Electronic Research Archive

http://www.aimspress.com/journal/ERA

ERA, 33(6): 3673–3698.
DOI: 10.3934/era.2025163
Received: 05 March 2025
Revised: 21 May 2025
Accepted: 05 June 2025
Published: 12 June 2025

Research article

An efficient improved YOLOv10 algorithm for detecting electric bikes in elevators

Tong Li[1], **Lanfang Lei**[2], **Zhong Wang**[1,3,*], **Peibei Shi**[1] **and Zhize Wu**[2]

[1] School of Computer and Artificial Intelligence, Hefei Normal University, Hefei 230601, China
[2] School of Artificial Intelligence and Big Data, Hefei University, Hefei 230601, China
[3] State Key Laboratory of Fire Science, University of Science and Technology of China, Hefei 230026, China

* **Correspondence:** Email: zhongw@ustc.edu.cn; Tel: +8618905607916.

Abstract: The presence of electric bicycles in elevators poses serious safety hazards, necessitating reliable automatic detection solutions. To address these issues, this paper proposes EBike-YOLO, an enhanced real-time detection model based on YOLOv10n, specifically optimized for elevator environments. EBike-YOLO introduces three key innovations: 1) the PIoU2_NWD loss function, combining adaptive penalty mechanisms, anchor-quality-aware gradient adjustments, non-monotonic attention, and Wasserstein distance for significantly improved object localization; 2) the EnhancedPSA structure, refining self-attention and feed-forward network architectures to enhance feature extraction; and 3) the C2f_Calibration module, incorporating self-calibration operations to robustly manage diverse orientations and scales of electric bicycles. Extensive experiments on a custom elevator dataset demonstrate substantial improvements over the YOLOv10n baseline, achieving notable enhancements in precision (4.1%), recall (0.3%), F1 score (2.0%), and mAP@50 (2.23%). Further validations on standard benchmark datasets (VisDrone2019, VOC2007, VOC2012) confirm the model's strong generalization capability, underscoring its applicability beyond specific elevator scenarios. These results clearly highlight the effectiveness, novelty, and practical value of the proposed model for diverse real-world detection tasks.

Keywords: Electric bicycles; YOLOv10n; PIoU2_NWD; EnhancedPSA; C2f_Calibration; electric bicycle

1. Introduction

With the acceleration of urbanization and changes in lifestyle, electric bicycles have become an integral part of urban transportation. However, electric bicycles and their batteries pose risks of overheating, spontaneous combustion, and even explosions, particularly when they are brought into enclosed spaces such as elevators, where these risks are significantly heightened [1]. In recent years, there have been frequent fire incidents caused by electric bicycles worldwide, which have not only resulted in casualties, but also sparked widespread concern about public safety [2,3]. To address this safety concern, multiple regions have already implemented relevant regulations, prohibiting electric bicycles from entering elevators or charging in public areas.

Identifying electric bicycles in elevator environments poses substantial challenges due to several inherent factors. First, elevator-mounted cameras typically have fixed positions, leading to limited viewing angles and consequently partial or incomplete capture of electric bicycles. Second, occlusions frequently occur, as bicycles may be partially blocked by individuals or other objects within the confined space, significantly complicating detection. Third, environmental factors—such as varying background textures, colors, and lighting conditions—often cause false positives and negatively impact detection reliability. Although regulations have been introduced to restrict electric bicycle usage in elevators, traditional surveillance and monitoring methods generally fall short in real-time detection accuracy and responsiveness, hindering effective prevention of safety hazards.

Recent advances in computer vision and artificial intelligence, particularly deep learning, have markedly improved object detection capabilities. Traditional detection methods such as the Viola-Jones (VJ) [4] and (HOG) [5] method, which depend heavily on handcrafted features, struggle to achieve the accuracy and real-time responsiveness required in practical scenarios. These methods frequently yield high false-positive rates due to the diverse shapes and appearances of electric bicycles, thus limiting their effectiveness.

Deep learning-based detection frameworks have significantly advanced detection performance, especially within challenging, constrained environments like elevators. Nevertheless, prevalent models continue to exhibit certain limitations. For instance, two-stage approaches, exemplified by Faster R-CNN [6,7], deliver high accuracy yet involve substantial computational overhead, making real-time operation difficult. Conversely, single-stage detectors, notably the YOLO family [8–11], provide more balanced speed and accuracy, but still face challenges in effectively managing occlusions, complex backgrounds, and detecting small-scale objects.

In addition, despite regulatory efforts to limit electric bicycle usage in elevators, traditional monitoring systems often fall short in real-time detection, accuracy, and responsiveness. Existing methods, such as mechanical barriers or sensor-based systems, suffer from high false positive rates, limited adaptability to diverse bicycle designs, and prohibitive implementation costs. Furthermore, cloud-based solutions face latency and data privacy challenges, while advanced deep learning models often demand excessive computational resources, making them unsuitable for embedded applications in constrained environments like elevators.

To address these challenges, this paper proposes a real-time detection algorithm, EBike-YOLO, which optimizes detection performance by introducing the PIoU2_NWD loss function, the EnhancedPSA structure, and the C2f_Calibration module. The PIoU2_NWD loss function combines an adaptive object size penalty factor, an anchor box quality-based gradient adjustment function, a non-monotonic attention layer, and the Wasserstein distance, overcoming the limitations of traditional IoU loss, and enhancing the model's localization ability and robustness. The EnhancedPSA structure improves the self-attention mechanism and feedforward network, significantly improving the feature map representation ability, especially in multi-scale object detection. The residual connection effectively retains the original information and further enhances the feature extraction ability, which is particularly suitable for multi-scale object detection in complex scenes. The C2f_Calibration module optimizes feature extraction and fusion through self-calibration operations, thereby improving the robustness of the model in complex environments, especially being able to handle angle changes and posture changes of electric bicycles. Our contributions are threefold:

1) The PIoU2_NWD loss function is proposed, which combines an adaptive penalty factor for object size, a gradient adjustment function based on anchor quality, a non-monotonic attention layer, and the Wasserstein distance. The modified loss function significantly improves the model's convergence speed and detection accuracy, with the mAP@50 reaching 87.097%, which is over 1.1% higher than that of the traditional IoU loss function.

2) Self-calibration operations are introduced into the C2f module to enhance feature extraction capabilities, addressing the issue of insufficient robustness when dealing with electric bicycles from different angles and forms. Experiments show that the improvement of this module enables the model to achieve an accuracy of 92.184% in complex scenarios.

3) An EnhancedPSA structure is proposed, which deeply processes each part of the input features through self-attention mechanisms and feedforward networks. This structure can more effectively capture the interactions between different features, enhancing the comprehensiveness of feature extraction and the learning ability of complex relationships.

This paper is organized as follows: Section 2 reviews related work, highlighting gaps in current electric bicycle detection technologies. Section 3 details the proposed EBike-YOLO methodology, including algorithmic design and implementation. Section 4 presents experimental results and discusses their implications. Finally, Section 5 concludes with key findings and potential future research directions.

2. Related works

In modern urban life, the safety risks posed by electric bicycles entering elevators have attracted widespread attention. However, traditional prevention methods face numerous issues. Manual monitoring and mechanical blocking measures are inefficient and unreliable, and existing elevator monitoring systems rely on manual patrols, which can lead to fatigue and lack real-time response capabilities. Mechanical blocking measures also have flaws, as traditional object detection methods struggle with the variety of electric bicycle types and their differing appearances, leading to high false positive rates, which affect the normal operation of elevators and reduce trust in the system. Long-term use may render the prevention measures ineffective, increasing safety risks. Therefore, there is an urgent need for more advanced and efficient electric bicycle detection technologies.

Deep learning-based two-stage object detection methods first involve identifying candidate regions, and then classifying and localizing these regions to detect electric bicycles. Li et al. [12] proposed an enhanced Faster R-CNN network for shared bicycle detection, which improved the detection efficiency by introducing feature fusion module and deformable convolution. On the shared bicycle dataset (SBD), this method improved the mean average precision (mAP) by 13%. John et al. [13] investigated the use of IoT and image classification technologies, employing RetinaNet and YOLO models for traffic pattern analysis. They found that RetinaNet outperformed YOLO in traffic object recognition, effectively reducing traffic accidents and automatically optimizing traffic routes. Miguel et al. [14] proposed a method utilizing computer vision technology to collect user information on hiking and cycling paths, identifying several promising computer vision techniques including

YOLOv3-Tiny, MobileNet-SSD V2, and Faster R-CNN with ResNet-50. These studies laid the foundation for future development of a solution that can be applied, tested, and demonstrated in real-world scenarios.

Single-stage object detection methods based on deep learning have achieved a balance between detection accuracy and real-time performance, making them widely applicable in engineering practices. These methods use global information to directly regress the bounding boxes and class information of target detections from the entire image. For instance, Wang et al. [15] proposed an improved YOLOv4 algorithm for real-time detection of electric bicycles in elevators. By reconstructing the feature pyramid and backbone network and introducing the attention mechanism in the residual network, the improved W_YOLOv4 algorithm significantly improved the detection accuracy and speed. Zhang et al. [16] introduced an electric bicycle detection method based on an improved YOLOv5 algorithm, aiming to improve the efficiency of parking and charging management. The improved YOLOv5 algorithm not only improves the detection accuracy, but also reduces the number of parameters, making it more suitable for deployment on mobile terminals. Sun et al. [17] proposed an improved electric bicycle detection algorithm based on YOLOv4, which enhances detection speed and accuracy by incorporating GhostNet, the ECA attention mechanism, and RFB modules. Su et al. [18] introduced an enhanced YOLOv5s model that enables real-time detection of electric bicycles in elevator environments, providing a feasible technical solution through deployment on edge devices. This model significantly reduces computational resource consumption while maintaining high performance, offering effective technical support for safety monitoring in practical applications. Zhao et al. [19] proposed an electric bicycle recognition method based on YOLOv5, which significantly improved the detection accuracy and recall by integrating the EIoU loss function, CBAM attention mechanism, and CARAFE operator. The model can stably and effectively identify electric bicycles on the Jetson TX2 NX platform, meeting the needs of fast detection in elevators. Liu et al. [20] developed an efficient electric bicycle tracking algorithm, EBTrack, which employs YOLOv7 as the object detector and incorporates the ResNetEB feature extraction network, adaptive modulation noise scale Kalman filter, and a specialized matching mechanism to improve detection and recognition efficiency and accuracy. Zhang et al. [21] introduced a novel fire monitoring system for electric bicycle sheds (FMS-EBS) based on YOLOv8, which utilizes deep learning technology to achieve real-time monitoring and automatic alarm functions, demonstrating high accuracy and practicality.

