Measurement 222 (2023) 113615

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Deep learning-based autonomous real-time digital meter reading recognition method for natural scenes

Jianqing Peng [a,b], Wei Zhou [a], Yu Han [a,b], Mengtang Li [a,*], Wanquan Liu [a]

[a] School of Intelligent Systems Engineering, Shenzhen Campus of Sun Yat-sen University, Shenzhen 518107, China
[b] Guangdong Provincial Key Laboratory of Fire Science and Technology, Guangzhou 510006, China

ARTICLE INFO

Keywords:
Digital meter reading
Autonomous recognition
Deep learning
Perspective transformation
Keypoint detection

ABSTRACT

Natural scenes with variable illumination, variable target scale and angular tilt pose significant challenges to the autonomous recognition of digital meter readings. Based on this, this paper proposes a deep learning-based autonomous real-time digital meter reading recognition method for natural scenes. First, the YOLO-style corner point detection method (YOLO-CPDM) for the reading area is proposed by reconstructing the detection heads and incorporating the corner detection loss function. Its localization accuracy is further refined by embedding attention mechanism module, implementing dynamic loss function and enhancing training data diversity through offline augmentation techniques like image rotation and flipping. Then, the detected corner points are used to geometrically correct the distorted reading area by perspective transformation to mitigate the interference caused by the shooting angle. Next, the YOLO-style end-to-end reading recognition method (YOLO-EERRM) is proposed to accurately extract the characters in the reading area. Finally, the validity of the YOLO-CPDM and YOLO-EERRM was verified on a produced dataset named SYSU-DM and 2 public datasets. Compared with the State of the arts (SOTA) keypoint detection model, the mean Average Precision @ 50:95 scores of the YOLO-CPDM improved by 2.8, 4.1, and 1.1 points, respectively, while the inference latency was only 5.3 ms, and YOLO-EERRM achieved 100 % accuracy and 3.1 ms inference latency on the SYSU-DM dataset. Statistically, the complete digital meter reading recognition method has 99.6 % accuracy and 8.6 ms inference latency, indicating that the system has high recognition accuracy and practicality.

1. Introduction

Automatic and accurate recognition of digital meter readings in natural scenarios is important for establishing intelligent systems such as smart billing [12] and smart monitoring [34]. Although traditional meters are gradually being replaced by smart meters, they may bring about security issues in data transmission [5], and the public has expressed concerns about privacy leakage [6]. In addition, traditional old mechanical water meters [7], gas meters [8], and electricity meters [9,10] are still widely used in relatively underdeveloped regions [11], requiring staff to visit the site for manual reading and recording [9,12]. Manual meter reading suffers from factors such as subjective readings and visual fatigue that led to error-prone, as well as low efficiency, so there is an urgent need for an intelligent and effective means to improve efficiency.

Machine vision technology is gradually receiving wide attention in the fields of smart agriculture [13], smart manufacturing [14], smart medical [15], and autonomous driving [16], and is becoming an important research direction in the field of meter reading recognition. There are two main approaches to automatic meter reading based on machine vision, one is fixed, which involves installing an embedded terminal device on the meter for optical character recognition (OCR) [17], however, this requires each meter to deploy an embedded device with a camera, which is relatively expensive. Another approach is mobile, utilizing handheld terminals for photography [912] or robots for telemetry [3], providing greater flexibility and traceability.

Machine vision-based digital meter reading recognition algorithm in natural scenes consists of two key steps: reading area detection and reading recognition. Traditional image processing methods (i.e., by color features [18], edge detection[19], and template matching [20]) can be applied to detect reading regions, and characters can be segmented using projection methods [21,22], and template matching or

* Corresponding author.
E-mail addresses: pengjq7@mail.sysu.edu.cn (J. Peng), zhouw68@mail2.sysu.edu.cn (W. Zhou), hanyu25@mail.sysu.edu.cn (Y. Han), limt29@mail.sysu.edu.cn (M. Li), liuwq63@mail.sysu.edu.cn (W. Liu).

https://doi.org/10.1016/j.measurement.2023.113615
Received 26 July 2023; Received in revised form 5 September 2023; Accepted 24 September 2023
Available online 4 October 2023

support vector machine (SVM) [20] can recognize characters. These methods are only applicable to well-designed scenes and are not suitable for natural scenes due to their susceptibility to environmental factors such as overexposure or reflection.

Compared with traditional methods, deep learning-based algorithms have obvious advantages in terms of accuracy [23], especially the YOLO series models [24–26], region-based convolutional neural network (R-CNN) series models [27–29], anchor free models [30,31] and classical OCR [32] algorithms, which have facilitated the development of autonomous recognition techniques for digital meter readings. Several scholars have used object detection models to achieve highly robust reading region detection in complex scenes [33–35]. However, general object detection models are not applicable to tilted shooting views. Zhang *et al.* [36] used stacked hourglass networks (SHN) keypoint detection model to achieve corner point detection in the reading region, but the real-time performance of the SHN is poor. The extraction of reading regions can reduce the difficulty of reading recognition, Laroca *et al.* [12] achieved end-to-end reading detection and recognition by Fast-OCR.

In addition, collecting and labeling digital instrument reading dataset requires intensive labor, making it difficult to obtain large datasets. Therefore, input images typically facing uncertainty and noise, such as rotation, translation, image distortion, and uneven lighting. There exist some latest studies on uncertainty theory [37–39], but they are not applicable to the object of this research. Inspired by [12,40], this paper synthesizes additional data through offline data augmentation algorithm, such as image rotation and flipping, to introduce random perturbations into the training data.

In order to solve the problems of image distortion caused by tilting the shooting angle in-plane and out-of-plane and poor real-time due to the complicated process of traditional detection and extraction methods, this study proposes an autonomous real-time digital meter reading recognition method based on deep learning for natural scenes. Corner point detection for the reading region is achieved by reconstructing the detection head and incorporating the corner detection loss function, its localization accuracy is further improved by embedding attention mechanism modules, dynamic loss functions, and offline enhancement techniques. Geometric correction of the coordinates of corner points in the distorted reading area is performed by the perspective transformation matrix to reduce the interference caused by the variable shooting angle. Further, the characters in the reading area are extracted by the YOLO-style end-to-end reading recognition method (YOLO-EERRM). Finally, all processes are integrated into an end-to-end pipeline to obtain complete readings identification results.

