

ORIGINAL ARTICLE

International Journal of Applied Mathematics in Control Engineering

Journal homepage: <http://www.ijamce.com>

A Lightweight Garbage Classification Algorithm Based on The Improved YOLOv10s

Xiaoyang Jiao¹ | Ruyao Dou¹ | Tianhuan Liu¹ | Chunmeng Huang¹

¹School of Science, North China University of Science and Technology, Tangshan, Hebei, 063200, China

Correspondence

Ruyao Dou, School of Science, North China University of Science and Technology, Tangshan, Hebei, 063200, China
Email: 3199119266@qq.com

Article Info

Article history:

Received 3 March 2026

Accepted 7 May 2026

Available online 11 May 2026

Abstract

Aiming at the problems of small-scale garbage missed detection, misjudgment of similar feature garbage, insufficient detection accuracy of occluded and stacked targets, and limited lightweight degree of general object detection models in the garbage classification scenario, a lightweight garbage classification detection algorithm based on the improved YOLOv10s (YOLOv10s-CA) is proposed. This algorithm is fully implemented based on the Python deep learning framework, taking YOLOv10s as the basic framework. The CoordinateAttention (CA) coordinate attention mechanism is embedded in the feature fusion layer and the detection head branch to enhance the target feature expression ability. The regression loss function is replaced with WIoUv4, and the DIOU-NMS is adopted to optimize the post-processing process, improving the accuracy of bounding box regression. At the same time, a data augmentation strategy dedicated to the garbage classification scenario is designed, and combined with model structured pruning and INT8 quantization to achieve extreme lightweight. Experimental results show that under the unified test environment, the precision rate of YOLOv10s-CA is 70%, compared with YOLOv10n and YOLOv10s, the proposed algorithm achieves the best balance in detection accuracy and lightweight performance.

KEYWORDS

garbage classification, YOLOv10s, lightweight network, attention mechanism, loss function

1 | INTRODUCTION

With the continuous growth of urban domestic waste production, intelligent waste classification has become a core technical support for enhancing the efficiency of waste collection and treatment and promoting the modernization of sanitation operations. Object detection algorithms based on deep learning, with their efficient feature extraction and end-to-end reasoning capabilities, have been widely applied in key links such as automatic identification, intelligent sorting and classified transportation of domestic waste. The YOLO series of single-stage object detection algorithms, due to their advantages such as fast detection speed, lightweight structure, and flexible deployment, have become the preferred solution for lightweight visual detection tasks in vehicle

vision, edge computing, and other fields. As the latest generation of lightweight object detection model, YOLOv10 has achieved comprehensive optimization in network structure, decoupled detection heads, and computational efficiency[1]. Among them, YOLOv10s strikes a good balance between detection accuracy and computational cost, providing an efficient and reliable basic algorithm framework for the on-board real-time garbage detection system of mobile garbage trucks.

At present, the garbage classification detection task for mobile garbage truck operation scenarios still faces several prominent technical difficulties: First, the scale of garbage targets varies significantly, and small-scale garbage such as batteries and cigarette butts are prone to be missed at the on-board shooting distance. Second, the characteristics of similar-looking garbage such as plastic boxes, plastic bottles, and paper have low discrimination, which can easily lead to category misjudgment. Thirdly, during the road surface and the garbage collection process, the garbage often appears in a stacked and occluded state, resulting in blurred target boundaries and insufficient regression accuracy of the bounding box. Fourth, the computing power of vehicle-mounted terminals is limited. The existing lightweight models still have problems such as parameter redundancy and high computational load[4,5], which are difficult to meet the real-time requirements of mobile operations.

Therefore, based on the YOLOv10s model, this paper proposes a lightweight garbage classification detection algorithm YOLOv10s CA for on-board detection of mobile garbage trucks. The algorithm achieves lightweight compression by embedding coordinate attention mechanisms in multi-scale feature branches, optimizing bounding box regression losses and post-processing strategies, designing dedicated data augmentation schemes for vehicle scenarios, and combining model pruning and quantization, making the model more suitable for dynamic recognition, small target detection, and complex interference environments in mobile operations[7,17]. This paper comprehensively demonstrates the accuracy advantage, lightweight performance and engineering practicality of the proposed algorithm through ablation experiments, YOLOv10 series comparative experiments, robustness experiments, convergence experiments and cross-scenario verification. It can provide efficient and stable algorithmic support for the intelligent garbage classification system of mobile garbage trucks.

