IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 12 July 2024, accepted 24 July 2024, date of publication 29 July 2024, date of current version 9 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3435146

RESEARCH ARTICLE

Loader Bucket Working Angle Identification Method Based on YOLOv5s and EMA Attention Mechanism

XUEDONG ZHANG[1], BO CUI[1,2], ZHAOXU WANG[1], AND WANGTING ZENG[1]
[1]College of Artificial Intelligence, North China University of Science and Technology, Tangshan 063210, China
[2]Hebei Key Laboratory of Industrial Intelligent Perception, Tangshan 063210, China

Corresponding author: Bo Cui (mikecui@ncst.edu.cn)

This work was supported in part by the Research Project on Major Issues of Industrial Technology Innovation and Economic Development in Hebei Province under Grant 20557605D; in part by the Ministry of Education Project for the Teaching Steering Committee of Electronic Information Majors in Higher Education Institutions under Grant 2021-JG-04; in part by the Ministry of Education Industry-Education Cooperation Collaborative Education Project 202101138019; and in part by the North China University of Science and Technology Graduate Innovation Project, in 2023, under Grant 2023047.

ABSTRACT In response to the issues of low recognition efficiency and large errors encountered in the process of identifying the working angle of the bucket during current automated loader construction operations, a method based on YOLOv5s and the EMA attention mechanism for loader bucket working angle identification is proposed. Initially, a small target detection head, utilizing YOLOv5s, was designed to enhance sensitivity towards target recognition. The EMA attention mechanism was introduced to increase the recognition rate of the target area and the positioning accuracy of the target frame, effectively differentiating the background area from the target area. The Focal-EIOU Loss function was added to address the slow convergence speed of YOLOv5. Subsequently, Depth Separable Convolution was employed to replace the standard convolution in the C3 module of the Backbone, improving the model's accuracy in identifying target deformation caused by changes in the bucket angle, reducing the computational load, and enhancing the model's operational speed. Experimental results demonstrate that the model's mean Average Precision (mAP) value reached 99.3%, a 3.0% increase over the benchmark model YOLOv5s. The GFLOPs reached 58.5, an increase of 42, with a growth rate of 254.55%. This method effectively enhances the precision and intelligence of loader construction operations.

INDEX TERMS Angle identification, YOLOv5s, small object detection, attention mechanism, depth separable convolution.

I. INTRODUCTION

Modern intelligent construction machinery has successfully integrated artificial intelligence, digital technology, robotics, and Cyber-Physical Systems, leading to significant technological innovations and upgrades of traditional construction equipment. Numerous construction machinery enterprises are continuously exploring and making breakthroughs in new technological fields [1]. The pace of market research and development has notably accelerated, with intelligent assistance driving, autonomous driving, and new energy technologies playing crucial roles in enhancing the safety of loader operations [2].

The shovel loading phase is where loaders consume the most energy and where operators engage most frequently [3]. The materials handled by loaders during shovel operations vary greatly in composition and complexity. As the shovel penetrates deeper, material accumulation at the tip of the shovel increases, and the work resistance escalates geometrically [4], [5]. Timely adjustment of the working device's posture can effectively break through the material in front of the shovel, facilitating reduced resistance penetration.

The associate editor coordinating the review of this manuscript and approving it for publication was Rajeeb Dey.

© 2024 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

Acquiring accurate information on the loader bucket's angle is crucial for automated and energy-efficient operations during the shovel loading phase.

Recent years have seen significant progress in the field of construction machinery equipment angle recognition, with methods mainly divided into contact and non-contact categories. Contact measurement techniques utilize built-in sensor systems in loaders to gather posture data. For example, a method proposed by Huang [6] utilizes a sensor network to achieve loader translation control by monitoring the angles of the boom and the arm to control the shovel's translation. Additionally, Qiu et al. [7] from XCMG Group developed an auxiliary operation control system based on the CAN bus system, enabling automatic positioning and one button lifting and leveling of the loader. The Mechatronics Engineering team [8] from Hanyang University chose low-cost sensors to identify the position of the excavator's bucket tip by measuring the cylinder lengths and direct movements of the boom, arm, and bucket. However, these studies did not fully explore the adaptability and stability of sensors under complex environmental conditions, such as extreme weather or terrain changes. While feasible under specific conditions, the accuracy of these contact methods can be affected by the working status of the machine body, and they are limited by high maintenance costs and short lifespans.

In the realm of non-contact measurement, Aldekoa et al. [9] have demonstrated groundbreaking achievements. They effectively monitored tool wear utilizing servo motor data, significantly reducing the complexity and cost of the monitoring system while ensuring the timeliness and accuracy of the monitoring outcomes, thus infusing new vitality into the development of non-contact measurement technologies. Machine vision technology, as a type of non-contact measurement technique, has also exhibited distinct advantages. For instance, Zhao et al. [10] introduced a method based on a monocular vision marker system for real-time automatic monitoring of the excavator joystick's posture, encompassing its orientation and position. However, the accuracy of this method hinges critically on the visibility and proper recognition of the markers. This reliability can be jeopardized under several conditions: when the markers are obscured or damaged, or when the camera's angle fails to align correctly with the mechanical arm. Cao et al. [11] designed a system that combines angle sensors and the CAN network to analyze multisource data, thereby achieving posture adjustment of the bucket at various stages. Machine vision technology, with its ability to continuously monitor target postures without fatigue, is particularly suited for long-term operations in harsh environments, offering a new avenue for enhancing the intelligence and problem-solving capabilities of loaders.

To address the issues of low accuracy and high error margins in bucket working angle recognition, and to effectively improve the safety and intelligence level of loaders, this article opts for the YOLOv5s object detection algorithm, known for its superior precision and speed compared to traditional detection methods. While newer versions like YOLOv7 and YOLOv8 offer easier deployment and smaller, less resource-intensive models, they still face challenges with small and densely packed object detection. This paper aims to improve the YOLOv5s algorithm for better performance in detecting small and dense objects.