Despite notable progress, existing methods for electric bicycle recognition still face significant challenges in detection speed, accuracy, and real-time performance, especially within complex scenarios. Traditional methods typically suffer from high computational complexity and limited robustness, hindering their practical applications. Two-stage deep learning methods, while accurate, involve substantial computational overhead and struggle with real-time requirements. Single-stage approaches achieve a better balance between speed and accuracy, but remain limited in effectively managing occlusions and complex scene variations.

To overcome these limitations, this paper introduces an enhanced real-time detection algorithm specifically designed for identifying electric bicycles in elevator environments. Our approach integrates three novel improvements: the PIoU2_NWD loss function, Enhanced PSA structure, and C2f_Calibration module. Specifically, the PIoU2_NWD loss function combines an adaptive penalty factor, gradient adjustment based on anchor quality, non-monotonic attention layers, and Wasserstein distance, improving localization precision and robustness against position deviations. The enhanced PSA structure enhances the self-attention and feed-forward mechanisms, significantly strengthening multi-scale feature extraction capabilities. Additionally, the C2f_Calibration module employs self-calibration operations to dynamically optimize feature extraction, thus increasing model robustness against varying angles and postures of electric bicycles. Together, these innovations effectively address existing shortcomings, providing a more accurate, efficient, and reliable detection solution. Experimental results on a custom elevator dataset demonstrate significant performance gains over the YOLOv10n baseline: precision improved by 4.1%, recall by 0.3%, F1 score by 2.0%, and mAP@50 by 2.23%. Additionally, the model achieved strong generalization on public benchmarks such as VisDrone2019, VOC2007, and VOC2012. These findings clearly demonstrate the practical effectiveness and applicability of our method for real-world detection tasks beyond elevator-specific conditions.

3. Research method

The YOLOv10 model represents the latest research achievement in the YOLO series of object detection algorithms both domestically and internationally. It introduces a model design strategy driven by holistic efficiency-accuracy. Specifically, this strategy encompasses a lightweight classification head, spatial-channel separation downsampling, and a structure based on intrinsic rank design. On the COCO image dataset, YOLOv10 demonstrates outstanding performance, maintaining high precision even in complex backgrounds while also addressing the need for model lightweightness, making it highly suitable for deployment on embedded devices.

To meet the requirements of electric bicycle object detection tasks in specific scenarios within elevators, and considering the dual demands of model efficiency and size in edge computing environments, this study selects YOLOv10n as the baseline model for exploration. YOLOv10n is composed of four main parts: input, backbone, neck, and head.

The input part is responsible for performing a series of image enhancement operations, such as HSV color space transformation, translation, scaling, flipping, and mosaic processing.

The backbone adopts the C2f and SPPF architectures from YOLOv8 and introduces three innovative components: SCDown, C2fCIB, and PSA. SCDown achieves efficient spatial-channel separation dimensionality reduction through the combination of point convolution and depthwise convolution. As the foundational building block, C2fCIB leverages low-cost depthwise convolutions to mix spatial information, saving computational resources while effectively enhancing feature extraction capabilities. PSA, on the other hand, addresses the high computational cost of traditional self-attention mechanisms, aiming to enhance the overall performance of the model at a lower cost.

The neck section continues to use the FPN + PAN combination to strengthen multi-scale feature fusion capabilities. For the head, lightweight detection head configurations, including One-to-one head and One-to-many head, are employed.

Finally, the entire system's prediction results are optimized through a consistency dual-allocation strategy. This design not only ensures high recognition accuracy, but also significantly reduces the resource consumption required during actual operation, making the solution particularly suitable for applications like electric bicycle tracking and localization in elevators, where space is limited and computational power is constrained.

In this paper, we utilize YOLOv10n as the baseline model and design an improved algorithm for electric bicycle detection within elevators, named EBike-YOLO, with its structure illustrated in Figure 1. The specific contributions of this work are as follows:

1) We first propose the PIoU2_NWD loss function, which integrates adaptive penalty factors for object size, gradient adjustment functions based on anchor quality, non-monotonic attention layers, and Wasserstein distance. This enhances the accuracy and convergence speed of object detection in

complex environments.

2) Considering the need for shallow feature maps by some small targets, we introduce an additional P2 small object detection head in the head network to more effectively capture the details and local features of electric bikes in elevators, thereby improving the accuracy of object detection.

3) We incorporate a self-calibration operation in the C2f module to enhance feature extraction capabilities, addressing the issue of insufficient robustness when dealing with electric bikes from different angles and shapes.

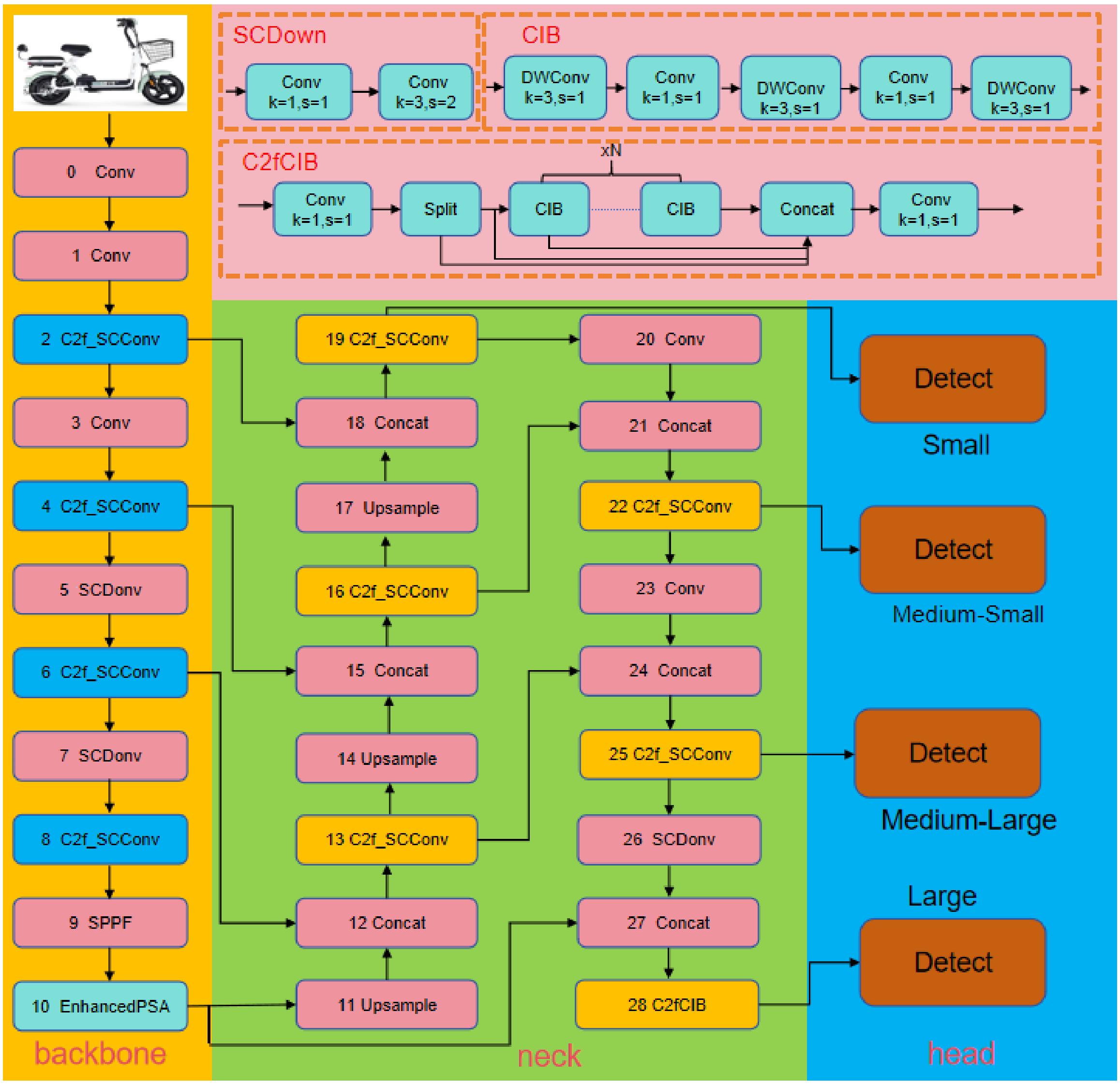


Figure 1. EBike-YOLO structure diagram. This figure illustrates the architecture of the EBike-YOLO model, which includes the backbone and neck sections based on YOLOv10. The head network incorporates an additional P2 small object detection head, a self-calibration operation within the C2f module, and the EnhancedPSA structure. These components are designed to enhance the accuracy and robustness of electric bike detection within elevators.

4) The proposed EnhancedPSA structure, by deeply processing each part of the input features

through self-attention mechanisms and feedforward networks, can more effectively capture the interactions between different features. This design significantly enhances the comprehensiveness of feature extraction and the learning ability for complex relationships, enabling the model to more deeply understand the diverse features and complex backgrounds of electric bikes in elevators, thereby improving overall detection accuracy and robustness.

3.1. PIoU2_NDW loss function

The PIoU loss function [22] solves the problems of anchor box expansion and slow convergence in the traditional IoU loss function by introducing an object size adaptive penalty and gradient adjustment based on anchor box quality. It guides the anchor box to move along an efficient regression path, thereby accelerating the convergence process.

PIoU is defined by Eq (1), where p is defined by Eq (2):

$$PIoU = IoU - 1 + e^{-p^2} \tag{1}$$

$$p = \frac{\left(\frac{dw_1}{w_{gt}}+\frac{dw_2}{w_{gt}}+\frac{dh_1}{h_{gt}}+\frac{dh_2}{h_{gt}}\right)}{4} \tag{2}$$

Here, (dw_1), (dw_2), (dh_1), and (dh_2) represent the absolute distances between the corresponding edges of the predicted bounding box and the ground truth bounding box, while (w_{gt}) and (h_{gt}) denote the width and height of the ground truth bounding box.