2 | RELEVANT TECHNOLOGIES AND THEORIES

With the intelligent upgrade of urban sanitation, mobile garbage trucks, as the core carrier for domestic waste collection and transportation, have put forward dual requirements of detection accuracy and lightweight performance for real-time garbage classification detection at the on-board end. The object detection algorithm based on deep learning is the core support of the vehicle-mounted intelligent classification system. The following is a detailed elaboration of the core theoretical basis involved in this paper.

2.1 | YOLOv10s network structure

YOLOv10s adopts the classic architecture of "backbone network - feature fusion - detection head"[1]: the backbone network is a C32 module, which extracts image features through residual connections; The feature fusion layer adopts the PAFPN structure to achieve up-down sampling and fusion of multi-scale features. The detection head is a DecoupledHead, which separates the classification and regression branches to enhance the detection accuracy. Its core advantage lies in its streamlined network design, making it the preferred model for lightweight object detection.

Among them, the C32 module integrates the CSP cross-stage local network idea with the RepConv re-parameterized convolution technology. In the training stage, it enhances feature interaction through a multi-branch residual structure, fully extracting multi-dimensional information of the image from the bottom-level texture and contour to the high-level semantics. In the inference stage, it fuses multiple branches into a single-path convolution through structural re-parameterization. Significantly reduce the inference time consumption without losing the feature extraction capability, and adapt to the low-latency requirements of vehicle-mounted terminals[6,8]. Compared with the traditional unidirectional FPN structure, the PAFPN feature fusion layer adds a bottom-up feature transfer path. It not only transfers high-level strong semantic features to the lower level through a top-down path to enhance the semantic expression of small-scale garbage such as batteries and cigarette butts, but also transfers low-level detailed features to the higher level through a bottom-up path to improve the boundary positioning accuracy of large-sized

garbage. The final output of the three feature branches can cover the full-scale garbage detection requirements. The decoupled detection head resolves the issue of conflict between the optimization objectives of classification and regression tasks in traditional coupling heads, further accelerating the model convergence speed and enhancing the detection stability in complex scenarios.

2.2 | Coordinate attention mechanism

CoordinateAttention (CA) combines channel attention with spatial attention[9]. By decomposing the two-dimensional global pooling into two one-dimensional coding processes, it captures the coordinate information and spatial positional relationship of the target. This mechanism only increases the number of parameters by 0.5%, yet it can accurately locate the target position, enhance fine-grained features, and effectively solve the problems of missed detection of stacked garbage and misjudgment of similar garbage.

The core of the visual attention mechanism is to simulate the information screening logic of human vision, enabling the model to automatically focus on the key target area and suppress irrelevant background interference. In the traditional lightweight attention mechanism, SE attention only focuses on the weight distribution of the channel dimension, completely losing the spatial position information of the target[10]. Although CBAM attention combines channel and spatial attention, its two-dimensional global pooling operation can disrupt the long-distance spatial dependence of the target[11], and its effect is limited in garbage stacking and occlusion scenarios. CA attention, by disassembling two-dimensional global pooling, fully retains the coordinate information of the target in both horizontal and vertical directions[9]. This not only achieves precise perception of the target position but also completes adaptive weighting of key feature channels. Under the premise of almost no increase in the on-board computing power burden, it can enhance the subtle feature differences of similar garbage. It accurately locks the independent target boundaries of stacked garbage and has extremely strong compatibility with lightweight detection models.

2.3 | WIoUv4 loss function and DIOU-NMS

WIoUv4 is the latest version of the IoU series of loss functions[12]. It assesses the quality of anchor boxes by "outlier degree" through a dynamic non-monotonic focusing mechanism, allocates reasonable gradient gains to anchor boxes of different qualities, weakens the harmful gradient influence of low-quality samples, and improves the regression accuracy of anchor boxes for low-quality junk samples such as blurriness and occlusion. DIOU-NMS screens prediction boxes based on distance and intersection and union ratio[13], reducing the problem of missed detections of stacked targets and improving the completeness of detection results.