The effectiveness of the YOLO-style corner point detection method (YOLO-CPDM) was confirmed on three datasets containing mechanical and electronic digital meters, scoring 2.8, 4.1, and 1.1 points higher than the SOTA keypoint detection model's $mAP50^{:95}$ (i.e., the mean of mAP scores at 10 positions, OKS = 0.50, 0.55,…, 0.90, 0.95) score at an inference delay of only 5.3 ms, respectively. In addition, YOLO-EERRM obtained nearly 100 % accuracy and 3.1 ms inference latency on the SYSU-DM dataset. The accuracy of the whole reading recognition is 99.6 % and the inference delay is 8.6 ms, indicating the high accuracy and practicality of this smart meter reading system. Therefore, the proposed method can contribute to the development of intelligent industry, smart cities and the Internet of Things by providing accurate and real-time data for decision-making.

In summary, the main contribution of this study are as follows:

1. Real-time digital meter reading recognition system in natural scenes is proposed to solve the problem of misrecognition caused by variable lighting, variable target scale and angular tilt of meter images in natural scenes. The system can be applied to different types of digital meters such as gas meters, water meters and electric energy meters, and has the characteristics of high real-time and robustness, which is important for establishing intelligent systems such as smart billing and smart monitoring.
2. A YOLO-CPDM for reading area is proposed, which reconstructs the detection head and optimization target of YOLOv5-C3CBAM model to solve the image distortion problem caused by tilting the shooting angle in-plane and out-of-plane. YOLO-CPDM has the advantages of box-and-point integration, unconstrained corner point position and real-time operation.
3. A YOLO-EERRM for reading recognition is proposed, which embeds the CBAM attention mechanism module in the C3 structure of YOLOv5, solving the problems of inaccurate reading recognition due to errors in character segmentation under dark light, exposure and uneven illumination, and improving the feature extraction capability and recognition accuracy.
4. A challenging dataset of digital meter readings under multiple shooting distances and angles (named SYSU-DM) was produced. It contains 1255 raw images of electronic digital multimeter readings taken under complex environments such as in-plane and out-of-plane with tilt angles, various shooting distances, and multiple lighting conditions, and precisely labeled corner points of the reading area. This dataset helps to more comprehensively simulate and solve the challenges of meter reading in practical applications.

The rest of the paper is organized as follows. Chapter 2 contains a review of related studies. The 3rd chapter proposes YOLO-CPDM for reading area detection and gives the YOLO-EERRM for reading recognition. The 4th chapter is the experimental part. The last chapter is a summary and outlook.

2. Related work

Automatic digital meter recognition is usually divided into two main subtasks: reading region detection, which is to accurately locate the reading region of a digital meter in an image; reading recognition, which is to generate a string related to the meter reading.

Subsequently, existing research on these two subtasks of digital meter reading recognition and related fields is reviewed and reveals the connection between existing methods and the approach of this paper, and gives the difficulties faced by the current research.

2.1. Reading area detection

Reading region detection essentially belongs to the category of object detection and aims to pinpoint reading regions in images, and similar research areas include text localization of metal surfaces [40] and code localization of containers [41,42], and so on. With the continuous development of computer vision technology, deep learning methods applied to reading region detection have been significantly improved in terms of detection accuracy compared with traditional image processing methods. In this paper, the reading region detection techniques are divided into traditional image processing-based methods and deep learning-based methods.

2.2. Traditional detection methods

Conventional reading region detection is achieved by traditional image processing methods for object detection. Anis *et al.* [18] converted RGB images to YCbCr color space and extracted reading regions by thresholding Cb and Cr to overcome the effects of illumination variations. Kanagarathinam *et al.* [9] implemented tilted position, blurred background, and reading area detection in day and night light by MSER algorithm, but these methods are susceptible to ambient glare, reflections, and dark light. Wu *et al.* [43] extracted the reading region by image differencing. Bai *et al.* [22] acquired binary images by global threshold binarization methods and located the reading area by horizontal and vertical projection, but these methods were performed in well

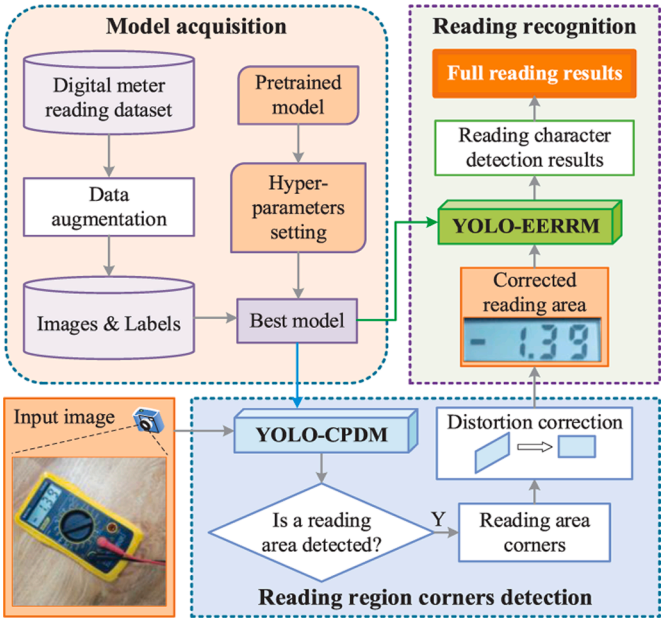

Fig. 1. The flowchart for autonomous recognition of digital meter readings.

controlled scenes and could not be applied to images taken under tilted view.

In short, limited by the low robustness of traditional image processing technology, these methods can only work in a fixed camera view or well controlled environment, which is very demanding on the working environment and therefore cannot be applied to natural scenes such as variable lighting conditions, multi-angle shooting and multi-scale shooting.