To enhance the focus on medium-quality anchor boxes, we introduce the PIoU2 loss function, which incorporates a non-monotonic attention layer combined with PIoU, to improve our paper. PIoU2 is defined as follows, where λ is set to 1.3:

$$q = e^{-P}, q \in (0,1] \tag{3}$$

$$u(x) = 3x * e^{-x^2} \tag{4}$$

$$PIoU2 = 1 - u(\lambda q) * \mathrm{e}^{-(\lambda q)^2} * PIoU \tag{5}$$

Considering that IoU-based metrics (such as IoU itself and its extensions) are highly sensitive to object position deviations in complex environments, the application of anchor-free detectors can lead to a sharp decline in detection performance. To avoid this, we combine PIoU2 [22] with the NWD (normalized Wasserstein distance) [23] to calculate the localization loss. NWD models the bounding box (BBox) as a two-dimensional Gaussian distribution and introduces a new metric that replaces the traditional IoU metric. For two two-dimensional Gaussian distributions, $\mu_1 = N(m_1, \varepsilon_1)$ and $\mu_2 = N(m_2, \varepsilon_2)$, the second-order Wasserstein distance between μ_1 and μ_2 is given by Eq (6).

$$W_2^2(\mu_1, \mu_2) = \|m_1 - m_2\|_2^2 + \left\|\varepsilon_1^{1/2} - \varepsilon_2^{1/2}\right\|_F^2 \tag{6}$$

Here, m_1 and m_2 represent the coordinates of the Gaussian distributions, ε_1 and ε_2 are the two different covariance matrices, and $\|\cdot\|_F$ denotes the Frobenius norm. The bounding boxes $A = (cx_a, cy_a, w_a, h_a)$ and $B = (cx_b, cy_b, w_b, h_b)$ model the Gaussian distributions N_a and N_b, respectively, which can be further simplified as

$$W_2^2(\mu_1, \mu_2) = \left\| \begin{pmatrix} \left[cx_a, cx_a, \frac{w_a}{2}, \frac{h_a}{2}\right]^T, \\ \left[cx_b, cx_b, \frac{w_b}{2}, \frac{h_b}{2}\right]^T \end{pmatrix} \right\|_2^2 \tag{7}$$

However, $W_2^2(N_a, N_b)$ is a distance metric and cannot be directly used as a similarity measure. Therefore, it is normalized in its exponential form to obtain the new metric, NWD, with the following Eq (8):

$$\mathrm{NWD}(\mathrm{N}_a, \mathrm{N}_b) = \exp\left(-\frac{\sqrt{W_2^2(\mathrm{N}_a, \mathrm{N}_b)}}{C}\right) \tag{8}$$

Here, C is a constant related to the dataset, which we set to 12.8 in this paper. Finally, the PIoU2-NWD loss function we propose is as follows:

$$L_{LOC} = (1-\beta)[1 - \mathrm{NWD}(\mathrm{N}_a, \mathrm{N}_b)] + \beta(1 - PIoU2) \tag{9}$$

where β is the weight coefficient, which we set to 0.5 in the experiments. The structural diagrams of PIoU2-NWD are shown in Figure 2.

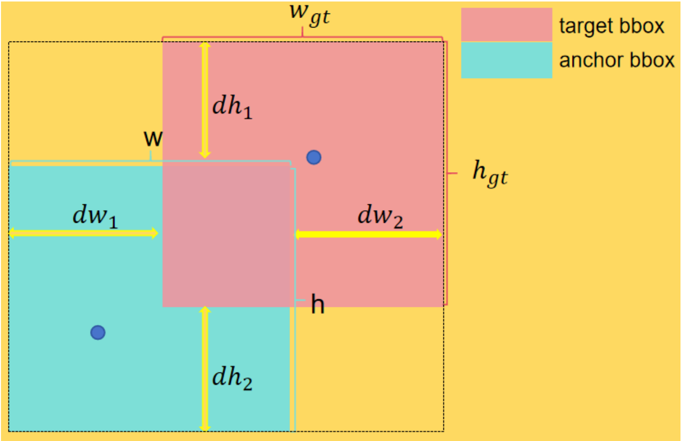


Figure 2. The structural diagrams of PIoU2-NWD.

3.2. C2f_Calibration module

The C2f module is a feature extraction module used in YOLOv10. It enhances the model's expressive capability and detection accuracy by cascading features. The core idea is to obtain richer feature representations through two forward propagations and fuse these features through concatenation operations. However, in the context of electric bicycle detection tasks, the C2f module has certain limitations. For instance, electric bicycles may appear in various angles and forms during actual detection, and the C2f module's robustness in handling such deformations and rotations is insufficient, thereby affecting the grading effect.

We refer to the self-calibrated convolution proposed by Liu et al. [24] and introduce a self-calibration operation into the C2f structure, which significantly enhances the feature extraction capability and improves the object detection performance of the model in complex environments. The C2f and C2f_Calibration module diagrams are shown in Figure 3. This improvement will help achieve more accurate object detection, especially for multi-scale objects and complex scenes.

First, a set of inputs $X \in R^{C\times H\times W}$ is provided. Given three parts of filters, denoted as $\{K_i\}_{i=1}^{3}$, where each part has the shape (C, C, K_h, K_w), we use $\{K_{1,}K_2, K_3\}$ to perform the self-calibration operation on X, producing Y.

The execution of the self-calibration operation is described as follows. For the input X, an average pooling operation with a filter size of r × r and a stride of r is applied, as shown in Eq (10).

$$T_1 = AvgPool_r(X) \tag{10}$$

The feature transformation on T_1 is based on K_1 as follows:

$$X' = Up\big(F_1(T_1)\big) = Up(T_1 * K_1) \tag{11}$$

Here, Up(·) denotes the bilinear interpolation operator, which maps intermediate reference values from a small-scale space back to the original feature space. Now, the calibration operation can be expressed in Eq (12) as

$$Y' = F_2(X) \cdot \sigma(X + X') \tag{12}$$

where $F_2(X) = X * K_2$, σ is the sigmoid function, and · represents element-wise multiplication. Using X' as the residual to form the weights for calibration, the final calibrated output can be expressed as in Eq (13):

$$Y = F_3(Y') = Y' * K_3 \tag{13}$$

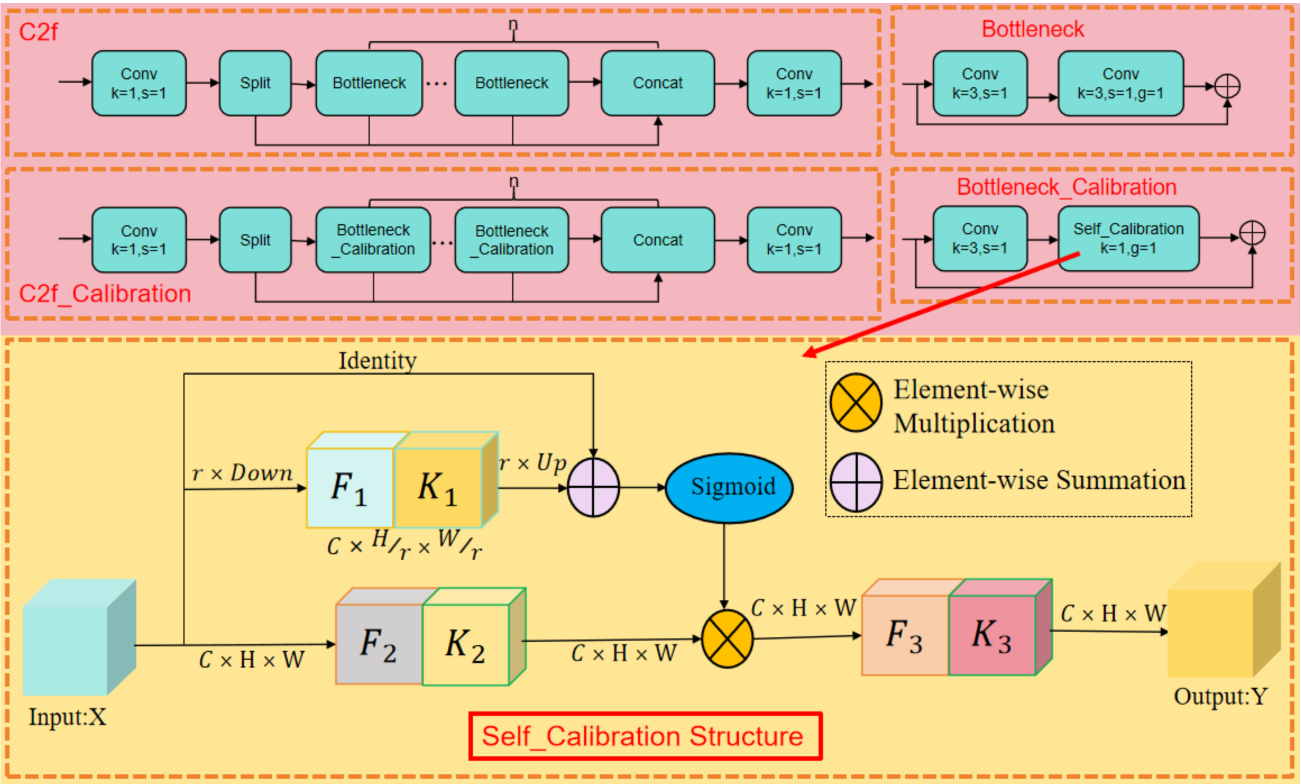


Figure 3. Structural diagrams of the C2f and C2f_Calibration modules.

3.3. EnhancedPSA

The partial self-attention (PSA) module integrates the self-attention mechanism to capture long-range dependencies and enhance feature representation. It is divided into module a and module b, with its structure illustrated in Figure 4. Module a consists of a multi-head self-attention (MHSA) layer and a feed-forward network (FFN) layer, as detailed in Eq (14).

$$\mathrm{PSA} = [\mathrm{MHSA} \rightarrow \mathrm{FFN}] \times N_{psa} \tag{14}$$

Each MHSA layer computes attention scores across the entire spatial dimension, and then refines the features through the FFN. The PSA module uses the residual connections from module b to maintain gradient flow and enable efficient training. The output of the PSA module is concatenated to aggregate the features extracted by module a and module b. Before passing the features to the prediction layer, a final 1×1 convolution is applied to adjust the output dimensions.

The MHSA operation within the PSA module can be described by Eq (15), where each attention head is computed as shown in Eq (16), and the attention function is defined by Eq (17).

$$\mathrm{MHSA}(Q, K, V) = \mathrm{Concat}(\mathrm{Head_1}, \mathrm{Head_2}, \mathrm{Head_3}, \cdots, \mathrm{Head_n})W^o \tag{15}$$

$$Head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right) \tag{16}$$

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{17}$$

Here, Q, K, and V represent the query, key, and value matrices, respectively. QW_i^Q, KW_i^K, VW_i^V, and W^o are the learned projection matrices. The FFN within the PSA module is a two-layer MLP with ReLU activation, as shown in Eq (18), where W_1 and W_2 are the weights, and b_1 and b_2 are the biases of the two linear transformations.

$$\mathrm{FFN}(X) = \max(0, XW_1 + b_1)W_2 + b_2 \tag{18}$$

However, the PSA primarily focuses on processing module a, which limits its ability to comprehensively extract and retain information. This simplicity may hinder its capacity to learn complex relationships in electric bicycle features. To address these limitations, we propose a new EnhancedPSA structure, as illustrated in Figure 4.

In the EnhancedPSA, both modules a and b undergo similar self-attention and feed-forward network processing, enabling more thorough extraction and enhancement of input features. The original input is also added back to the output, helping preserve more original information and reducing gradient vanishing.

Furthermore, submodules a and b represent different feature channels, and this thorough processing can better facilitate the recognition of various objects or textures in the image. This approach is particularly effective for complex tasks such as e-bike detection in elevator environments.

