Meter-YOLOv8n: A Lightweight and Efficient Algorithm for Word-Wheel Water Meter Reading Recognition

Shichao Qiao, Yuying Yuan*, Ruijie Qi
School of Computer Science and Technology, Shandong University of Technology, Zibo 255000, Shandong, China

***Abstract*—To address the issues of low efficiency and large parameters in the current word-wheel water meter reading recognition algorithms, this paper proposes a Meter-YOLOv8n algorithm based on YOLOv8n. Firstly, the C2f component of YOLOv8n is improved by introducing an enhanced inverted residual mobile block (iRMB). It enables the model to efficiently capture global features and fully extract the key information of the water meter characters. Secondly, the Slim-Neck feature fusion structure is employed in the neck network. By replacing the original convolutional kernels with GSConv, the model's ability to express the features of small object characters is enhanced, and the number of parameters in the model is reduced. Finally, Inner-EIoU is used to optimize the bounding box loss function. This simplifies the calculation process of the loss function and improves the model's ability to locate dense bounding boxes. The experimental results show that, compared with the original model, the precision, recall, mAP@0.5, and mAP@0.5:0.95 of the improved model have increased by 1.7%, 1.2%, 3.4%, and 3.3% respectively. Meanwhile, the parameters, FLOPs, and model size have decreased by 0.56M, 2.6G, and 0.7MB respectively. The improved model can better balance the relationship between detection performance and computational complexity. It is suitable for the task of recognizing word-wheel water meter readings and has practical application value.**

***Keywords*—Word-wheel water meter; YOLOv8n; global features; slim-neck; loss function**

I. Introduction

The word-wheel water meter is a common flow measurement instrument widely used in tap water metering, playing a crucial role in helping water utility companies monitor users' water consumption. Due to its simple structure, low cost, and strong anti-interference capability, it is widely used in residential communities and industrial workshops. Traditionally, meter reading is performed manually by staff who visually observe the meter readings and record them by hand. This method is labor-intensive and highly repetitive, consuming significant human and material resources [1], [2]. Additionally, it is susceptible to environmental factors and psychological variations of the staff, leading to errors such as misreadings and omissions. In recent years, object detection algorithms have continuously advanced, especially deep learning-based methods, which have significantly improved accuracy. The use of computer vision technology for fast and accurate reading of word-wheel water meters has become a research hotspot [3].

The methods for recognizing the readings of word-wheel water meters can be divided into two types. One is the optical character recognition (OCR) method. This method first locates the reading area of the water meter, then segments the reading area using image segmentation technology, and finally employs a convolutional neural network to recognize the readings in the segmented area. However, this method has high requirements for image quality and a slow processing speed, making it difficult to meet the reading recognition needs of character wheel water meters in complex scenarios. Another method is to directly perform character detection on the water meter image to read the water meter reading. It omits the intermediate positioning and segmentation steps, making the overall recognition process more concise and efficient. When dealing with a large number of water meter images, direct detection can save more time and computational resources and improve the reading efficiency. Considering the two methods comprehensively, in order to enable the model to complete rapid detection in an environment with limited resources, this paper uses an object detection algorithm to directly recognize the readings of the word-wheel water meters [4].

Object detection algorithms can be categorized into single-stage and two-stage algorithms. Common two-stage algorithms include R-CNN, Fast R-CNN [5], and Faster R-CNN [6]. In two-stage algorithms, a large number of candidate boxes are first generated, followed by object classification and bounding box regression. Since classification and regression need to be performed on numerous candidate regions, the training and inference speed of these algorithms is relatively slow. In practical applications, especially in scenarios requiring real-time detection, this slower detection speed can become a limiting factor.

Single-stage detection algorithms include SSD [7] and the YOLO series [8]. These algorithms perform dense predictions on the input image through a single network, achieving both object detection and localization in one step. Compared to two-stage algorithms, single-stage algorithms require lower computational costs and offer better real-time performance. The YOLO algorithm features an integrated design, enabling it to process images at dozens of frames per second or even higher speeds. This allows for fast localization of water meters and reading recognition, making it particularly suitable for applications with high-speed processing requirements.

In order to achieve accurate and rapid reading of the readings of word-wheel water meters, this paper proposes a

lightweight word-wheel water meter reading recognition algorithm based on YOLOv8n. The main contributions of this paper are as follows:

- Based on the publicly available water meter dataset, a word-wheel water meter dataset in real-world scenarios has been constructed. It includes complete characters and transitional characters, totaling twenty categories. This dataset takes into account various unfavorable factors in the real environment, such as light and noise, providing a data foundation for future research.
- To address the issues of low accuracy and a large number of parameters in the reading recognition of word-wheel water meters, this paper proposes the Meter-YOLOv8n algorithm for word-wheel water meter reading recognition. The detection accuracy and efficiency are improved through the use of C2f-iRMBS, Slim-Neck, and Inner-EIoU.
- Comprehensive experiments have been carried out on the word-wheel water meter dataset, and a visual analysis of the test results has been conducted. It proves the detection performance and generalization ability of the improved algorithm, which is more suitable for the task of reading recognition of word-wheel water meters.

This paper is structured as follows: Section II introduces the related work in the field of word-wheel water meter reading. Section III describes the improved algorithm for reading of word-wheel water meters. Section IV introduces the dataset of word-wheel water meters, the evaluation metrics, and the experimental environment. Section V describes the experimental results and visual analysis. Section VI analyzes the deficiencies of the existing work and the directions for future exploration.

II. Related Work

A. YOLOv8 Algorithm

As one of the most renowned algorithms in the YOLO series, YOLOv8 is particularly suitable for application scenarios that require fast real-time object detection, such as autonomous driving [9], video surveillance [10], and drone monitoring [11]. YOLOv8 inherits the structural concept of YOLOv5 [12] and is mainly composed of three parts: the backbone network, the neck network, and the head network.

YOLOv8 preprocesses images using mosaic augmentation, adaptive anchor box calculation, and adaptive grayscale padding. It employs an anchor-free approach to directly predict object centers, thereby improving the speed of Non-Maximum Suppression (NMS) [13]. The backbone is responsible for extracting image features and primarily consists of modules such as Conv, C2f, and SPPF. The neck section integrates and transmits features using a Path Aggregation Network (PAN) structure [14]. The head is responsible for object detection and classification tasks, including a detection head and a classification head. Loss computation is divided into two parts: classification loss and regression loss. The classification loss is trained using Binary Cross-Entropy Loss (BCE Loss), while the regression loss combines Distribution Focal Loss (DF Loss) and CIoU Loss.

YOLOv8 is divided into five different sizes, namely n, s, m, l, and x, according to the depth and width of the network. Taking into comprehensive consideration the size and complexity of the algorithm, this paper selects YOLOv8n as the baseline algorithm for improvement.

B. Reading Recognition of Word-Wheel Water Meters

To improve the reading recognition accuracy of word-wheel water meters and enhance the representation of transitional characters, Cai et al. [15], proposed an efficient automatic meter localization and recognition method. First, they performed a coarse-to-fine detection of the entire meter to locate the reading region. Then, they used a projection-based method to segment the reading area, and finally, a BP neural network was employed to recognize the segmented meter region. Jawas et al. [16], localized the meter using contour information, segmented the reading region, and then applied OCR technology to recognize the meter characters. Chen et al. [17], used an improved U-Net network to locate the dial's reading region in large-scale water meter images. They then segmented individual characters based on the structural features of the dial and finally performed reading recognition using an improved VGG16 network. Men et al. [18], proposed a water meter reading recognition region segmentation method based on an improved U~2-Net. They designed a modified Double-RSU module based on the RSU module, which increases the depth and complexity of the network, thereby enhancing the model's generalization ability and robustness.