2.3. Deep Learning-Based detection methods

Thanks to the powerful nonlinear fitting ability of deep learning techniques, it has a wide application prospect in the field of reading region detection. The deep learning-based object detection techniques are mainly divided into two-stage R-CNN series models [27–29] and one-stage YOLO series models [24–26], and the model accuracy and computational efficiency are continuously improved with version iterations. However, they serve as general-purpose object detection models, and the output is a horizontal rectangular box, so they are not applicable to tilted shooting views.

Several researchers have directly utilized a general-purpose object detection model for reading region detection. Shuo *et al.* [34] performed digital region localization by a lightweight object detection model (i.e., MobileNetv2-SSD). Chong *et al.* Iqbal *et al.* [44] achieved high accuracy and reliability using an inception model containing a single shot detector.

In response to the problem of difficult detection of tilted perspectives, some researchers have conducted studies on this issue. Hong *et al.* [7] proposed a reading angle adjustment method based on orientation alignment network, which can locate the digit box by YOLOv3. However, in-plane rotation correction cannot completely solve the problem of image perspective transformation. Laroca *et al.* [12] first detected the reading area by Fast-YOLOv4-SmallObj and then used CDCC-Net to detect the position of corner points. Furthermore, perspective transformation was applied to correct the reading area. However, cascading two models resulted in image features being repeatedly computed, which increases the unnecessary time overhead. Carvalho *et al.* [45] developed a real-time digital meter reading system. The pipeline of the system was executed by serially integrating into three neural network models: screen border detection, screen keypoint detection, and text detection. However, the more steps, the more probability of system error increases subsequently. Zhang *et al.* [36] proposed a water meter reading area detection method based on keypoint localization and geometric correction, but the corners detection model based on SHN is hard to achieve real-time monitoring. Although these methods take into account the image distortion caused by tilted views, their steps are cumbersome, and the models are too large to be deployed in practice.

Therefore, considering the complex lighting conditions and multiple scales for digital meter imaging captured in natural scenes, a deep learning-based detection method is the optimal choice. However, there are in-plane and out-of-plane tilts in the shooting angle, which causes image distortion and failure of general object detection models. Moreover, the inference of existing corner detection models is time-consuming. Therefore, it is necessary to design a real-time, high-accuracy and high-stability reading area corner point detection model.

2.4. Reading recognition

Reading recognition is the recognition of meter indications in the detected structured reading area and essentially falls under the category of OCR. In this paper, we divide the existing reading recognition techniques into two-step methods and end-to-end methods that does not require segmentation.

2.5. Two-step methods

Anis *et al.* [18] segmented individual characters in edge images using connected domain filtering, morphological processing and vertical projection, and recognized the corresponding characters based on the encoded values. However, the study has more steps and is error-prone. Bin *et al.* [46] located the digits by morphological operations, connected domain filtering, and horizontal-vertical projection. Then, the feature vectors are obtained using the image midline scan for character recognition. However, the method of character segmentation is only suitable for well controlled scenes. Zhou *et al.* [35] separated the characters by calculating the line equations on both sides of the characters, however, this method is not applicable to multiple types of instruments. Wu *et al.* [43] used a connected domain approach for character segmentation of binary images and used K nearest neighbors algorithm for character recognition, but this method can only be used in well controlled scenarios. Shuo *et al.* [34] used a projection histogram for character segmentation and used an SVM classifier to recognize each segmented character. The projection method is susceptible to environmental factors such as overexposure, dark light, reflection and glare. Hong *et al.* [7] proposed a space layout guidance algorithm to locate digits and use CNN for digit classification.

In summary, the two-step method reduces the difficulty of recognizing individual characters by classifying the segmented characters, but errors in intermediate processing processes such as character segmentation can lead to errors in character recognition, which in turn can lead to reading errors.

2.6. End-to-end methods

In order to avoid the problem of erroneous recognition of readings due to errors in intermediate processing processes such as character segmentation, some researchers have investigated end-to-end readings recognition methods. Gómez et al. [47] proposed a segmentation-free system for text recognition in natural scenes, which is trained in an end-to-end manner in a CNN architecture and is able to output readings directly without any explicit text localization. However, the method has no text localization and is not effective in recognizing perspective distorted images. Laroca *et al.* [12] built a new lightweight detection network, called Fast-OCR, which treats reading recognition as an object detection problem. Carvalho *et al.* [45] used pre-trained Rosetta and CRNN models for text recognition, respectively. Li *et al.* [8] used a character recognition model based on a multi-attention mechanism and

