.; Pastor, C.; Lin, T.Y. Design and Experimental Validation of a 3D-Printed Hybrid Soft Robotic Gripper for Delicate Object Manipulation. *Electronics* **2026**, *15*, 848. [CrossRef]
12. Hassan, M.; Anik, T.M.; Ahmed, S.; Talha, A. Design, Modeling and Control of a Manipulator with Bio-Inspired Soft Robotic Gripper. Bachelor's Thesis, BRAC University, Dhaka, Bangladesh, 2026.
13. Mohammadi, A.; Lavranos, J.; Zhou, H.; Mutlu, R.; Alici, G.; Tan, Y.; Choong, P.; Oetomo, D. A practical 3D-printed soft robotic prosthetic hand with multi-articulating capabilities. *PLoS ONE* **2020**, *15*, e0232766. [CrossRef]
14. Liu, C.H.; Chung, F.M.; Chen, Y.; Chiu, C.H.; Chen, T.L. Optimal Design of a Motor-Driven Three-Finger Soft Robotic Gripper. *IEEE/ASME Trans. Mechatron.* **2020**, *25*, 1830–1840. [CrossRef]
15. Hao, Y.; Zhou, P.; Zhou, W.; Zeng, T.; Liu, Z.; Ren, Z.; Zhang, J. Adjusting the Interdigital Space and Hybrid Fingers of an Adaptive Gripper Utilizing a Single Motor. *IEEE/ASME Trans. Mechatron.* **2025**, *30*, 5005–5016. [CrossRef]
16. Li, W.; Huang, L.; Li, Y. Bioinspired Dual-layered Soft-rigid Gripper for Reduced Damage and Improved Grasping Stability with Real-time Classification. *J. Bionic Eng.* **2026**, *23*, 192–224. [CrossRef]
17. Wu, X.; Yu, Y. Research on flexible control method of robot based on visual servoing. *J. Comb. Math. Comb. Comput.* **2025**, *127a*, 1005–1028. [CrossRef]
18. Jocher, G.; Chaurasia, A.; Qiu, J. *Ultralytics YOLO*; Zenodo: Genève, Switzerland, 2024.
19. Wu, T.; Dong, Y. YOLO-SE: Improved YOLOv8 for Remote Sensing Object Detection and Recognition. *Appl. Sci.* **2023**, *13*, 12977. [CrossRef]
20. Talib, M.; Al-Noori, A.H.Y.; Suad, J. YOLOv8-CAB: Improved YOLOv8 for Real-time object detection. *Karbala Int. J. Mod. Sci.* **2024**, *10*, 5. [CrossRef]
21. Yi, B.; Liu, B.; Zhao, L.; Liu, E. Small Object Detection Algorithm Based on Improved YOLOv8 for Remote Sensing. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 1734–1747. [CrossRef]
22. Wang, C.; Liu, Y.; Cai, L.; Chen, L.; Li, Y. YOLOv8-QSD: An Improved Small Object Detection Algorithm for Autonomous Vehicles Based on YOLOv8. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 1–16. [CrossRef]
23. Shen, B.; Lang, S.; Song, Z. DS-YOLOv8-Based Object Detection Method for Remote Sensing Images. *IEEE Access* **2023**, *11*, 125122–125137. [CrossRef]

24. Liu, H.; Su, W.; Li, B.; Hong, J. Bio-Inspired Topology Optimization Framework for Flexible Robotic Grippers. *J. Mech. Des.* **2026**, *148*, 013302. [CrossRef]
25. MathWorks. *ROS Toolbox Release Notes: Co-Simulation Between Gazebo and Simulink & ROS 2 Node Generation*; MathWorks Documentation 2025, R2025a; MathWorks: Natick, MA, USA, 2025.
26. Zhang, D.; Hao, Z.; Wang, M.; Zeng, S.; Huan, S.; She, J. Improved DETR-based visual servoing for robotic arm satellite tracking. *J. Aerosp. Eng.* **2026**, *39*, 04025112. [CrossRef]
27. Zhou, H. Design and Kinematic Simulation Analysis of a Vision-Guided 5-DOF Robotic Arm System. *ICCK Trans. Intell. Cyber-Phys. Syst.* **2026**, *1*, 51–59.
28. Borgerink, D.J.; Brouwer, D.M.; Stegenga, J.; Stramigioli, S. Kinematic Design Method for Rail-Guided Robotic Arms. *J. Mech. Robot.* **2017**, *9*, 011010. [CrossRef]
29. Moran, M.E. Evolution of robotic arms. *J. Robot. Surg.* **2007**, *1*, 103–111. [CrossRef] [PubMed]
30. BASF Polyurethanes GmbH. *Elastollan® 1185 A Datasheet*; Material Data Center: Würzburg, Germany; Available online: https://www.materialdatacenter.com/ms/en/tradenames/Elastollan/BASF%2BPolyurethanes%2BGmbH/Elastollan%C2%AE%2B1185%2BA/65a16011/906 (accessed on 8 April 2026).
31. Jiang, Y.; Mohammadpour Velni, J. A Soft Robotic Gripper for Crop Harvesting: Prototyping, Imaging, and Model-Based Control. *AgriEngineering* **2025**, *7*, 378. [CrossRef]
32. Wang, X.; Kang, H.; Zhou, H.; Au, W.; Wang, M.Y.; Chen, C. Development and evaluation of a robust soft robotic gripper for apple harvesting. *Comput. Electron. Agric.* **2023**, *204*, 107552. [CrossRef]
33. Wang, C.-Y.; Liao, H.-Y.M.; Wu, Y.-H.; Chen, P.-Y.; Hsieh, J.-W.; Yeh, I.-H. CSPNet: A New Backbone that Can Enhance Learning Capability of CNN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) 2020, Seattle, WA, USA, 14–19 June 2020; pp. 1571–1580. [CrossRef]
34. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In *Computer Vision—ECCV 2018; Lecture Notes in Computer Science*; Springer: Cham, Switzerland, 2018; Volume 11211, pp. 3–19. [CrossRef]
35. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934. [CrossRef]
36. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141. [CrossRef]
37. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020. [CrossRef]
38. Hou, Q.; Zhou, D.; Feng, J. Coordinate Attention for Efficient Mobile Network Design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 13708–13717. [CrossRef]
39. Zhang, Z. A Flexible New Technique for Camera Calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334. [CrossRef]
40. Flores Gonzalez, J.M.; Coronado, E.; Yamanobe, N. ROS-Compatible Robotics Simulators for Industry 4.0 and Industry 5.0: A Systematic Review of Trends and Technologies. *Appl. Sci.* **2025**, *15*, 8637. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.