Article | Received 2 October 2025; Revised 12 December 2025; Accepted 25 December 2025; Published 16 January 2026
https://doi.org/10.55092/rl20260002

Scale-adaptive visual perception and sequence planning for coordinated bimanual robotic disassembly retired power battery

Jiangming Zhang[1,2], Shaohua Kong[1,3], Junyang Wang[1,3], Wu Yan[1], Xuefeng Zhou[1] and Hongmin Wu[1,*]

[1] Guangdong Key Laboratory of Modern Control Technology, Institute of Intelligent Manufacturing, Guangdong Academy of Sciences, Guangzhou, China
[2] School of Mechanical Engineering, Guangdong Polytechnic Normal University, Guangzhou, China
[3] School of Mechanical and Automation Engineering, Wuyi University, Jiangmen, China

* Correspondence author; E-mail: hm.wu@giim.ac.cn.

Highlights:

- Scale-adaptive visual perception for coordinated bimanual battery disassembly.
- Enhanced YOLOv8 enables robust multi-scale detection of irregular components.
- ACO-based dual-arm planning achieves efficient disassembly on real battery packs.

Abstract: The rapid growth of retired power batteries has created an urgent demand for efficient, safe, and automated disassembly solutions. Existing processes still rely heavily on repetitive manual operations, resulting in low efficiency and potential safety hazards. To address these challenges, this paper presents a novel framework for scale-adaptive visual perception and sequence planning of heterogeneous structural components (e.g., fastener, battery management systems, conductor wires) to enable coordinated bimanual robotic disassembly of retired power batteries. First, a Cell–Module–Pack (CMP) dataset is constructed and augmented to enhance the generalization capability of semantic recognition and pose estimation. Second, an improved YOLOv8-based adaptive visual perception network is developed to robustly recognize and precisely localize irregularly shaped and scale-varying components. On the battery pack dataset, the detection results are: overall Mask Precision 0.961, Mask Recall 0.971, and mAP@50:95 of 0.810; for individual classes, fastener 0.584, copperbar 0.926, structural parts 0.835, and Battery Management System (BMS) 0.957. Third, a sequence planning algorithm based on ant colony optimization is introduced to generate efficient disassembly orders while minimizing task conflicts. The present paper employs dual-arm robotic collaboration for the purpose of disassembly, with a total of 56 planning points and an overall planning length of 5.83 m. The lengths of the left and right arms have been meticulously planned, with the former set at 2.91 m and the latter at 2.92 m. Finally, coordinated bimanual disassembly is realized through precise hand–eye calibration. Experimental validation on real retired battery packs demonstrates that the proposed framework achieves robust visual perception and effective disassembly sequence planning, showing strong potential for industrial application in automated battery recycling.

Copyright©2026 by the authors. Published by ELSP. This work is licensed under Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium provided the original work is properly cited.

Keywords: retired power battery disassembly; bimanual robotic manipulation; scale-adaptive visual perception; disassembly sequence planning

1. Introduction

In the context of the accelerating global energy transition, the development of electric vehicles has emerged as a prominent trend within the global automotive industry. According to aggregated data from the International Energy Agency (IEA) and other industry analyses, cumulative global power battery installations are projected to exceed 2.8 terawatt-hours (2800 gigawatt-hours) by 2025. Current sales rates of new energy vehicles and projected lifespans of battery technology are being used to predict the total volume of retired batteries that will exceed 12 million tonnes by 2030.

However, given the substantial volume of retired battery packs currently on the market, the dismantling of retired power battery packs remains predominantly manual, involving high labour intensity, as shown in Figure 1. Improper handling of these materials has been demonstrated to result in a wasteful consumption of resources, in addition to the potential for environmental pollution and safety incidents. Consequently, the development of automated precision dismantling technologies that are both highly secure and flexible has become a significant trend in the industry's advancement. Figure 2 shows the comparison between manual and robotic disassembly of irregular components.

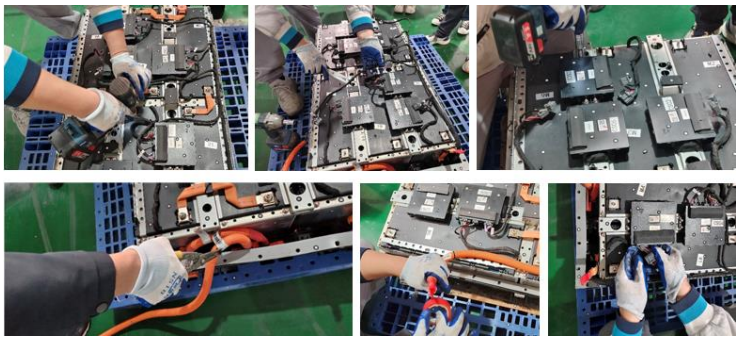


Figure 1. Manual disassembly of retired battery packs, which is still labor-intensive and prone to safety hazards.

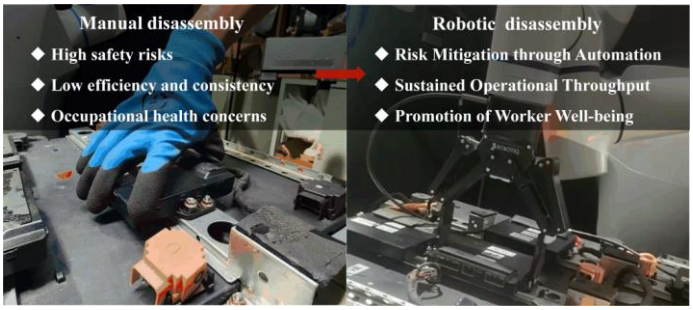


Figure 2. Comparison between manual and robotic disassembly of irregular components: manual disassembly suffers from safety risks, low efficiency, and health concerns, whereas robotic disassembly offers higher efficiency, precise repeatability, and standardized performance, making it more suitable for flexible disassembly of battery packs.

Currently, the two main challenges in automating end-of-life battery pack dismantling are the following. First, it is critical to accurately identify non-standard components within battery packs that exhibit complex assembly relationships. Second, once the semantic information of these non-standard

components has been obtained, an efficient disassembly sequence must be determined. As indicated in the existing literature, previous research has addressed aspects of visual perception [1] and sequence analysis [2] in this context.

In the field of visual recognition, the rapid development of deep learning has enabled object detection and semantic segmentation to achieve high accuracy in industrial scenarios. DiFilippo and Jouaneh used traditional vision methods [3] such as Scale-Invariant Feature Transform (SIFT) [4], Speeded Up Robust Features (SURF) [5], Canny edge detection, and template matching, combined with force sensors for screw detection, but these approaches are fragile under lighting variation, reflection, and occlusion. Zorn *et al.* applied high-resolution 3D depth cameras with deep learning networks such as PointNet, PointNet++, and Pose Vector Network for 3D Object Pose Estimation (PVN3D) [6], enabling precise 3D recognition and 6D pose estimation of irregular components in EV battery packs, though point cloud registration still lacks robustness for certain parts. Zhou *et al.* concentrated on screw detection, evaluating deep learning-based detectors including Region-Based Convolutional Neural Network (R-CNN), You Only Look Once (YOLO), and Single Shot MultiBox Detector (SSD) [7], and proposed modules to enhance multimodal fusion under challenging conditions, yet the physical condition of fasteners, such as wear, corrosion, or foreign attachments, continues to limit disassembly success beyond the scope of visual methods.