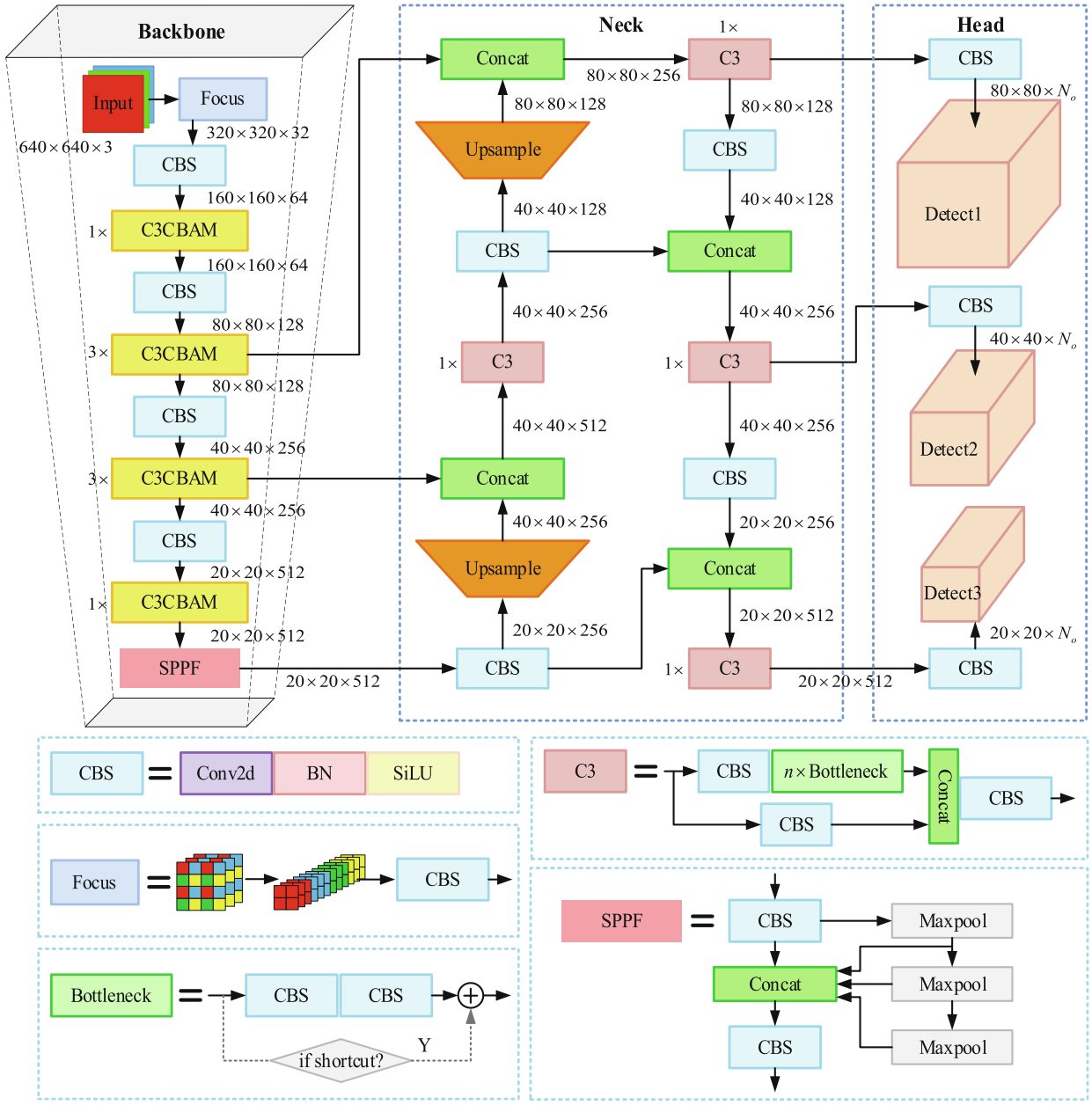

Fig. 2. Model structure diagram of YOLOv5-C3CBAM.

codec (MAEDR + CTC), which performs well in extreme scenarios such as overexposure, blurring and image occlusion. Imran *et al.* [10] recognized the numbers of meters directly by YOLOv3, which outperformed the traditional hand-made SVM-based classifier. However, the method tends to increase the difficulty of recognition due to its own small scale, distorted text distortion, and the presence of other text in the image.

In summary, end-to-end reading recognition methods are mainly divided into character detection methods based on object detection and text recognition methods based on indefinite length sequences, which do not cause final reading errors due to errors in intermediate processing processes such as character segmentation. The difficulty of reading recognition is greatly simplified by region extraction and distortion correction of readings.

3. Method

The flowchart of the deep learning-based autonomous real-time digital meter reading recognition method for natural scenes is shown in Fig. 1, which mainly includes the model acquisition module, the corner point detection module of the reading area and the reading recognition module. The model acquisition module includes manually labeling the dataset images of digital meters, acquiring the corresponding labels, and fine-tuning the pre-trained model acquired on the large-scale dataset. The corner point detection module of the reading area mainly locates the four corner point coordinates of the reading area through the YOLO-CPDM model based on the digital meter images taken from natural scenes, and then uses the corner point coordinates for perspective transformation to realize the distortion correction of the reading area. The reading recognition module uses the YOLO-EERRM model to detect each character in the reading area based on the distortion-corrected reading area image, and then obtains the reading.

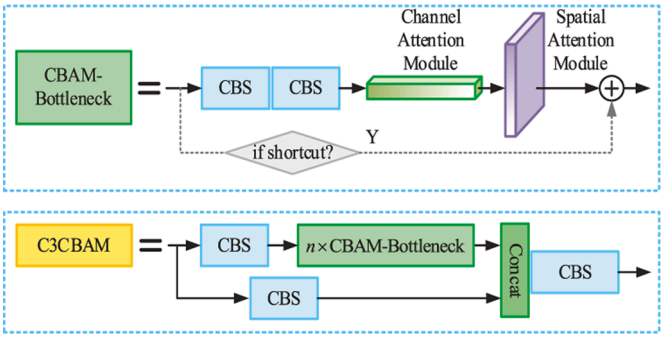

Fig. 3. The schematic diagram of the C3CBAM module.

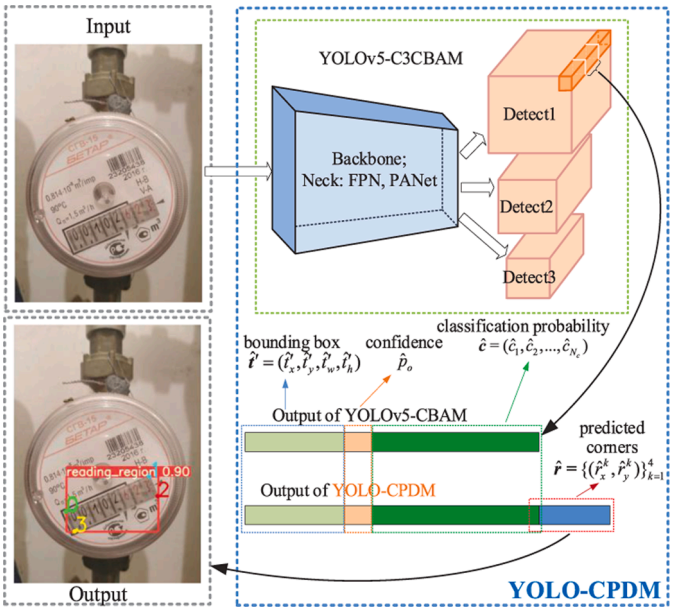

Fig. 4. The schematic diagram of the YOLO-CPDM model and prediction results.

