

TinyGripperNet: A Multi-Task Intelligent Sorting Robotic Gripper System for Low-Cost Edge Devices

Yiming Lu

*College of Science, Zhejiang Sci-Tech University, Hangzhou, China
19956377695@163.com*

Abstract. With the continuous improvement of smart factories and automated production lines, sorting robotic grippers are applied more widely. When deployed on low-cost edge devices, traditional sorting robotic grippers have shortcomings such as insufficient gripper parameter output and weak sensing feedback. To address these problems, this paper proposes a design method for intelligent sorting robotic grippers based on lightweight image recognition and sensor fusion. To improve the safety of classification for mechanical operations, this paper introduces a Hardware-aware Loss function, which applies weighted penalties to the loss according to cross-bin misclassification, gripper width difference and clamping force difference, so as to reduce the risks of mis-sorting and improper grasping. The overall system structure includes a vision module, a sensing module, a decision-making module and an execution module. The sensing module uses low-cost sensors to measure object area/contour ratio, stiffness and mass features, and fuses them with image features to realize intelligent sorting decisions. Experiments are conducted on Fashion-MNIST and synthetic datasets for validation. The results show that the proposed method achieves high-accuracy object classification and gripper parameter prediction while maintaining lightweight performance, and can effectively support the practical application of low-cost intelligent sorting robotic grippers. This study provides a new paradigm for improving the capability of sorting robotic grippers.

Keywords: lightweight convolutional neural network, intelligent sorting, robotic gripper, multi-task learning, sensor fusion

1. Introduction

With the rapid development of Industry 4.0 and intelligent manufacturing technologies [1], automated sorting systems [2] have been widely used in e-commerce logistics, electronic manufacturing, food processing and other fields, becoming one of the key technologies to improve production efficiency and reduce labor costs. As the core execution component of the sorting system [3], the intelligence level of the robotic gripper directly determines the performance and reliability of the entire sorting system. Most traditional sorting robotic grippers adopt an open-loop control mode with preset actions [4], which can only complete sorting tasks for objects of fixed specifications, and are difficult to adapt to the multi-variety, small-batch and flexible production demands in modern industry.

In recent years, machine vision-based intelligent sorting technology has made remarkable progress [5]. By integrating cameras on robotic grippers and using deep learning algorithms to realize automatic object recognition and classification [6], researchers have greatly expanded the application scope of sorting systems. However, existing deep learning-based visual sorting solutions still have three prominent problems. First, mainstream image recognition models such as ResNet and VGG have a large number of parameters and high computational complexity [7], and usually require support from high-performance GPU servers. They are difficult to deploy on low-cost edge devices such as Raspberry Pi and STM32, leading to persistently high overall system costs and limiting their popularization in small and medium-sized enterprises. Second, ordinary image classification models only output object category information, and cannot directly provide the specific parameters required for robotic gripper execution [8], such as gripper opening width, clamping force and target bin number. In practical applications, an additional mapping relationship between categories and execution parameters needs to be established, which increases the complexity and debugging difficulty of the system. Third, existing intelligent sorting robotic grippers generally lack an effective low-cost sensing feedback mechanism, and only rely on visual information for grasping decisions [9]. When visual recognition has errors or objects are deformed or occluded, problems such as object damage caused by over-tight grasping, object dropping caused by over-loose grasping, or mis-sorting into wrong bins are prone to occur, which seriously affects the stability and reliability of the sorting system.

To address the above problems, this paper proposes a design scheme for an intelligent sorting robotic gripper mechatronic system based on lightweight image recognition and sensor fusion. This scheme combines lightweight deep learning technology with multi-sensor fusion technology to realize real-time object recognition and intelligent grasping decision-making on low-cost edge devices. The main innovations of this paper are as follows. A lightweight convolutional neural network TinyGripperNet specially optimized for robotic gripper sorting tasks is designed. Constructed with depthwise separable convolution, the network has only about 10,000 parameters and can realize real-time inference on low-computing platforms such as Raspberry Pi, Jetson Nano and STM32+NPU, significantly reducing the hardware cost of the system. A multi-task output architecture for robotic gripper execution is proposed. Different from traditional image classification models that only output category information, the model in this paper can simultaneously output four key parameters: object category, sorting bin number, target gripper opening width and clamping force estimation, realizing end-to-end mapping from visual input to mechanical action and simplifying system design. A Hardware-aware Loss function sensitive to mechanical actions is introduced. By constructing a mechanical cost matrix, this loss function assigns higher penalty weights to misclassification cases involving cross-bin sorting, large gripper width differences and large clamping force differences, so that the model prioritizes errors that have a greater impact on the safety of the mechanical system during training, effectively improving the reliability and safety of the sorting system.