Determining an efficient disassembly sequence is a classic combinatorial optimization problem, and recent studies have introduced intelligent planning approaches. Liu *et al.* integrated visual recognition into a digital twin framework [8], but their method lacked robustness in detecting missing components and was constrained to dynamic programming scenarios with only one absent part, limiting real-world applicability. Wang and Li developed a parallel robotic planning model optimizing time and energy [9], which improved efficiency but incurred high computational costs for large-scale tasks. Feng *et al.* incorporated remanufacturing value via improved algorithms and expert consensus [10], enhancing decision reliability but increasing planning time and reducing agility in dynamic environments.

In the processing of retired battery packs and complex equipment, disassembly is a critical stage for resource recovery and component remanufacturing. Manual disassembly is inefficient and hazardous, making it unsuitable for large-scale battery retirement. Advances in robotics have thus spurred interest in automated and flexible disassembly. Mengling Chu and Weida Chen proposed an integrated optimization approach for human-robot collaborative disassembly, establishing mathematical models to optimize scheduling, component disassembly, and task allocation [11], though the large solution space and complexity limit scalability. Wansong Liu *et al*. developed a quantitative uncertainty neural network for human motion prediction to enable task-constrained, collision-free robotic motion during collaboration [12], but the method relies on costly specialized infrastructure, restricting applicability in broader industrial environments.

The objective of the present research is to sequentially dismantle irregular components within battery packs, including but not limited to fastener, BMS units, structural components, and copper busbars. To this end, we propose a comprehensive framework for the coordinated bimanual robotic disassembly of retired battery packs, aiming to enhance flexibility and efficiency. Within this framework, the existing battery pack dataset is first augmented through spatial geometric transformations [13], pixel-level alterations [14], and the Copy-Paste method [15]. To enable robust recognition of irregular components, an enhanced YOLOv8 [16] network is employed to perform semantic segmentation and obtain pose

information of the components. Regarding sequence planning, the disassembly of irregular components is reformulated as a Travelling Salesman Problem (TSP). An ant colony optimization algorithm [17] is applied to generate efficient disassembly sequences based on the identified component poses. Finally, the framework outputs are integrated into robotic systems to validate automated coordinated disassembly of irregular components in battery packs. Figure 3 provides a framework diagram of the system.

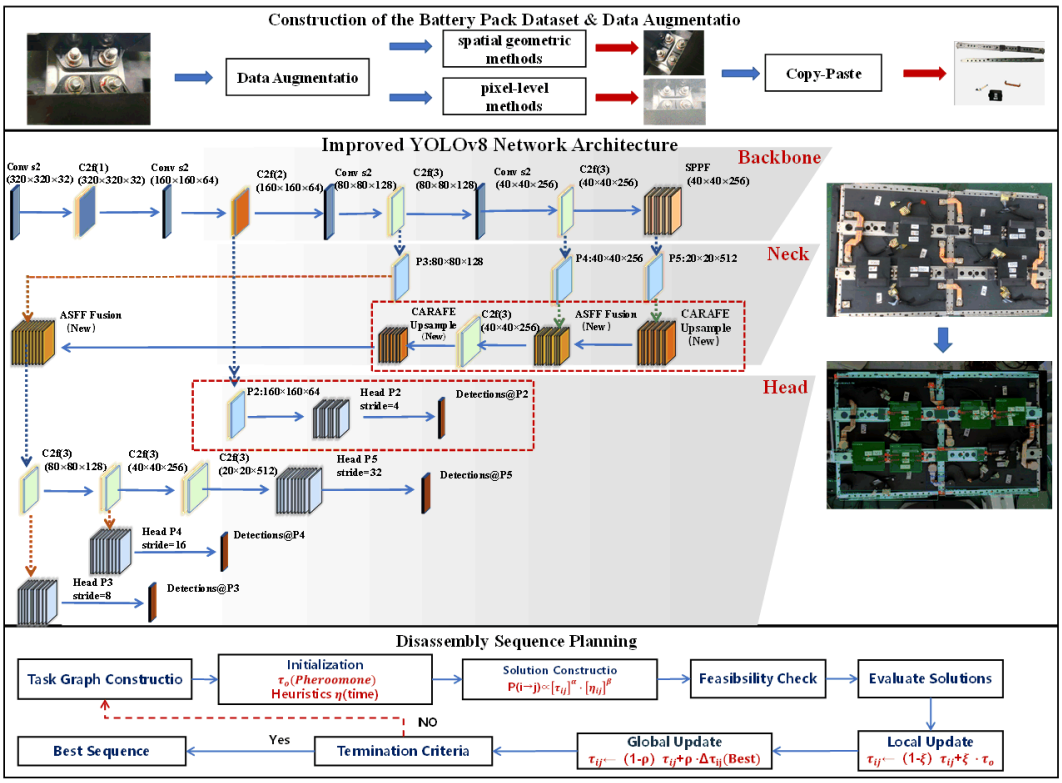


Figure 3. During the construction of the dataset, data augmentation techniques were employed on battery pack component images, including spatial geometric transformations, pixel-level enhancement, and Copy-Paste methods. Subsequently, during the object detection phase, an enhanced YOLOv8 network architecture was proposed. This incorporates Adaptive Feature Fusion (ASFF) fusion and CARAFE upsampling modules within the Backbone, Neck, and Head components to improve part detection and segmentation performance. In the final phase of sequence planning, an algorithm based on ant colony optimization was employed in order to construct the task graph, perform a solution space search, conduct feasibility checks, and execute global and local updates. This process yielded the optimal disassembly sequence for battery pack components.

2. Semantic dataset and augmentation for battery pack components

2.1. Construction of the battery pack dataset

We constructed a dataset of retired CMP battery packs sourced from CATL-equipped public buses, specifically the CATL-172.8V120AH-B model with dimensions of a 103.8 cm in length, 60.8 cm in width, and 25.5 cm in height. The primary irregular components requiring disassembly comprise four types: hexagonal fastener (62 units), BMS units (6 units), structural components (9 units), and copper bar (5 units). Data acquisition was carried out using Mech-Eye LSR-L vision sensors and mobile phones,

capturing RGB images at resolutions of 4000 × 3000 pixels and 3216 × 1440 pixels respectively. The dataset of various irregular components collected is shown in Figure 4.

The collected dataset comprises a total of 1547 images, including: The dataset under consideration comprises 474 images of screws, 419 images of BMS, 363 images of structural components, and 291 images of copper busbars. To ensure both diversity and comprehensiveness, the images were captured from a variety of angles, lighting conditions, and backgrounds. Moreover, we arranged team of researchers conducted rigorous, sequential inspections of the dataset, evaluating file integrity, image clarity and resolution, and annotation quality. This process enhances the generalization capability of subsequent visual recognition and localization model training.