Considering the high real-time, high accuracy and multi-scale features of YOLO series models, they have significant advantages in reading area detection [7,12] and reading recognition [10]. In this paper, YOLOv5-C3CBAM model is proposed based on the open-source object detection model YOLOv5 R5.5 from Ultralytics, and the structure diagram of the model is shown in Fig. 2. The neck network utilizes the feature pyramid network (FPN) and path aggregation network (PAN) [48] to fuse the features extracted from the backbone. Finally, three types of detection heads are used to predict outputs of muti-scales objects.

Specifically, the CBAM [49] attention mechanism is embedded in the C3 module of its backbone to form the C3CBAM attention module. We replace the C3 module in the YOLOv5 model backbone with the C3CBAM module, i.e., the Bottleneck component in C3 is replaced with the CBAM-Bottleneck component, as shown in Fig. 3. The introduction of the attention mechanism allows the model to focus on the information that is more critical to the current task and reduces the attention to other information, thus improving the accuracy of corner point localization and character detection.

Besides, Fig. 2 lists the variations in size and channels of a 640 × 640 RGB image during the model's forward propagation. The input image size must be divisible by 32. The model's module stack count is also annotated, remaining fixed as YOLOv5-C3CBAM is based on the yolov5s model. These parameters are applicable to various datasets and serve as fixed parameters for general detection problems.

3.1. Reading area detection method in multi-angle

3.1.1. Corner point detection method for digital meter reading area

Since the YOLOv5-C3CBAM model is designed for universal object detection, the detection box includes only the coordinates of the center point and the width and height, while the actual captured images often have in-plane and out-of-plane tilt angles, which can make the rectangular detection box of the YOLOv5-C3CBAM have redundant background information. Furthermore, the tilt or inversion of the readings increases the difficulty of subsequent readings recognition. Therefore, in order to realize the area detection of digital meter readings under multiple shooting views and meet the requirements of high real-time and accuracy, the YOLOv5-C3CBAM object detection model is improved and the YOLO-CPDM is proposed, as shown in Fig. 4. It reconstructs the detection head and add the loss function of corner point detection, so that it can detect the positions of four key corner points in the upper left, upper right, lower left and lower right of the reading area, and then perform horizontal correction through perspective transformation to recognize the corrected horizontal reading area, and its accuracy is greatly improved.

YOLO-CPDM is an improved dense prediction network based on anchors, whose prediction object $\{\hat{o} \in \hat{\mathbf{O}}\}$ includes the location of the rectangular box in the reading area $\hat{\boldsymbol{b}} = (\hat{b}_x, \hat{b}_y, \hat{b}_w, \hat{b}_h)$, the prediction category $\hat{c}$ and the location of the four corner points $\hat{\boldsymbol{r}} = \{(\hat{x}_k, \hat{y}_k)\}_{k=1}^{4}$. The actual target $\{\boldsymbol{o} \in \mathbf{O}\}$ is obtained by extending YOLOv5 to include the bounding box $\boldsymbol{b} = (b_x, b_y, b_w, b_h)$, the category c, and the set of 4 corner points $\boldsymbol{r} = \{(x_k, y_k)\}_{k=1}^{4}$ in the reading area. $\boldsymbol{b}$, $\hat{\boldsymbol{b}}$, $\boldsymbol{r}$, and $\hat{\boldsymbol{r}}$ are all measured based on the dimensions of the original figure.

The input and output of the YOLO-CPDM model $\mathcal{N}$ prediction process can be expressed as:

$$\hat{\mathbf{G}} = \mathcal{N}(\mathbf{I}) \tag{1}$$

where the input of model $\mathcal{N}$ is a three-channel RGB image $\mathbf{I} \in \mathbb{R}^{h \times w \times 3}$, and the height and width of the image are h and w, respectively. Model $\mathcal{N}$ has three detection heads, so its output $\hat{\mathbf{G}}$ can be represented as a dense grid of three scales, i.e.:

$$\hat{\mathbf{G}} = \{\widehat{\mathcal{G}}^s | s \in \{8, 16, 32\}\} \tag{2}$$

where s denotes the downsampling multiplicity of the model $\mathcal{N}$, $\widehat{\mathcal{G}}^s \in \mathbb{R}^{\frac{h}{s} \times \frac{w}{s} \times N_a \times N_o}$ denotes the output of the detection head with downsampling multiplicity s, N_a is the number of preset anchor boxes for each output grid, $N_o = 3 \times (13 + N_c)$ is the number of output channels for each prediction grid, and N_c denotes the number of object classes.

Since the convolution process of the model is to downsample the input image, the coordinates (i, j) of the output grid cell $\widehat{\mathcal{G}}_{i,j}^s$ can be mapped to a small block $\mathbf{I}_{i,j}^s$ of the original image by multiplying it with the downsampling multiplier s, i.e.:

$$\mathbf{I}_{i,j}^s = \mathbf{I}_{s \times i:s \times (i+1), s \times j:s \times (j+1)} \tag{3}$$

Each output grid cell $\widehat{\mathcal{G}}_{i,j}^s$ has N_a different scales of anchor boxes preset, so:

$$\mathbf{A}^s = \{(A_{w_a}^s, A_{h_a}^s)\}_{a=1}^{N_a} \tag{4}$$

where $A_{w_a}^s$, $A_{h_a}^s$ are the width and height of each anchor measured based on the size of the grid graph, which is obtained by computing the width and height of the labeled boxes of the training set by the K-means [50] algorithm. Typically, $N_a = 3$.