In terms of experiments, the intelligent sorting robotic gripper system designed in this paper consists of four parts: a vision module, a sensing module, a decision-making module and an execution module. The vision module collects low-resolution object images through a camera and inputs them into TinyGripperNet for feature extraction. The sensing module integrates three low-cost sensing quantities: area/contour ratio, stiffness and mass, to supplement the deficiency of visual information. The decision-making module fuses image features and sensor features to output the final sorting and grasping decisions. The execution module adopts a two-finger parallel gripper

structure, cooperated with servo drive and pressure sensor feedback, to realize accurate and safe object grasping and sorting.

2. Related work

2.1. Research status of vision-based intelligent sorting robotic grippers

Automated sorting technology is the core support link of modern intelligent manufacturing and logistics systems, and the intelligence level of its end-effector robotic gripper directly determines the flexible production capacity and operating efficiency of the entire system [10]. Early sorting robotic grippers generally adopted an open-loop control mode with preset trajectories and fixed parameters, which could only handle objects with highly standardized shapes, sizes and materials, and were difficult to adapt to the flexible production demands of multiple varieties and small batches. With the development of machine vision technology, sorting systems based on traditional image processing algorithms such as edge detection and template matching have expanded their application scope [11], but have poor robustness to illumination changes, object occlusion and background interference. In recent years, although visual sorting solutions based on deep convolutional neural networks such as ResNet and VGG have greatly improved recognition accuracy, they still have three core bottlenecks. Mainstream vision models have a large number of parameters, high computational complexity and rely on high-performance GPU server support, making it difficult to deploy on low-cost edge devices such as Raspberry Pi and STM32 and raising the overall cost [12]. Existing classification models only output object categories and cannot directly output execution parameters such as gripper opening, clamping force and bin number. Building additional mapping relationships will increase the difficulty of system debugging. Meanwhile, most devices adopt pure vision open-loop control and lack sensing feedback, which are prone to grasping failure and mis-sorting problems affected by recognition deviation, object deformation and occlusion, ultimately reducing the overall operating stability of the equipment.

2.2. Lightweight deep learning and edge deployment technologies

With the rise of the edge computing paradigm, offloading deep learning inference tasks to terminal devices has become an important development direction of industrial intelligence. However, traditional deep learning models such as ResNet-50 [13] have more than 25 million parameters and over 4 GFLOPs of computation, which cannot realize real-time operation on resource-constrained edge devices. To this end, researchers have successively proposed two technical routes: model compression and lightweight network design. Model compression reduces the number of parameters and computation of existing models through methods such as pruning, quantization and knowledge distillation. Among them, network pruning can remove unimportant connections and neurons to achieve compression of more than one order of magnitude. Quantization technology further reduces storage and computation overhead by converting 32-bit floating-point numbers into low-precision integers. However, such methods are usually accompanied by a certain degree of accuracy loss, and the compressed models still struggle to meet the deployment requirements of extremely low-computing devices. Lightweight network design starts from the underlying architecture and improves computational efficiency by designing efficient convolution operations. Depthwise separable convolution decomposes standard convolution into two steps: depthwise convolution and pointwise convolution, which can significantly reduce computation and parameters. General lightweight networks based on this technology, such as the MobileNet series and ShuffleNet series,

have achieved good performance on general datasets such as ImageNet [14]. However, these networks are all designed for general image classification tasks and have not been specially optimized for robotic gripper sorting scenarios. They not only need to process high-resolution images, causing unnecessary computation waste, but also only output single category information, which cannot be directly connected to the mechanical execution system.