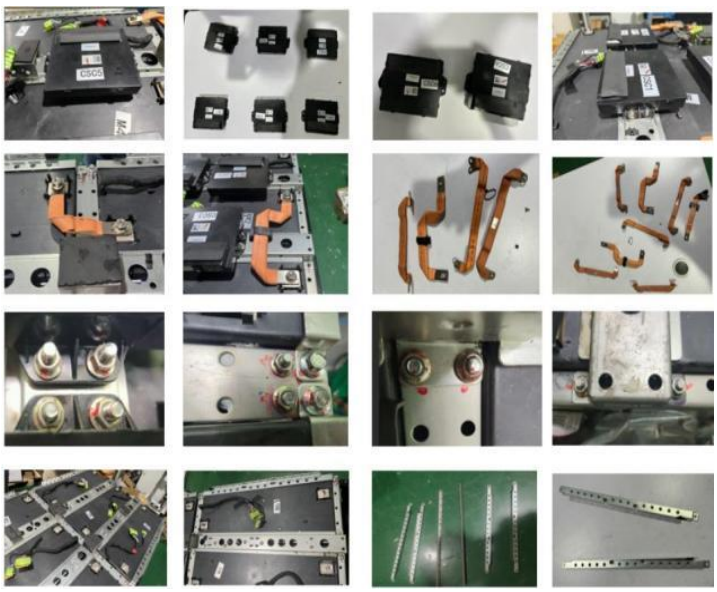


Figure 4. CMP battery pack dataset (CATL-172.8V120AH-B) for semantic recognition and pose estimation.

2.2. Data augmentation strategies

Due to the complex assembly relationships, the non-standard components within the retired battery packs present the following issues:

• The imbalance in the quantity of special-shaped components, with fastener significantly outnumbering BMS, structural parts and copper bar, may lead to inadequate recognition capabilities for the remaining special-shaped components.

• Target complexity and detection interference: The battery pack contains diverse non-standard components made from various materials. Metallic parts such as fastener and copper bar may exhibit reflections. Interference between these irregularly shaped components can cause feature loss during identification, thereby reducing the success rate of detection and segmentation processes.

• Varied scene interference arises when identical non-standard components exhibit differing geometric configurations and spatial relationships within the battery pack. Should the dataset fail to adequately cover such spatial variations, it will consequently diminish the success rate of visual inspection.

Given the aforementioned challenges, this paper employs traditional data augmentation techniques involving spatial geometric transformations and pixel-level alterations, such as flipping,

rotation, and scaling (spatial geometric methods), alongside color dithering, noise augmentation, and Gaussian blurring (pixel-level methods). To better address the complex distribution of irregularly shaped components within battery pack scenarios, this paper employs the Copy-Paste data augmentation method alongside conventional approaches. During model training, this technique directly replicates an image from the dataset and pastes it onto another scene image to generate novel training samples. The newly created data significantly enhances segmentation models trained under both supervised and semi-supervised conditions [18]. Figure 5 shows the dataset after data augmentation as mentioned above.

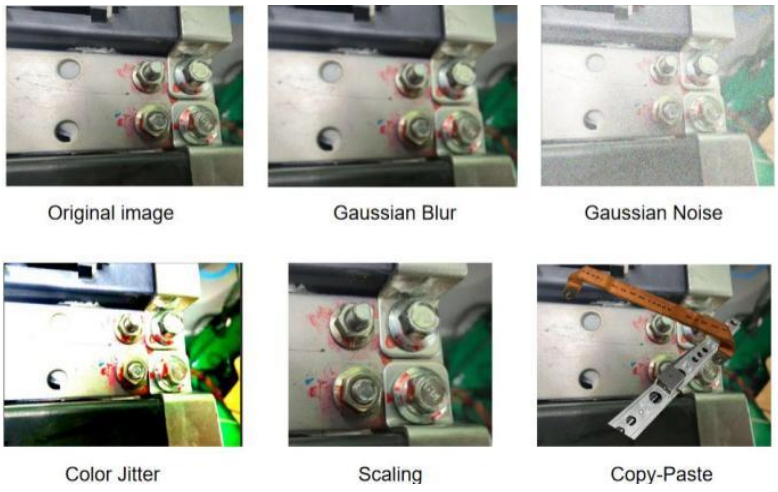


Figure 5. The data augmentation results obtained through Gaussian Blur, Gaussian Noise, Color Jitter, Scaling and Copy-Paste.

3. Recognition and localization of irregular components

3.1. Method for semantic recognition

The present work employs the YOLOv8 model, a method for object detection and instance segmentation that combines high accuracy with high efficiency. The system demonstrates a high level of performance in detecting small, densely packed, and multi-scale objects. The model's architecture is adept at balancing computational complexity and the number of parameters, while maintaining both inference speed and detection precision. Consequently, YOLOv8 has seen widespread adoption in various everyday and industrial scenarios in recent years.

The network structure of YOLOv8 mainly consists of three parts:

• Backbone: The C2f module [19] (CSPNet with two convolutions and a fusion of multiple bottlenecks) has been integrated and optimized based on the CSPDarknet architecture. While retaining the cross-stage feature allocation philosophy of CSPNet, the C2f module introduces additional residual branches and two convolutional layers with integrated bottlenecks. This improves the reuse of features while effectively controlling computational complexity without compromising expressiveness.

• Neck: The enhanced Path Aggregation Network-Feature Pyramid Network (PAN-FPN) [20] architecture effectively integrates the deep semantic information extracted by the backbone network with shallow spatial detail information through bidirectional feature pyramid fusion. This enables objects of various types to be detected at the corresponding feature scale.

• Head: The Decoupled head and Anchor-free mechanisms operate independently within the decoupled head, with classification and regression working separately. The Anchor-free mechanism [21] streamlines the design process and enhances generalization capabilities across diverse datasets.

3.2. Adaptive irregular component inspection network

In order to adapt to the complex assembly structures, non-standard component material properties, and variable recognition environments within retired battery packs, partial modifications have been implemented in the YOLOv8 network architecture.

In practical environments for detecting irregularly shaped objects, small targets such as fastener occupy only a dozen or so pixels in the captured image. Within the YOLOv8 network architecture, detection heads are only available at three scales: P3 (80 × 80), P4 (40 × 40) and P5 (20 × 20). Small objects such as fastener often shrink to 3–5 pixels on P3, which leads to feature loss and insufficient gradient contribution during training and causes them to be overlooked. Therefore, we have introduced a new detection head, P2, on the high-resolution feature map of the shallow backbone layer (stride = 4, size = 160 × 160 × 64). This dedicated head handles small object detection and produces classification and regression predictions [22] that are identical to those of P3–P5. During training, the loss of P2 is combined with that of the other three heads to describe the morphology of small fastener, significantly improving recall for small objects and reducing missed detections.

During the process of feature fusion in the Neck layer, YOLOv8 typically utilizes the following two methodologies:

• Nearest Neighbor Interpolation [23]: He process entails the direct mapping of each pixel point from a low-resolution feature map to a high-resolution space, thereby replicating the pixel's value within the extended region.

• Bilinear Interpolation [24]: The estimation of high-resolution features is achieved through the calculation of the weighted average of the target pixel and its adjacent 2 × 2 pixels.

When Nearest Neighbor Interpolation (NNI) is employed for small targets, such as fastener with diameters of merely a dozen pixels, the resultant boundaries exhibit jagged edges, which can cause morphological [25]. This phenomenon is characterized by pronounced "blocking" effects along the edge regions, impeding the detection head to accurately regress bounding boxes. Furthermore, within battery packs where the fasteners are densely distributed, replicated block-like features may cause adjacent target boundaries to coalesce, increasing false detection rates. While Bilinear Interpolation [26] has been demonstrated to effectively mitigate the "blocking" issue associated with NNI, its weighted averaging operation overlook subtle local structural variations. In the context of screw targets within battery packs, this method refines edges and textures during upscaling, reducing contrast between objects and their backgrounds. This blurring becomes particularly pronounced when metallic fastener adjoin reflective battery casing backgrounds, diminishing the detector's ability to distinguish small objects. This has been shown to increase the likelihood of missed detections, and cause adjacent small targets' boundaries to merge, resulting in multiple objects being erroneously identified as a single entity.