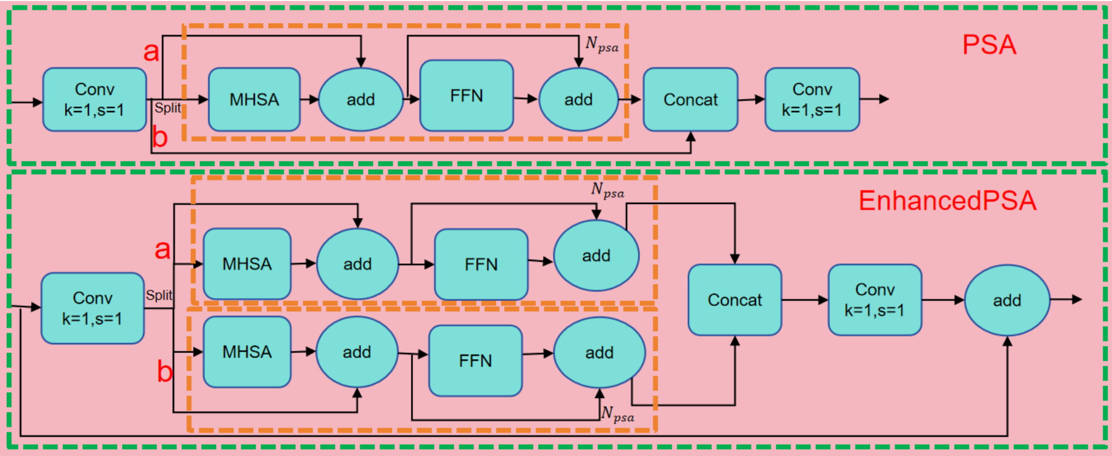


Figure 4. Structural diagrams of the PSA and EnhancedPSA modules.

4. Experimental results

4.1. Experimental environment and performance metrics

To validate the effectiveness of the proposed method, an experimental platform was set up utilizing Ubuntu 18.04 as the operating system and PyTorch as the deep learning framework. YOLOv10n was employed as the baseline network model. The specific configuration of the experimental environment is detailed in Table 1.

Table 1. Experimental environment configuration table.

Environmental Parameter	Value
Operating system	Ubuntu 18.04
Deep learning framework	PyTorch
Programming language	Python 3.8
CPU	Intel Xeon Scale 8358
GPU	NVIDIA A100(SXM4,80 GB)
RAM	256 GB

Throughout all training processes in our experiments, we maintained consistent hyperparameters. Table 2 lists the specific hyperparameter settings used during training.

Table 2. Training hyperparameters.

Hyperparameters	Value
Learning Rate	0.01
Image Size	640 × 640
Momentum	SGD
Batch Size	16
Epoch	200
Weight Decay	0.0005

To evaluate the performance of our proposed smoke detection model, we employ standard metrics commonly used in object detection tasks: precision, recall, and mean average precision (mAP).

$$IoU = \frac{A \cup B}{A \cap B} \tag{19}$$

$$Precision = \frac{TP}{TP+FP} \tag{20}$$

$$Recall = \frac{TP}{TP+FN} \tag{21}$$

$$AP = \sum_{i=1}^{n-1}(r_{i+1} - r_i)\, P_{inter}(r_i + 1) \tag{22}$$

$$\text{F1 Score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{23}$$

$$mAP = \frac{\sum_{i=1}^{k} AP_i}{k} \tag{24}$$

In Eq (19), A represents the ground truth box, B represents the predicted box, $A \cap B$ denotes the area of overlap between A and B, and $A \cup B$ represents the union of their areas.

In Eqs (20) and (21), P represents the number of true positives (positive class correctly predicted as positive), FP represents the number of false positives (negative class incorrectly predicted as positive), and FN represents the number of false negatives (positive class incorrectly predicted as negative).

In Eq (22), $r_{1,}r_2, \cdots, r_n$ are the values of recall R, arranged in ascending order based on the first interpolation segment of precision P.

In Eq (23), the F1 score represents the harmonic mean of precision and recall, which is used to evaluate the performance of a classification model on imbalanced datasets.

In Eq (24), the mean average precision (mAP) is the average of the AP values across all categories, commonly used to assess the overall performance of a model across all categories. Here, k denotes the number of categories, which is set to 1 in this study.

4.2. Dataset description

To address the issues of insufficient quantity, incomplete categories, and inaccurate annotations in existing public datasets, we constructed a custom dataset specifically for detecting electric bicycles in elevators. By collecting surveillance images from various monitoring perspectives inside elevators and supplementing them with data gathered from online sources, we obtained a total of 4567 images.

We used the LabelImg tool for manual annotation, generating XML files containing image information and bounding box coordinates. These annotations were subsequently converted into the YOLO-compatible TXT format. The dataset was divided into training and testing sets in an 8: 2 ratio, resulting in 3653 images for training and 914 images for testing. The dataset contains a single label type: “Electric-bicycle”.

The richness and completeness of this dataset effectively support the training and evaluation of electric bicycle detection models in elevator scenarios. It ensures a sufficient number of electric bicycle image samples for training. An example from the dataset is shown in Figure 5.

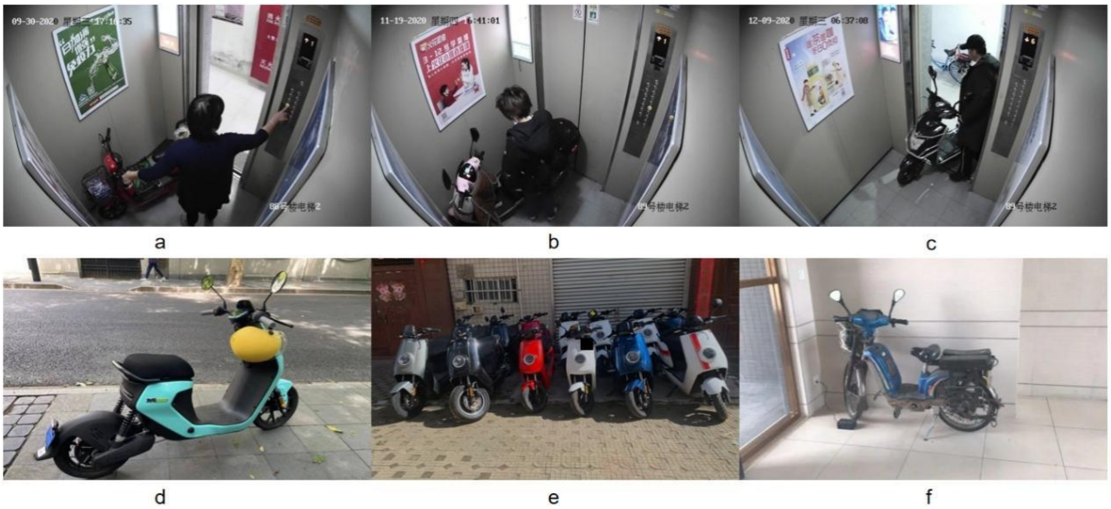


Figure 5. Sample images from the electric bicycle detection dataset used for training: (a) Electric bicycles inside elevators; (b) Electric bicycles partially occluded in elevators; (c) Electric bicycles entering elevators; (d) Electric bicycles parked outdoors; (e) Multiple electric bicycles densely parked; (f) A single electric bicycle parked indoors.

4.3. Experimental results

4.3.1. Comparison of loss functions

To delve into the impact of different loss functions on object detection performance, we selected the YOLOv10n model as the baseline and experimented with eleven loss functions, including CIoU [25], WIoU [26], and PIoU2 [22], combined with the NWD loss function to compare mAP50, as shown in Table 3. mAP50 is a critical metric for evaluating the performance of object detection models, effectively reflecting the model's accuracy in detecting objects in real-world applications.

Among all the loss functions, our PIoU2_NWD demonstrated outstanding performance, achieving the highest mAP@50 value of 0.87097. Additionally, a comparison of mAP@50 and mAP@50 + NWD across different models is presented in Figure 6. This figure clearly illustrates the differences in efficiency and accuracy among the models, highlighting their performance across various metrics. The red markers in the figure emphasize the superiority of the EBike-YOLO model, showcasing its effectiveness in the task of detecting electric bicycles in elevators.

Conventional IoU-based loss functions are often influenced by inappropriate penalty factors, causing anchor boxes to expand excessively during the regression process. This significantly slows down convergence and leads to an enlargement of the guided anchor boxes. In contrast, the PIoU2 loss function incorporates a size-adaptive penalty factor and a gradient adjustment function based on anchor box quality. This approach effectively guides anchor boxes to regress along an efficient path, resulting in faster convergence compared to traditional IoU loss functions.

In summary, the introduction of PIoU2_NWD not only addresses the limitations of existing loss functions, but also enhances object detection capabilities in complex environments by incorporating the NWD. This advancement enables superior performance in detecting electric bicycles within elevators.

Table 3. Comparison of loss functions

Model	mAP50↑	Model	mAP50↑
CIoU [25]	0.85914	CIoU + NWD	0.86467
WIoU [26]	0.86528	WIoU + NWD	0.86222
PIoU [22]	0.86478	PIoU + NWD	0.86443
DIoU [25]	0.86356	DIoU + NWD	0.86253
SIoU [27]	0.86706	SIoU + NWD	0.86409
ShapeIoU [28]	0.86205	ShapeIoU + NWD	0.86163
GIOU [29]	0.86637	GIoU + NWD	0.86137
EIoU [30]	0.86495	EIoU + NWD	0.85951
MDPIoU [31]	0.8554	MPDIoU + NWD	0.85977
InnerCIoU [32]	0.85587	InnerCIoU + NWD	0.86085
PIoU2 [22]	0.86729	PIoU2_NWD	0.87097

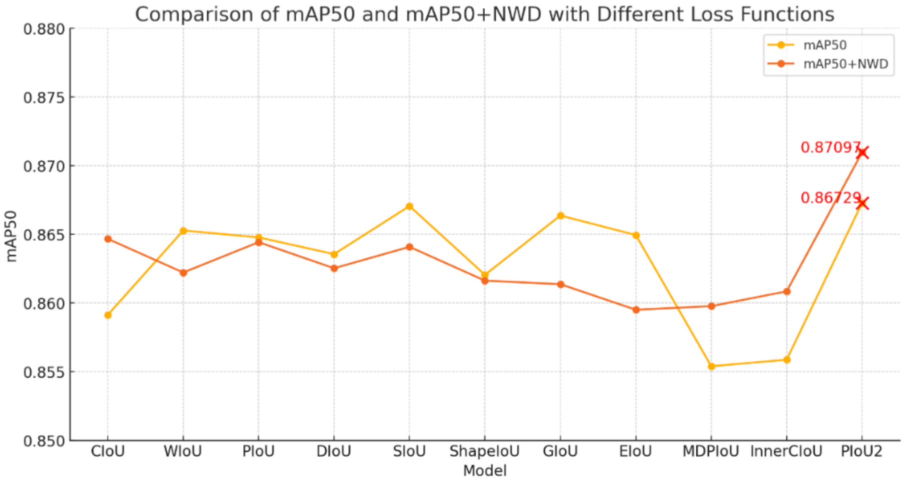


Figure 6. Comparison of mAP@50 and mAP@50 + NWD across different models. The values for PIoU2 and PIoU2_NWD are highlighted in the figure, showcasing the superior performance of PIoU2_NWD compared to other loss functions.