II. MATERIALS AND METHOD

A. TARGET DESIGN

Due to the indistinct features of the bucket itself, the accuracy of object detection using YOLOv5s is relatively low. By arranging characteristic images on the bucket, the detection accuracy of YOLOv5s can be improved for angle recognition [12]. For this purpose, a type of target that is easily recognizable by stereo cameras and can accurately reflect the angular posture of mechanical equipment was designed. As shown in Figure 1, each target uses binary encoding, with the basic shapes including squares and circles. The circular design ensures that the line connecting the center of the circle is parallel to the component, while the square is used for encoding. By dividing each side of the square into three parts and representing binary "0" or "1" through the presence or absence of different edges, this design allows for the target's deformation to remain minimal in the image, even when the contour captured by stereo cameras is hard to recognize due to deformation, thus facilitating easy detection.

FIGURE 1. Target design: A target designed using binary coding, where each side of a square is divided into three equal parts, and the absence of a different side is indicated as "0" and the presence of a different side is indicated as "1."

B. ANGLE CALCULATION METHOD

1) 3D RECONSTRUCTION

Given that the real world is three-dimensional while computer images capture two-dimensional data, performing 3D reconstruction based on the captured 2D information is necessary. This is achieved using stereo vision technology, where points matched between the left and right views facilitate the reconstruction. Pixel coordinates of corresponding points in the left and right views are captured. Assuming that the matched points in the left and right views are $D_l\ (x_l, y_l)$ and $D_r\ (x_r, y_r)$ respectively, their 3D coordinates in the camera coordinate system are calculated using triangulation. Assuming the camera's focal length is f, the 3D coordinates in the camera coordinate system are $D_d\ (x_l, y_l, f)$ and $D_{\mathrm{cr}}\ (x_r, y_r, f)$. The left camera coordinate system is selected as the world coordinate system. Ideally, the intersection of the rays O_dD_l and $O_{cr}D_r$ would pinpoint the position of the 3D point. However, due to matching errors, these rays typically do not intersect. To address this, the perpendicular bisector method is used,

setting vector w perpendicular to both $O_{dl}D_l$ and $O_{dl}D_l$, computed as $\mathrm{w} = D_{dl} \times R_{r2l}D_{cr}$. Using triangulation:

$$aD_{dl} - bR_{r2l}D_{cr} + cw = t_{r2l} \tag{1}$$

The 3D point coordinates can be expressed as:

$$D_w = aD_{dl} + \frac{cw}{2} \tag{2}$$

2) LOADER POSE ANGLE CALCULATION

Square-coded targets are affixed to the boom and arm to detect their planar coordinates $\mathrm{D}_1\left(\mathrm{x}_1, \mathrm{y}_1\right)$, $\mathrm{D}_2\left(\mathrm{x}_2, \mathrm{y}_2\right)$, $\mathrm{D}_3\left(\mathrm{x}_3, \mathrm{y}_3\right)$. The inclination angle between each pair of coded points is computed as:

$$\theta_{12} = \arctan\frac{y_1 - y_2}{x_1 - x_2} \tag{3}$$

$$\theta_{13} = \arctan\frac{y_1 - y_3}{x_1 - x_3} \tag{4}$$

$$\theta_{23} = \arctan\frac{y_2 - y_3}{x_2 - x_3} \tag{5}$$

The average of these angles is used as the pose angle:

$$\theta_2 = \frac{\theta_{12} + \theta_{13} + \theta_{23}}{3} \tag{6}$$

3) KINEMATIC MODELING

The Denavit-Hartenberg (D-H) parameter method is used to model the kinematics of the loader's working mechanism, describing the motion relations between each link and joint. D-H parameters define the geometric relationships between two adjacent links, including:

Link length a_i: Distance along the i link axis from the i joint axis to the $i+1$ joint axis.

Link offset d_i : Distance along the i joint axis from the i joint axis to the $i+1$ joint axis.

Link twist angle α_i: Angle between the i joint axis and the $i+1$ joint axis, rotated around the i link axis.

Joint angle θ_i: Angle between the i joint axis and the $i+1$ link axis, rotated around the i joint axis.

To describe the motion among different components of the loader, coordinate systems are established at each joint and link. The loader's mechanism consists of a base, boom, arm, and bucket, all connected by rotary joints:

Base coordinate system $x_0y_0z_0o_0$: Fixed and centered at the base.

Boom coordinate system $x_1y_1z_1a_1$: Connected to the boom, linked to the base via a rotary joint.

Arm coordinate system $x_2y_2z_2O_2$: Connected to the arm, linked to the boom via a rotary joint. Bucket coordinate system $x_3y_3z_3o_3$:

Connected to the bucket, linked to the arm via a rotary joint. Each joint's coordinate transformation is represented by a 4×4 homogeneous transformation matrix. By multiplying these matrices, a comprehensive transformation matrix from the base to any link end is obtained, describing the position and orientation of the end-effector in the base coordinate system:

$$T_i^{-1} = \begin{vmatrix} C_i & -\cos\alpha_i S_i & \sin\alpha_i S_i & \alpha_i C_i \\ S_i & \cos\alpha_i C_i & -\sin\alpha_i S_i & \alpha_i S_i \\ 0 & S_i & \cos\alpha_i & d_i \\ 0 & 0 & 0 & 0 \end{vmatrix} \tag{7}$$