In the Neck section, the original YOLOv8 framework's upsampling pathways are substituted with CARAFE [27] (Content-Aware Reassembly of Features) employing adaptive convolutional

kernel upsampling. These pathways extend from deep features (P5, stride = 32) to intermediate features (P4, stride = 16), and from intermediate features (P4, stride = 16) to shallow features (P3, stride = 8). A specialized upsampling kernel is trained based on the local feature content. For each position, a "reassembly weight" [28] is predicted, and these weights used utilized to reassemble neighboring pixels from the low-resolution feature map into a high-resolution feature map. This process preserves more edge and texture details during upsampling. And significantly enhance the clarity and discriminative power of small targets, such as fastener, during the feature fusion stage, reducing missed detections and false positives caused by upsampling blurring.

The PAN-FPN architecture facilitates multi-scale feature fusion at the Neck layer within YOLOv8. This structure is characterized by the concatenation or summation of feature maps from disparate scales, prior to convolutional fusion. However, objects of varying sizes exhibit differing dependencies on feature layers: smaller objects rely more on high-resolution, shallow-layer features, whilst larger objects depend more on low-resolution, deep-layer features. Merely concatenating or summing features during the initial training phase is inadequate for adjusting the fusion ratio for differing locations or objects. In the context of disassembling fastener within battery packs, this approach can exacerbate background noise from metallic reflections or obscure screw features beneath deeper semantic layers, compromising the accuracy of screw recognition.

In order to facilitate the network's capacity for adaptive selection of the proportion of features from disparate levels during fusion, Neck replaces the original concatenating or summing method with ASFF [29]. ASFF generates learnable fusion weights at each spatial location, adaptively determining the blending ratio between shallow detail features and deep semantic features. The fusion process can be represented as follows:

$$F_{fused}(x,y) = [\alpha(x,y)\ \ \beta(x,y)\ \ \gamma(x,y)] \cdot \begin{bmatrix} F_s(x,y) \\ F_m(x,y) \\ F_l(x,y) \end{bmatrix} \quad (1)$$

The values $\alpha(x, y)$, $\beta(x, y)$, and $\gamma(x, y)$ are dynamically generated through the use of lightweight convolutions and SoftMax operations, satisfying the constraint $\alpha + \beta + \gamma = 1$. This mechanism facilitates the model's capacity to place greater reliance on high-resolution features within small target regions and on low-resolution features within large target regions. Consequently, it effectively suppresses false detections caused by complex backgrounds while significantly enhancing the ability to distinguish high-density small objects, such as fastener.

4. Disassembly sequence planning

Following the identification and positioning of irregular components within the battery pack, a suitable disassembly sequence is required to dismantle these parts. The presence of non-standard components other than fastener gives rise to occlusion issues, arising from their geometric and spatial relationships, specific sequence constraints during the disassembly. Conversely, the removing fastener does not exhibit the same level of sequential dependencies; it is sufficient to address all points. Optimizing the screw removal can be cast as a TSP [30], where the objective is to plan a path that minimizes either total distance or time, given a set of discrete nodes (screw locations). The focus was on the sequence for dismantling fastener. Following the completion of the screw removal

process, manual dexterity and prior knowledge were utilized to determine the dismantling sequence for the remaining non-standard components. These components were then systematically dismantled.

Ant Colony Optimization (ACO) [31] is a bio-inspired optimization algorithm that emulates the path-optimization behavior of ant colonies foraging in nature. The core principle is to mimic the mechanism by which ant colonies depend on pheromones for path selection and coordination of foraging activities. In the context of battery pack disassembly, many fasteners are distributed throughout the pack. The robot must select an optimal sequence for removing fastener from identified positions to minimize total path length and operational time. This problem is analogous to the TSP, which can be abstracted as finding an approximate optimal path through a set of discrete nodes. Each screw location can be regarded as a node in a graph, with the robot's initial position corresponding to the ants' departure point. During the execution of the algorithm, artificial ants traverse these nodes, continuously testing different disassembly sequences. The action of these organisms is determined by the concentration of pheromones in the environment, in addition to a heuristic function, which can be defined as the inverse of the distance between fastener. It has been demonstrated that superior paths-defined as those with shorter disassembly distances and reduced time-release greater pheromone concentrations. This gradual attraction of ants to select this solution during iterations ultimately forms a positive feedback mechanism. This enables the entire colony to rapidly converge upon an approximate optimal screw disassembly sequence.

The ACO algorithm is initiated with an initialization phase in which the pheromone matrix $\tau_{ij}(0) = \tau_0$ is established. The ant colony size, m, maximum iteration count T_{max}, and parameters α (pheromone importance coefficient), β (heuristic factor weight), ρ (pheromone decay coefficient), and Q (pheromone release intensity) are specified. Concurrently, a distance matrix d_{ij} is constructed based on the task scenario, with the heuristic function defined as $\eta_{ij} = 1/d_{ij}$. In the ensuing phase of solution construction, each ant transitions from its current node (i) to the next node (j), in accordance with the state transition probabilities. This transition can be expressed as follows:

$$P_{ij}^{k}(t) = \begin{cases} \dfrac{[\tau_{ij}(t)]^{\alpha} \cdot [\eta_{ij}]^{\beta}}{\sum_{l \in allowed_k} [\tau_{ij}(t)]^{\alpha} \cdot [\eta_{ij}]^{\beta}}, & j \in \text{allowed}_k \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

In this context, the term "$allowed_k$" is employed to denote the set of nodes that have not yet been visited by ant k, thereby ensuring the validity of the path. Following the completion of a traversal by the ant, the length of its path can be expressed as follows:

$$L_k = \sum_{(i,j) \in Tour_k} d_{ij} \tag{3}$$

In this study, L_k is used to denote the total path length of ant k, and $Tour_k$ is employed to represent the set of paths traversed by ant k. This value is utilized as a metric for assessing path quality. Subsequently, the pheromone update phase is initiated, initially considering the mechanisms of pheromone decay and enhancement. The following update rule, in its totality, is hereby presented:

$$(t+1) = (1-\rho) \cdot \tau_{ij}(t) + \Delta\tau_{ij}^{k} \tag{4}$$

The symbol $(1-\rho) \cdot \tau_{ij}(t)$ denotes the natural dissipation of the pheromone, thereby preventing infinite accumulation.

The symbol$\Delta\tau_{ij}^{k}$ represents the pheromone released by k ants along edge i→j, as defined as

$$\Delta\tau_{ij}^{k} = \begin{cases} \frac{Q}{L_k}, & if\ ant\ k\ traverses\ edge\ ij \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

Where Q is a constant controlling pheromone release; L_k denotes the ant trail length, with shorter trails contributing more significantly. It has been demonstrated that superior trails accumulate greater pheromone, making them more readily selected by subsequent generations of ants. This phenomenon guides the solution towards convergence to the global optimum.

The process is iterated repeatedly until either the maximum iteration count T_{max} is reached, or the optimal path ceases to improve after multiple iterations. At this juncture, the length of the globally optimal path is outputted as follows:

$$L^{*} = \min_{k \in \{1,\ldots,m\}, t \in \{1,\ldots,T_{max}\}} L_k(t) \quad (6)$$

The objective is to ultimately arrive at the optimal or near-optimal solution to the problem.