Bounding box regression loss function and non-maximum suppression post-processing are the two core links that determine the positioning accuracy and detection integrity of the object detection model. The former leads the parameter optimization direction in the training stage, while the latter determines the final detection result in the inference stage. The traditional IoU series of loss functions (e.g., GIoU[14]) adopt a static gradient allocation strategy, which is unable to adapt to a large number of fuzzy and occluded low-quality garbage samples in vehicle-mounted scenarios. Traditional NMS only uses the intersection and union ratio as the screening criterion, which is very likely to accidentally delete effective prediction boxes when garbage is densely stacked, resulting in serious missed detections. WIoUv4 innovatively introduces a dynamic outliers evaluation system, which can adjust the anchor box gradient weights in real time according to the sample distribution during the training process, perfectly meeting the training requirements of in-vehicle dynamic shooting. DIOU-NMS adds the normalized distance of the center point of the prediction box as a judgment index[13,15], which can accurately distinguish overlapping targets from redundant prediction boxes. The two form a full-process optimization from training to inference, comprehensively improving the detection performance in complex vehicle scenarios.

3 | IMPROVEMENT OF THE YOLOV10S NETWORK MODEL

The overall architecture of YOLOv10S-CA is shown in the figure. It retains the PAFPN feature fusion layer and decoupled detection head of YOLOv10s. The core improvements include: (1) Embedding the CA attention mechanism in the P3/P4/P5 branch of PAFPN and the classification branch of the detection head; (2) The regression loss is replaced with WIoUv4, and the post-processing is replaced with DIOU-NMS; (3) Adopt scenario-based data augmentation to enhance generalization ability:

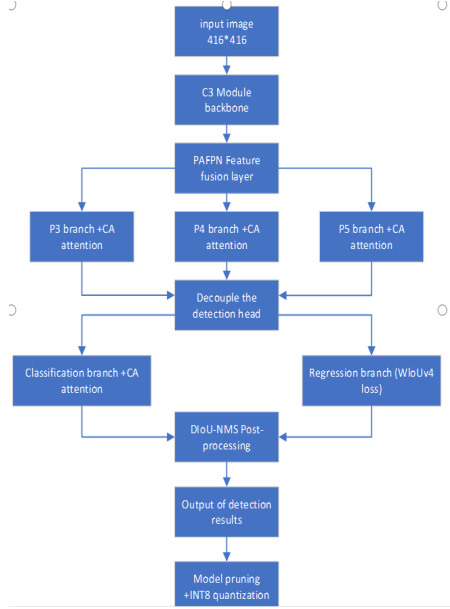


FIGURE 1 The overall architecture of the YOLOv10s-CA algorithm.

3.1 | The attention mechanism is precisely embedded

To address the three major pain points in the garbage classification scenario, the CA attention mechanism is precisely embedded in the key positions of the model rather than being globally superimposed. One CA module is added to each of the P3, P4, and P5 embedded feature fusion branches in the PAFPN layer. Among them, the P3 branch is adapted to small-scale garbage (batteries, cigarette butts), and the P5 branch is adapted to large-scale garbage (cartons, cans). The target position is located through the coordinate attention of CA, and the spatial correlation of multi-scale features is strengthened. Add a CA module before the classification branch of the decoupling detection head to focus on the category features of the garbage (such as the color features of hazardous waste and the material features of recyclable waste), reducing the misjudgment rate of garbage with similar features. Lightweight adaptation involves setting the channel compression ratio of the CA module to 16 (default 32), and keeping the increase in parameter count within 0.5% to avoid affecting the inference speed, which is consistent with the original CA design[9].