By multiplying the transformation matrices for each joint, the overall transformation matrix from the base to the bucket end is derived, representing the bucket's position and orientation relative to the base:

$$\begin{aligned} T_4^0 &= T_1^0 T_2^1 T_3^2 T_4^3 \\ &= \begin{vmatrix} C_1C_{234} & -C_1C_{234} & S_1 & C_1\left(a_4C_{234}+a_3C_{23}+a_2C_2+a_1\right) \\ S_1C_{234} & -S_1C_{244} & C_2 & S_1\left(a_4C_{234}+a_3C_{23}+a_2C_2+a_1\right) \\ S_{234} & C_{234} & 0 & a_4S_{234}+a_3S_{23}+a_2S_2+d_1) \\ 0 & 0 & 0 & 1 \end{vmatrix} \end{aligned} \tag{8}$$

The pose space of the loader is determined by the tip's position and orientation angles in the base coordinate system. Assuming each joint angle is positive in the counterclockwise direction, the coordinate transformation of $x_iy_iz_io_i$ relative to $x_{i-1}y_{-1}z_{i-1}o_{i-1}$ can be expressed in matrix form:

$$\begin{cases} x = C_1\left(a_4C_{234} + a_3C_{23} + a_2C_2 + a_1\right) \\ y = S_1\left(a_4C_{234} + a_3C_{23} + a_2C_2 + a_1\right) \\ z = a_4S_{234} + a_3S_{23} + a_2S_2 + d_1) \\ \theta_w = \theta_2 + \theta_3 + \theta_4 \end{cases} \tag{9}$$

4) BUCKET POSE ANGLE CALCULATION

Targets are placed on the bucket linkage, forming vectors $\overline{AB}$ and $\overline{AO_3}$, with directional angles θ and $\theta_2 + \theta_3$ respectively. The directional angles φ_1 and φ_2 are calculated as:

$$\varphi_2 = \arctan\frac{|\overline{AB}|\sin\theta - \left|\overline{AO_3}\right|\sin\left(\theta_2+\theta_3\right)}{|\overline{AB}|\cos\theta - \left|\overline{AO_3}\right|\cos\left(\theta_2+\theta_3\right)} \tag{10}$$

$$\varphi_1 = \arccos\frac{|O_3B|^2 + |\mathrm{O_3C}|^2 + |BC|^2}{2\,|\mathrm{O_3C}||\mathrm{BC}|} \tag{11}$$

Thus, the pose angle θ_w is:

$$\theta_w = \varphi_2 - \varphi_1 - \varphi_0 \tag{12}$$

C. YOLOv5s BASE MODEL

The YOLOv5 (You Only Look Once version 5) [13], launched by Ultralytics in 2020, has been highly acclaimed for its exceptional real-time object detection capabilities. It is a Python-based lightweight detection model that comes in four versions: YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x, with each version progressively increasing in model weight and structure. This paper focuses on the YOLOv5s version, which achieves rapid and precise object detection through a simplified algorithm structure, demonstrating the core advantages of the YOLO series in terms of speed and accuracy. The framework of this version is illustrated in Figure 2.

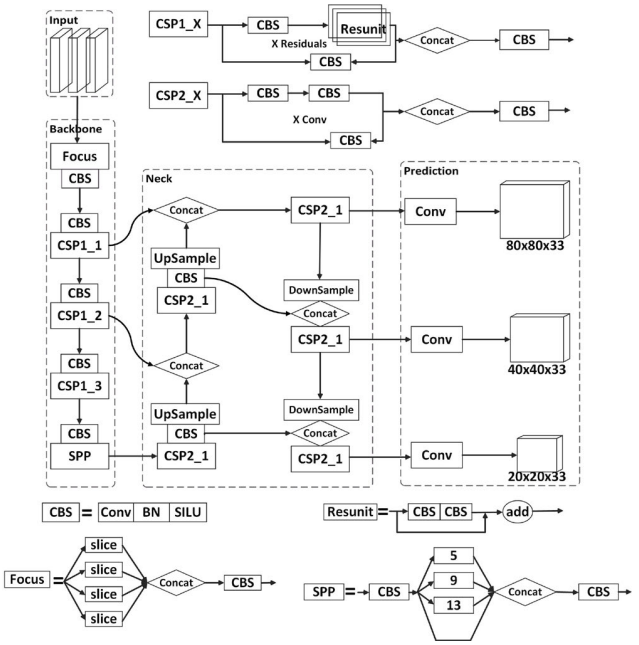

FIGURE 2. The basic framework for YOLOv5s based on Python lightweight detection models.

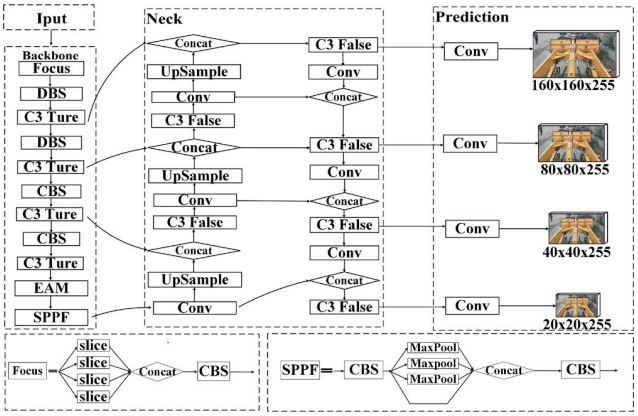

FIGURE 3. Add a small target detection header on top of the original detection header, add EMA attention mechanism in front of backbone's SPPF module, introduce Focal-EIOU Loss to replace YOLOv5s' GIoU, introduce Depth Separable Convolution module DCN to integrate the new module C3_True, and replace part of C3 module.