In the battery pack disassembly task, all screw coordinates identified by visual recognition are first abstracted as nodes in a graph. The robot's initial position is designated as the origin of the search, thereby converting the problem of planning screw sequences into a combinatorial optimization problem: namely, identifying the optimal or near-optimal path combination among all nodes. During the algorithm's initial phase, a distance matrix between screw locations is established and a heuristic function defined. Concurrently, preliminary pheromone values are established along all edges, and parameters such as ant population size, maximum iteration count, pheromone decay coefficient, and pheromone-to-heuristic weighting ratios are determined. In the subsequent phase of path construction, each ant determines the probability of selecting the next screw position based on pheromone concentration and the heuristic function value. The heuristic function [32] is generally established as the inverse of the distance, thereby ensuring that closer fastener is prioritized.

In conclusion, it can be posited that each ant traverses all screw nodes to form a complete disassembly sequence, with task allocation between the two arms achievable through zone partitioning or nearest neighbor strategies [33]. Subsequent to the construction of the pathway, the algorithm will undertake an evaluation of the total length or disassembly time of each path. In scenarios involving two arms, the objective is to minimize maximum operational time to ensure load balancing. Subsequently, the pheromone update phase is initiated. Superior paths, defined by their reduced total lengths or minimized operation times, have been observed to accumulate greater pheromone concentrations. This gradual attraction of ants to select them during iterations is a key finding of the study. Concurrently, older pheromones are released at a controlled rate to prevent premature convergence. The process under discussion involves cyclical iteration through path construction, path evaluation, and pheromone update. The purpose of this iteration is to progressively reinforce superior sequences while eliminating inferior ones. The termination of the algorithm occurs either when the maximum iteration count is reached, or when the optimal solution no longer improves after multiple iterations. The final output screw removal path has been developed to optimize overall disassembly efficiency while ensuring balanced collaboration between the robot's dual arms. The disassembly sequence of double-arm screws based on ACO is shown in Algorithm 1.

Algorithm 1: ACO-based Dual-Arm Screws Disassembly Sequencing

Input: Screw set $V = \{1, \dots, N\}$; dual-arm start poses s_L, s_R; distance matrix $D = [d_{ij}]$
Output: Disassembly sequences $(\boldsymbol{\Pi_L}, \boldsymbol{\Pi_R})$

Initialize: number of ants m, max iterations T_{max}, parameters α, β, ρ, Q; pheromone $\tau_{ij}! \leftarrow ! \tau_0$ for all $i \neq j$; heuristic $\eta_{ij}! \leftarrow ! 1/d_{ij}$.
for $t = 1$ **to** T_{max} **do**
  $(V_L, V_R) \leftarrow \backslash Partition(V, s_L, s_R)$;
  // Optional split (e.g., spatial/KMeans)
  for $k = 1$ **to** m **do**
    $(\boldsymbol{\Pi_L^k}, \boldsymbol{\Pi_R^k}) \leftarrow \backslash Construct(s_L, s_R, \boldsymbol{V_L}, \boldsymbol{V_R}, \tau, \eta, \alpha, \beta)$;
    $(T_L^k, T_R^k, L^k) \leftarrow \backslash Evaluate(\boldsymbol{\Pi_L^k}, \boldsymbol{\Pi_R^k}, \boldsymbol{D})$;
  end
  $k^\star \leftarrow \arg\min \max(\boldsymbol{T_L^k}, \boldsymbol{T_R^k})\backslash$;
  If $\max(T_L^{k^\star}, T_R^{k^\star}) < \max(T_L^{best}, T_R^{best})$
    $(\boldsymbol{\Pi_L^{best}}, \boldsymbol{\Pi_R^{best}}) \leftarrow (\boldsymbol{\Pi_L^{k^\star}}, \boldsymbol{\Pi_R^{k^\star}})$;
    $(T_L^{best}, T_R^{best}) \leftarrow (T_L^{k^\star}, T_R^{k^\star})$;
  end
  Update Pheromone $\{\tau, \{\sqcap_L^k, \sqcap_R^k\}_{k=1}^m, \rho, Q, \{L^k\}\}$;
end
return $(\boldsymbol{\Pi_L^{best}}, \boldsymbol{\Pi_R^{best}})$
Function
  Construct $(s_L, s_R, V_L, V_R, \tau, \eta, \alpha, \beta)$;
  $\sqcap_L \leftarrow [s_L]$, $\sqcap_R \leftarrow [s_R]$; $U \leftarrow V_L \cup V_R$;
  while $U \neq \emptyset$ **do**
    choose arm $a \in \{\mathrm{L}, \mathrm{R}\}$ by nearest-next or load balancing;
    let i be the end of arm a; $C \leftarrow U\backslash$;
    foreach $j \in C$ **do**
      compute weight $w_{ij} = \tau_{ij}^\alpha \eta_{ij}^\beta$
    end
    sample $j^\star$ from normalized w_{ij}; append $j^\star$ to $\boldsymbol{\Pi_a}$; $U \leftarrow U \setminus \{j^\star\}\backslash$;
  end
  return $(\boldsymbol{\Pi_L}, \boldsymbol{\Pi_R})$
end
Function Evaluate Objective $(\boldsymbol{\Pi_L}, \boldsymbol{\Pi_R}, D)$:
  compute travel + operation time $\rightarrow (T_L, T_R)$; $L \leftarrow$ total path length (may include safety costs)\;
  return (T_L, T_R, L)
end
Function
  Update Pheromone$(\tau, \{\boldsymbol{\Pi_L^k}, \boldsymbol{\Pi_R^k}\}, \rho, Q, \{L^k\})$:
  $\tau_{ij} \leftarrow (1 - \rho)\tau_{ij}$ for all $i \neq j$;
  for $k = 1$ **to** m **do**
    foreach edge (i, j) in $\boldsymbol{\Pi_L^k}$ or $\boldsymbol{\Pi_R^k}$ **do**
      $\tau_{ij} \leftarrow \tau_{ij} + Q/L^k$;
    end
  end
end

5. Experimental setup and results

5.1. Visual perception platform setup

The selection of the Mech-Eye LSR-L camera (operating distance 1.3–3.0m) as the visual sensor was made on the basis of the actual dimensions of the battery pack and the operational scenario. The apparatus's substantial measurement range and high depth accuracy satisfy the inspection requirements for the battery pack's structural components and fasteners. In order to guarantee comprehensive coverage of the battery pack's upper surface working area, a fixed aluminum profile support frame measuring 2.05 m in length and 1.98 m in height was constructed. The camera was positioned in such a manner that it was directly above the battery pack, thereby maintaining a vertical imaging perspective relative to the working plane. This approach serves to minimize point cloud gaps caused by obstructions and parallax. At this installation height, the vision system covers a 103.8 × 60.8 cm working area on the battery pack, ensuring reliable identification of fastener and key component geometries.