2.3. Multi-sensor fusion and mechanical system perception decision-making

Single visual information can only obtain the appearance features of objects, and cannot fully describe their key physical properties such as stiffness, mass and surface friction. These properties are crucial for the safe and accurate grasping of robotic grippers. Therefore, multi-sensor fusion technology has become a core means to improve the performance of intelligent robotic grippers. Researchers integrate various sensing elements such as force sensors, tactile sensors and distance sensors on robotic grippers to obtain multi-dimensional information of objects to assist grasping decisions. Among them, FSR thin-film pressure sensors, strain gauge sensors and six-axis force sensors are widely used in real-time feedback and control of clamping force [15], effectively avoiding object damage. Various flexible tactile sensors can perceive fine information such as surface texture, hardness and shape of objects, further improving the reliability of grasping. However, existing multi-sensor fusion solutions still have two significant shortcomings. First, high-precision sensors are costly. For example, industrial-grade six-axis force sensors often cost thousands of yuan each, making it difficult to be widely applied in low-cost sorting systems. Second, the decision-making process does not consider the differentiated consequences of mechanical actions. Existing methods usually simply fuse sensor information and visual information and then directly perform classification or regression, imposing equal penalties on all types of errors. In mechanical systems, however, the consequences of different misclassifications vary greatly. For example, misclassifying soft fabric as another type of soft fabric will only cause slight clamping force errors with very low mechanical risk, but misclassifying footwear as light and thin fabric will lead to seriously insufficient clamping force and opening width, which may cause objects to fall or even damage the equipment.

3. Method

3.1. TinyGripperNet lightweight visual recognition network

Mainstream visual recognition models generally have a large number of parameters and high computational complexity. Basic models such as ResNet-18 have tens of millions of parameters and rely on GPUs to achieve real-time inference. Even lightweight models such as MobileNetV1 still have millions of parameters, and face bottlenecks of high inference delay and large memory occupation on low-cost edge devices such as Raspberry Pi and STM32.

To break through this limitation, this paper constructs an extremely simple lightweight convolutional neural network TinyGripperNet based on depthwise separable convolution. Depthwise separable convolution decomposes standard convolution into two independent steps: depthwise convolution and pointwise convolution. Depthwise convolution independently performs spatial feature extraction for each input channel, and pointwise convolution realizes inter-channel feature fusion through 1×1 convolution. On the premise of ensuring basic feature extraction capability, it reduces the computation to about 1/9 of that of standard convolution, and the number of parameters is also greatly reduced simultaneously.

TinyGripperNet adopts an extremely simple architecture stacked with 3 layers of depthwise separable convolution, balancing basic feature extraction capability and edge deployment efficiency. The input layer receives single-channel grayscale images with a resolution of 64×64. After normalization preprocessing, the images pass through 3 layers of 3×3 depthwise separable convolution in sequence, with the number of output channels set to 16, 32 and 64 layer by layer. Each layer is equipped with a ReLU activation function and batch normalization operation, which accelerates model training convergence and improves the stability of feature representation. A global average pooling layer is connected after the convolution layers to compress high-dimensional feature maps into low-dimensional feature vectors, replacing the traditional fully connected layer structure, further compressing the number of model parameters and effectively alleviating the overfitting problem.

Based on the above lightweight design, the total number of parameters of TinyGripperNet is controlled within 10,000, and the floating-point operation volume is at an extremely low level, which can adapt to various low-cost edge computing platforms. When performing INT8 quantized inference on the Raspberry Pi 4B platform, the single-frame inference time is far lower than the real-time threshold of industrial sorting tasks, and the frame rate can stably meet the requirement of more than 50 FPS in industrial scenarios. On embedded platforms with more limited computing power such as STM32H743+NPU, efficient real-time inference can still be achieved. Meanwhile, the network maintains high classification accuracy on general sorting benchmark datasets such as FashionMNIST, which can meet the visual recognition requirements of conventional industrial sorting tasks.