$\widehat{\mathcal{G}}_{i,j}^s$ is responsible for detecting target objects $\boldsymbol{o}$ with center location (b_x, b_y) on $\mathbf{I}_{i,j}^s$, and assigns target $\boldsymbol{o}$ to an anchor $\widehat{\mathcal{G}}_{i,j,a}^s$ for training based

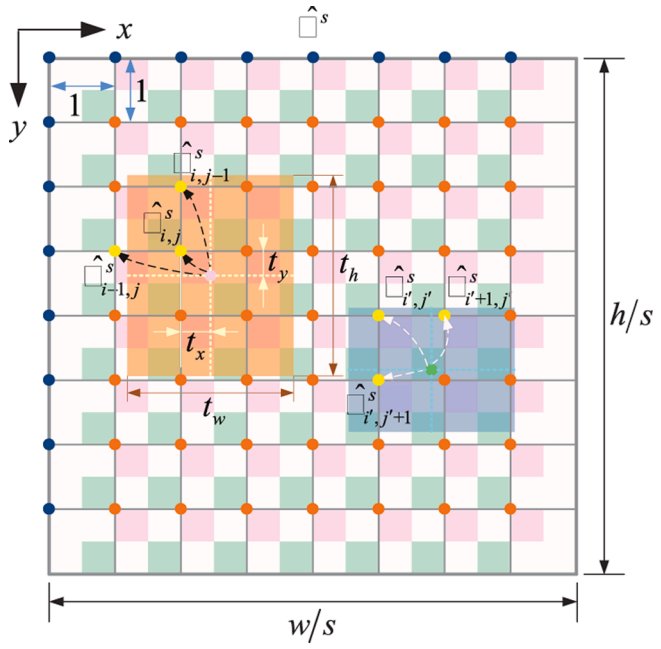

Fig. 5. YOLOv5 dense grid of augmentation training objectives.

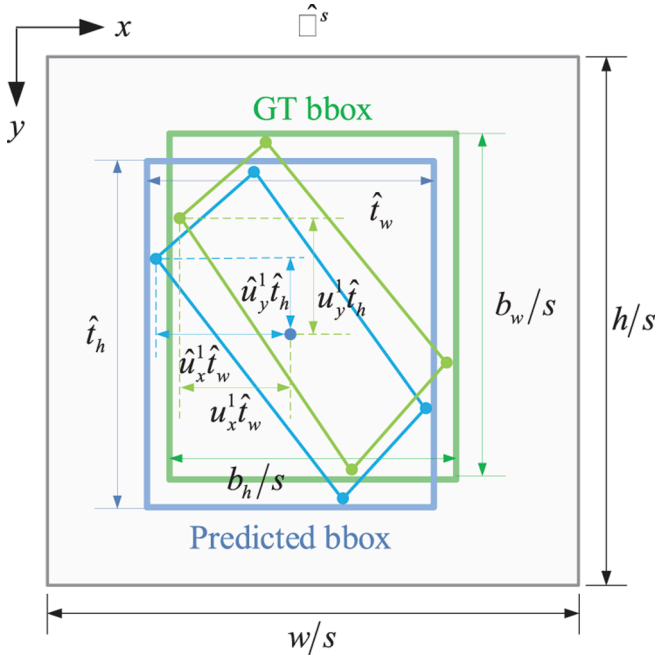

Fig. 6. Variable relationship of reconstructed corner detection model.

on the width-height matching tolerance of target o to the anchor. Since each grid (i,j) of $\widehat{\mathscr{G}}^s_{i,j}$ has N_a anchors of different widths and heights, it has the ability to detect targets of different types and scales. As shown in Fig. 5, we assign additional grids $\widehat{\mathscr{G}}^s_{i-1,j}$, $\widehat{\mathscr{G}}^s_{ij-1}$ or $\widehat{\mathscr{G}}^s_{i+1j}$, $\widehat{\mathscr{G}}^s_{ij+1}$ to $\widehat{\mathscr{G}}^s_{ij}$ and use the same anchor box to predict the target object with center position (b_x, b_y) on $\mathbf{I}_p$ to increase the number of training targets.

As shown in Fig. 6, the N_o-dimensional prediction result for each anchor box $\widehat{\mathscr{G}}^s_{i,j,a}$ is $(\hat{t}, \hat{p}_o, \hat{c}, \hat{u})$, where $\hat{t} = (\hat{t}_x, \hat{t}_y, \hat{t}_w, \hat{t}_h)$ is an unnormalized prediction bounding box, $(\hat{t}_x, \hat{t}_y)$ is the center coordinate of the prediction box relative to the grid coordinate (i,j), which is an unnormalized offset, $(\hat{t}_w, \hat{t}_h)$ is the unnormalized prediction box width and height, $\hat{p}_o$ is the confidence level that an object exists in the prediction box, $\hat{c} = (\hat{c}_1, \hat{c}_2, ..., \hat{c}_{N_c})$ is the score vector for object classification, and $\hat{u}' = \{(\hat{u}'^k_x, \hat{u}'^k_y)\}^4_{k=1}$ is the corner point coordinate of the unnormalized prediction.

According to Eqs. (5) ~ (8), normalizing $\hat{t}' = (\hat{t}'_x, \hat{t}'_y, \hat{t}'_w, \hat{t}'_h)$, we can obtain the prediction box $\hat{t} = (\hat{t}_x, \hat{t}_y, \hat{t}_w, \hat{t}_h)$ with respect to the grid (i,j), i.e.:

$$\hat{t}_x = 2\sigma(\hat{t}'_x) - 0.5 \tag{5}$$

$$\hat{t}_y = 2\sigma(\hat{t}'_y) - 0.5 \tag{6}$$

$$\hat{t}_w = (2\sigma(\hat{t}'_w))^2 \times A^s_{w_a} \tag{7}$$

$$\hat{t}_h = (2\sigma(\hat{t}'_h))^2 \times A^s_{h_a} \tag{8}$$

Based on the downsampling multiplier s and the grid coordinates (i,j), the prediction box $\hat{t} = (\hat{t}_x, \hat{t}_y, \hat{t}_w, \hat{t}_h)$ can be mapped to the corresponding position $\hat{b} = (\hat{b}_x, \hat{b}_y, \hat{b}_w, \hat{b}_h)$ on the original map, i.e.:

$$\hat{b}_x = s \times (\hat{t}_x + i) \tag{9}$$

$$\hat{b}_y = s \times (\hat{t}_y + j) \tag{10}$$

$$\hat{b}_w = s \times \hat{t}_w \tag{11}$$

$$\hat{b}_h = s \times \hat{t}_h \tag{12}$$

By normalizing $\hat{u}' = \{(\hat{u}'^k_x, \hat{u}'^k_y)\}^4_{k=1}$ according to Eqs. (13) ~ (14), the offsets $\hat{v} = \{(\hat{v}^k_x, \hat{v}^k_y)\}^4_{k=1}$ in the x- and y-axis directions relative to the center coordinates of the prediction box can be obtained, i.e.:

$$\hat{v}^k_x = (2\sigma(\hat{u}'^k_x) - 1) \times \hat{t}_w, k = 1,2,3,4 \tag{13}$$

$$\hat{v}^k_y = (2\sigma(\hat{u}'^k_y) - 1) \times \hat{t}_h, k = 1,2,3,4 \tag{14}$$