D. IMPROVEMENTS TO THE YOLOv5s FRAMEWORK

In the process of recognizing targets on the loader bucket, the motion of the bucket of ten leads to the deformation of the targets, and the existing YOLOv5 network structure cannot accurately identify target information. To address this issue, this paper adopts YOLOv5s as the base model and optimizes the model from aspects such as small object detection heads, attention mechanisms, loss functions, and Depth Separable Convolution. Firstly, an additional small object detection head is added on top of the original detection heads to better capture detailed information, thereby optimizing the detection performance for small targets. Secondly, an EMA attention mechanism is introduced before the SPPF module in the Backbone to effectively solve the interference problem of the background environment on target feature extraction. Then, the Focal-EIOU Loss function is introduced, replacing the GIoU of YOLOv5s as the positioning regression loss function, to speed up the model's convergence. Finally, a Depth Separable Convolution module, DCN, is introduced as a new module, C3_True, replacing some of the C3 modules in the Backbone, to adapt to the deformation of targets and reduce computational load, thereby enhancing the model's operational efficiency. The overall framework of the improved model is shown in Figure 3.

1) IMPROVED NETWORK FOR SMALL OBJECT DETECTION LAYER

In analyzing the dataset of shovel unloading states, the original YOLOv5s model outputs feature maps of different sizes (20×20, 40×40 and 80×80 pixels) through its three detection heads to accommodate targets of various sizes [14]. However, the dimensions of these feature maps are not sufficiently fine for detecting binary targets of loaders, especially when targets undergo deformation due to motion.

To address this issue, an innovative approach is proposed in this paper: adding a small object detection head with a 160×160 pixel size. This new detection head is realized by upsampling the feature map generated from the first C3 module in the backbone network, producing a finer granularity 160×160 pixel feature map. Subsequently, these upsampled feature maps are fused to form the final output of the small object detection head.

The advantage of this method lies in the new detection head's ability to utilize information from shallower layers in the backbone network. Because these shallower layers have smaller receptive fields, they can capture more low-level details, which is particularly important for the recognition of small targets. In this way, the new detection head can not only detect the presence of small objects but also adapt better to deformations caused by object motion or other external factors to some extent.

2) EFFICIENT MULTI-SCALE ATTENTION (EMA) MECHANISM NETWORK

During the unloading process of loaders, excessive tilting of the bucket may cause severe target deformation, making it difficult for the algorithm model to distinguish between interference information and target features in the image, affecting recognition accuracy and leading to missed or false detections. To reduce the interference of background noise on feature extraction [15], as is shown in Figure 4, this study introduces an attention mechanism to optimize the model's weight distribution for image regions, thereby enhancing focus on targets and efficiency in feature extraction. Widely used attention mechanisms include Spatial Transformer Networks (STN) [16], Squeeze-and-Excitation (SE) [17], Convolutional Block Attention Module (CBAM)

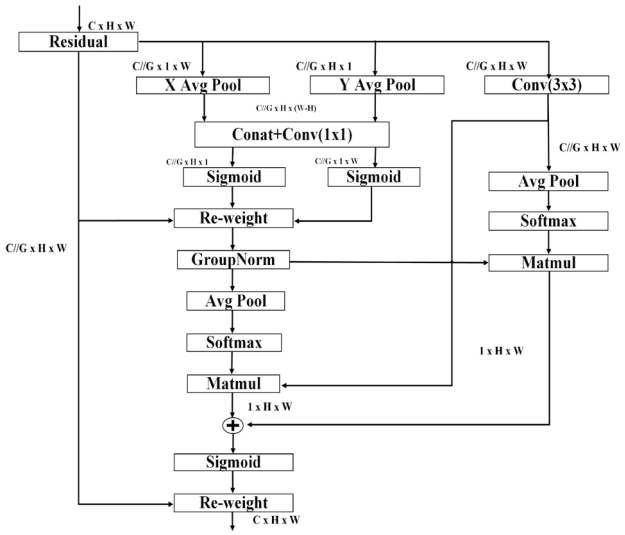

FIGURE 4. EMA attention mechanism.

[18], and Exponential Moving Average (EMA) [19]. Among these, STN and SE modules have made progress in spatial structure adjustment and channel-wise attention enhancement, but STN has limitations in extracting long-distance relationships, and the SE module does not fully address spatial dimension attention. CBAM enhances focus on target areas through a combination of spatial and channel attention, but it incurs a high computational cost. The EMA attention mechanism, by dynamically adjusting weights, enhances focus on key features, especially in processing large-scale datasets and real-time computing scenarios, demonstrating superior performance due to its lower computational complexity.

3) IMPROVED LOSS FUNCTION: FOCAL-EIoU LOSS

The YOLOv5s network employs the CIoU bounding box loss function for confidence and class loss. CIoU is primarily used for gradient descent, but it focuses only on accuracy for the correct label prediction, neglecting the differences among other incorrect labels. This leads to scattered feature learning, causing some discrete losses that cannot be directly differentiated, thus inhibiting the model's regression optimization speed.

To address the issues of slow model convergence and inaccurate regression results, this paper introduces a loss function Focal-EIoU Loss [20]. This function comprehensively considers three key geometric dimensions: overlap area, center point difference, and edge length difference, to accurately assess the discrepancy between predicted and actual boxes. Focal-EIoU Loss aims to enhance the model's precision in bounding box localization, especially in cases of low overlap between the predicted and actual boxes, by increasing the penalty in these situations to promote faster convergence and higher accuracy. Moreover, this loss function integrates the principle of Focal Loss, adjusting the weights of anchor boxes with different overlap degrees, effectively addressing class imbalance and enhancing the detection capability for small targets and complex background targets. This method not only improves the effectiveness of the loss function in bounding box regression description but also significantly enhances the model's robustness and accuracy in handling imbalanced datasets.

4) DEPTH SEPARABLE CONVOLUTION

In the practical loading and unloading process of loaders, challenges such as target deformation and significant differences between the before and after states are encountered. Traditional convolution kernels, typically fixed in size, exhibit poor adaptability to targets of varying sizes, thus diminishing the model's ability to recognize targets. The YOLOv5s model primarily utilizes the C3 (context convolutional module) to extract features from targets. However, due to the introduction of a large number of parameters, the C3 module is prone to overfitting and is limited in its ability to perceive information from distant layers [21]. To address these issues, this paper employs Depth Separable Convolution for improved feature extraction.