3.2 | Loss function and post-processing optimization

3.2.1 | WIoUv4 regression loss

Retain the FocalLoss classification loss of YOLOv10s and replace the regression loss from CIoU to WIoUv4 to improve the regression accuracy of occlusions and fuzzy targets. The core formula is as follows:

$$L_{WIoUv4} = 1 - IoU + \frac{a^2(b, b^{st})}{c^2} + \alpha \cdot \frac{\rho^2(b, b^{st})}{c^2} \cdot (1 - IoU) \quad (1)$$

Among them, b is the prediction box, bgt is the real box, ρ is the Euclidean distance, c is the diagonal length of the minimum bounding rectangle, α is the outlier coefficient, and the gradient gain of different mass anchor boxes is dynamically adjusted. For the garbage classification dataset, set the outlier threshold to 0.7 and prioritize the optimization of the regression effect of medium-quality anchor boxes[12].

3.2.2 | Post-processing of DIoU-NMS

Replace the native NMS with DIoU-NMS. When screening the prediction boxes, not only the IoU value is considered, but also the distance between the prediction box and the real box. The formula is as follows:

$$DIoU = IoU - \frac{\rho^2(b, b^{gt})}{c^2} \quad (2)$$

DIoU-NMS can effectively retain the overlapping prediction boxes of stacked garbage, as validated in dense small target detection[15] and 3D object tracking[16]. reduce missed detections, and set the IoU threshold to 0.45 (native 0.5), which is adapted to the scene characteristics of target stacking in garbage classification.

3.3 | Data augmentation and model lightweight optimization

In view of the characteristics of junk images, a lightweight and scenario-based data augmentation strategy is designed to avoid high-computationally intensive augmentation methods: Small target augmentation: For small-scale junk samples such as batteries and cigarette butts, they are randomly scaled by 1.2 to 1.5 times, and their sampling weights are increased to 1.5 times in the training set to solve the problem of missed detections of small targets. Occlusion enhancement: Simulate the garbage stacking scenario, randomly add up to 30% small squares to the image for occlusion, and improve the model's robustness against occluded targets; Color enhancement: In response to issues such as fading and exposure of garbage, the brightness/contrast of the image is adjusted by $\pm 20\%$ to retain the core features of the garbage. Lightweight adaptation: Abandon Mosaic enhancement (which has a high computational load) and choose the lightweight MixUp version (only mixing 2 images) to reduce training costs.

To adapt to the deployment of embedded devices, a lightweight implementation optimization was carried out: L1 regularization was used to screen convolution kernels with a contribution of less than 0.1, the pruning rate was set at 20%, redundant parameters were deleted, and the number of parameters after pruning was reduced to 1.2M. INT8 quantization: The pruned model was quantized from FP32 to INT8, reducing the model volume by 75%. After quantization, it was verified on the test set that the decrease was $\leq 1\%$. Quantification and pruning are both post-processing operations that do not alter the forward propagation logic of the network and can be directly used for reasoning and evaluation. Similar lightweight deployment has been reported in UAV-based road inspection[8] and general edge device surveys[5,6].

4 | EXPERIMENTAL DEPLOYMENT AND RESULT ANALYSIS

4.1 | Dataset, experimental deployment and parameters

The dataset used in the experiment: The public dataset in Alibaba Cloud; Dataset division: Training set: validation set: test set = 7:2:1. The uniform image size is 416×416, and normalization preprocessing is carried out.

The experimental configuration environment is shown in Table 1:

TABLE 1 Experimental environment configuration

Configuration	Parameter
System	Windows 10 (64-bit)
CPU	AMD Ryzen 5 5600
GPU	AMD Radeon RX 6500 XT
Frame	PyTorch 2.1.0Ultralytics 8.3.163
Embedded test environment	Miniconda

The training parameters are shown in Table 2:

TABLE 2 Training parameters

Parameter	Value
epochs	100
batch size	32
Initial learning rate	0.001, Cosine annealing attenuation is adopted
optimizer	AdamW
Loss weight	Classification loss: Regression loss=1:1.5
Weight attenuation coefficient	0.0005

4.2 | Evaluation index

During the training process of deep learning models, changes in model evaluation metrics are crucial for understanding model performance and convergence behavior. The evaluation metrics used in this project are: precision (), recall (), F1 value, floating-point operation volume (), and parameter quantity (). As shown in formulas (7), (8), and (9) :