To improve recognition efficiency and achieve rapid acquisition of water consumption data, Li et al. [19], proposed a novel lightweight concatenated convolutional network. This network replaces a certain number of standard 3×3 convolution operations with 1×1 convolutions, resulting in a more efficient and lightweight network with better overall performance. Zou et al. [20], utilized a geometric method to perform rotational correction on water meter images and then employed the WDPDet network to identify the reading region of the water meter. This network is capable of handling complex and variable scenes. Zhang et al. [21], applied homography transformation to geometrically correct the deformed reading region. In the transformation stage, new recognition markers and probability vectors were added between each digit to address the issue of digit rotation. Li et al. [22], adopted an improved YOLOv4 object detection algorithm for reading recognition, expanding the receptive field and reducing the loss of original information by introducing a focus structure. Additionally, they enhanced the network's ability to fuse multi-scale features and improved the representation of transitional characters by constructing cross-stage partial connection modules. Wang et al. [23], proposed the GMS-YOLO algorithm for water meter reading recognition, replacing standard convolutions in the C2f module with Grouped Multi-Scale Convolution (GMSC) to enable the model to acquire receptive fields at different scales, thereby enhancing its feature extraction capability. Moreover, they integrated the Large Separable Kernel Attention (LSKA) mechanism into the SPPF module to improve the perception of small-scale features. Finally, the SIoU bounding box loss function was used instead of CIoU, strengthening the model's object localization ability and accelerating convergence speed.

Although the above-mentioned methods have made significant progress in the research of word-wheel water meter reading recognition, they still face the following challenges.

Due to environmental factors and changes in shooting angles, the proportion of effective characters in water meter images is relatively small, and image distortion may occur. This leads to a decrease in recognition accuracy, and sometimes false detections and missed detections may also occur. In order to improve the detection ability of small object characters and distorted characters, we have integrated shift-wise conv based on the inverted residual mobile block (iRMB) [24] and designed a lightweight and efficient C2f-iRMBS module. It can extract the feature information of water meter characters more efficiently.

Deep learning algorithms rely on stacked multi-layer convolutional networks, which have complex structures and require a large amount of computational resources. To reduce the size of the model, we have introduced the Slim-Neck [25] structure into the neck network. The GSConv module and the VoVGSCSP module can simplify the network structure, reduce the computational complexity of the network, and improve the reading efficiency of word-wheel water meters.

In order to solve the problem of overlapping detection boxes of water meter characters, accelerate the training efficiency of the model and improve the convergence speed of the model, the Inner-EIoU loss function is used to replace the complete intersection over union loss (CIoU loss). This enhances the model's ability to locate dense characters.

III. Proposed Algorithm

A. Meter-YOLOv8n

Aiming at the problems of low accuracy, false detections, and missed detections that occur in the practical application of deep learning algorithms, this paper proposes a detection algorithm named Meter-YOLOv8n, which is specifically designed for the task of identifying the readings of word-wheel water meters.

The design objectives of the Meter-YOLOv8n algorithm are twofold: firstly, in real-world scenarios, it aims to improve the detection accuracy of both the complete characters and transitional characters on water meters. Secondly, it seeks to reduce the computational complexity of the algorithm so that it can be deployed in more environments. The network structure of Meter-YOLOv8n is shown in Fig. 1.

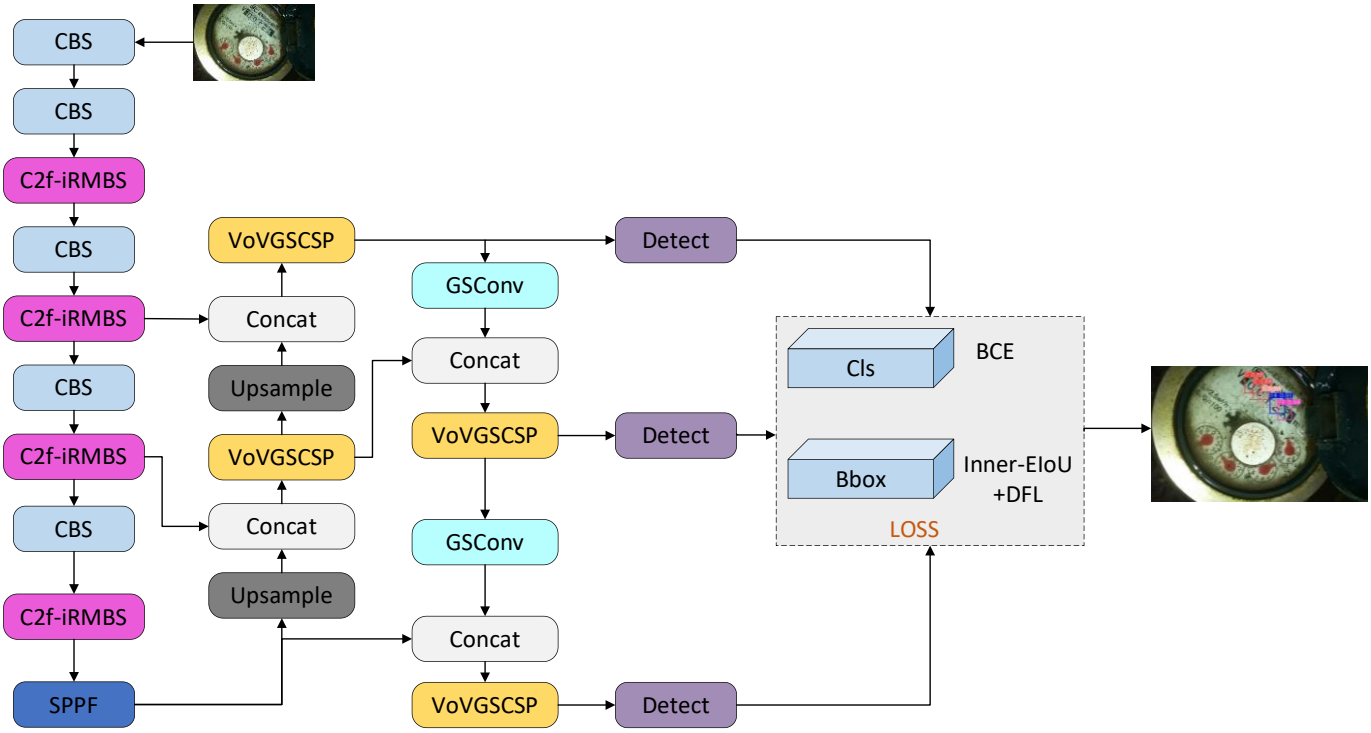


Fig. 1. Network structure of Meter-YOLOv8n.

B. C2f-iRMBS

1) iRMB: In the task of object detection, the attention mechanism can help the algorithm improve its efficiency and accuracy when dealing with complex data. In the images of word-wheel water meters, the background occupies a large proportion, and there are relatively many small object characters that need to be detected. iRMB is a lightweight attention mechanism designed for small object tasks, taking into account the advantages of both dynamic global modeling and static local information fusion. iRMB can better capture the feature information of the objects to adapt to objects of different scales. Moreover, it can effectively increase the receptive field of the model and enhance the model's ability for downstream tasks. The structure of iRMB is shown in Fig. 2.

Firstly, the combined multi-layer perception (CMLP) is used to generate the attention matrices Q and K. A dilated

convolution is employed to generate the attention matrix *V*. Then, the window self-attention mechanism is applied to *Q* and *K* for long-range interaction. Immediately afterwards, a depthwise separable convolution (DWConv) is utilized to model the local features. Finally, a compression convolution is used to restore the number of channels, which is then added to the input to obtain the final result.