Since its potential was highlighted by Sifre in 2013, Depth Separable Convolution has become a key technique for achieving model lightweight in many efficient neural network architectures [22], [23]. This technique comprises two parts: depth convolution and pointwise convolution. Depth convolution performs convolutions independently across each channel of the input, achieving separation of the conventional convolution in the spatial dimension. Pointwise convolution then transforms the feature map resulting from depth convolution into a new feature representation through 1×1 convolutions, effectively integrating channel information. The process is illustrated as follows:

$$S(x, y) = P\left(F_p, D\left(F_d, G\right)\right) \tag{13}$$

$$D(x, y) = \sum_{k,l} F(k, l)G(x - k, y - l) \tag{14}$$

$$P(x, y) = \sum_{m} F_p G(x, y, m) \tag{15}$$

In the formula: $D(x, y)$ is the depth convolution operation, $P(x, y)$ is the pointwise convolution operation, F_p represents a convolution kernel of size 1×1, F_d is a convolution kernel of size $k \times l$, and G is an input matrix of size $u \times n$, with m being the number of channels.

Compared to traditional convolution, when a standard convolution operation with a kernel size of $k \times k$ and depth d is applied to a feature map of size $H \times W \times N$, the amount of computational parameters is: $H \times W \times N \times k \times k \times d$, whereas for Depth Separable Convolution, it is: $H \times W \times N \times (k \times k + d)$. The number of parameters for standard convolution is $k \times k \times d/(k \times k + d)$ times that of Depth Separable Convolution. As shown in Figure 5, the Depth Separable Convolution module DCN is integrated into an innovative module C3_True, which replaces some of the C3 modules in Backbone, aiming to better adapt to the deformations of the

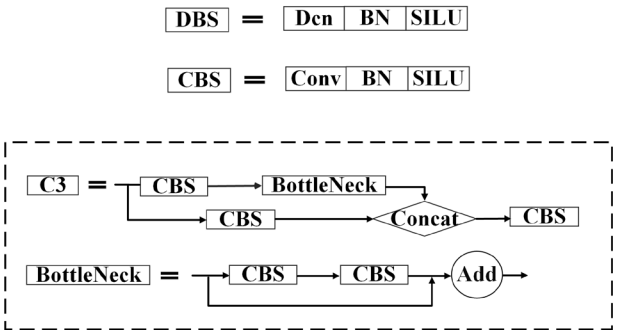

FIGURE 5. Depth Separable Convolution replacing standard convolution in the C3 module.

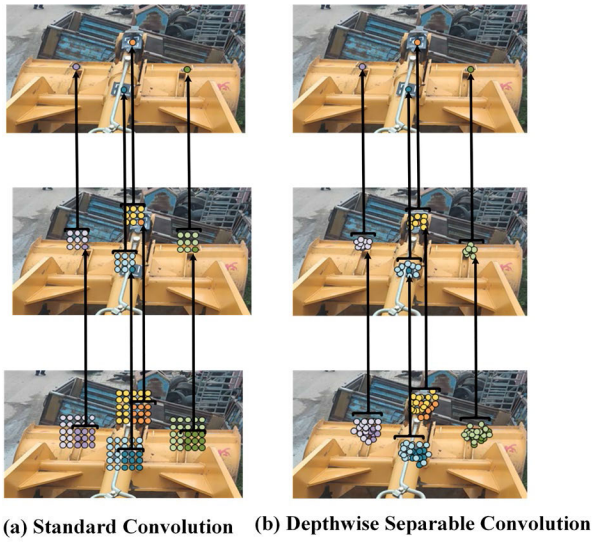

FIGURE 6. Convolution process: (a) Standard convolution, (b) Depth Separable Convolution, where Depth Separable Convolution allows the convolution kernel to dynamically adjust the sampling position on the feature map.

target while significantly reducing the computational effort, and thus significantly improving the operational efficiency and performance of the model.

Depth Separable convolution optimizes the traditional convolution operation by introducing learnable offsets, allowing the convolution kernels to dynamically adjust their sampling positions on the feature map. This mechanism enables the kernels to adapt according to the feature extraction needs at different positions. During the feature extraction process, offsets are first calculated using convolution operations and then combined with the original input feature map and these offsets to generate the final output feature map. The extraction process, as shown in Figure 6, effectively enhances the model's ability to recognize deformed or variably scaled targets.

E. EXPERIMENTAL ENVIRONMENT AND DATASET

The experiment was conducted using the Ubuntu 20.04 operating system, with an Nvidia GeForce RTX 4090 GPU

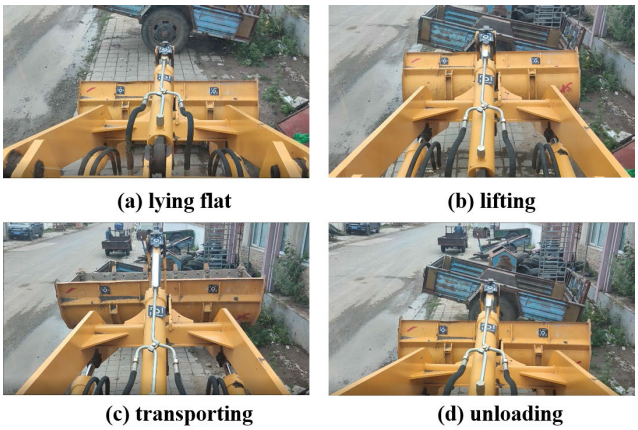

FIGURE 7. Sample experimental data. (a) lying flat, (b) lifting, (c) transporting, and (d) unloading.