Combining the offsets $\hat{v} = \{(\hat{v}^k_x, \hat{v}^k_y)\}^4_{k=1}$, downsampling multiplicity s, prediction box center offset $(\hat{t}_x, \hat{t}_y)$ and grid coordinates (i,j), the predicted coordinates $\hat{r} = \{(\hat{x}_k, \hat{y}_k)\}^4_{k=1}$ of corner points on the original map can be calculated, i.e.:

$$\hat{x}_k = s \times (\hat{t}_x + i + \hat{v}^k_x), k = 1,2,3,4 \tag{15}$$

$$\hat{y}_k = s \times (\hat{t}_y + j + \hat{v}^k_y), k = 1,2,3,4 \tag{16}$$

3.2. Loss function of the YOLO-CPDM

Map the target boxes in the original image onto a dense grid, construct the set of targets $\boldsymbol{G}$, and use the joint optimization target $\mathscr{L}^{CD}(\hat{\boldsymbol{G}}, \boldsymbol{G})$ to learn the objects confidence $\hat{p}_o(\mathscr{L}^{CD}_{obj})$, predicted bounding boxes $\hat{t}(\mathscr{L}^{CD}_{box})$, classification scores $\hat{c}(\mathscr{L}^{CD}_{cls})$ and predicted corners $\hat{u}(\mathscr{L}^{CD}_{cor})$. Thus, the loss of a single image can be written as:

$$\mathscr{L}^{CD}_{obj} = \sum_{s\in\{8,16,32\}} \frac{w_s}{N(\mathscr{G}^s_{i,j,a} \in \widehat{\mathscr{G}}^s)} \sum_{\mathscr{G}^s_{i,j,a} \in \hat{\mathscr{G}}^s} \mathrm{BCE}(\hat{p}_o, p_o \cdot \mathrm{CIoU}(\hat{t}, t)) \tag{17}$$

$$\mathscr{L}^{CD}_{box} = \sum_{s\in\{8,16,32\}} \frac{1}{N(\boldsymbol{o} \in \widehat{\mathscr{G}}^s)} \sum_{\boldsymbol{o}\in\hat{\mathscr{G}}^s} 1 - \mathrm{CIoU}(\hat{t}, t) \tag{18}$$

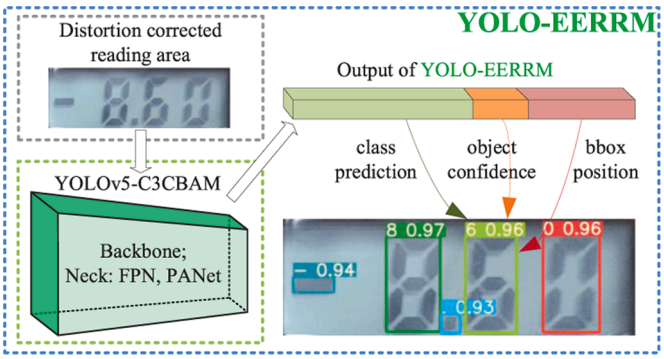

Fig. 7. Prediction results based on YOLO-EERRM.

Table 1
Software and hardware environment for experiments.

System and Hardware Configurations	Software and Environment Configurations
Operating System: Windows 11	Pycharm Version: 2022.3
CPU: 12th Gen Intel® Core™ i9-12900H	Python Version: 3.9.16
GPU: NVIDIA GeForce RTX 3070 Laptop GPU	CUDA Version: 11.3
RAM: 32 GB DDR5 4800 MHz	cuDNN Version: 8.3.2

$$\mathscr{L}_{cls}^{CD} = \sum_{s\in\{8,16,32\}} \frac{1}{N(\boldsymbol{o}\in\widehat{\mathscr{G}}^s)} \sum_{\boldsymbol{o}\in\widehat{\mathscr{G}}^s} \mathrm{BCE}(\hat{\boldsymbol{c}},\boldsymbol{c}) \tag{19}$$

$$\mathscr{L}_{cor}^{CD} = \sum_{s\in\{8,16,32\}} \frac{1}{N(\boldsymbol{o}\in\widehat{\mathscr{G}}^s)} \sum_{\boldsymbol{o}\in\widehat{\mathscr{G}}^s} smooth_{L_1}(\hat{\boldsymbol{u}},\boldsymbol{u}) \tag{20}$$

where w_s is the weight corresponding to the three dense grids, $N(*)$ is the number of conditions $*$ satisfied, BCE is the binary cross-entropy function, CIoU is the complete intersection over union [51], and $smooth_{L_1}$ is the same as that of [28].

When the anchor $\mathscr{G}_{i,j,a}^s$ is assigned to predict an object $\boldsymbol{o}$, $p_o = 1\cdot \mathrm{CIoU}(\hat{\boldsymbol{t}},\boldsymbol{t})$, when the anchor $\mathscr{G}_{i,j,a}^s$ is not assigned to an object $\boldsymbol{o}$, $p_o = 0$. Thus, the total loss $\mathscr{L}^{CD}$ is:

$$\mathscr{L}^{CD} = \lambda_{obj}^{CD}\mathscr{L}_{obj}^{CD} + \lambda_{box}^{CD}\mathscr{L}_{box}^{CD} + \lambda_{cls}^{CD}\mathscr{L}_{cls}^{CD} + \lambda_{cor}^{CD}\mathscr{L}_{cor}^{CD} \tag{21}$$