To achieve precise manipulation of irregular components within battery packs by dual-arm robots, the eye-to-hand and hand-eye calibration method are employed. The system employs a calibration board as an external reference benchmark. The process of acquiring image data of the calibration board across a range of poses serves to establish the spatial relationship between the camera coordinate system and the robot base coordinate system. The core challenge in dual-arm robotic hand-eye calibration lies in determining the transformation matrix between the camera coordinate system T_{Camera} and the robot base coordinate system T_{Base}. The mathematical model can be expressed as follows:

$$T_{End}^{Base} \cdot (T_{Base}^{Camera})^{-1} = T_{End}^{Camera} \cdot T_{Object}^{Camera} \tag{7}$$

The following definitions are employed: $T_{\mathrm{End}}^{\mathrm{Base}}$ the end-effector's pose in the base coordinate system; $T_{\mathrm{Base}}^{\mathrm{Camera}}$ represents the rigid-body transformation from base to camera to be solved; $T_{\mathrm{End}}^{\mathrm{Camera}}$ indicates the end-effector's pose in the camera coordinate system; T_{Object}^{Camera} signifies the target pose obtained via the calibration plate. In practical solution, the calibration process is transformed into:

$$A \cdot X = X \cdot B \tag{8}$$

Where $A = \left(T_{End1}^{Base}\right)^{-1}$ denotes the relative pose between the two end-effector movements; $B = T_{Camera1}^{Object} \cdot \left(T_{Camera2}^{Object}\right)^{-1}$ denotes the relative pose of the calibration plate as observed by the camera; $X = T_{\mathrm{Base}}^{\mathrm{Camera}}$ represents the external parameter matrix from the base to the camera that is to be solved for.

The determination of the parameters T is crucial for the completion of hand-eye calibration. Employing optimization algorithms such as the Least Squares Method or Singular Value Decomposition (SVD) and collecting multiple sets of data from the robot end-effectors and camera observations facilitates the robust determination of the parameters T_{Base}^{Camera}.

This method ensures operational consistency for dual-arm robots within a unified coordinate system, providing an accurate spatial reference for subsequent collaborative disassembly tasks. The calibration relationship is illustrated in Figure 6.

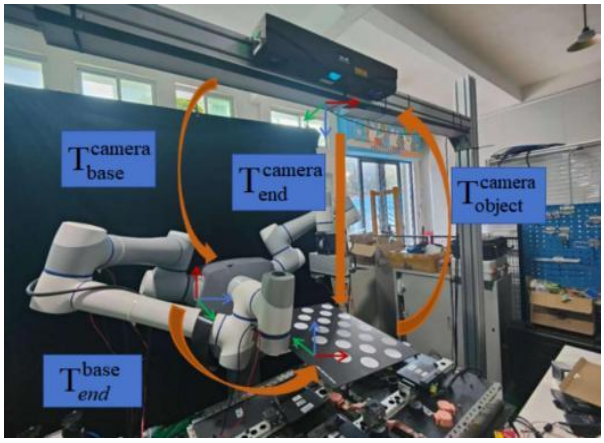


Figure 6. Illustration of coordinate systems and transformations in hand-eye calibration. The figure shows the main coordinate frames and their transformations, including $T_{base \rightarrow end}$ (robot base to end-effector), $T_{camera \rightarrow base}$ (camera to robot base), $T_{camera \rightarrow end}$ (camera to end-effector), and $T_{camera \rightarrow object}$ (camera to object). These transformations constitute the fundamental mathematical relations in hand-eye calibration and pose estimation.

5.2. Platform setup: dual-arm robot with end-effectors

The objective of this study was to validate the proposed method for identifying and dismantling irregular components within battery packs. To this end, an experimental platform was established utilizing the Dobot CR10A dual-arm collaborative robot. The robot exhibits six degrees of freedom, with a single-arm rated payload of 10 kg and a working radius of 1.3 m. It is capable of fulfilling the disassembly requirements for battery pack fastener and structural components. In order to facilitate the automated removal of fastener, specially designed screw extraction devices were installed at the end effectors of the robotic arms. This enabled the disassembly of fastener which were densely distributed on the battery pack. The configuration of the robotic platform is illustrated in Figure 7.

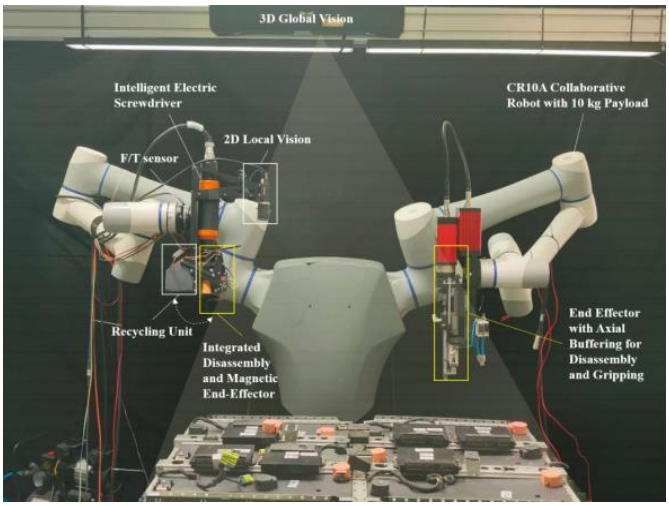


Figure 7. Robotic disassembly system integrating global vision and local visual fusion for precise, adaptive bimanual operations. Dual-arm setup with specialized end effectors enables handling of diverse screw specifications, supporting accurate recognition, fine manipulation, and coordinated disassembly of complex structures.

5.3. Experimental results and analysis

The experiments were conducted on a high-performance tower workstation (Model: AMAXTS40-X3), equipped with an Intel Xeon Silver 4310 CPU, a GeForce RTX 4090 GPU, running on Ubuntu 22.04 with CUDA version 12.4. The dataset was split into training and validation sets, with 80% allocated for training and 20% for validation. The training parameters were as follows: optimizer: Adam W, batch size = 24, epochs = 300, and learning rate = 1×10^{-3}. The components of the battery pack, including their type, material, quantity, and whether they are standard components, are summarized in Table 1. The disassembly task in this experiment focused on M8 nuts, with a total of 56 fasteners involved in the process.

Table 1. Irregular components requiring disassembly within the battery pack.

Component	Material	Number	Standard Part
Fastener (M8, M13)	Carbon Steel	76	Yes
Copper Bar	Copper	5	No
Structural Parts	Carbon Steel	9	No
BMS	Plastic	6	No

The enhanced YOLOv8 demonstrates solid performance in irregular component recognition, with an average precision of 0.961, recall of 0.971, and mAP@50:95 of 0.810. Compared to the baseline YOLOv8 model (mAP@50:95 = 0.749), the improved model shows modest improvements across all metrics, particularly in the fastener category. Specifically, Mask Precision for fasteners increases from 0.814 to 0.907, while Mask Recall improves from 0.772 to 0.927, resulting in a mAP of 0.584. The copperbar category shows strong performance with a recall of 0.993, precision of 0.979, and mAP of 0.926. The model also performs consistently well with structural components (precision: 0.963, recall: 0.965, mAP: 0.835) and BMS components (precision: 0.991, recall: 0.981, mAP: 0.957), demonstrating reliability in recognizing both small and large components.

In addition to these performance improvements, the enhanced model also achieves an overall inference speed of 13.9 ms per image, with the breakdown as follows: 0.5 ms for preprocessing, 8.5 ms for inference, and 4.9 ms for postprocessing. This demonstrates that the model is capable of performing fast and efficient object detection in real-time applications, offering a balance between high accuracy and computational efficiency. The comparison of recognition results between the improved model and the baseline YOLOv8 is shown in Table 2. The results of visual perception are shown in Figure 8.