4.3.2. Model comparison experiments

To validate the efficiency of the EBike-YOLO model, we conducted comprehensive comparative experiments using several mainstream YOLO-series models, including YOLOv3-tiny and YOLOv5n, as well as Transformer-based models such as RTDETR. The performance of these models was evaluated across multiple metrics, including number of parameters, GFLOPs, precision, recall, F1 score, and mAP@50. The results are summarized in Table 4.

Additionally, as shown in Figure 7, we highlighted the performance variations of the EBike-YOLO model under different parameter settings using red markers. The figure clearly demonstrates the model's performance in terms of both efficiency and accuracy.

The EBike-YOLO model demonstrates outstanding performance across all metrics, particularly in key indicators such as precision, recall, F1 score, and mAP@50, outperforming other models. Specifically, EBike-YOLO achieved a 1.7% improvement in mAP@50 compared to the YOLOv10l model, despite YOLOv10l having parameter and GFLOPs values that are 7 times and 8 times higher, respectively.

In terms of precision, EBike-YOLO performs comparably to the RTDETR-X model, but requires significantly fewer parameters and computational resources, while exceeding RTDETR-X in mAP@50 by 3.122%. Additionally, although EBike-YOLO has a recall similar to YOLOv9t, it surpasses YOLOv9t in F1 score and mAP@50 by 1.504% and 2.229%, respectively. Furthermore, EBike-YOLO achieves an F1 score that is 0.825% higher than that of RTDETR-X, showcasing its significant performance advantages.

Notably, EBike-YOLO features fewer parameters and GFLOPs, giving it a clear edge in computational efficiency and resource consumption. This combination of low computational complexity and excellent detection performance makes EBike-YOLO highly suitable for real-time scenarios such as detecting electric bicycles in elevators, contributing to enhanced safety in such environments.

In summary, the EBike-YOLO model demonstrates significant advantages in both performance and cost-effectiveness. It achieves outstanding detection results while maintaining low model complexity, making it particularly well-suited for resource-constrained applications such as detecting electric bicycles in elevators.

Table 4. Comparison of different models.

Model	Parameters	GFLOPs	Precision	Recall	F1 Score	mAP50
YOLOv3tiny [10]	12,128,178	18.9	0.8666	0.77956	0.82078	0.84265
YOLOv5n	2,649,200	7.7	0.8891	0.81189	0.84874	0.855
YOLOv6n [33]	4,233,843	11.8	0.88854	0.81378	0.84952	0.860
YOLOv7tiny [34]	6,014,988	13.2	0.8940	0.7662	0.82518	0.8392
YOLOv8n	3,005,843	8.1	0.89002	0.81361	0.85010	0.86126
YOLOv9t [35]	1,970,979	7.6	0.89014	0.81422	0.85049	0.85919
YOLOv10n [36]	2,265,363	6.5	0.88145	0.81189	0.84524	0.86093
YOLOv10x [36]	29,397,491	160	0.8618	0.80224	0.83095	0.85448
YOLOv10l [36]	24,310,099	120.0	0.90152	0.80166	0.84866	0.86381
YOLOv10b [36]	19,004,883	91.6	0.88926	0.78919	0.83624	0.85794
RTDETR-ResNet18 [37]	19,873,044	56.9	0.9160	0.803	0.85579	0.84
RTDETR-ResNet50 [37]	41,936,739	125.6	0.9205	0.79784	0.85479	0.84267
RTDETR-ResNet101 [37]	60,902,755	186.2	0.91062	0.78772	0.84472	0.84318
RTDETR-X [37]	65,469,491	222.5	0.92197	0.80108	0.85728	0.85026
RTDETR-L [37]	28,445,315	100.6	0.92118	0.8	0.85632	0.85094
EBike-YOLO	3,058,452	13.2	0.92255	0.81514	0.86553	0.88148

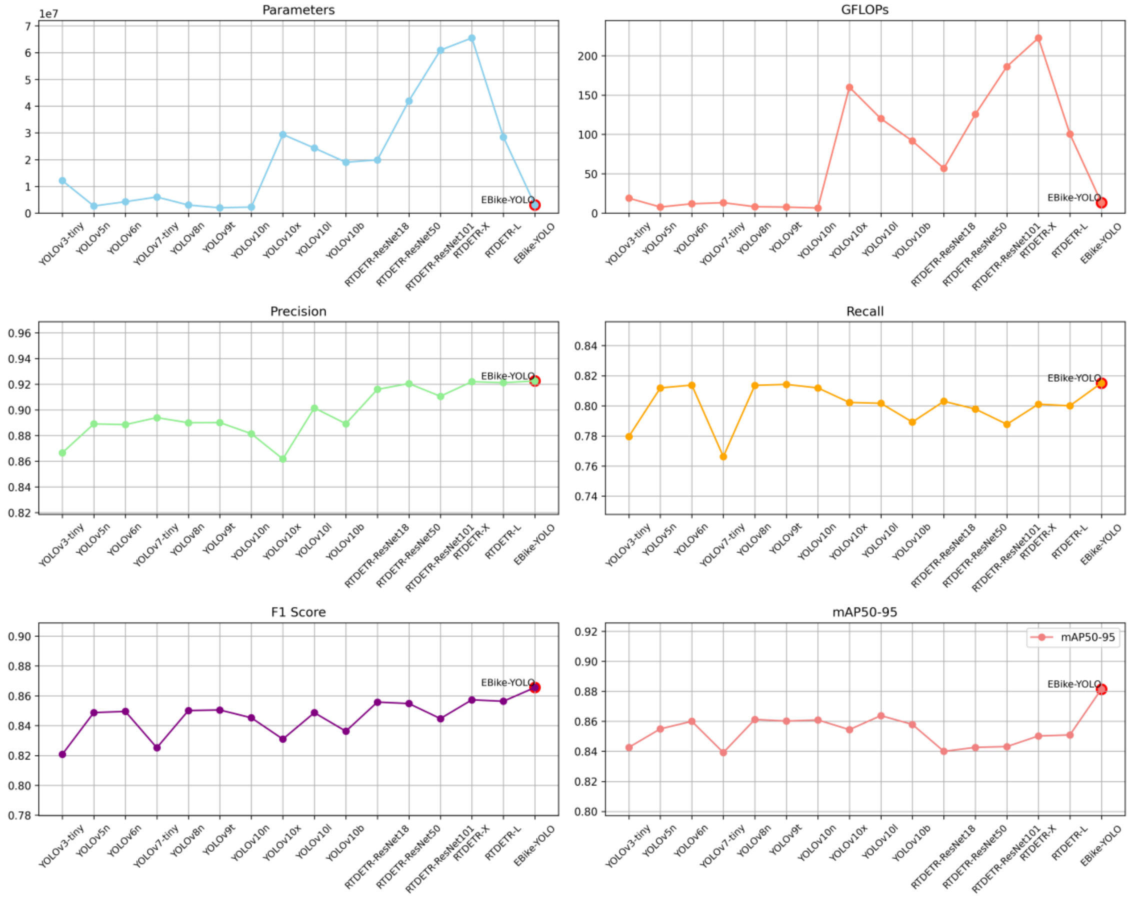


Figure 7. Performance comparison of different YOLO-series models, including parameters, GFLOPs, precision, recall, and mAP (mean average precision). The figure clearly illustrates the differences in efficiency and accuracy across the models, highlighting their performance in various metrics. The EBike-YOLO model is emphasized with red markers.

4.3.3. Ablation experiment

To assess the impact of our optimization modules, we conducted ablation experiments using a controlled variable method. Training and testing were performed on the same dataset with identical parameters. The results are presented in Table 5. Figures 8 and 9 show the convergence curves of each optimization module throughout the training process. The results demonstrate that our model achieves convergence after only 150 iterations, which ensures stable optimal performance within a shorter training time, enhancing the efficiency of object detection, particularly in complex environments like elevator-based electric bicycle identification.

By incorporating the PIoU2_NWD loss function, which combines a size-adaptive penalty factor, a gradient adjustment based on anchor box quality, a non-monotonic attention layer, and the Wasserstein distance, our model shows significant improvements in precision, F1 score, and mAP@50 (increases of 3.005%, 1.166%, and 1.183%, respectively), despite a slight decrease in recall. This loss function refines the training process by more accurately quantifying discrepancies between model predictions and ground truth annotations, encouraging the model to learn better feature representations.

To enhance small-object detection, we introduced a P2 detection head in the network's head module. This modification led to a 1.257% increase in recall and a 0.257% improvement in mAP@50, effectively improving detection accuracy for electric bicycles in elevators.

In response to the limitations of the PSA module in feature extraction and retention, we proposed the EnhancedPSA structure, which improves information extraction by applying a more robust self-attention and feed-forward network processing. This results in improvements of 2.805% in precision, 1.012% in F1 score, and 0.291% in mAP@50.

The C2f module, although effective, faces challenges with handling the deformation and rotation of electric bicycles from various angles. To address these limitations, we introduced self-calibration operations, which dynamically adjust feature weights and apply multiple filters, significantly enhancing feature extraction. This improvement resulted in a 0.503% increase in mAP@50 and maintained high precision levels.

Figures 8 and 9 show that our model surpasses YOLOv10n both in convergence speed and performance, demonstrating the superior accuracy of EBike-YOLO in detecting objects in complex environments, especially for electric bicycles in elevators.

In summary, the combination of self-calibration in the C2f module, the PIoU2_NWD loss function, the EnhancedPSA structure, and the P2 detection head greatly enhanced EBike-YOLO's performance, providing more accurate detection, faster convergence, and improved handling of diverse electric bicycle features in complex elevator environments.

Table 5. Ablation experiment results.