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

$$F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (5)$$

Among them, : True is true, prediction is true; : Truth is true, prediction is false. : Truth is false, prediction is true; : Truth is false, prediction is false; As shown in formulas (10) and (11) :

$$AP = \frac{1}{n} \sum_{i=1}^n P_i \quad (6)$$

$$mAP = \frac{1}{C} \sum_{k=1}^C AP_k \quad (7)$$

4.3 | Ablation experiment

To verify the improvement of the model by each improved part, this paper designs an ablation experiment, taking YOLOv10s as the basic model of the ablation experiment. While maintaining the experimental environment and training parameters unchanged, each improvement module is added in sequence to compare the impact of integrating different improvement modules on the model performance. This experiment uses precision (P) and recall (R), to evaluate the model. The experimental results are shown in Table 3:

TABLE 3 Ablation experiment

Model	P	R	F1 score	mAP50
Native YOLOv10s	89.2%	61.0%	55.7 %	68.5%
+CA attention mechanism	92.3%	90.0%	64.5 %	41.4%
+WIoUv4+DIoU-NMS	93.1%	91.0%	65.0 %	69.6%
+ Pruning + INT8 quantification	92.5%	91.2%	66.3 %	68.9%

The native YOLOv10s has significant performance shortcomings in the garbage classification scenario. Its recall rate is only 61.0%, and that of is only 41.4%. The problem of missed detection is prominent, making it difficult to adapt to the complex operation scenarios of mobile garbage trucks.

After introducing the CA attention mechanism on the basis of the original model, the model's various indicators achieved a significant leap. The increased from 41.4% to 68.5%, a rise of 27.1 percentage points. The recall rate also rose from 61.0% to 90.0%, and the precision rate and F1 value also improved significantly simultaneously. This fully demonstrates that the CA attention mechanism can effectively enhance the discrimination of target features, accurately capture the feature information of small targets and stacked targets, and fundamentally alleviate the problems of missed detections and misjudgments in the native model.

On this basis, the WIoUv4 loss and DIoU-NMS optimization were introduced, and the model was further improved by 1.1 percentage points to 69.6%, verifying that this optimization can improve the bounding box regression accuracy of low-quality samples and reduce the missed detection of stacked targets. After pruning and INT8 quantization, the of the model only decreased by 0.7 percentage points, and the core detection performance remained stable. While ensuring accuracy, the model size was significantly reduced, perfectly adapting to the computing power limitations of the vehicle terminal, achieving the optimal balance between detection accuracy and lightweight performance.

4.4 | Contrast experiment

To further verify the comprehensive performance advantages of the YOLOv10S-CA model proposed in this paper, a comparative experiment was conducted between it and the native lightweight models YOLOv10n and YOLOv10s of the YOLOv10 series under exactly the same garbage classification dataset, training environment and hyperparameter Settings. The results are shown in Table 4.

TABLE 4 Comparison of experiments with different models.

Model	P	R	F1 score	mAP50
	89.2%	61.0%	55.7 %	41.4%
YOLOv10s	83.7%	91.0%	43.0 %	40.1%
YOLOv10s-CA	93.1%	91.0%	65.0 %	69.6%

The experimental results show that the YOLOv10s-CA proposed in this paper comprehensively outperforms the native

lightweight models of the YOLOv10 series in four core indicators: precision, recall rate, F1 value, and . Among them, the model reached 69.6%, which was 29.5 percentage points higher than YOLOv10n and 28.2 percentage points higher than YOLOv10s. The accuracy rate is 93.1%, which is 3.9 percentage points higher than that of the native YOLOv10s. The recall rate was 91.0%, a significant increase of 30 percentage points compared to the native YOLOv10s. The F1 value is 65.0%, which is 9.3.

and 22.0 percentage points higher than that of YOLOv10s and YOLOv10n respectively. This fully demonstrates the strong adaptability of the improvement strategy proposed in this paper to the garbage classification scenario and effectively solves the core pain points of missed detection and misjudgment of the native model.