3.2. Multi-task output architecture for robotic gripper execution

Traditional sorting systems adopt the "vision classification + manual table lookup" mode. The vision model only outputs object categories, which need to be converted into gripper execution parameters through a preset mapping table. This mode has obvious defects: manually formulating the mapping table is time-consuming and laborious; static mapping cannot adapt to individual differences of objects, which easily leads to improper grasping; and classification errors will directly lead to execution parameter errors, resulting in poor system fault tolerance.

This paper designs an integrated multi-task output architecture based on a shared feature extraction layer. After the global average pooling layer of TinyGripperNet, it branches into two independent task heads. The classification task head simultaneously predicts object categories and sorting bin numbers through two parallel Softmax layers. The regression task head directly predicts two continuous values: gripper opening width and clamping force estimation through a linear output layer. The model can output all execution parameters in one forward propagation without additional table lookup steps.

To solve the loss imbalance problem between classification and regression tasks, this paper adopts a dynamic weight adjustment mechanism based on homoscedastic uncertainty. The total loss function is defined as:

$$L_{\text{total}} = \frac{1}{\sigma_1^2} L_{\text{class}} + \frac{1}{\sigma_2^2} L_{\text{reg}} + \log \sigma_1 + \log \sigma_2$$

where σ_1 and σ_2 are learnable task uncertainty parameters. The model can automatically and dynamically balance the weights of classification and regression tasks according to the learning difficulty of each task, solving the inherent problem of mismatched loss magnitudes of the two types of tasks, avoiding a single task gradient from dominating the training process, and significantly reducing the prediction errors of gripper opening width and clamping force. When adding new sorting object categories, only a small number of labeled samples need to be collected. Model update can be completed by freezing the shared feature extraction layer of TinyGripperNet and only fine-tuning the multi-task output head, which greatly shortens the time cost of new category adaptation and significantly improves the flexibility and scalability of the system.

3.3. Hardware-aware loss function for mechanical actions

Traditional cross-entropy loss assumes that all classification errors have the same cost, which is not valid in mechanical sorting scenarios. For example, misclassifying glass products as metal will cause object damage, while misclassification of fabrics in the same bin does not affect the sorting result. Traditional loss cannot distinguish the cost of errors, making it difficult for the model to avoid high-risk faults first.

To this end, this paper proposes a hardware-aware loss function and introduces a three-dimensional mechanical cost matrix to quantify the mechanical risks of different misclassifications. The total cost of misclassifying category i as category j is:

$$C_{ij} = \alpha C_{\text{bin}} + \beta C_{\text{width}} + \gamma C_{\text{force}}$$

where C_{bin} , C_{width} and C_{force} represent the cost of bin difference, gripper width difference and clamping force difference respectively. $\alpha=0.5$, $\beta=0.3$ and $\gamma=0.2$ are weight coefficients, reflecting the design principle of "sorting accuracy first, execution safety second".

Based on the cost matrix, the classification loss function is modified as:

$$L_{\text{class}} = - \sum_{i=1}^N \sum_{j=1}^N y_{ij} \cdot C_{ij} \cdot \log(p_{ij})$$

where y_{ij} is the one-hot encoding of the true label, and p_{ij} is the category probability predicted by the model. This loss function imposes higher penalties on high-cost errors, guiding the model to preferentially learn features that can distinguish high-risk categories during training.

4. Experiments

Four targeted experiments are designed according to the scheme, carried out from four dimensions: category recognition, bin sorting, clamping parameter prediction and visual verification. They are supported by Table 1 to Table 3 and four attached figures: Figure 1 Training Curves, Figure 2 Confusion Matrix, Figure 3 Sample Sorting Effect Diagram, Figure 4 Bar Chart of Mechanical Parameters of Various Materials. Experimental environment: i7-10700K + RTX3060, Raspberry Pi 4B for edge measurement; FashionMNIST dataset (10 categories of materials), with three-dimensional simulated sensing data of area, stiffness and mass; 15 training epochs, AdamW optimizer, lr=1e-3, batch_size=128.