Model	yolov10n	PIoU2_NWD	p2	EnhancedPSA	C2f_ Calibration	Precision	Recall	F1	mAP50
1	√					0.88145	0.81189	0.84524	0.85914
2	√	√				0.91213	0.80797	0.85690	0.87097
3	√	√	√			0.89379	0.82054	0.85560	0.87354
4	√	√	√	√		0.92184	0.81604	0.86572	0.87645
Ours	√	√	√	√	√	0.92255	0.81514	0.86553	0.88148

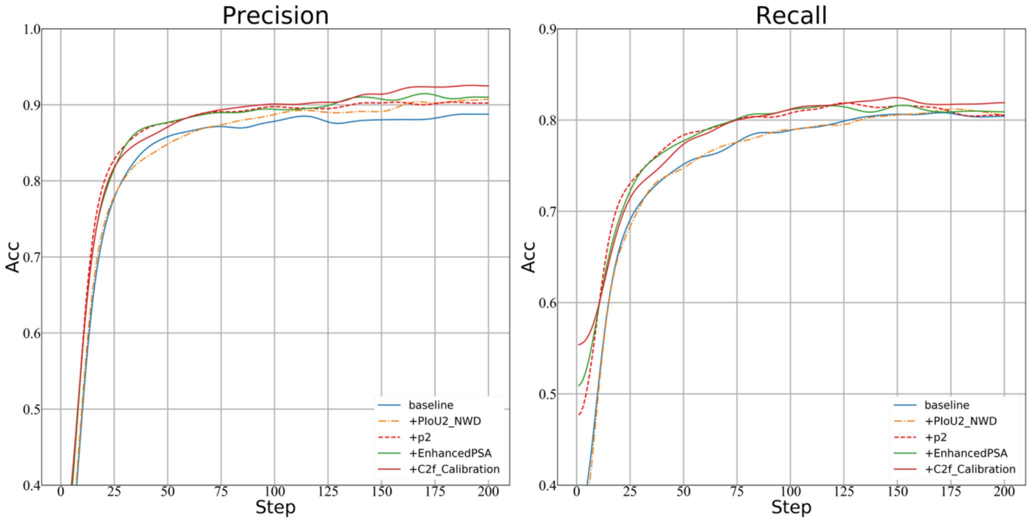


Figure 8. Precision and recall curves throughout the entire iteration process for the model.

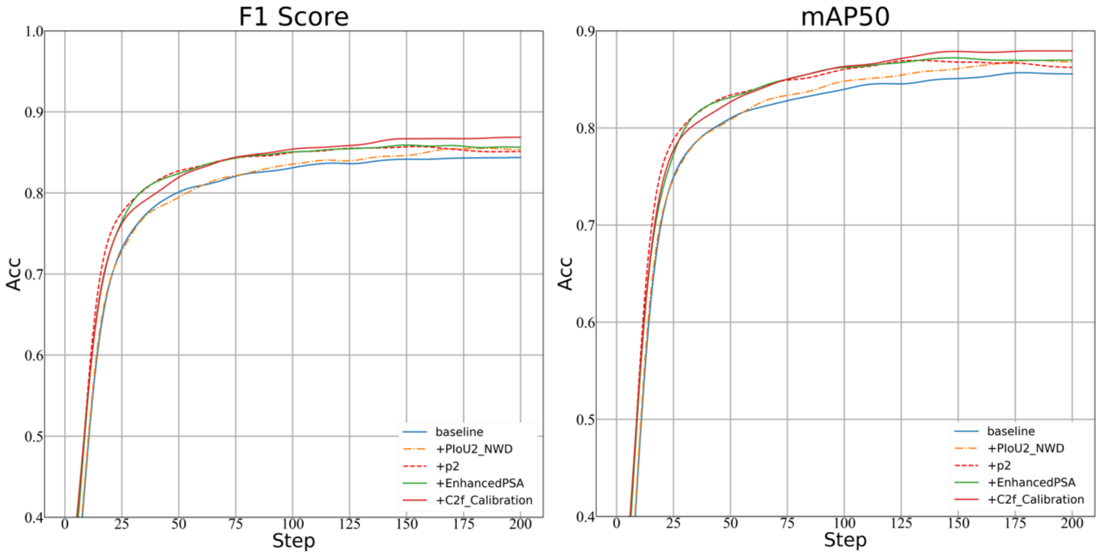


Figure 9. F1 score and mAP@50 curves throughout the entire iteration process for the model.

Using YOLOv10n as the baseline, this study incrementally integrates several key optimization modules to enhance its detection performance. We conducted comprehensive evaluations of various models regarding parameters, computational complexity (GFLOPs), frame rate (FPS), latency, and model size, with detailed experimental results presented in Table 6.

The baseline YOLOv10n itself exhibits low computational complexity (6.5 GFLOPs) and a high frame rate (109.1 FPS), laying a solid foundation for real-time detection applications. Incorporating the NWD module further improves the real-time processing capability, increasing the frame rate to 113.5 FPS and reducing latency. Subsequently, introducing the P2 module slightly decreases the frame rate to 95.4 FPS but significantly enhances feature extraction capability by increasing the model's parameters and computational cost. This improvement enables the model to handle more complex detection tasks, thereby broadening its potential application scenarios.

Further integration of the EnhancedPSA module significantly improves detection accuracy at the cost of reduced frame rate (82.3 FPS) and increased latency. This trade-off is particularly beneficial for applications where high detection precision is paramount. Finally, combining all proposed modules results in a model with 3,058,452 parameters and 13.2 GFLOPs of computational complexity. Despite increased computational demands, the model maintains an acceptable frame rate of 70.7 FPS, satisfying real-time detection criteria. Although latency increases to 0.01807 s, this slight compromise is fully justified by the substantial improvements in accuracy and robustness.

Compared with other existing models, our proposed model effectively balances accuracy and computational complexity, while preserving strong real-time performance. Through careful design and optimization of each module, the model demonstrates considerable flexibility and adaptability, meeting the diverse requirements of practical applications. It represents a robust solution for real-time monitoring systems and high-precision industrial detection tasks, marking a notable advancement in object detection methodologies and making it a suitable choice for scenarios demanding both high accuracy and real-time responsiveness.

Table 6. Ablation experiment and performance analysis.

Model	YOLOv10n	PIoU2_ NWD	p2	EnhancedPSA	C2f_ Calibration	Parameters	GFLOPs	FPS	Latency	Size
1	√					2,265,363	6.5	109.1	0.00917 s +-0.00648 s	5.5M
2	√	√				2,265,363	6.5	113.5	0.00881 s +-0.00467 s	5.5M
3	√	√	√			2,452,372	11.7	95.4	0.01048 s +-0.00339 s	5.1M
4	√	√	√	√		2,452,372	11.8	82.3	0.01215 s +- 0.00304 s	5.1M
Ours	√	√	√	√	√	3,058,452	13.2	70.7	0.01807 s +- 0.00518 s	6.3M

4.3.4. Experimental comparison based on public datasets

In this study, the EBike-YOLO model is compared against several advanced YOLO series models, including YOLOv5s, YOLOv6s, YOLOv8n, and YOLOv10n. Evaluations were conducted on three datasets: VisDrone2019, VOC2007, and VOC2012, using key performance metrics such as precision, recall, mAP50, and mAP50-95. Table 7 presents the detailed performance results of each model across the different datasets.

On the VisDrone2019 dataset, EBike-YOLO was systematically compared with mainstream detection models such as YOLOv5s, YOLOv6s, YOLOv8n, and YOLOv10n. The experimental results demonstrate that EBike-YOLO achieved the best performance across all metrics. Specifically, EBike-YOLO attained a precision of 0.489, recall of 0.358, mAP50 of 0.377, and mAP50-95 of 0.221, consistently outperforming other models. Compared to YOLOv10n, precision improved by 4.2%, recall by 2.7%, mAP50 by 4.4%, and mAP50-95 by 3.1%. Furthermore, compared with YOLOv5s, YOLOv6s, and YOLOv8n, EBike-YOLO achieved comprehensive improvements in both detection accuracy and localization precision, fully demonstrating its superior adaptability in complex scenarios such as small object detection, occlusions, and multi-scale variations.

On the VOC2007 dataset, EBike-YOLO also exhibited outstanding performance. Comparisons with models such as YOLOv3-Tiny, YOLOv5n, YOLOv6n, YOLOv8n, YOLOv9t, and YOLOv10n revealed that EBike-YOLO achieved higher scores across precision, recall, mAP50, and mAP50-95. Specifically, EBike-YOLO achieved a precision of 0.768, recall of 0.616, mAP50 of 0.703, and mAP50-95 of 0.489. Compared to YOLOv10n, precision increased by 1.5%, recall by 2.9%, mAP50 by 3.0%, and mAP50-95 by 2.6%. In addition, EBike-YOLO maintained a clear advantage over new-generation lightweight models such as YOLOv8n and YOLOv9t, further validating its robustness and generalization capabilities in multi-scenario object detection tasks.

On the VOC2012 dataset, EBike-YOLO's performance improvements were even more pronounced. Compared with models such as YOLOv3-Tiny, YOLOv5n, YOLOv8n, and YOLOv10n, EBike-YOLO achieved leading results across all evaluation metrics. Specifically, it achieved a precision of 0.664, recall of 0.540, mAP50 of 0.589, and mAP50-95 of 0.423. Compared to YOLOv10n, precision improved by 7.0%, recall by 3.4%, mAP50 by 4.4%, and mAP50-95 by 2.9%.

Meanwhile, compared to YOLOv8n, EBike-YOLO further optimized detection precision for small objects and complex environments. This indicates that EBike-YOLO not only performs excellently on standard datasets but also maintains stable and efficient detection performance across a broader range of IoU thresholds.

In summary, EBike-YOLO consistently outperforms existing YOLO series models across multiple datasets and evaluation metrics, especially excelling in object detection accuracy and recall in complex scenarios, providing a more reliable and efficient solution for real-world object detection tasks.

Table 7. Model comparison study on public datasets.

Datasets	Model	Precision	Recall	mAP50	mAP50-95
VisDrone2019 [38]	YOLOv5s	0.413	0.316	0.318	0.176
	YOLOv6s	0.456	0.356	0.352	0.208
	YOLOv8n	0.450	0.338	0.339	0.196
	YOLOv10n	0.447	0.331	0.333	0.190
	EBike-YOLO	0.489	0.358	0.377	0.221
VOC2007 [39]	YOLOv3-Tiny	0.717	0.541	0.618	0.361
	YOLOv5n	0.722	0.588	0.668	0.439
	YOLOv6n	0.742	0.599	0.681	0.470
	YOLOv8n	0.753	0.604	0.687	0.471
	YOLOv9t	0.723	0.630	0.693	0.487
	YOLOv10n	0.753	0.587	0.673	0.463
	EBike-YOLO	0.768	0.616	0.703	0.489
VOC2012 [40]	YOLOv3-Tiny	0.493	69.03	0.426	0.200
	YOLOV5n	0.608	0.509	0.541	0.374
	YOLOV8n	0.634	0.517	0.558	0.395
	YOLOv10n	0.620	0.506	0.545	0.393
	EBike-YOLO	0.664	0.540	0.589	0.423

4.3.5. Visualization analysis

To comprehensively demonstrate the significant improvements of our EBike-YOLO model in detecting electric bicycles within elevators, we conducted a series of multi-scenario tests to evaluate its performance and robustness. Figure 10 illustrates the detection results across eight groups (a-h). Each group contains three columns: the first column presents the original image, the second shows the detection result from the baseline YOLOv10n model, and the third displays the result from our EBike-YOLO model.