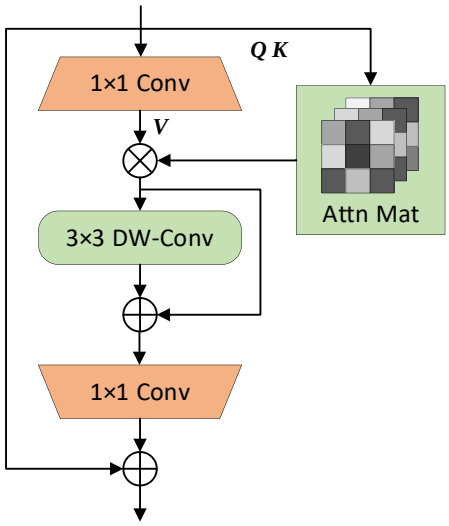
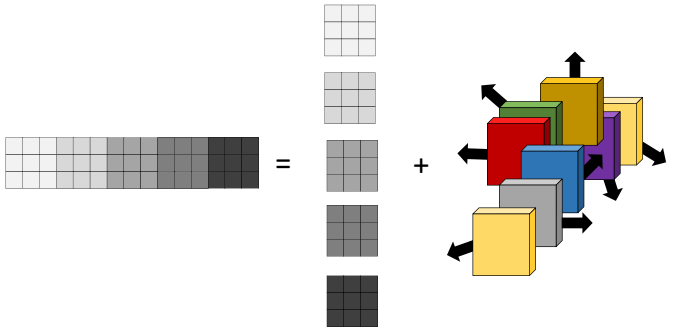
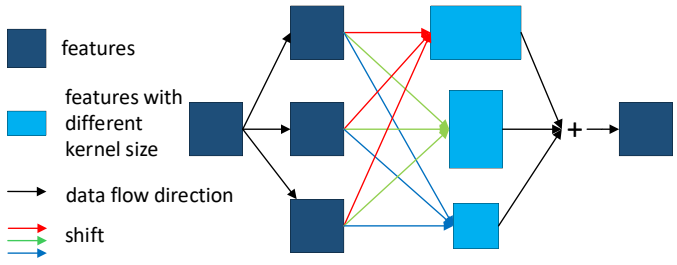


Fig. 2. Structure of iRMB.

2) Shift-wise conv: The process of the shift-wise operation is shown in Fig. 3.



(a) The equivalent operation of large-kernel convolution



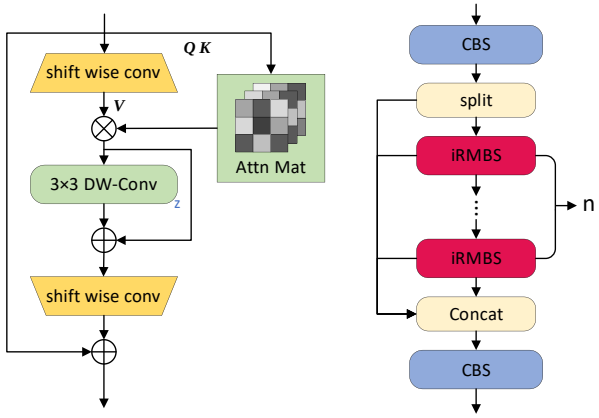
(b) The overall structure of the module

Fig. 3. Schematic diagram of shift-wise conv.

Large-kernel convolution can efficiently capture features in a larger scope, which helps understand the global structure and relationships in the input data. However, the large convolution kernel of large-kernel convolution results in poor ability of the algorithm to process detailed information. Moreover, large-kernel convolution needs to process a larger input data range, so the computational complexity increases, leading to longer training and inference times. To address the above issues, this paper adopts a shift-wise operation. By means of the sparsity mechanism, it ensures that the convolutional neural network (CNN) can capture both long-and short-range dependencies, enabling small convolution kernels to capture global features more efficiently.

3) C2f-iRMBS: Word-wheel water meters are often installed underground or in remote corners, and they are affected by unfavorable factors such as dust, light, and water mist. When reading the word-wheel water meters in real-world scenarios, there are problems such as image deformation and the easy loss of object information. Based on iRMB, this paper integrates shift-wise conv to construct a new module, iRMBS, and combines it with C2f to design the novel C2f-iRMBS module, as shown in Fig. 4. In the iRMBS module, shift-wise conv is used to optimize the ordinary convolution, enabling the model to capture global features more efficiently, better learn information such as the scale of the object and the background, and reduce the number of parameters while improving the detection performance.

The improved C2f-iRMBS module is used to replace the C2f in the backbone network, so that the network can correctly capture the key information of the features such as the small size and occlusion of the water meter characters even in complex and changeable real-world scenarios, and it has stronger robustness and generalization ability.



(a) iRMBS (b) C2f-iRMBS

Fig. 4. Structure of C2f-iRMBS.

C. Slim-Neck Structure

YOLOv8n employs FPN and PAN structures for feature fusion. However, generating multi-scale feature maps using FPN and PAN requires multiple convolution and upsampling operations, which increases computational cost and demands more memory to store these feature maps, ultimately slowing down inference speed. Additionally, FPN and PAN structures can lead to incomplete transmission of feature information across different levels, particularly causing information loss or

blurring during cross-layer information aggregation, which negatively impacts the detection performance for small objects. To reduce model complexity while maintaining accuracy, this paper introduces the Slim-Neck structure in the neck network, replacing the Conv and C2f modules with GSConv and VoVGSCSP modules.

1) GSConv: In the neck layer, the GSConv is used. At this stage, the channel dimension C has reached its maximum value, while the height H and width W have reached their minimum values. As a result, there is minimal redundant information, and no compression is needed. The structural diagram of the GSConv is shown in Fig. 5.

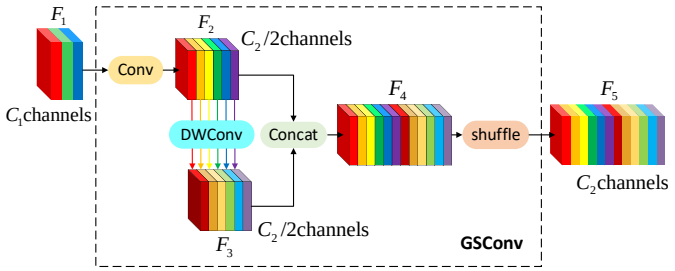


Fig. 5. Structure of GSConv.

In the GSConv, the input feature map F_1 undergoes downsampling through a 3×3 convolution to obtain the feature map F_2. Then, F_2 passes through a depthwise convolution (DWConv) to produce the feature map F_3. Next, F_2 and F_3 are concatenated along the channel dimension to form a new feature map F_4. Finally, a Shuffle operation is performed, returning an output F_5 with reordered channels. The computation formula for GSConv is shown in Eq. (1).

$$F_{GCS} = Shuffle(Cat(\delta(F_1)_{C_2/2,}\varepsilon(\delta(F_1)_{C_2/2}))_{C_2} \quad (1)$$

In Eq. (1), F_1 represents the input feature map with C_1 channels, δ denotes the convolution operation, ε represents the depth wise convolution operation, and F_{GCS} represents the output feature map with C_2 channels obtained through the GSConv.

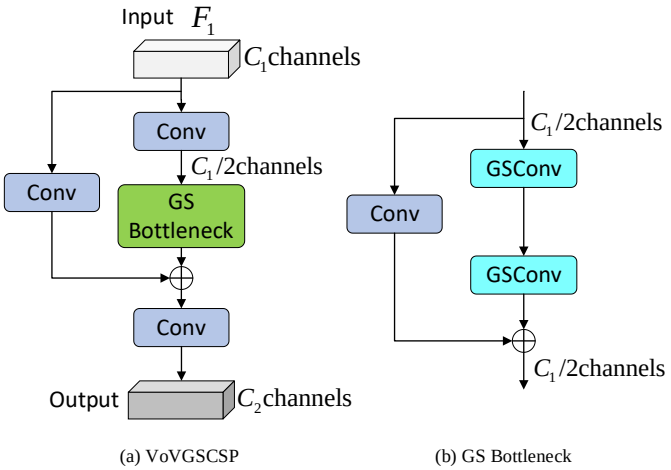
2) VoVGSCSP: VoVGSCSP utilizes grouped spatial context supervision to better capture character information at different scales in water meter images, thereby improving detection accuracy. The structure of VoVGSCSP is shown in Fig. 6.