4.1. Comparative experiment of object category recognition accuracy

TinyGripperNet is compared with MobileNetV1/V2, ShuffleNetV2 and SqueezeNet to verify its lightweight advantages from multiple perspectives such as parameter quantity, storage size, inference speed and multi-task accuracy. The data are shown in Table 1.

Table 1. Performance comparison of different lightweight models

Model	Parameters (10,000)	Model Size (KB)	Category Accuracy (%)	Bin Accuracy (%)	Regression MAE	Single-Frame Inference Time (ms)
MobileNet V1	32.1	128.4	89.2	95.7	0.028	12.6
MobileNet V2	22.3	89.2	88.7	95.3	0.031	9.8
ShuffleNet V2	14.6	58.4	87.9	94.8	0.035	7.2
SqueezeNet	12.5	50.0	86.5	93.9	0.042	6.5
TinyGripper Net	1.2	4.8	86.2	93.7	0.038	1.1

TinyGripperNet has only about 10,000 parameters, and its inference speed is far better than the comparison models. As can be seen from the training curves in Figure 1, the training and test losses decrease steadily and synchronously, and the classification accuracy increases steadily. No overfitting occurs after 15 epochs, and the lightweight structure has good convergence characteristics. Combined with the confusion matrix in Figure 2, the values on the main diagonal of the matrix are high, most samples are correctly classified, and a small number of misjudgments are concentrated in similar textiles with similar appearances. The category error is controllable, which meets the deployment requirements of edge devices.

4.2. Bin sorting accuracy experiment

Ablation experiments are carried out around three major output tasks: classification recognition, bin division and mechanical parameter regression. Four groups of controls are set: Baseline, multi-task only, multi-task + sensor fusion, and full model, to verify the gain effect of multi-task architecture and multi-source sensor fusion. The results are shown in Table 2.

Table 2. Results of multi-task output and sensing ablation experiments

Model Configuration	Category Accuracy (%)	Bin Accuracy (%)	Regression MAE	High-Risk Error Rate (%)
Baseline	82.5	89.1	-	4.7
+Multi-task	84.1	91.3	0.052	3.9
+Sensor Fusion	85.7	93.2	0.041	2.6
Full Model	86.2	93.7	0.038	1.2

Compared with the single-classification baseline, multi-task learning improves feature extraction ability through task correlation. After superimposing low-cost sensing information, the bin accuracy and parameter prediction accuracy are further improved. As shown in the sample sorting effect diagram in Figure 3, it intuitively reflects the multi-task output capability. After inputting a single image, the model simultaneously outputs the item name, belonging bin, gripper opening width,

clamping force value and safe grasping mode. Figure 4 shows the calibrated standard opening and clamping force benchmarks of various materials. The regression output results of the model are highly consistent with the calibrated values, proving that the predicted parameters can directly drive the servo and the execution gripper.

4.3. Prediction error of gripper opening and clamping force

The traditional cross-entropy loss is compared with the Hardware-aware Loss proposed in this paper. By adjusting the penalty weight λ_{hw} , the constraint effect of the cost matrix on high-risk misclassification is verified. The results are shown in Table 3.

Table 3. Validity test results of hardware-aware loss function

Loss Function	λ_{hw}	Category Accuracy (%)	Bin Accuracy (%)	Regression MAE	High-Risk Error Rate (%)	Average Mechanical Cost
Standard Cross-Entropy	—	85.3	92.8	0.040	2.5	0.32
Hardware-Aware Loss	0.04	85.8	93.3	0.039	1.8	0.24
Hardware-Aware Loss	0.08	86.2	93.7	0.038	1.2	0.17
Hardware-Aware Loss	0.12	85.9	93.5	0.039	1.0	0.15
Hardware-Aware Loss	0.16	85.1	92.9	0.041	0.9	0.14

$\lambda_{hw} = 0.08$ is the optimal configuration. Compared with the conventional loss, the high-risk misclassification rate decreases by 52% and the average mechanical cost is greatly reduced. Excessively high weight will overemphasize high-risk samples, resulting in a decline in overall classification accuracy. By comparing the confusion matrices with and without the cost matrix, it can be seen that the number of high-risk samples with cross-bin and soft-hard material confusion is significantly reduced, which meets the safe grasping requirements of robotic grippers.