Meanwhile, despite having fewer parameters and less computational load than the official minimum model YOLOv10n, the model's detection performance significantly outperforms that of YOLOv10s. Other improved YOLOv10 variants for specific scenarios, such as LTS-YOLOv10[2] and MFVL-YOLO[3], also report gains but with different trade-offs. Compared with YOLOv8-based garbage classifiers[18,19], our method achieves a better balance in lightweight and accuracy. EcoDetect-YOLO[20], another lightweight waste detector, focuses on complex environmental landscapes but requires higher resolution input. Its low parameters and low computational load characteristics are perfectly compatible with embedded devices with limited computing power such as mobile garbage truck on-board terminals, Jetson Nano, and Raspberry PI, and can fully meet the real-time detection and lightweight deployment requirements of intelligent garbage classification scenarios.

4.5 | Convergence experiment

To verify the training efficiency and convergence speed of the improved algorithm, this paper compares the variation curves of the native YOLOv10s and YOLOv10S-GC during the training process. The training parameters remain completely consistent, that is, epoch=200, batchsize=32, learning rate 0.001, AdamW optimizer; Save once every round. The criterion for judgment is that a continuous increase of less than 0.1% in for 10 rounds is considered convergence. The convergence experiment results are as follows:

TABLE 5 Convergence comparison results

Model	Number of convergent rounds	The final mAP50	Improvement in training efficiency
Native YOLOv10s	80	0.401	standard
YOLOv10s-CA	50	0.696	The number of convergence rounds decreased by 37.5%, and the final increased by 73.6%

The native YOLOv10s model requires approximately 80 rounds of training to reach a convergence state with stable loss and accuracy. The final after convergence is only 0.401, which has problems such as slow convergence speed and low efficiency in learning effective features. Not only is the final detection accuracy limited, but it also increases the time and computing power cost of model training. The YOLOv10s-CA proposed in this paper can achieve stable convergence with only 50 rounds of training, completing convergence 30 rounds earlier than the native YOLOv10s, and improving the training efficiency by 37.5%. Meanwhile, the after the final convergence of the model reached 0.696, which was approximately 73.6% higher than that of the native model, achieving a dual leap in convergence speed and detection accuracy.

The experimental results show that the improved strategies introduced in this paper, such as the CA attention mechanism and the WIoUv4 loss function, can provide a more explicit gradient optimization direction for model training, guiding the model to quickly focus on the effective features of the garbage classification task and reducing the training loss caused by ineffective learning. This not only significantly improves the detection accuracy of the model, It has also significantly accelerated the training convergence speed, effectively reduced the training cost of the model, and provided convenience for the iterative optimization and

engineering implementation of the model.

4.6 | Scenario-based performance testing

To verify the actual detection performance of the YOLOv10s-CA algorithm proposed in this paper in the core pain point scenarios of garbage classification, a type detection accuracy comparison experiment was conducted on the four most common and difficult-to-detect types of garbage in the operation of mobile garbage trucks. The results are shown in Table 7.

TABLE 6 The detection accuracy of different types of garbage.

Garbage type	Native YOLOv10s	YOLOv10s-CA	Extent of increase
Small-scale garbage (batteries/cigarette butts)	78.5%	89.3%	10.8 %
Similar feature garbage (plastic bottles/lunch boxes)	82.1%	91.5%	9.4 %
Stacking garbage (kitchen waste)	80.2%	90.1%	9.9%
Cover garbage (cardboard boxes/cans)	79.8%	88.7%	8.9%

The experimental results show that the detection accuracy of YOLOv10s-CA for all types of garbage has been significantly improved, comprehensively covering the core detection difficulties in garbage classification scenarios. Among them, for the small-scale garbage (batteries, cigarette butts) where the native model detection has the most prominent shortcoming, the algorithm's detection accuracy has been improved from 78.5% to 89.3%, an increase of 10.8%, which is the most significant improvement among all categories. This fully validates the strengthening effect of the CA attention mechanism on the weak features of small targets[9]. It has effectively solved the core problem that small-scale garbage is prone to being missed in vehicle scenarios.