Firstly, a 1×1 conv is applied to the input feature map with a channel size of C_1 for feature extraction, reducing the channel dimension to half of the original input. The resulting feature map is then fed into the GS Bottleneck. The GS Bottleneck follows the residual network concept, where the input feature map undergoes two GSConv operations. The output is then concatenated with the feature map obtained through a 1×1 convolution, producing the module's output. Finally, VoVGSCSP concatenates the branch output with the GS Bottleneck output and applies a 1×1 convolution to obtain a feature map with a channel dimension of C_2. The computation formula for VoVGSCSP are shown in Eq. (2) and Eq. (3).

$$GSB_{out} = F_{GSC}(F_{GSC}(\alpha(F_1)_{C_2})) + \alpha(F_1)_{C_1/2} \quad (2)$$

$$VoVGSCSP_{out} = \alpha(Concat(GSB_{out}, \alpha(F_1))) \quad (3)$$

In Eq. (2) and Eq. (3), F_1 represents the input feature map with C_1 channels, α represents the convolution operation, GSB_{out} denotes the output of the GS Bottleneck module, and $VoVGSCSP_{out}$ represents the final output of this module.



(a) VoVGSCSP (b) GS Bottleneck

Fig. 6. Structure of VoVGSCSP.

D. Inner-EIoU

In the task of recognizing readings from word-wheel water meters, accurately locating water meter characters is essential, including predicting the coordinates of bounding boxes and the positions of center points. Choosing an appropriate bounding box loss function ensures that the model can precisely predict character locations, thereby improving detection accuracy and localization precision.

The YOLOv8n model uses CIoU as the bounding box regression loss function. However, CIoU involves numerous parameters and has a high computational cost, making it less effective in precisely localizing highly overlapping detection boxes. To enhance training efficiency and convergence accuracy, this paper adopts Inner-EIoU as the bounding box regression loss function. Inner-EIoU introduces a scale factor ratio, which controls the size of an auxiliary bounding box for loss computation. The auxiliary bounding box is shown in Fig. 7.

b^{gt} and b represent the ground truth box and the predicted box, respectively. (x_c^{gt}, y_c^{gt}) denotes the centroid coordinates of the ground truth box, (x_c, y_c) denotes the centroid coordinates of the predicted box. w^{gt} and h^{gt} represent the width and height of the ground truth box, respectively, while w and h represent the width and height of the predicted box, respectively.

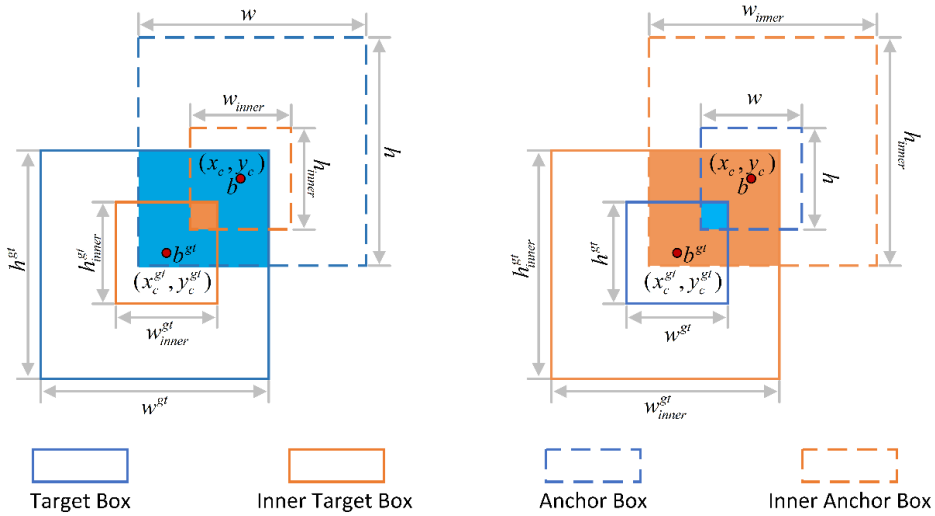


Fig. 7. Auxiliary bounding box drawing.

The calculation formula of Inner-EIoU is shown in Eq. (4).

$$L_{Inner-EIoU} = L_{EIoU} + IoU - IoU^{inner} \tag{4}$$

In Eq. (4), L_{EIoU} represents the EIoU loss function, IoU represents the intersection over union ratio between the predicted bounding box and the ground truth bounding box, and the definition of IoU^{inner} is shown in Eq. (5).

$$IoU^{inner} = \frac{inter}{union} \tag{5}$$

The ratio is the scale factor for generating the auxiliary bounding box, and its value range is usually [0.5, 1.5]. The definitions of *inter* and *union* are shown in Eq. (6) and Eq. (7).

$$\begin{aligned} inter = &\left(\min\left(b_r^{gt}, b_r\right) - \max\left(b_l^{gt}, b_l\right)\right) \\ &*\left(\min\left(b_b^{gt}, b_b\right) - \max\left(b_t^{gt}, b_t\right)\right) \end{aligned} \tag{6}$$

$$\begin{aligned} union = &\\ &\left(w^{gt} * h^{gt}\right) * (ratio)^2 + (w * h) * (ratio)^2 - inter \end{aligned} \tag{7}$$

The definitions of b_l^{gt}, b_r^{gt}, b_t^{gt}, b_b^{gt}, b_l, b_r, b_t and b_b are shown in Eq. (8) to Eq. (11).

$$b_l^{gt} = x_c^{gt} - \frac{w^{gt} * ratio}{2}, b_r^{gt} = x_c^{gt} + \frac{w^{gt} * ratio}{2} \tag{8}$$

$$b_t^{gt} = y_c^{gt} - \frac{h^{gt} * ratio}{2}, b_b^{gt} = y_c^{gt} + \frac{h^{gt} * ratio}{2} \tag{9}$$

$$b_l = x_c - \frac{w * ratio}{2}, b_r = x_c + \frac{w * ratio}{2} \tag{10}$$

$$b_t = y_c - \frac{h * ratio}{2}, b_b = y_c + \frac{h * ratio}{2} \tag{11}$$

Inner-EIoU takes into account not only the overlapping area but also geometric information such as the distance and angle between the predicted bounding box and the ground truth bounding box. This enables the model to adjust the position and orientation of the predicted bounding box more precisely during the training process, improving the detection accuracy for small objects, dense objects, and objects with irregular shapes.

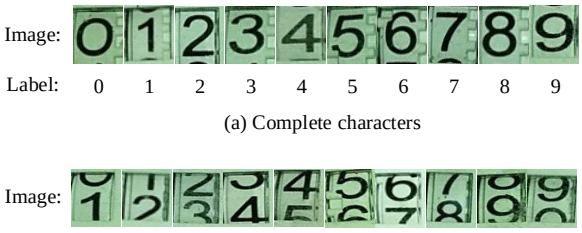
IV. Dataset and Experimental Design

A. Dataset

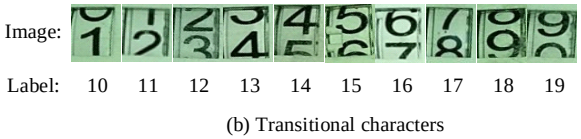
In this experiment, the dataset publicly available in the CCF Big Data and Computational Intelligence Contest is adopted. It is named the Automatic Water Meter Reading dataset in Real-world Scenarios. To improve the generalization ability of the dataset, and while ensuring that the water meter reading area can be recognized, this paper uses seven different data augmentation methods combined randomly to enhance the images in the dataset. The seven data augmentation methods used in the experiment are: 1) adding Gaussian noise; 2) changing the color temperature; 3) setting random brightness; 4) applying Gaussian blur; 5) random cropping and padding; 6) random rotation between -45° and +45°; 7) random scaling. The augmented dataset consists of a total of 3680 images with a size of 960×540 pixels. Some images from the dataset are shown in Fig. 8.

We used the LabelImg image annotation software to annotate the valid characters in water meter images. During the annotation process, we observed that the characters in the reading area might be in a transitional state. To achieve more precise readings, we introduced 10 new labels to represent transitional characters based on the complete characters. The annotated content includes 20 labels, ranging from "0" to "19", where "0 to 9" represent complete characters and "10 to 19" represent transitional characters. The individual characters of the water meter and their corresponding labels are shown in Fig. 9. The dataset was divided into training, testing, and validation sets in a 7:1:2 ratio for the experiments.

Fig. 8. Example diagram of the dataset.



(a) Complete characters



(b) Transitional characters

Fig. 9. The label corresponding to a single character in the dataset.