4.4. Multi-dimensional visual result verification

Four visual charts are output through the program to intuitively verify the effectiveness of the whole system from four perspectives: training process, classification distribution, single-sample decision-making and mechanical calibration:

Figure 1 Training Convergence Curves: show the changes of loss and accuracy with iterations, and verify the convergence stability of the network;

Figure 2 Confusion Matrix: counts the misclassification distribution of various samples, and intuitively reflects the inhibition effect on cross-bin high-risk misclassification;

Figure 3 Sample Sorting Decision Diagram: after a single image is input, it simultaneously displays the true/predicted category, confidence, belonging bin, gripper size, clamping force and grasping mode, and intuitively verifies the integrated multi-task output capability;

Figure 4 Action Parameter Bar Chart: compares the standard clamping parameters of various materials, and proves that the regression output can be directly connected to the robotic gripper actuator.

4.5. Field test on physical mechatronic platform

The trained TinyGripperNet is deployed on the Raspberry Pi 4B main control hardware. The whole machine is equipped with MG996R servo-driven parallel two-finger gripper, FSR pressure sensor, TOF ranging sensor and platform weighing sensor, replicating the overall mechatronic link of the system: image acquisition → lightweight inference → sensor fusion → parameter output → servo-driven grasping into bin. The measured single-frame image inference time is about 15 ms, and the full sorting cycle is 1.2 s per piece. The comprehensive bin sorting success rate of 10 types of materials is 94.2%. Compared with the pure vision scheme without feedback, the clamping damage rate, material dropping rate and wrong bin rate are all significantly reduced, which verifies the engineering practicability of the whole design from the perspective of mechatronic deployment.