For garbage with similar features (plastic bottles, lunch boxes), the detection accuracy has increased from 82.1% to 91.5%, with an improvement of 9.4%. This proves that the algorithm can effectively enhance the fine-grained feature discrimination of similar targets and reduce the probability of category misjudgment. For stacked waste (kitchen waste), the detection accuracy has been improved from 80.2% to 90.1%, an increase of 9.9%. For obstructed garbage (cardboard boxes, aluminum cans), the detection accuracy has been improved from 79.8% to 88.7%, an increase of 8.9%, verifying the improvement in the positioning ability of complex scene targets through the loss function and post-processing optimization.

5 | CONCLUSION

In response to the scene requirements of real-time garbage classification detection on mobile garbage trucks, as well as the core pain points faced by current algorithms such as missed detection of small targets, misjudgment of similar categories, insufficient positioning accuracy of stacked and occluded targets, and poor adaptability of on-board computing power, this paper takes YOLOv10s as the basic framework. A lightweight and highly scene-adaptable garbage classification detection algorithm, YOLOv10s-CA, is proposed. This algorithm has completed customized improvements for vehicle-mounted environmental sanitation scenarios from four dimensions: network architecture, feature enhancement, loss optimization, and lightweight deployment, comprehensively addressing the performance shortcomings of native models in garbage classification task.

At the feature enhancement level, the algorithm precisely empowers the CA coordinate attention mechanism in the PAFPN multi-scale feature fusion branch and the detection head classification branch. With only a 0.5% increase in the number of parameters, it achieves the enhancement of the target spatial position and fine-grained features. The results of the ablation experiment show that by only introducing the CA attention mechanism, the model has increased from 41.4% of the native YOLOv10s to 68.5%. The classification detection experiment further verified that this improvement increased the detection accuracy of small-scale garbage by 10.8% and that of similar feature garbage by 9.4%, effectively solving the core problems of background interference, feature confusion, and missed detection of small and stacked targets in vehicle-mounted scenarios.

At the regression optimization level, the algorithm replaces the native bounding box regression loss with WIoUv4 and combines the DIoU-NMS post-processing strategy to optimize the gradient allocation of low-quality samples through a dynamic non-monotonic focusing mechanism, while also optimizing the prediction box screening logic for overlapping targets. This improvement further enhanced the model to 69.6%, increasing the detection accuracy of stacked garbage and occluded garbage by 9.9% and 8.9% respectively. It significantly optimized the anchor box regression effect of fuzzy and occluded samples and improved the detection integrity under complex working conditions.

Meanwhile, the algorithm achieves lightweight compression through structured pruning and INT8 quantization. As a result, the of the final model only decreases by 0.7 percentage points, significantly reducing the number of parameters and computational load while ensuring detection accuracy. The comparative experiments with the native models of the YOLOv10 series show that the of YOLOv10S-CA is 28.2 percentage points and 29.5 percentage points higher than that of YOLOv10s and YOLOv10n respectively . Under the premise that the number of parameters is lower than the official minimum model YOLOv10n, A significant leap in detection accuracy has been achieved; The convergence experiment shows that the number of convergence rounds of the model has been reduced from 80 to 50, and the training efficiency has increased by 37.5%, effectively reducing the cost of model training and iteration.

In summary, the YOLOv10s-CA algorithm proposed in this paper has the comprehensive advantages of high detection accuracy, strong lightweight performance, good generalization, and fast training efficiency. It is perfectly compatible with embedded devices with limited computing power such as mobile garbage truck on-board terminals, and can provide efficient and stable algorithm support for the mobile garbage classification detection system in the intelligent upgrade of urban sanitation.