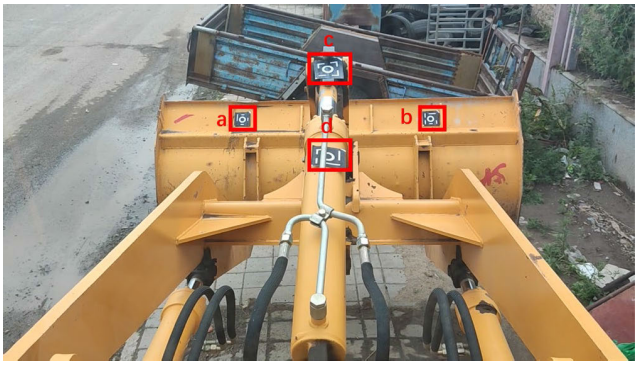

FIGURE 8. Targets a-d. Targets a-d are designated as follows: starting from the far-left front end of the bucket, and moving rightward, we identify target-a and target-b. Moving from the bucket to the boom, the designations continue as target-c and target-d.

featuring 24GB of VRAM, CUDA 12.1, the deep learning framework PyTorch 2.2.0, and the programming language environment Python 3.9.

The model training dataset consists of 2008 images, divided into training, validation, and test sets in an 8:1:1 ratio for experimentation. A portion of the dataset images is illustrated in Figure 7, with data collected from four scenarios: (a) lying flat, (b) lifting, (c) transporting, and (d) unloading. Figure 8 labels target a to d. Specifically, starting from the far-left front end of the excavator bucket and moving rightward, we can identify targets a and b. Then, moving from the bucket along the boom, targets c and d can be distinguished. Such labeling helps us identify and differentiate between the positions of the targets, facilitating further analysis and discussion.

F. EVALUATION METRICS

In this paper, Recall, mean Average Precision (mAP), parameters, and GFLOPs are used as the evaluation indexes. mAP is the sum of the AP values of each target category, and then averaged according to the number of categories, where AP is the area formed by the P-R curve with Precision (P)

TABLE 1. Results of incorporating different attention mechanisms.

Model	AP of Target a	AP of Target b	AP of Target c	AP of Target d	mAP/%
YOLOv5s	96.3	95.6	97.3	96.2	96.3
+STN	97.5	97.3	98.3	96.5	97.4
+CBAM	97.3	96.4	97.5	97.2	97.1
+SE	96.4	98.2	97.2	96.3	97.0
+EMA	95.8	99.5	99.5	99.5	99.3

TABLE 2. Model performance comparison.

Model	Recall	mAP/%	parameter(mb)	GFLOPs
YOLO v5s	96.3	96.3	7.24	16.5
YOLO v7-tiny	96.7	96.1	6.23	13.8
YOLO v8	96.5	96.5	11.2	28.6
Ours	98.6	98.3	10.7	58.5

and Recall (R) as the vertical and horizontal coordinates, respectively. AP is the area enclosed by the P-R curve with Precision (P) and Recall (R) as vertical and horizontal coordinates, respectively. Precision, Recall, AP, and mAP values are respectively:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{16}$$

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i \tag{17}$$

where TP is the number of correctly classified positive samples, FP is the number of incorrectly classified negative samples, FN is the number of incorrectly classified positive samples, and N is the total number of detections classified. Where AP is $AP = \int_0^1 p(r)dr$, p(r) is Precision $= \frac{TP}{TP+FP}$.

III. RESULTS AND DISCUSSIONS

We conducted experiments from various perspectives, including the model's attention mechanism, comparisons with other models, ablation studies of the model, and comparisons with contact sensors.

A. ATTENTION MECHANISM EXPERIMENT

To verify the superiority of the combination of YOLOv5s and EMA attentional mechanisms in the loader work scenario.

Different attentional mechanisms of STN, CBAM, and SE were selected for the comparison experiments to compare the AP values of each target category.

As shown in Table 1, the introduction of STN and CBAM attention mechanisms improved the model's mAP by 0.9% and 0.8%, respectively, although these improvements are present, they are not significant. The addition of the SE attention mechanism only increased the mAP by 0.7%. In contrast, after integrating the EMA attention mechanism, the model achieved significant improvements in detection performance across all types of targets, with a mAP increase of 3.0% over the original model, clearly surpassing the gains from the other three attention mechanisms.

B. COMPARATIVE EXPERIMENT

To prove the superiority and effectiveness of the improved algorithmic model of this paper, comparative experiments are conducted. The model was compared with the YOLO series of algorithms.

From Table 2, it is clear that the algorithm proposed in this study surpasses mainstream models such as YOLOv5s, YOLOv7-tiny, and YOLOv8 in terms of accuracy, parameter size, and computational speed. Recall and mAP serve as indicators of the model's accuracy, parameter size gauges the model complexity, and GFLOPs represent an increase in computational speed. In terms of accuracy, the improved model shows a 2.3% increase in Recall and a 2.0% improvement in mAP compared to the original YOLOv5s model, indicating a significant enhancement in the model's detection precision. Although this study's model has more parameters than YOLOv5s and YOLOv7-tiny [24], it has fewer than YOLOv8 [25], yet its GFLOPs are considerably higher than those of the other models, suggesting the improved model has enhanced operational speed and robustness in loader angle recognition. Regarding computational speed, the proposed model demonstrates substantial improvements over current mainstream models, primarily due to the deployment of Depth separable convolutions. This approach significantly reduces the model's parameter count and computational demands, making the model lighter and more suitable for resource-constrained devices. With reduced parameter numbers, the model's training and inference speeds are enhanced. The GFLOPs reached 58.5, marking an increase of 42 GFLOPs, with a growth rate of 254.55%.