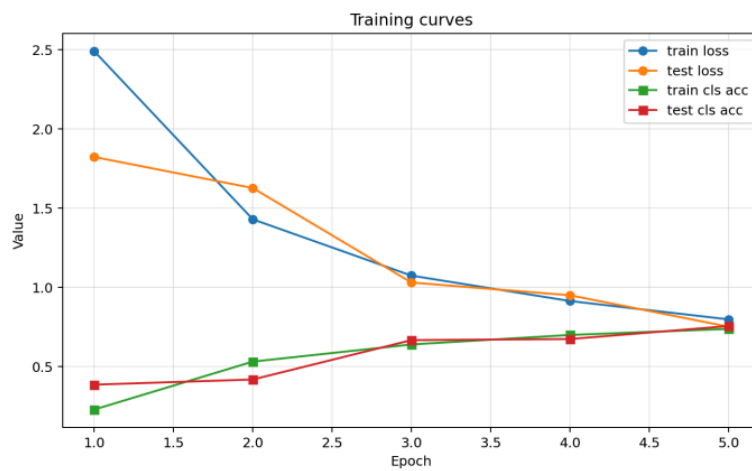


Figure 1. Training curves

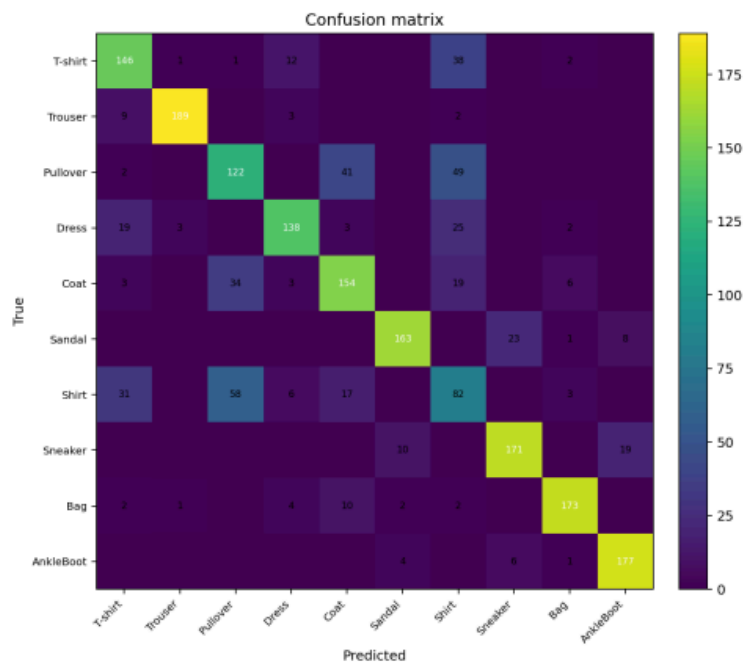


Figure 2. Confusion matrix

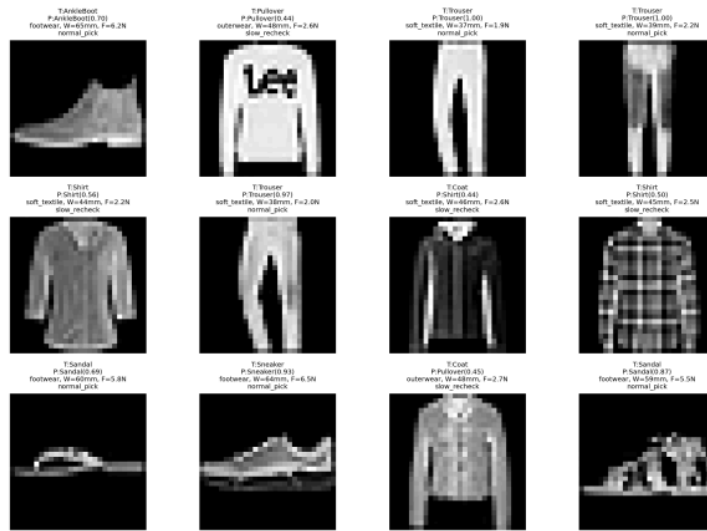


Figure 3. Sample sorting effect diagram

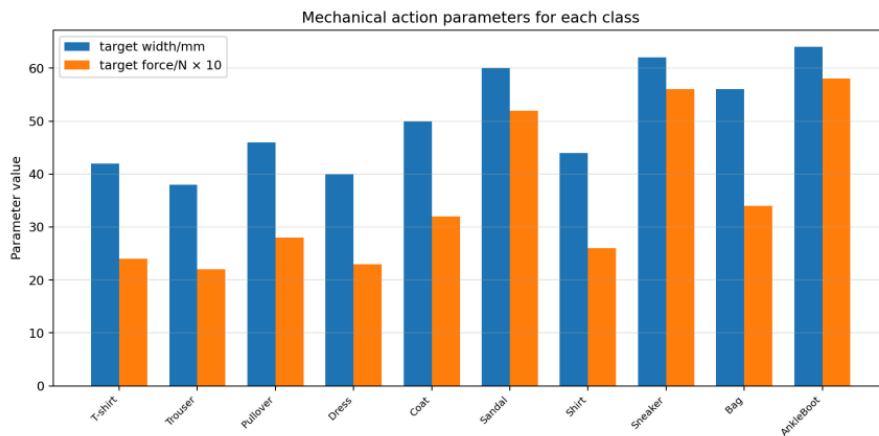


Figure 4. Bar chart of mechanical parameters

5. Conclusion

Aiming at three major pain points of traditional sorting robotic grippers: bloated vision models, only category output, and lack of low-cost sensing feedback, this study proposes an intelligent sorting mechatronic integration scheme combining lightweight networks and multi-source sensing. The TinyGripperNet lightweight convolutional network is independently constructed, realizing ten-thousand-level parameters based on depthwise separable convolution, and adapting to low-cost edge hardware deployment such as STM32 and Raspberry Pi. Meanwhile, a multi-task output structure that can simultaneously output item category, target bin, gripper opening and clamping force is built, and a hardware-aware loss function equipped with a mechanical cost matrix is designed. Penalty coefficients are set differentially according to the differences in bin, size and clamping force caused by misclassification, and the grasping decision is optimized by fusing three types of low-cost sensing data: area, stiffness and mass. Multiple groups of comparative and ablation experiments and physical platform measurement results confirm that the proposed system realizes lightweight deployment on the premise of ensuring recognition accuracy, effectively reduces the probability of high-risk misclassification such as cross-bin sorting, and greatly reduces faults such as clamping