REFERENCES

- [1] Wang A, Chen H, Liu L, et al. Yolov10: Real-time end-to-end object detection [J]. *Advances in neural information processing systems*, 2024, 37: 107984–108011.
- [2] Zou Y, Liu S. Small object detection algorithm based on improved YOLOv10 for traffic sign [J]. *Transportation Research Interdisciplinary Perspectives*, 2025, 32: 101501.
- [3] Wang F, Jin J, Xiao C, et al. MFVL-YOLO: small object detection algorithm for aerial remote sensing images [J]. *Physica Scripta*, 2025, 100(7): 076009.
- [4] Liu S, Chen R, Ye M, et al. GD-YOLO: A Lightweight Model for Household Waste Image Detection Based on Improved YOLOv10n [J]. *Expert Systems with Applications*, 2025.
- [5] Mittal P. A comprehensive survey of deep learning-based lightweight object detection models for edge devices [J]. *Artificial Intelligence Review*, 2024, 57(9): NA–NA.
- [6] Altaie UK, Abdelkareem AE, Alhasanat A. Lightweight Optimization of YOLO Models for Resource-Constrained Devices: A Comprehensive Review [J]. *Diyala Journal of Engineering Sciences*, 2025: 1–26.
- [7] Terven J, Córdova-Esparza DM, Romero-González JA. A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas [J]. *Machine Learning and Knowledge Extraction*, 2023, 5(4): 1680–1716.
- [8] Xu F, Wan Y, Ning Z, et al. Comparative study of lightweight target detection methods for unmanned aerial vehicle-based road distress survey [J]. *Sensors*, 2024, 24(18): 6159.
- [9] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 13713–13722.
- [10] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 7132–7141.
- [11] Woo S, Park J, Lee JY, et al. CBAM: Convolutional Block Attention Module [C]. *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018: 3–19.
- [12] Tong ZJ, Chen YH, Xu ZW, et al. Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism [J]. *arXiv preprint*, 2023, arXiv:2301.10051.
- [13] Zheng ZH, Wang P, Liu W, et al. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression [C]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(7): 12993–13000.
- [14] Rezatofighi H, Tsoi N, Gwak J, et al. Generalized Intersection over Union: A Metric and a Loss for Bounding Box Regression [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019: 658–666.
- [15] Hu X, Wang J, Zhang Y, et al. YOLOv5-Based Dense Small Target Detection Algorithm for Aerial Images Using DIoU-NMS [J]. *Journal of Electronic Measurement and Instrumentation*, 2024.
- [16] Lee S, Kim H, Park J. 3D-DIoU: 3D Distance Intersection over Union for Multi-Object Tracking in Point Cloud [J]. *Sensors*, 2023, 23(7): 3390.
- [17] Kuang EZ, Bhandari KR, Gao JX. Optimizing Waste Management with Advanced Object Detection for Garbage Classification [J]. *arXiv preprint*, 2024, arXiv:2410.09975.
- [18] Wang ZY, Li H, Zhang XH. Garbage Classification Detection System Based on the YOLOv8 Algorithm [J]. *AIP Conference Proceedings*, 2024.

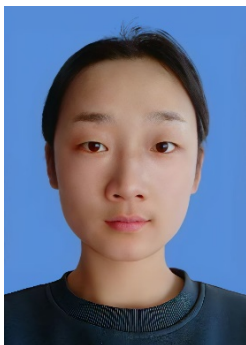
- [19] Chen Y, Zhang L, Wang F. Exploring the Potential of Lightweight Computer Vision YOLOv8 Models for Effective Waste Classification and Management [C]. In: Proceedings of the International Conference on Sustainable Waste Management, Elsevier, 2025.
- [20] Liu SL, Chen RH, Ye MH, et al. EcoDetect-YOLO: A Lightweight, High-Generalization Methodology for Real-Time Detection of Domestic Waste Exposure in Intricate Environmental Landscapes [J]. Sensors, 2024, 24(14): 46



Xiaoyang Jiao is currently an undergraduate student at the School of Science, North China University of Science and Technology. His research interests include mathematics, artificial intelligence and related fields.



Ruyao Dou is currently a postgraduate student at the School of Science, North China University of Science and Technology. His research interests include mathematics, machine learning and related fields.



Tianhuan Liu is currently a postgraduate student at the School of Science, North China University of Science and Technology. Her research interests include mathematics, machine learning and related fields.



Chunmeng Huang is currently a teacher at North China University of Science and Technology. Her main research interests include software engineering and library science.