C. ABLATION EXPERIMENT

To verify the effectiveness of each improvement strategy proposed in this paper, ablation experiments are conducted on the benchmark model YOLOv5s, and the detection performance of small targets is tested on the validation set, and the experimental results are shown in Table 3. The model in this paper has the following four main improvement parts: small target detection head, EMA attention mechanism, Focal-EIOU Loss function, and Depth Separable Convolution instead of ordinary convolution of C3 module in Backbone.

Table 3 further shows that the first addition of the small object detection head increased the mAP by 0.6%. Then, the addition of the EMA attention mechanism increased the mAP by 1.2% compared to the original model. After

TABLE 3. Ablation experiment.

Model	AP of Target a	AP of Target b	AP of Target c	AP of Target d	mAP/%
YOLOv5s	96.3	95.6	97.3	96.2	96.3
YOLOv5s + Small Object Detection Head	96.6	95.0	97.7	98.5	96.9
YOLOv5s + Small Object Detection Head+EMA	95.8	99.3	97.5	97.5	97.5
YOLOv5s + Small Object Detection Head+EMA+ Focal-EIOU Loss	97.8	99.0	98.7	96.9	98.1
YOLOv5s + Small Object Detection Head+EMA+ Focal-EIOU Loss +Depth Separable Convolution	98.6	99.5	99.5	99.5	99.3

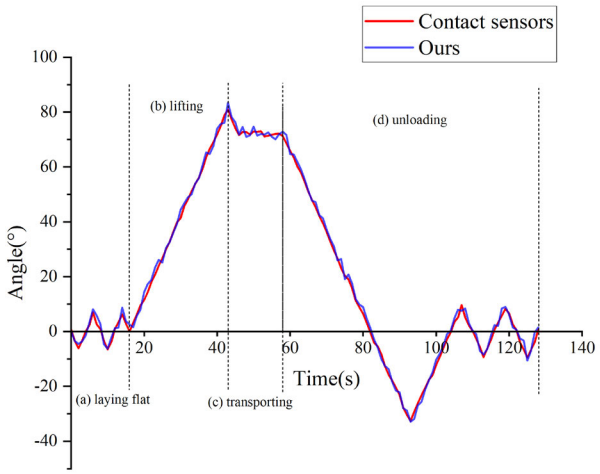

FIGURE 9. Real-time testing of model with contact sensors.

changing the loss function to Focal-EIOU Loss, the mAP increased by 1.8%. The inclusion of Depth separable convolution, due to its integration into the C3 module, increased the mAP by 3.0%, effectively addressing the challenge of target deformation and solving the problem of multiple target types and large deformation differences in loaders [26]. The research results support our initial hypothesis: the integration of advanced attention mechanisms and Depth separable convolution significantly improved the detection performance of small objects. The introduction of the EMA attention mechanism, Focal-EIOU loss function, and Depth Separable Convolution collectively enhanced the model's ability to recognize small objects. This demonstrates the model's adaptability and robustness in facing diverse targets in complex scenes, highlighting its potential value in practical applications.

As shown in Figure 9, during a test period of 128 seconds, we conducted experiments using contact sensors to validate the effectiveness of our model algorithm. In this process, we carried out extensive testing on the four different stages mentioned in Figure 7—laying flat, lifting, transporting, and unloading. Comparing the data with the results from the contact sensors, we found a very high level of consistency between the two. This finding not only confirms the accuracy of our model but also demonstrates its reliability and effectiveness in practical applications.

IV. CONCLUSION

This article, grounded in the YOLOv5s model, has significantly advanced the state-of-the-art in loader shovel angle recognition, demonstrating remarkable enhancements in both accuracy and computational efficiency. The tailored small object detection head has markedly heightened the model's sensitivity to subtle angle variations, while the incorporation of the EMA attention mechanism has optimized recognition rates and refined bounding box accuracy. Furthermore, the adoption of the Focal-EIoU Loss function has expedited model convergence, and the integration of DCN has bolstered the model's capability to effectively handle deformations. These cumulative enhancements have yielded a substantial 3.0% accuracy boost and a notable improvement in computational efficiency, evidenced by a substantial reduction in computational load by 254.55% in terms of GFLOPs.

Despite promising results, high computational demands hinder edge deployment. Future work should focus on model compression and pruning for wider applicability. Additionally, investigating robustness under varying conditions and integrating non-contact technologies could further refine precision.

The advancements in loader shovel angle recognition capabilities have profound practical ramifications, enhancing operational precision and safety in construction machinery. As the industry embarks on a journey towards heightened automation and intelligence, the integration of these improvements holds the potential to transform the sector, propelling it towards greater efficiency and sustainability. This study represents a pivotal step towards realizing a future where construction equipment operates with unparalleled precision and efficiency, thereby laying a solid foundation for smarter, safer, and more environmentally friendly construction practices.

REFERENCES

[1] X. Y. Wang, Y. X. Hao, and L. Quan, "Research on the lifting system characteristics of electro-hydraulic dual-source hybrid drive loaders," *J. Mech. Eng.*, vol. 59, no. 7, pp. 41–51, 2023.

[2] T. H. Zhao, M. S. Wang, and W. G. Pan, "Design of intelligent assistance construction system for excavators," *J. Shandong Univ. Eng. Sci.*, vol. 53, no. 4, pp. 163–172, 2023.