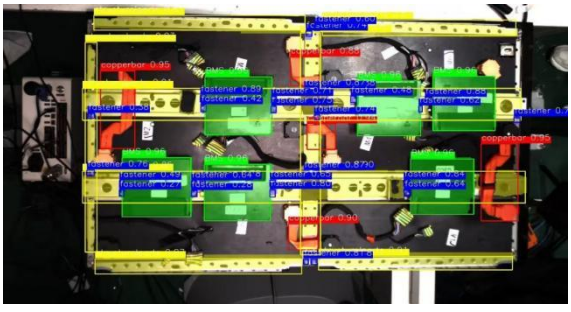


Figure 8. Results of the visual perception achieved by our proposed method, which is capable of recognizing four distinct types of heterogeneous structural components.

Table 2. Component-wise segmentation performance of the baseline and enhanced models.

Model Variant	Class	Mask Precision	Mask Recall	Mask mAP@50:95
YOLOv8	Fastener	0.814	0.772	0.369
	Copperbar	0.973	0.989	0.881
	Structuralparts	0.945	0.964	0.802
	BM	0.972	0.978	0.943
	All types	0.925	0.929	0.749
YOLOv8 (Enhanced)	Fastener	0.907	0.927	0.584
	Copperbar	0.979	0.993	0.926
	Structuralparts	0.963	0.965	0.835
	BMS	0.991	0.981	0.957
	All types	0.961	0.971	0.810

The CATL-172.8V120AH-B battery pack contains a total of 56 M8 fasteners. The dismantling sequence was devised using an ant colony optimization algorithm. The total planned length of both arms is 5.83 m. The planned length of the left arm is 2.91 m, while the planned length of the right arm is 2.93 m. The results of the planning process are presented in Table 3 and Figure 9.

Table 3. Disassembly sequence planning results for dual-arm robot.

Algorithm	Nodes (fastener)	Total Length(m)	Left Arm(m)	Right Arm(m)	Disassembly time(s)
ACO	56	5.83	2.91	2.93	742.19

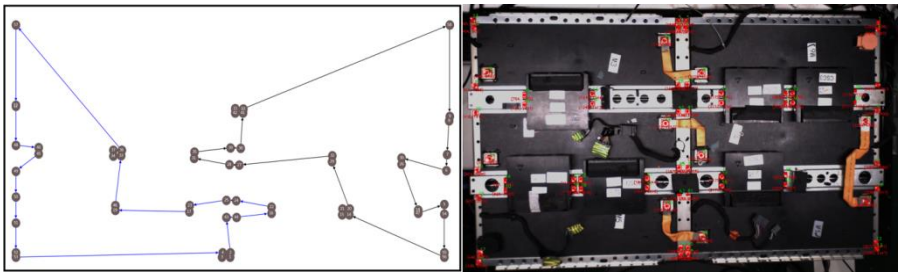


Figure 9. The screw removal sequence obtained via the ant colony algorithm is as follows: nodes 57 and 58 represent the two arms of the dual-arm robot and the remaining nodes denote the fastener on the battery pack. Planning begins at nodes 57 and 58, the starting points, and proceeds sequentially to remove the fastener.

The objective of this operation is to dismantle all non-standard components within the battery pack. The disassembly process is considered to be complete once all components have been removed, with the exception of the battery modules. The final disassembly outcome is illustrated in the Figure 10.



Figure 10. This schematic diagram is to be completed upon the final stage of disassembly, which indicates that all non-standard components have been removed, leaving only six battery modules.

6. Discussion

In this study, we selected a normally retired battery pack featuring internal irregularly shaped components as the primary research object to preliminarily validate the feasibility of the proposed method. Consequently, we have not yet addressed or analyzed potential challenges inherent in real-world dismantling environments—such as missing fasteners, insufficient lighting, or corroded screws. These practical considerations will constitute key directions for our future research.

The two intelligent electric screwdriver end-effectors currently developed are specifically designed for M8 screws; thus, sequence planning in the current disassembly framework is limited to M8 fasteners and does not extend to other types of irregular components. In subsequent work, we aim to design specialized end-effectors tailored to the diverse geometries of components within battery packs, enabling comprehensive disassembly and grasping operations across all part types, along with integrated sequence planning for the entire dismantling process.

The present study focuses exclusively on the identification and disassembly planning for a single type of CMP battery pack. As part of our ongoing research, we intend to generalize and evaluate the proposed methodology across a broader range of retired battery pack architectures to enhance its adaptability and practical applicability.

7. Conclusion

The present paper establishes a dataset comprising retired CMP battery packs from CATL-172.8V120AH-B model buses. The dataset is augmented using traditional spatial geometric transformations, pixel-level transformations, and Copy-Paste data augmentation techniques. A total of three modifications were implemented within the YOLOv8 framework, with the objective of enabling adaptive detection of irregularly shaped components within battery packs. Firstly, a novel P2 detection head was introduced at shallow feature maps to enhance recall for small targets. Secondly, the CARAFE adaptive upscaling technique superseded conventional nearest-neighbor and bilinear interpolation methods, thereby ensuring superior preservation of edge and texture details. Thirdly, the original simple concatenation/additive feature fusion was replaced with ASFF, enabling the model to dynamically balance shallow details with deep semantic features. The experimental results demonstrate the effectiveness of these enhancements in reducing both missed detections and false positives of fastener against complex reflective backgrounds. This, in turn, serves to strengthen the reliability and precision of visual perception in irregular component disassembly tasks.

In order to address the sequencing challenge of disassembling irregular components within battery packs, the paper proposes a model of the task as a Travelling Salesman Problem and employs an ant colony algorithm to plan screw removal sequences. The subsequent process involves the sequencing and dismantling of the remaining irregular components, including structural parts, copper bars, and BMS components. This is undertaken utilizing manual techniques informed by prior knowledge. This approach ensures balanced task allocation in dual-arm scenarios.

Data availability statement

No supplementary or additional data were generated in this study.

Acknowledgments

This work is supported by the National Key Research and Development Program of China (2024YFB4709700), Guangdong Basic and Applied Basic Research Foundation (2025A1515011114, 2024A1515240007), National Natural Science Foundation of China (62203126, 62473102), Outstanding Young Scientist Project of Guangdong Academy of Sciences (2025GDASQNRC-0106),Youth S&T Talent Support Programmed of Guangdong Provincial Association for Science and Technology (SKXRC2025094), Natural Science Foundation of Shandong Province (ZR2024QF013), Jiangmen Science and Technology Program Project (202378020005000921).

Authors' contribution

Conceptualization, Jiangming Zhang and Hongmin Wu; methodology, Jiangming Zhang; software, Jiangming Zhang and Wu Yan; validation, Jiangming Zhang, Junyang Wang and Shaohua Kong; formal analysis, Jiangming Zhang; investigation, Jiangming Zhang, Hongmin Wu, Junyang Wang and Shaohua Kong; resources, Hongmin Wu ,Wu Yan and Xuefeng Zhou; data curation, Jiangming Zhang; writing—original draft preparation, Jiangming Zhang; writing—review and editing, Jiangming Zhang, Hongmin Wu and Xuefeng Zhou; visualization, Jiangming Zhang; supervision, Hongmin Wu; project administration, Hongmin Wu and Xuefeng Zhou; funding acquisition, Hongmin Wu and Xuefeng Zhou. All authors have read and agreed to the published version of the manuscript.

Conflicts of interests

The authors declare no conflict of interest.