B. Evaluation Metrics

In the experiment of this paper, precision (P), recall (R), mAP@0.5 and mAP@0.5:0.95 are used to measure the reading recognition performance of the improved algorithm for the word-wheel water meter.

Precision refers to the proportion of actual positive samples among all the samples predicted as positive. Recall refers to the proportion of samples that are correctly predicted as positive among all the actual positive samples. The calculation formulas are shown in Eq. (12) and Eq. (13).

$$P = \frac{TP}{TP + FP} \tag{12}$$

$$R = \frac{TP}{TP + FN} \tag{13}$$

TP(True positive) represents the number of samples that are actually positive and predicted as positive, FP(false positive) represents the number of samples that are actually negative but predicted as positive, and FN(false negative) represents the number of samples that are actually positive but predicted as negative.

mAP represents the mean average precision. mAP@0.5 is calculated by computing the average precision of each category when the Intersection over Union (IoU) threshold is 0.5, and then taking the average of the average precisions of all categories. mAP@0.5:0.95 represents the average mAP at different IoU thresholds (ranging from 0.5 to 0.95 with a step size of 0.05), which is used to evaluate the detection performance of the model. The higher the mAP value is, the better the performance of the algorithm in the task of identifying the word-wheel water meter. The calculation formula are shown in Eq. (14) and Eq. (15).

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{14}$$

$$AP_i = \int_0^1 P_i(R)d(R) \tag{15}$$

FLOPs (Floating-Point Operations) refer to the number of floating-point operations, which is an important indicator for measuring the computational complexity and computational workload of deep learning models. Through the value of FLOPs, one can intuitively understand the computational complexity of the model. The larger the FLOPs value is, it indicates that the model needs to perform more floating-point operations during operation, the computational cost is higher, and it may require more computational resources and longer computation time.

C. Experimental Environment

The CPU of the experimental operating environment is an Intel(R) Xeon(R) Platinum 8255C 2.50GHz 12-core processor, with 43GB of memory. The GPU is an NVIDIA RTX 3090, and the video memory is 24GB. The operating system is Ubuntu 20.04, and the acceleration environment is CUDA 11.3. The programming language is Python 3.8, and the deep learning framework is Pytorch 1.11.0. The experimental parameter settings are shown in Table I:

TABLE I. PARAMETERS OF THE EXPERIMENT

Parameter Name	Parameter Value
Input resolution	640×640
Epochs	230
Batch size	8
Initial learning rate (Lr0)	0.01
Weight_decay	0.0005

V. EXPERIMENTAL RESULTS AND ANALYSIS

In order to verify the effectiveness of the improved algorithm, this paper conducts comparative experiments and ablation experiments on the dataset of word-wheel water meters. During the training process, the same experimental equipment and experimental parameters are adopted, and the obtained experimental results are compared and analyzed.

A. Performance Comparison Before and After the Algorithm Improvement

In this paper, the original YOLOv8n algorithm and the improved algorithm Meter-YOLOv8n were used to train on the same dataset of word-wheel water meters, resulting in the YOLOv8n model and the Meter-YOLOv8n model. The

changing trends of various evaluation indicators during the training process are shown in Fig. 10. Compared with the YOLOv8n model, the improved Meter-YOLOv8n model has improvements in multiple evaluation indicators, which proves the superiority of the Meter-YOLOv8n model.

As can be seen from Fig. 10, the improved model converges faster, indicating that the improved module effectively reduces the computational cost and resource consumption. The design of the improved model structure is more reasonable, enabling the model to perform better in the task of identifying the readings of word-wheel water meters. Compared with the original model, the improved model has improvements in precision, recall, and mAP, and the improved model also converges faster.

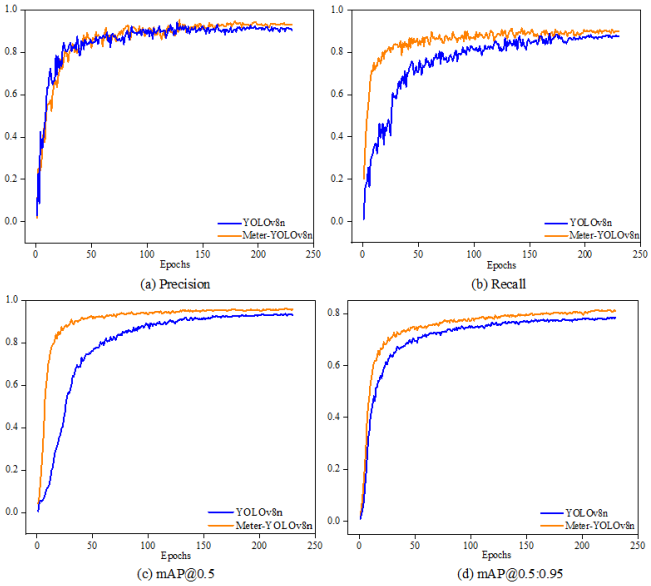


Fig. 10. The changing trends of various indicators before and after the algorithm improvement.

In order to explore the specific improvements of the improved model on the complete characters and transitional characters in the dataset, an experimental comparison is conducted between the improved model and the original YOLOv8n model. The experimental results of these two models on the test dataset of word-wheel water meters are shown in Table II.

TABLE II. TEST RESULTS OF THE MODEL BEFORE AND AFTER THE IMPROVEMENT

Model	Class	P(%)	R(%)	mAP@ 0.5(%)	mAP@ 0.5:0.95(%)
YOLOv8n	Complete characters	95.5	88.6	95.3	81.5
	Transitional characters	91.9	85.8	90.5	74.1
	All	93.7	87.2	92.9	77.8
Meter-YOLOv8n	Complete characters	96.1	90.2	97.0	83.5
	Transitional characters	94.7	88.0	95.6	76.7
	All	95.4	89.1	96.3	80.1

Table II compares the performance of the model before and after the improvement. Based on the experimental results of the two categories, namely complete characters and transitional characters, it can be concluded that the improved model has improvements in all indicators. The mAP@0.5 and mAP@0.5:0.95 of complete characters have increased by 1.7% and 2.0% respectively, while the mAP@0.5 and mAP@0.5:0.95 of transitional characters have increased by 5.1% and 2.6% respectively. The improvement range of various indicators of transitional characters is relatively large, indicating that the improved model has enhanced the expressive ability for transitional characters.

B. Comparative Experiments of the C2f Module

In order to improve the feature extraction ability of the backbone network and simplify the feature extraction process, this paper proposes several different improvement schemes for the C2f module, which are as follows: 1) Use the original C2f module. 2) Replace the convolution module in the Bottleneck structure of C2f with ODConv, named C2f-ODConv. 3) Add the attention mechanism in CloFormer that fuses global and local features to the Bottleneck of C2f, named C2f-CloAtt. 4) Integrate SCConv into the C2f module, named C2f-SCConv. 5) Replace the Bottleneck in C2f with Dilated Re-parameterized Block module from UniRepLKNet, named C2f-DRB. 6) Replace the Bottleneck in C2f with the Diverse Branch Block, named C2f-DBB. 7) Introduce the Faster Block module into the C2f module, named C2f-Faster. 8) Combine the iRMB attention mechanism and shift-wise conv to obtain the lightweight C2f-iRMBS module, and replace the Bottleneck structure in C2f with it, named C2f-iRMBS. These improvement strategies are trained using the same dataset under the same experimental environment. The experimental results are shown in Table III.