Group a-Panoramic Entry Detection (Figure 10(a)): This group tested panoramic images of electric bicycles as they entered the elevator. EBike-YOLO achieved an 8% improvement in accuracy compared to the YOLOv10n model, demonstrating its enhanced ability to detect newly appearing objects.

Group b-Frontal Detection (Figure 10(b)): This scenario focused on the detection of electric bicycle front sections. Our model showed a 10% increase in accuracy, confirming its superior capability in capturing local feature details.

Group c-Enhanced Localization Precision (Figure 10(c)): In this test, we further assessed front-section recognition. EBike-YOLO improved detection accuracy by 50% and significantly enhanced bounding box precision, demonstrating the reliability of the proposed model modifications.

Group d-Occlusion Handling (Figure 10(d)): Electric bicycles partially obscured by raincoats were tested in this group. Despite significant occlusion, our model maintained high detection accuracy and precise localization, highlighting its robustness in complex environments.

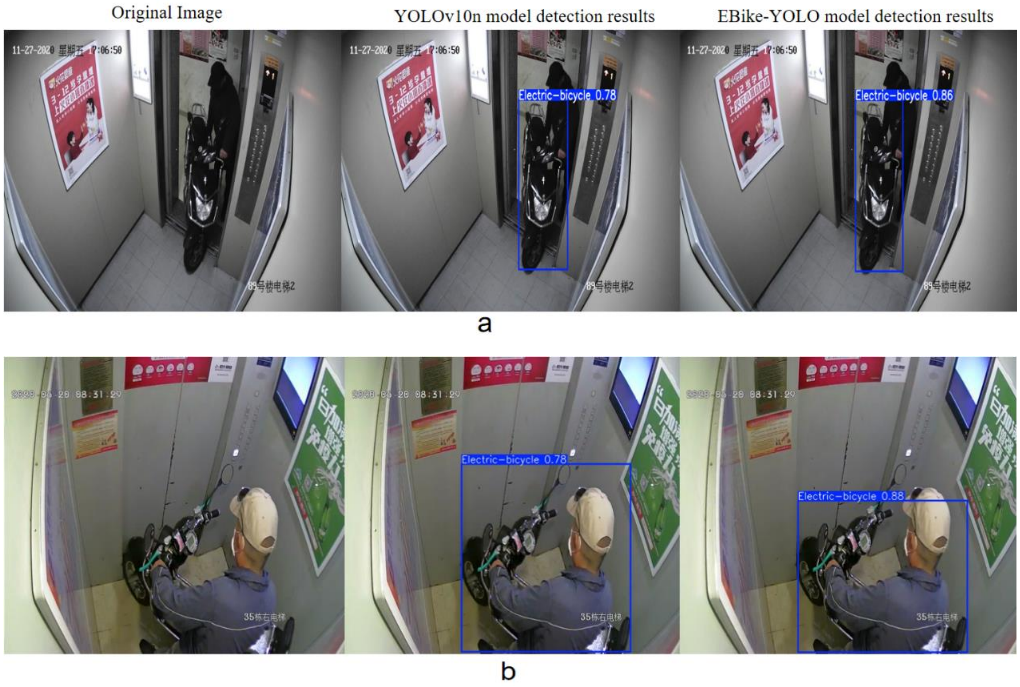
Group e-Low-light Frontal Detection (Figure 10(e)): This group evaluated detection in a dark elevator, with windshields covering the front of the electric bicycles. EBike-YOLO outperformed the baseline with a 13% accuracy increase, showcasing its strength in low-light conditions.

Group f-Rear Detection in Low Light (Figure 10(f)): Here, rear sections of bicycles in dim environments were evaluated. EBike-YOLO delivered a 15% increase in detection accuracy, further proving its adaptability to varying lighting conditions.

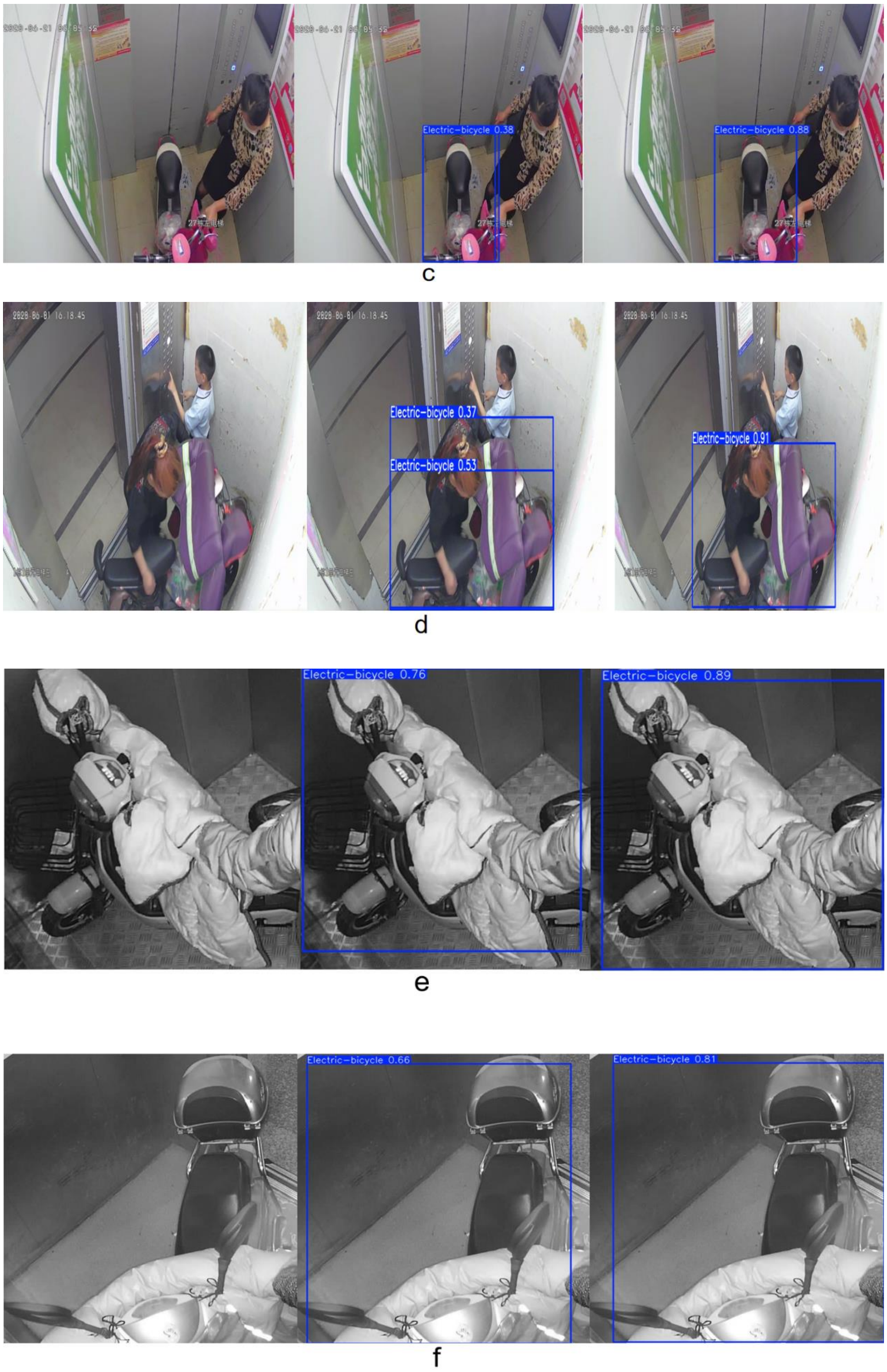
Group g-Missed Detection Recovery (Figure 10(g)): This scenario simulated a missed detection case. While YOLOv10n failed to identify the bicycle, EBike-YOLO successfully detected it with a confidence score of 0.72, underscoring its superior detection reliability in challenging elevator scenarios.

Group h-Outdoor Detection (Figure 10(h)): In outdoor environments, the baseline model missed one e-bike, whereas EBike-YOLO detected all the e-bikes with improved confidence scores. This validates the model's generalization ability beyond elevator-specific conditions.

In conclusion, our algorithm demonstrates exceptional classification accuracy and localization precision in the task of detecting electric bicycles within elevators, fully meeting the stringent requirements of industrial equipment for high-precision detection. By significantly enhancing detection accuracy and reliability, our approach effectively improves elevator safety and ensures robust risk control during charging processes. Furthermore, this innovative technology exhibits great potential in practical applications, offering a new direction for the development of intelligent monitoring systems.



continued on next page



c

d

e

f

continued on next page

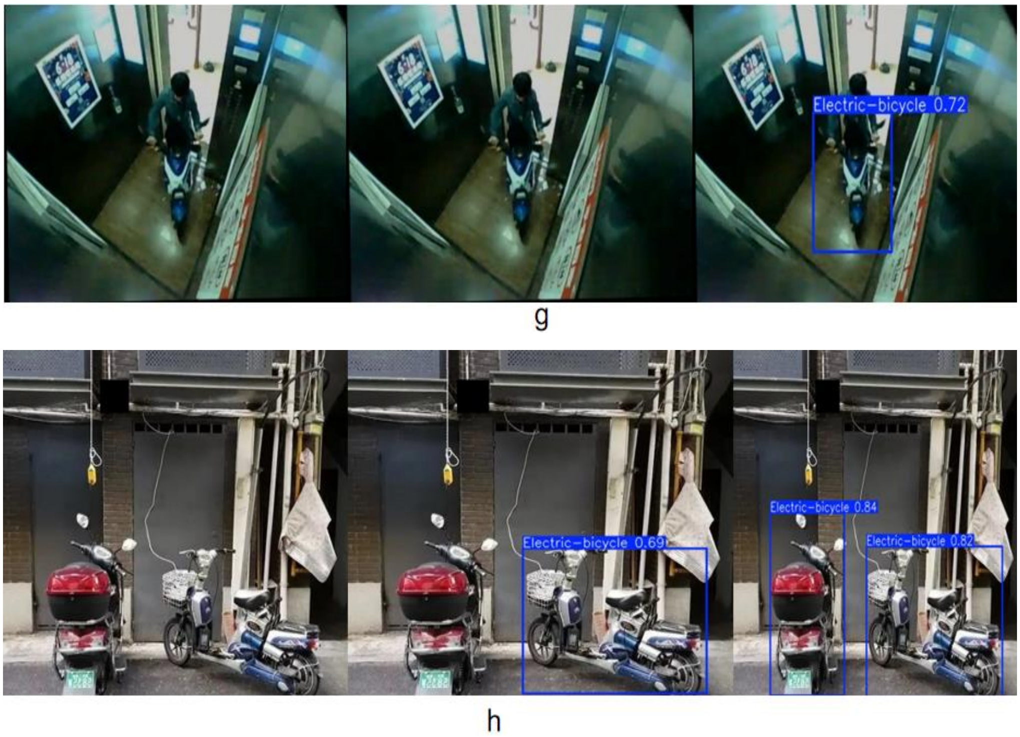


Figure 10. Detection results on the test dataset comparing the original YOLOv10n and our EBike-YOLO model. Group a: Panoramic images of electric bicycles entering the elevator. Group b: Detection of bicycle front sections within the elevator. Group c: Detection of bicycle rear sections within the elevator. Group d: Detection of bicycles partially obscured within the elevator. Group e: Detection of bicycle fronts in dark elevator environments. Group f: Detection of bicycle rears in dark elevator environments. Group g: Detection of electric bicycles under missed detection conditions in an elevator scenario. Group h: Detection of bicycle rears in an outdoor environment.