[3] B. Cao, X. Liu, W. Chen, K. Yang, and P. Tan, "Skid-proof operation of wheel loader based on model prediction and electro-hydraulic proportional control technology," *IEEE Access*, vol. 8, pp. 81–92, 2020, doi: 10.1109/ACCESS.2019.2961364.
[4] B. Hong and X. Ma, "Path optimization for a wheel loader considering construction site terrain," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 2098–2103, doi: 10.1109/IVS.2018.8500447.
[5] B. Liu, J. Huang, L. W. Guo, and L. Liu, "Research on automatic digging control and algorithm of the working device of skid steer loader," *Appl. Mech. Mater.*, vols. 66–68, pp. 2046–2051, Jul. 2011, doi: 10.4028/www.scientific.net/amm.66-68.2046.
[6] C. Y. Huang, "Research on bucket translational control of loaders based on sensor networks," Ph.D. thesis, Soochow Univ., 2021.
[7] C. R. Qiu, S. D. Wang, and P. P. Ma, "An auxiliary operation control system for loaders," *Construct. Machinery Technol. Manage.*, vol. 35, no. 3, pp. 50–51, 2022.
[8] D. Sun, C. Ji, S. Jang, S. Lee, J. No, C. Han, J. Han, and M. Kang, "Analysis of the position recognition of the bucket tip according to the motion measurement method of excavator boom, stick and bucket," *Sensors*, vol. 20, no. 10, p. 2881, May 2020, doi: 10.3390/s20102881.
[9] I. Aldekoa, A. del Olmo, L. Sastoque-Pinilla, S. Sendino-Mouliet, U. Lopez-Novoa, and L. N. L. de Lacalle, "Early detection of tool wear in electromechanical broaching machines by monitoring main stroke servomotors," *Mech. Syst. Signal Process.*, vol. 204, Mar. 2023, Art. no. 110773, doi: 10.1016/j.ymssp.2022.110773.
[10] J. Zhao, Y. Hu, and M. Tian, "Pose estimation of excavator manipulator based on monocular vision marker system," *Sensors*, vol. 21, no. 13, p. 4478, Jun. 2021, doi: 10.3390/s21134478.
[11] B. Cao, X. Liu, W. Chen, H. Li, and X. Wang, "Intelligentization of wheel loader shoveling system based on multi-source data acquisition," *Autom. Construct.*, vol. 147, Mar. 2023, Art. no. 104733.
[12] H. B. Wang, H. L. Zou, and R. Z. Zhang, "Measurement system for excavator working device posture based on vision measurement," *J. Agricult. Machinery*, vol. 46, no. 4, pp. 302–308, 2015.
[13] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.
[14] P. F. Wang, H. M. Huang, and M. Q. Wang, "Improved YOLOv5 algorithm for complex road object detection," *Comput. Eng. Appl.*, vol. 2022, vol. 58, no. 17, pp. 81–92, 2022.
[15] T. C. Gu, W. B. Xu, and B. Li, "Optimization algorithm for loader material fine-grained detection based on YOLOv5," *Computer Integr. Manuf. Syst.*, vol. 30, no. 1, p. 239, 2023, doi: 10.13196/j.cims.2023.09.005.
[16] C. Xu, Y. Makihara, X. Li, Y. Yagi, and J. Lu, "Cross-view gait recognition using pairwise spatial transformer networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 1, pp. 260–274, Jan. 2021.
[17] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," 2017, *arXiv:1709.01507*.
[18] S. Woo, J. Park, J.-Y. Lee, and I. So Kweon, "CBAM: Convolutional block attention module," 2018, *arXiv:1807.06521*.
[19] D. Ouyang, S. He, G. Zhang, M. Luo, H. Guo, J. Zhan, and Z. Huang, "Efficient multi-scale attention module with cross-spatial learning," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2023, pp. 1–5.
[20] Y.-F. Zhang, W. Ren, Z. Zhang, Z. Jia, L. Wang, and T. Tan, "Focal and efficient IOU loss for accurate bounding box regression," *Neurocomputing*, vol. 506, pp. 146–157, Sep. 2022.
[21] F. R. Kou, W. Xiao, and H. Y. He, "Study on undermine target detection based on improved YOLOv5," *J. Electron. Inf. Technol.*, vol. 45, no. 7, pp. 2642–2649, 2023.
[22] C. L. Wang, Q. Zhao, and Y. Zhao, "Real-time remote sensing object detection algorithm based on depth separable convolution," *Electron. Control*, vol. 29, no. 8, pp. 45–49, 2022.
[23] C. H. Li and Y. Lu, "Facial expression recognition based on depth separable convolution," *Comput. Eng. Des.*, vol. 42, no. 5, pp. 1448–1454, 2021, doi: 10.16208/j.issn1000-7024.2021.05.034.
[24] V. A. Kulyukin and A. V. Kulyukin, "Accuracy vs. energy: An assessment of bee object inference in videos from on-hive video loggers with YOLOv3, YOLOv4-tiny, and YOLOv7-tiny," *Sensors*, vol. 23, no. 15, p. 6791, Jul. 2023, doi: 10.3390/s23156791.
[25] F. M. Talaat and H. ZainEldin, "An improved fire detection approach based on YOLO-v8 for smart cities," *Neural Comput. Appl.*, vol. 35, no. 28, pp. 20939–20954, 2023.
[26] M. González, A. Rodríguez, U. López-Saratxaga, O. Pereira, and L. N. L. De Lacalle, "Adaptive edge finishing process on distorted features through robot-assisted computer vision," *J. Manuf. Syst.*, vol. 74, pp. 41–54, Jun. 2024, doi: 10.1016/j.jmsy.2024.02.014.

XUEDONG ZHANG is currently pursuing the degree in electronic information engineering with the North China University of Technology. His main research interests include machine vision, image processing, and intelligent signal processing.

BO CUI is currently pursuing the Ph.D. degree. He is also an Associate Professor and a Master's Supervisor. His main research interests include signal and information intelligence processing, simulation methods, and algorithm applications.

ZHAOXU WANG is currently pursuing the degree in communication engineering with the North China University of Technology. His main research interests include signal and information intelligent processing and image processing.

WANGTING ZENG is currently pursuing the degree in computer science and technology with the North China University of Science and Technology. Her main research interests include image processing and intelligent signal processing.