TABLE III. EXPERIMENTAL RESULTS OF DIFFERENT C2F MODULES

Model	mAP@0.5(%)	Params(M)	FLOPs(G)
C2f	92.9	3.15	8.7
C2f-ODConv [26]	92.3	3.07	5.8
C2f-CloAtt [27]	94.7	3.77	9.3
C2f-SCConv [28]	93.6	3.96	9.5
C2f-DRB [29]	90.6	2.60	6.4
C2f-DBB [30]	93.1	4.28	11.2
C2f-Faster [31]	94.2	2.71	6.7
C2f-iRMBS	94.4	2.72	6.7

As can be seen from Table III, C2f-ODConv significantly reduces the computational complexity and improves the computational efficiency. However, the mAP@0.5 drops by 0.6%, indicating that C2f-ODConv needs to strike a balance between computational complexity and mAP. C2f-CloAtt increases the mAP@0.5 to 94.7%, but the FLOPs increase significantly, raising the requirements for computational resources. C2f-SCConv improves the mAP of the model, but it also suffers from the problem of excessive reliance on computational resources, which is not conducive to practical

applications. C2f-DRB reduces the FLOPs, but the mAP@0.5 also drops, degrading the detection performance of the model. C2f-DBB has excessively high computational complexity and FLOPs, with mediocre overall performance. Both C2f-Faster and C2f-iRMBS improve the mAP of the model and reduce the number of parameters and FLOPs. This proves that C2f-Faster and C2f-iRMBS can well balance the relationship between computational complexity and mAP. Compared with other improvement strategies, C2f-iRMBS increases the mAP by 1.5% and reduces the number of parameters and FLOPs by 0.43M and 2.0G respectively. It can improve the detection accuracy and efficiency of word-wheel water meters and is more suitable for the task of word-wheel water meter reading recognition.

C. Comparative Experiments of different Loss Functions

In order to verify the impact of different loss functions on the model's detection performance, this paper conducts comparative experiments using models with several bounding box loss functions, namely CIoU, SIoU, EIoU, and Inner-EIoU. The ratio values of Inner-EIoU are set to 0.7, 0.8, 0.9, and 1.1 respectively. The experimental results are shown in Table IV.

TABLE IV. EXPERIMENTAL RESULTS OF DIFFERENT LOSS FUNCTIONS

Loss function	ratio	P/%	R/%	mAP@0.5/%
CIoU	-	93.7	87.9	92.9
SIoU	-	93.6	86.5	92.6
EIoU	-	94.0	88.1	93.1
Inner-EIoU	0.7	94.3	88.2	93.2
Inner-EIoU	0.8	94.6	88.3	93.3
Inner-EIoU	0.9	94.7	88.6	93.7
Inner-EIoU	1.1	94.4	88.2	93.4

From the experimental results in Table IV, it can be seen that when Inner-EIoU is selected as the loss function of the model, the detection performance is the best. When the ratio is set to 0.9, precision, recall, and mAP@0.50 achieve the optimal values. Therefore, Inner-EIoU is selected as the bounding box loss function in this paper, and the ratio is set to 0.9.

D. Ablation Experiment

In order to verify the contribution of the improved module to the improved model, an ablation experiment was conducted on the dataset of word-wheel water meters. By gradually introducing the improved module and evaluating the performance of the model, the results of the ablation experiment are shown in detail in Table V. YOLOv8n represents the baseline model. YOLOv8n-C represents YOLOv8n + C2f-iRMBS. YOLOv8n-S represents YOLOv8n + Slim-Neck. YOLOv8n-CS means YOLOv8n + C2f-iRMBS + Slim-Neck. Meter-YOLOv8n represents YOLOv8n + C2f-iRMBS + Slim-Neck + Inner-EIoU.

The baseline model, YOLOv8n, achieves a precision of 93.7%, a recall of 87.9%, a mAP@0.5 of 92.9%, and a mAP@0.5:0.95 of 76.8%. It has 3.15M parameters, 8.7G FLOPs, a model size of 6.3M, and an FPS of 83 F/S. Overall, its performance is average. After introducing C2f-iRMBS, mAP@0.5 and mAP@0.5:0.95 improved by 1.5% and 0.7%, respectively, while the number of parameters and FLOPs decreased by 0.43M and 2.0G, respectively. This indicates that C2f-iRMBS enhances the detection performance by capturing richer character feature maps through branch structures and shift operations while reducing model parameters and computational cost.

With the Slim-Neck structure improving the neck network, mAP@0.5 and mAP@0.5:0.95 increased by 1.2% and 0.4%, respectively, while parameters and FLOPs decreased by 0.14M and 1.6G, respectively. This suggests that Slim-Neck effectively integrates features across different scales and levels, improving detection efficiency.

When both C2f-iRMBS and Slim-Neck are introduced simultaneously, mAP@0.5 and mAP@0.5:0.95 increased by 3.0% and 2.8%, respectively, while parameters and FLOPs decreased by 0.57M and 2.6G, demonstrating their strong compatibility. This enables the model to better balance detection performance and computational cost.

By integrating all modules, the resulting Meter-YOLOv8n model achieves an accuracy of 95.4%, a recall of 89.1%, a mAP@0.5 of 96.3%, and a mAP@0.5:0.95 of 80.1%. It has 2.59M parameters, 6.1G FLOPs, a model size of 5.6M, and an FPS of 87 F/S. Experimental results confirm that Meter-YOLOv8n enhances the detection performance of word-wheel water meters while maintaining a lightweight design, with each improvement module playing a positive role.

TABLE V. RESULTS OF ABLATION EXPERIMENT

Model	P (%)	R (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)	Params (M)	FLOPs (G)	Size (MB)	FPS (F/S)
YOLOv8n	93.7	87.9	92.9	76.8	3.15	8.7	6.3	83
YOLOv8-C	94.7	88.2	94.4	77.5	2.72	6.7	5.8	85
YOLOv8n-S	94.1	88.0	94.1	77.2	3.05	8.1	6.1	84
YOLOv8n-CS	95.3	89.1	95.9	79.6	2.58	6.1	5.6	87
Meter-YOLOv8n	95.4	89.1	96.3	80.1	2.59	6.1	5.6	87

E. Comparative Experiment

To verify the detection performance of the improved algorithm on the word-wheel water meter, the improved algorithm was compared with the currently popular single-stage and two-stage algorithms, specifically including Faster R-CNN, DETR, YOLOv3, YOLOv4, YOLOv5s, YOLOv7-tiny, YOLOv8s, YOLOv8n, YOLOv10n and YOLOv11n. The comparison results are shown in Table VI.

According to the experimental results in Table VI, the number of parameters, FLOPs, and model size of the Faster R-CNN, YOLOv3, and DETR algorithms are too large, and their detection performance is mediocre. Therefore, their application in real-world scenarios is somewhat limited. The YOLOv4 algorithm has relatively high computational complexity and requires more computational resources, so it is not suitable for the task of word-wheel water meter reading recognition. The YOLOv7-tiny algorithm strikes a balance among performance, computational load, and model size, but its overall performance is just average. YOLOv5s and YOLOv8n have similar performance, but the model size and computational load of YOLOv5s are larger than those of YOLOv8n. The YOLOv10n and YOLOv11n algorithms have fewer parameters and FLOPs, but their detection performance is not satisfactory. The recognition accuracy and mAP of YOLOv8s are slightly higher than those of YOLOv8n, but its model size is more than three times that of YOLOv8n, leading to difficulties in deployment. Compared with other models, Meter-YOLOv8n has the highest mAP. Although its number of parameters is slightly higher than that of YOLOv10n, its FLOPs and model size are lower. Thus, it can quickly and accurately read the readings of word-wheel water meters in resource - constrained environments.

TABLE VI. RESULTS OF COMPARATIVE EXPERIMENT

Model	P (%)	R(%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)	Params (M)	FLOPs (G)	Size (MB)	FPS (F/S)
Faster R-CNN	92.8	86.1	90.9	75.7	43.95	133.6	106.2	37
DETR [32]	93.5	86.9	92.3	76.8	42.30	122.3	113.3	38
YOLOv3 [33]	82.3	73.1	85.6	69.7	10.33	45.6	78.6	45
YOLOv4 [34]	92.9	86.1	91.1	76.2	9.96	33.7	32.9	65
YOLOv5s	93.8	87.7	92.9	77.1	9.12	23.2	18.6	72
YOLOv7-tiny	92.6	86.0	90.7	75.6	6.01	12.8	14.1	75
YOLOv8s	94.0	88.1	93.1	77.1	11.1	28.5	42.1	62
YOLOv8n	93.7	87.9	92.9	76.8	3.15	8.7	6.3	83
YOLOv10n [35]	92.1	85.6	90.3	75.3	2.58	7.8	5.9	85
YOLOv11n [36]	92.0	85.3	90.1	74.9	2.60	6.3	5.8	85
Meter-YOLOv8n	95.4	89.1	96.3	80.1	2.59	6.1	5.6	87

F. Visual Analysis

1) Dataset visualization: The label distribution diagram after training on the word-wheel dataset is shown in Fig. 11.