5. Conclusions

This paper introduces EBike-YOLO, an enhanced real-time detection algorithm based on YOLOv10n, incorporating the PIoU2_NWD loss function, self-calibration in the C2f module, and an EnhancedPSA structure. Experimental evaluations on a custom dataset of electric bicycles in elevator scenarios demonstrate that EBike-YOLO achieves notable improvements over the baseline YOLOv10n, specifically a 4.1% increase in detection accuracy, a 0.3% increase in recall, a 2.0% improvement in F1 score, and a 2.23% enhancement in mAP@50. These quantified performance gains clearly underscore the effectiveness and practical applicability of the proposed model, particularly emphasizing its robust detection capabilities in complex, real-world elevator environments.

Future research will focus on further refining EBike-YOLO to enhance its adaptability to diverse lighting conditions and complicated backgrounds. Additionally, lightweight optimization strategies will be explored to reduce computational demands, thereby facilitating broader and more efficient deployment across various hardware platforms. Moreover, we plan to extend the model’s applicability to other enclosed spaces, such as stairwells and underground garages, and evaluate its potential for detecting additional vehicle types, including motorcycles and shared bicycles, thus further expanding its practical value.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This research was funded The National Natural Science Foundation of China, grant number 61976198, The Natural Science Research Key Project for Colleges and University of Anhui Province, grant number 2022AH052141, 2022AH052142 and 2024AH051567, and The Hefei Key Technology R&D "Champion-Based Selection" Project, grant number 2023SGJ011.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. Y. Liu, Q. Xu, Y. Yang, W. Zhang, Detection of electric bicycle indoor charging for electrical safety: A NILM approach, *IEEE Trans. Smart Grid*, **14** (2023), 3862–3875. https://doi.org/10.1109/TSG.2023.3245636
2. Y. Li, L. Han, X. Ning, Y. Xu, Fire risk of electric bicycle based on fuzzy Bayesian network, *J. Phys.: Conf. Ser.*, **1578** (2020), 012153. https://doi.org/10.1088/1742-6596/1578/1/012153
3. W. Ying, Z. Yongping, X. Fang, X. Jian, Analysis model for fire accidents of electric bicycles based on principal component analysis, in *2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, **1** (2017), 760–762. https://doi.org/10.1109/CSE-EUC.2017.149
4. P. Viola, M. J. Jones, Robust real-time face detection, *Int. J. Comput. Vis.*, **57** (2004), 137–154. https://doi.org/10.1023/B:VISI.0000013087.49260.fb
5. N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Pecognition (CVPR'05)*, **1** (2005), 886–893. https://doi.org/10.1109/CVPR.2005.177
6. S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, *Adv. Neural Inf. Process. Syst.*, **28** (2015).
7. R. Girshick, Fast r-cnn, in *Proceedings of the IEEE International Conference on Computer Vision*, (2015), 1440–1448. https://doi.org/10.1109/ICCV.2015.169
8. J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2016), 779–788. https://doi.org/10.1109/CVPR.2016.91
9. J. Redmon, A. Farhadi, YOLO9000: Better, faster, stronger, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2017), 7263–7271. https://doi.org/10.1109/CVPR.2017.690
10. J. Redmon, A. Farhadi, YOLOv3: An incremental improvement, preprint, arXiv:1804.02767.

11. A. Bochkovskiy, C. Y. Wang, H. Y. M. Liao, YOLOv4: Optimal speed and accuracy of object detection, preprint, arXiv:2004.10934.
12. L. Li, X. Wang, M. Yang, H. Zhang, An accurate shared bicycle detection network based on faster R-CNN, *IET Image Process*., **17** (2023), 1919–1930. https://doi.org/10.1049/ipr2.12766
13. A. John, D. Meva, N. Arora, Deep learning based road traffic assessment for vehicle rerouting: An extensive experimental study of RetinaNet and YOLO models, *Int. Res. J. Multidiscip. Technovation*, **6** (2024), 134–152. https://doi.org/10.54392/irjmt2459
14. J. Miguel, P. Mendonça, A. Quelhas, J. M. Caldeira, V. N. Soares, Using computer vision to collect information on cycling and hiking trails users, *Future Internet*, **16** (2024), 104. https://doi.org/10.3390/fi16030104
15. W. Wang, Y. Xu, Z. Xu, C. Zhang, T. Li, J. Wang, et al., A detection method of electro-bicycle in elevators based on improved YOLO v4, in *2021 26th International Conference on Automation and Computing (ICAC)*, (2021), 1–6. https://doi.org/10.23919/ICAC50006.2021.9594217
16. C. Zhang, A. Xiong, X. Luo, C. Zhou, J. Liang, Electric bicycle detection based on improved YOLOv5, in *2022 4th International Conference on Advances in Computer Technology, Information Science and Communications (CTISC)*, (2022), 1–5. https://doi.org/10.1109/CTISC54888.2022.9849750
17. J. Sun, Y. Zhang, Electric bicycle detection based on deep learning, in *Proceedings of the 5th International Conference on Computer Science and Software Engineering*, (2022), 115–120. https://doi.org/10.1145/3569966.3570001
18. J. Su, M. Yang, X. Tang, Integration of ShuffleNet V2 and YOLOv5s networks for a lightweight object detection model of electric bikes within elevators, *Electronics*, **13** (2024), 394. https://doi.org/10.3390/electronics13020394
19. Z. Zhao, S. Li, C. Wu, X. Wei, Research on the rapid recognition method of electric bicycles in elevators based on machine vision, *Sustainability*, **15** (2023), 13550. https://doi.org/10.3390/su151813550
20. Z. Liu, C. Dai, X. Li, An electric bicycle tracking algorithm for improved traffic management, *Heliyon*, **10** (2024). https://doi.org/10.1016/j.heliyon.2024.e32708
21. H. Zhong, Z. Wang, Z. Chen, W. Chen, Y. Li, A novel fire monitoring system for electric bicycle shed based on YOLOv8, in *2023 IEEE 16th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC)*, (2023), 142–147. https://doi.org/10.1109/MCSoC60832.2023.00029
22. C. Liu, K. Wang, Q. Li, F. Zhao, K. Zhao, H. Ma, Powerful-IoU: More straightforward and faster bounding box regression loss with a nonmonotonic focusing mechanism, *Neural Networks*, **170** (2024), 276–284. https://doi.org/10.1016/j.neunet.2023.11.041
23. J. Wang, C. Xu, W. Yang, L. Yu, A normalized Gaussian Wasserstein distance for tiny object detection, preprint, arXiv:2110.13389.
24. J. J. Liu, Q. Hou, M. M. Cheng, C. Wang, J. Feng, Improving convolutional networks with self-calibrated convolutions, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2020), 10096–10105.
25. Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, D. Ren, Distance-IoU loss: Faster and better learning for bounding box regression, in *Proceedings of the AAAI Conference on Artificial Intelligence*, **34** (2020), 12993–13000. https://doi.org/10.1609/aaai.v34i07.6999

26. Z. Tong, Y. Chen, Z. Xu, R. Yu, Wise-IoU: Bounding box regression loss with dynamic focusing mechanism, preprint, arXiv:2301.10051.
27. Z. Gevorgyan, SIoU loss: More powerful learning for bounding box regression, preprint, arXiv:2205.12740.
28. H. Zhang, S. Zhang, Shape-IoU: More accurate metric considering bounding box shape and scale, preprint, arXiv:2312.17663.
29. H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, S. Savarese, Generalized intersection over union: A metric and a loss for bounding box regression, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2019), 658–666. https://doi.org/10.1109/CVPR.2019.00075
30. Y. F. Zhang, W. Ren, Z. Zhang, Z. Jia, L. Wang, T. Tan, Focal and efficient IOU loss for accurate bounding box regression, *Neurocomputing*, **506** (2022), 146–157. https://doi.org/10.1016/j.neucom.2022.07.042
31. S. Ma, Y. Xu, MPDIoU: A loss for efficient and accurate bounding box regression, preprint, arXiv:2307.07662.
32. H. Zhang, C. Xu, S. Zhang, Inner-IoU: More effective intersection over union loss with auxiliary bounding box, preprint, arXiv:2311.02877.
33. C. Li, L. Li, H. Jiang, K. Weng, Y. Geng, L. Li, et al., YOLOv6: A single-stage object detection framework for industrial applications, preprint, arXiv:2209.02976.
34. C. Y. Wang, A. Bochkovskiy, H. Y. M. Liao, YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2023), 7464–7475. https://doi.org/10.1109/CVPR52729.2023.00721
35. C. Y. Wang, I. H. Yeh, H. Y. Mark Liao, Yolov9: Learning what you want to learn using programmable gradient information, in *European Conference on Computer Vision*, (2024), 1–21. https://doi.org/10.1007/978-3-031-72751-1_1
36. A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, Yolov10: Real-time end-to-end object detection, *Adv. Neural Inf. Process. Syst.*, **37** (2024), 107984–108011.
37. Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, et al., DETRs beat YOLOs on real-time object detection, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2024), 16965–16974. https://doi.org/10.1109/CVPR52733.2024.01605
38. D. Du, P. Zhu, L. Wen, X. Bian, H. Lin, Q. Hu, et al., VisDrone-DET2019: The vision meets drone object detection in image challenge results, in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019.
39. M. Everingham, L. Van Gool, C. K. Williams, J. Winn, A. Zisserman, The pascal visual object classes (voc) challenge, *Int. J. Comput. Vis.*, **88** (2010), 303–338. https://doi.org/10.1007/s11263-009-0275-4
40. M. Everingham, J. Winn, The pascal visual object classes challenge 2012 (voc2012) development kit, *Pattern Anal. Stat. Model. Comput. Learn., Tech. Rep.*, **8** (2011), 2–5.

AIMS Press

©2025 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0)