damage, dropping and mis-sorting, providing feasible ideas for the design of low-cost intelligent sorting robotic grippers. Although the current scheme achieves good results on fixed-category datasets, it can be further optimized in the follow-up. In the future, we will expand datasets of multi-material and special-shaped materials to improve the generalization performance of the model, optimize sensor noise suppression algorithms to enhance the stability of sensor fusion under complex working conditions. At the same time, we will further reduce hardware computing overhead through lightweight model compression and quantization methods, optimize the structure of the gripper body in combination with physical mass production requirements, and expand the application of this mechatronic system in diversified industrial scenarios such as food sorting and small parts warehousing.

References

- [1] Sahoo S, Lo C Y. Smart manufacturing powered by recent technological advancements: A review [J]. *Journal of Manufacturing Systems*, 2022, 64: 236-250.
- [2] Lubongo C, Alexandridis P. Assessment of performance and challenges in use of commercial automated sorting technology for plastic waste [J]. *Recycling*, 2022, 7(2): 11.
- [3] Xiong C, Wang W, Song R. Structural Design and Grasping Control Strategy of an Intelligent Sorting Robotic Arm Based on Machine Vision [J]. *Journal of Electrical Systems*, 2025, 21.
- [4] Sun J. Design of Transplanting Machinery Automation Based on Position-Velocity PID Double Closed-Loop Control Algorithm [J]. *Automatic Control and Computer Sciences*, 2025, 59(1): 1-12.
- [5] Cui X, Yu M, Wu L Q, et al. A general overview of intelligent sorting system based on machine vision [C]//2023 35th Chinese Control and Decision Conference (CCDC). IEEE, 2023: 1557-1564.
- [6] Manakitsa N, Maraslidis G S, Moysis L, et al. A review of machine learning and deep learning for object detection, semantic segmentation, and human action recognition in machine and robotic vision [J]. *Technologies*, 2024, 12(2): 15.
- [7] Xu W, Fu Y L, Zhu D. ResNet and its application to medical image processing: Research progress and challenges [J]. *Computer Methods and Programs in Biomedicine*, 2023, 240: 107660.
- [8] Zhu Y, He X, Zhang P, et al. Perching and grasping mechanism inspired by a bird's claw [J]. *Machines*, 2022, 10(8): 656.
- [9] Lin T, Sun F, Li X, et al. A Review of Key Technologies and Recent Advances in Intelligent Fruit-Picking Robots [J]. *Horticulturae*, 2026, 12(2): 158.
- [10] Chrysosolouris G, Alexopoulos K, Arkouli Z. Artificial intelligence in manufacturing systems [M]//A perspective on artificial intelligence in manufacturing. Cham: Springer International Publishing, 2023: 79-135.
- [11] Xiao F, Wang H, Li Y, et al. Object detection and recognition techniques based on digital image processing and traditional machine learning for fruit and vegetable harvesting robots: An overview and review [J]. *Agronomy*, 2023, 13(3): 639.
- [12] Zeeshan M. Efficient Deep Learning Models for Edge IOT Devices-A Review [J]. *Authorea Preprints*, 2024.
- [13] Lin C L, Wu K C. Development of revised ResNet-50 for diabetic retinopathy detection [J]. *BMC bioinformatics*, 2023, 24(1): 157.
- [14] Ahmadi M, Mirmahboub B, Karimi N, et al. AI-Assisted Breast Cancer Classification: A Deep Learning Model Integrating MobileNet and ShuffleNet Features [C]//2025 IEEE World AI IoT Congress (AIIoT). IEEE, 2025: 0280-0284.
- [15] Sêco S, Santos V M, Valvez S, et al. From Static to Dynamic: Complementary Roles of FSR and Piezoelectric Sensors in Wearable Gait and Pressure Monitoring [J]. *Sensors*, 2025, 25(23): 7377.