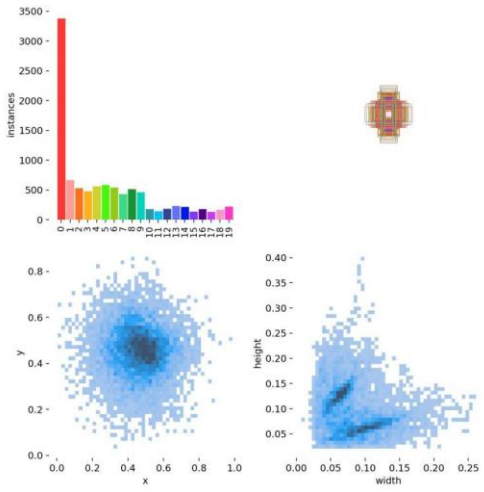


Fig. 11. Dataset visualization.

From the bar chart, it can be seen that the number of samples in the "0" category is the largest, exceeding 3000. The number of samples of complete characters "1 to 9" is around 500, and the number of samples of transitional characters "10 to 19" is relatively small, approximately 250. From the x-y density plot and the width - height density plot, it can be seen that most of the object center points are concentrated near (0.5, 0.5), forming an obvious dense area and exhibiting the characteristics of a Gaussian distribution. The width and height of most objects are relatively small, concentrated between 0.05 and 0.15. From the overlapping box plot, it can be seen that most of the objects are located in the central area of the image, and the distribution of the bounding boxes is relatively symmetric.

2) Heatmap visualization: In order to more precisely observe the degree of attention paid by the model to the effective characters of the water meter before and after the improvement, this paper uses Gradient-weighted Class Activation Mapping (Grad-CAM) to generate heatmaps for the YOLOv8n model and the improved model Meter-YOLOv8n. Grad-CAM plays a crucial role in the field of model interpretability. It enables researchers to determine whether the

model accurately focuses on the correct image features related to the recognition of water meter readings when processing images. By observing the different colored areas in the heatmaps, we can determine the contribution degree of different regions in the water meter image to the prediction results. The red and yellow areas indicate a higher contribution degree to the prediction results, the green areas indicate a lower contribution degree to the prediction results, and the blue areas indicate no contribution to the prediction results. The heatmaps before and after the model improvement are shown in Fig. 12.

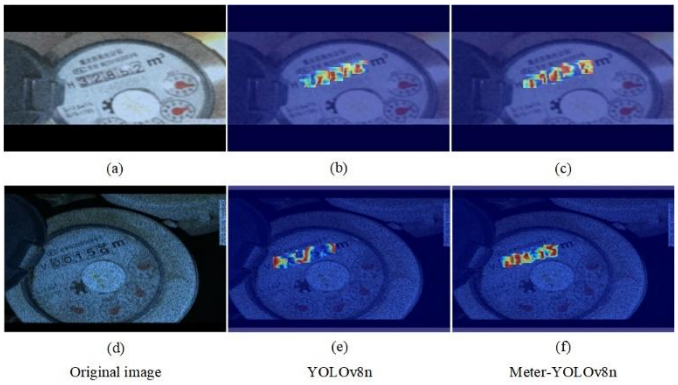


Fig. 12. Heatmap visualization.

By comparing Fig. 12(b) and Fig. 12(c), it can be seen that the YOLOv8n model pays more attention to the edge information of the effective characters, which leads to a decrease in the recognition accuracy of the water meter by the model. As shown in Fig. 12(e), when the water meter image is affected by noise, the YOLOv8n model reduces its attention to the effective characters, and the blue area accounts for a relatively large proportion, resulting in situations of missed detection and false detection by the model. As shown in Fig. 12(f), the improved model focuses more on the character information on the dial and suppresses the interference of other invalid characters on the dial. By observing the heatmap generated by Grad-CAM, it can be known that, compared with the YOLOv8n model, the image areas that the improved model focuses on when making decisions are more comprehensive, which improves its detection ability for complete characters and transitional characters.

3) Results visualization: In order to more intuitively demonstrate the detection performance and generalization ability of the improved model, models with relatively good comprehensive performance are used to conduct inference verification on the validation set of word-wheel water meters. These models include YOLOv7-tiny, YOLOv10n, YOLOv8n, YOLOv8s, and the improved model Meter-YOLOv8n. The visualization of the experimental results is shown in Fig. 13.

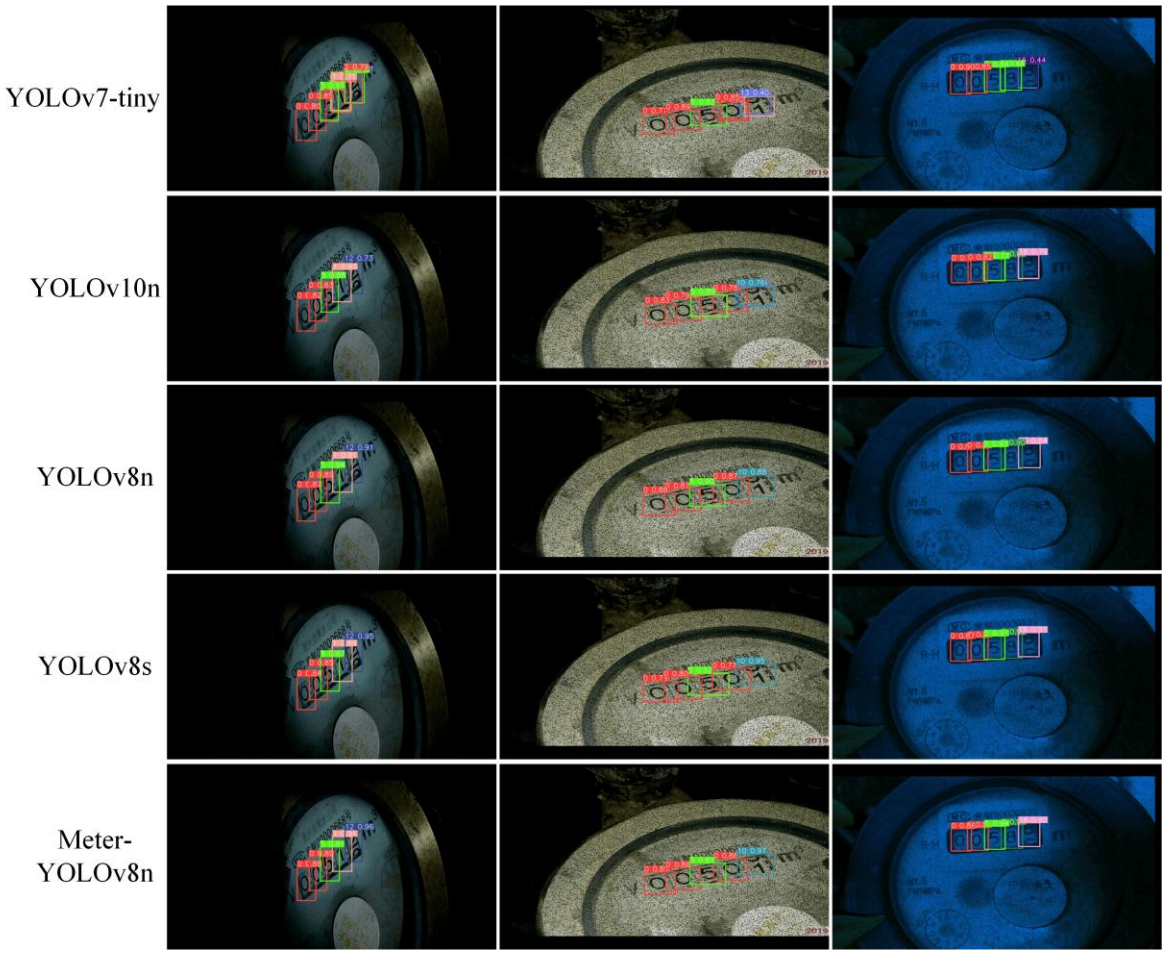


Fig. 13. Result visualization.