References

[1] Cai Y, He M, Chen B. Efficient deep learning based rail fastener screw detection method for fastener screw maintenance robot under complex lighting conditions. *Sci. Rep.* 2024, 14(1):28957.

[2] Chand M, Ravi C. A state-of-the-art literature survey on artificial intelligence techniques for disassembly sequence planning. *CIRP J. Manage. Sci. Rep.* 2023, 41:292–310.

[3] DiFilippo NM, Jouaneh MK, Jedson AD. Optimizing automated detection of cross-recessed screws in laptops using a neural network. *Appl. Sci.* 2024, 14(14):6301.

[4] Burger W, Burge MJ. Scale-invariant feature transform (SIFT). In *Digital Image Processing: An Algorithmic Introduction*. Cham: Springer, 2022. pp. 709–763,

[5] Chen L, Li L, Pan H, Wan Z, Chen J. A cylinder head identification and localization method using OFS-GMS feature matching and minimum surrounding contour. *IEEE Trans. Instrum. Meas.* 2025, 74:1–9.

[6] Zorn M, Ionescu C, Klohs D, Zähl K, Kisseler N, *et al*. An approach for automated disassembly of lithium-ion battery packs and high-quality recycling using computer vision, labeling, and material characterization. *Recycling* 2022, 7(4):48.

[7] Zhou C, Wu Y, Sterkens W, Piessens M, Vandewalle P, *et al*. Towards robotic disassembly: a comparison of coarse-to-fine and multimodal fusion screw detection methods. *J. Manuf. Syst.* 2024, 74:633–646.

[8] Liu J, Xu Z, Xiong H, Lin Q, Xu W, *et al*. Digital twin-driven robotic disassembly sequence dynamic planning under uncertain missing condition. *IEEE Trans. Ind. Inf.* 2023, 19(12):11846–11855.

[9] Wang K, Gao L, Li X, Li P. Energy-efficient robotic parallel disassembly sequence planning for end-of-life products. *IEEE Trans. Autom. Sci. Eng.* 2022, 19(2):1277–1285.

[10] Feng Y, Cui K, Hong Z, Li Z, Yan W, *et al*. Disassembly sequence planning of product structure with an improved qica considering expert consensus for remanufacturing. *IEEE Trans. Ind. Inf.* 2023, 19(5):7201–7213.

[11] Chu M, Chen W. Human-robot collaboration disassembly planning for end-of-life power batteries. *J. Manuf. Syst.* 2023, 69:271–291.

[12] Liu W, Liang X, Zheng M. Task-constrained motion planning considering uncertainty-informed human motion prediction for human-robot collaborative disassembly. *IEEE/ASME Trans. Mechatron.* 2023, 28(4):2056–2063.

[13] Hughes N, Chang Y, Hu S, Talak R, Abdulhai R, *et al*. Foundations of spatial perception for robotics: hierarchical representations and real-time systems. *Int. J. Rob. Res.* 2024, 43(10):1457–1505.

[14] Lyu C, Guo S, Zhou B, Xiong H, Zhou H. 3DOPFormer: 3D occupancy perception from multi-camera images with directional and distance enhancement. *IEEE Trans. Intell. Veh.* 2024, 9:5478–5489.

[15] Weng S, Zhu T, Zhang T, Zhang C. UCM-Net: a U-Net-like tampered-region-related framework for copy-move forgery detection. *IEEE Trans. Multimedia* 2024, 26:750–763.

[16] Terven J, Córdova-Esparza DM, Romero-González JA. A comprehensive review of YOLO architectures in computer vision: from yolov1 to YOLOv8 and YOLO-NAS. *Mach. Learn. Knowl. Extr.* 2023, 5(4):1680–1716.

[17] Junjie P, Dingwei W. An ant colony optimization algorithm for multiple travelling salesman problem. In *First International Conference on Innovative Computing, Information and Control-Volume I (ICI-CIC'06)*, Beijing, China, August 30–September 1, 2006, pp. 210–213.

[18] Grill JB, Strub F, Altché F, Tallec C, Richemond P, *et al*. Bootstrap your own latent: a new approach to self-supervised learning. *Adv. Neural Inf. Process. Syst.* 2020, 33:21271–21284.

[19] Wang CY, Liao HYM, Wu YH, Chen PY, Hsieh JW, *et al*. CSPNet: a new backbone that can enhance learning capability of CNN. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, USA, June 14–19, 2020, pp. 1571–1580.

[20] Chen Z, Ji H, Zhang Y, Zhu Z, Li Y. High-resolution feature pyramid network for small object detection on drone view. *IEEE Trans. Circuits Syst. Video Technol.* 2024, 34(1):475–489.

[21] Sun Y, Feng J. Fire and smoke precise detection method based on the attention mechanism and anchor-free mechanism. *Complex Intell. Syst.* 2023, 9:5185–5198.

[22] Wilcox B, Yip MC. SOLAR-GP: sparse online locally adaptive regression using gaussian processes for bayesian robot model learning and control. *IEEE Rob. Autom. Lett.* 2020, 5(2):2832–2839.

[23] Miranda-Vega JE, Rivas-López M, Fuentes WF. K-nearest neighbor classification for pattern recognition of a reference source light for machine vision system. *IEEE Sens. J.* 2020, 21(10):11514–11521.

[24] Zhou R, Hu W, Fan P, Ian H. Quantum realization of the bilinear interpolation method for NEQR. *Sci. Rep*. 2017, 7(1):2511.

[25] Zhang X, Yu C, Qiu Z, Li J, Lian C, *et al*. Noncontact high-precision aperture detection based on distortion compensation and adaptive mean blur algorithm. *IEEE Trans. Instrum. Meas.* 2023, 72:1–10.

[26] Molina A, Rajamani K, Azadet K. Concurrent dual-band digital predistortion using 2-D lookup tables with bilinear interpolation and extrapolation: Direct least squares coefficient adaptation. *IEEE Trans. Microwave Theory Tech.* 2017, 65(4):1381–1393.

[27] Wang J, Chen K, Xu R, Liu Z, Loy CC, *et al*. Carafe++: unified content-aware reassembly of features. *IEEE Trans. Pattern Anal. Mach. Intell.* 2022, 44(9):4674–4687.

[28] Lee S, Lee J, Lee J. Learning virtual chimeras by dynamic motion reassembly. *ACM Trans. Graphics* 2022, 41(6):1–13.

[29] Huang X, Zhu J, Huo Y. SSA-YOLO: an improved yolo for hot-rolled strip steel surface defect detection. *IEEE Trans. Instrum. Meas.* 2024, 73:1–17.

[30] Cheikhrouhou O, Khoufi I. A comprehensive survey on the multiple travelling salesman problem: Applications, approaches and taxonomy. *Comput. Sci. Rev.* 2021, 40:100369.

[31] Liu J, Anavatti S, Garratt M, Abbas HA. Modified continuous ant colony optimisation for multiple unmanned ground vehicle path planning. *Expert Syst. Appl.* 2022, 196:116605.

[32] Huang B, Zhou M, Abusorrah A, Sedraoui K. Scheduling robotic cellular manufacturing systems with timed petri net, A* search, and admissible heuristic function. *IEEE Trans. Autom. Sci. Eng.* 2022, 19(1):243–250.

[33] Pudaruth S, Soyjaudah KMS, Gunputh RP. An innovative multi-segment strategy for the classification of legal judgments using the k-nearest neighbour classifier. *Complex Intell. Syst.* 2018, 4(1):1–10.