In the absence of interference, YOLOv7-tiny and YOLOv10n can correctly recognize the complete characters on the word-wheel water meter, but they perform poorly in recognizing transitional characters. When the image is affected by unfavorable factors such as lighting and noise, the recognition accuracy of YOLOv7-tiny and YOLOv10n drops significantly, and YOLOv7-tiny experiences false detections and repeated detections. YOLOv10n, YOLOv8n, and YOLOv8s can accurately recognize both the complete characters and transitional characters of the water meter in an interference-free environment. However, when the image is affected by environmental interference, the recognition accuracy of these models for characters decreases, especially for transitional characters. The improved model Meter-YOLOv8n can accurately recognize both the complete characters and transitional characters of the word-wheel water meter in different environments, and it has the highest recognition accuracy, fully verifying the detection performance and generalization ability of the improved model Meter-YOLOv8n.

VI. Conclusion

This paper proposes a high-accuracy and lightweight word-wheel water meter reading recognition model, Meter-YOLOv8n, and introduces multiple improvements tailored to the characteristics of water meter images. The C2f-iRMBS module is introduced to replace the original C2f, simplifying the feature extraction network while enhancing the model's ability to extract character information from water meters. To address the challenges of small object features being indistinct and highly similar to the background in water meter images, a Slim-Neck module is incorporated to enhance multi-scale feature fusion of small objects, thereby improving detection accuracy. Additionally, the Inner-EIoU loss function replaces the original CIoU loss function, enhancing the performance of bounding box regression. Through comparative experiments, ablation studies, and visualization analysis, the improved model achieves a 3.4% increase in mAP@0.5 compared to the original model, while reducing the number of parameters by 0.56M and FLOPs by 2.6G. The proposed improvements achieve a better balance between detection accuracy and model complexity. The Meter-YOLOv8n model has improved the recognition accuracy of word-wheel water meters while reducing the computational load and model size. It is more in line with the characteristics of edge devices, which have limited resources but require high detection accuracy. This has laid a solid foundation for its deployment on edge devices.

In future work, we will collect more images of word-wheel water meters from different brands and environments, and increase the number of transitional characters in the dataset. We will take the actual deployment on low-power edge hardware such as Raspberry Pi or NVIDIA Jetson Nano.

Acknowledgment

This work was supported the National Natural Science Foundation of China (Grant No. 62076152).

References

[1] Bhushan D S. A review paper on automatic meter reading and instant billing[J]. International Journal of Advanced Research in Computer and Communication Engineering, 2015, 4(1).

[2] Zhao S, Lu Q, Zhang C, et al. Effective recognition of word-wheel water meter readings for smart urban infrastructure[J]. IEEE Internet of Things Journal, 2024, 11(10): 17283-17291.

[3] Rawat N, Rana S, Yadav B, et al. A review paper on automatic energy meter reading system[C]//2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom). IEEE, 2016: 3254-3257.

[4] Liang Y, Liao Y, Li S, et al. Research on water meter reading recognition based on deep learning[J]. Scientific Reports, 2022, 12(1): 12861.

[5] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[6] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 39(6): 1137-1149.

[7] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.

[8] Sapkota R, Qureshi R, Calero M F, et al. YOLOv10 to its genesis: a decadal and comprehensive review of the you only look once (YOLO) series[J]. arXiv preprint arXiv:2406.19407, 2024.

[9] Jiang H, Lu Y, Zhang D, et al. Deep learning-based fusion networks with high-order attention mechanism for 3D object detection in autonomous driving scenarios[J]. Applied Soft Computing, 2024, 152: 111253.

[10] Rezaee K, Rezakhani S M, Khosravi M R, et al. A survey on deep learning-based real-time crowd anomaly detection for secure distributed video surveillance[J]. Personal and Ubiquitous Computing, 2024, 28(1): 135-151.

[11] Gupta H, Verma O P. Monitoring and surveillance of urban road traffic using low altitude drone images: a deep learning approach[J]. Multimedia Tools and Applications, 2022, 81(14): 19683-19703.

[12] Jaiswal S K, Agrawal R. A Comprehensive Review of YOLOv5: Advances in Real-Time Object Detection[J]. International Journal of Innovative Research in Computer Science and Technology, 2024, 12(3): 75-80.

[13] Gong M, Wang D, Zhao X, et al. A review of non-maximum suppression algorithms for deep learning target detection[C]//Seventh Symposium on Novel Photoelectronic Detection Technology and Applications. SPIE, 2021, 11763: 821-828.

[14] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8759-8768.

[15] Cai Z, Wei C, Yuan Y. An efficient method for electric meter readings automatic location and recognition[J]. Procedia Engineering, 2011, 23: 565-571.

[16] Jawas N. Image based automatic water meter reader[C]//Journal of Physics: Conference Series. IOP Publishing, 2018, 953(1): 012027.

[17] Chen L, Sun W, Tang L, et al. Research on Automatic Reading Recognition of Wheel Mechanical Water Meter Based on Improved U-Net and VGG16[J]. WSEAS Transactions on Computers, 2022, 21: 283-293.

[18] Men Z. An Improved Method for Digital Water Meter Reading Area Segmentation Based on U~ 2-Net[J]. Academic Journal of Computing & Information Science, 2023, 6(12): 33-44.

[19] Li C, Su Y, Yuan R, et al. Light-weight spliced convolution network-based automatic water meter reading in smart city[J]. IEEE Access, 2019, 7: 174359-174367.

[20] Zou L, Xu L, Liang Y, et al. Robust water meter reading recognition method for complex scenes[J]. Procedia Computer Science, 2021, 183: 46-52.

[21] Zhang J, Liu W, Xu S, et al. Key point localization and recurrent neural network based water meter reading recognition[J]. Displays, 2022, 74: 102222.

[22] Li J, Shen J, Nie K, et al. Reading Recognition Method of Mechanical Water Meter Based on Convolutional Neural Network in Natural Scenes[J]. Journal of Advanced Computational Intelligence and Intelligent Informatics, 2024, 28(1): 206-215.

[23] Wang Y, Xiang X. GMS-YOLO: an enhanced algorithm for water meter reading recognition in complex environments[J]. Journal of Real-Time Image Processing, 2024, 21(5): 173.

[24] Chiang H Y, Frumkin N, Liang F, et al. Mobiletl: On-device transfer learning with inverted residual blocks[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2023, 37(6): 7166-7174.

[25] Li H, Li J, Wei H, et al. Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles[J]. arXiv preprint arXiv:2206.02424, 2022, 10.

[26] Li C, Zhou A, Yao A. Omni-dimensional dynamic convolution[J]. arXiv preprint arXiv:2209.07947, 2022.

[27] Fan Q, Huang H, Guan J, et al. Rethinking local perception in lightweight vision transformer[J]. arXiv preprint arXiv:2303.17803, 2023.

[28] Li J, Wen Y, He L. Scconv: Spatial and channel reconstruction convolution for feature redundancy[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 6153-6162.

[29] Ding X, Zhang Y, Ge Y, et al. Unireplknet: A universal perception large-kernel convnet for audio video point cloud time-series and image recognition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 5513-5524.

[30] Ding X, Zhang X, Han J, et al. Diverse branch block: Building a convolution as an inception-like unit[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 10886-10895.

[31] Chen J, Kao S, He H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 12021-12031.

[32] Zhu X, Su W, Lu L, et al. Deformable detr: Deformable transformers for end-to-end object detection[J]. arXiv preprint arXiv:2010.04159, 2020.

[33] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.

[34] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.

[35] Wang A, Chen H, Liu L, et al. Yolov10: Real-time end-to-end object detection[J]. Advances in Neural Information Processing Systems, 2024, 37: 107984-108011.

[36] Khanam R, Hussain M. Yolov11: An overview of the key architectural enhancements[J]. arXiv preprint arXiv:2410.17725, 2024.