where λ_{obj}^{CD}, λ_{box}^{CD}, λ_{cls}^{CD} and λ_{cor}^{CD} are the weights of $\mathscr{L}_{obj}^{CD}$, $\mathscr{L}_{box}^{CD}$, $\mathscr{L}_{cls}^{CD}$, $\mathscr{L}_{cor}^{CD}$, respectively. It is also the hyperparameters of the YOLO-CPDM model, which need to be adjusted according to different datasets to optimize the performance of the model. In this paper, the appropriate range of parameters are ($\lambda_{obj}^{CD} = 0.5\sim1.5$, $\lambda_{box}^{CD} = 0.03\sim0.1$, $\lambda_{cls}^{CD} = 0.3\sim1.0$, $\lambda_{cor}^{CD} = 2.0\sim10.0$). The performance is great and insensitive under the given range of parameters. Parameter settings within this range typically prevent gradient explosion and disappearance. By default, we set $\lambda_{obj}^{CD} = 1.0$, $\lambda_{box}^{CD} = 0.05$, $\lambda_{cls}^{CD} = 0.5$, and $\lambda_{cor}^{CD} = 5$.

3.3. Reading region distortion correction

Before reading recognition, the rotating box of the reading area predicted by the YOLO-CPDM needs to be perspective transformed and corrected to a horizontal rectangular box, where the perspective transformation matrix can be expressed as:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \boldsymbol{A}\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{22}$$

where $\boldsymbol{A}$ is the perspective transformation matrix, $[x,y]^{\mathrm{T}}$ is the pixel coordinate of the point on the original image, $[x,y,1]^{\mathrm{T}}$ is the source target point, and $[X,Y,Z]^{\mathrm{T}}$ is the shifted target point.

Projecting the 3-D target points onto the new 2-D plane, this relationship can be described as:

$$X' = \frac{X}{Z} = \frac{a_{11}x + a_{12}y + a_{13}}{a_{31}x + a_{32}y + a_{33}} \tag{23}$$

$$Y' = \frac{Y}{Z} = \frac{a_{21}x + a_{22}y + a_{23}}{a_{31}x + a_{32}y + a_{33}} \tag{24}$$

$$Z' = \frac{Z}{Z} = 1 \tag{25}$$

3.4 Reading recognition method

3.4.1. YOLO-EERRM for digital meter

As shown in Fig. 7, the redundant background outside the reading region can be removed by perspective transformation distortion correction, and a horizontal reading region image can be obtained. Further, the reading recognition problem can then be transformed into a character detection problem within the reading region. Therefore, YOLOv5-C3CBAM detection can be directly employed to recognize individual characters.

3.4.2. Loss function of the YOLO-EERRM

Similar to the loss function of YOLO-CPDM, all annotated characters in the original image are mapped on a dense grid of YOLO-EERRM detection heads, and the joint optimization objective $\mathscr{L}^{RR}(\hat{\mathbf{G}},\mathbf{G})$ is used to learn objects confidence $\hat{p}_o(\mathscr{L}_{obj}^{RR})$, predicted bounding boxes $\hat{\boldsymbol{t}}(\mathscr{L}_{box}^{RR})$, and classification scores $\hat{\boldsymbol{c}}(\mathscr{L}_{cls}^{RR})$, so that the loss value of a single image can be written as:

$$\mathscr{L}_{obj}^{RR} = \sum_{s\in\{8,16,32\}} \frac{w_s}{N(\mathscr{G}_{i,j,a}^s\in\widehat{\mathscr{G}}^s)} \sum_{\mathscr{G}_{i,j,a}^s\in\widehat{\mathscr{G}}^s} \mathrm{BCE}(\hat{p}_o, p_o\cdot\mathrm{CIoU}(\hat{\boldsymbol{t}},\boldsymbol{t})) \tag{26}$$

$$\mathscr{L}_{obj}^{RR} = \sum_{s\in\{8,16,32\}} \frac{w_s}{N(\mathscr{G}_{i,j,a}^s\in\widehat{\mathscr{G}}^s)} \sum_{\mathscr{G}_{i,j,a}^s\in\widehat{\mathscr{G}}^s} \mathrm{BCE}(\hat{p}_o, p_o\cdot\mathrm{CIoU}(\hat{\boldsymbol{t}},\boldsymbol{t})) \tag{27}$$

$$\mathscr{L}_{cls}^{RR} = \sum_{s\in\{8,16,32\}} \frac{1}{N(\boldsymbol{o}\in\widehat{\mathscr{G}}^s)} \sum_{\boldsymbol{o}\in\widehat{\mathscr{G}}^s} \mathrm{BCE}(\hat{\boldsymbol{c}},\boldsymbol{c}) \tag{28}$$

Therefore, the total loss $\mathscr{L}^{RR}$ is the weighted sum of the above components, i.e.:

$$\mathscr{L}^{RR} = \lambda_{obj}^{RR}\mathscr{L}_{obj}^{RR} + \lambda_{box}^{RR}\mathscr{L}_{box}^{RR} + \lambda_{cls}^{RR}\mathscr{L}_{cls}^{RR} \tag{29}$$

where λ_{obj}^{RR}, λ_{box}^{RR} and λ_{cls}^{RR} are the weights of $\mathscr{L}_{obj}^{RR}$, $\mathscr{L}_{box}^{RR}$, and $\mathscr{L}_{cls}^{RR}$, respectively. It is also the hyperparameters of the YOLO-EERRM model, which need to be adjusted to optimize the performance of the model. In this paper, the appropriate range of parameters are ($\lambda_{obj}^{RR} = 0.5\sim1.5$, $\lambda_{box}^{RR} = 0.03\sim0.1$, $\lambda_{cls}^{RR} = 0.3\sim1.0$). By default, we set $\lambda_{obj}^{RR} = 1.0$, $\lambda_{box}^{RR} = 0.05$, and $\lambda_{cls}^{RR} = 0.5$.

4. Experiments

The experimental environment of this study is using Windows 11 operating system, the laptop graphics card model is RTX 3070, which is equipped with CUDA acceleration of version 11.3, using Pycharm IDE for project development, the detailed configuration of the experimental environment is shown in TABLE 1.

4.1. Introduction to the dataset

Datasets are crucial for establishing an important result in the

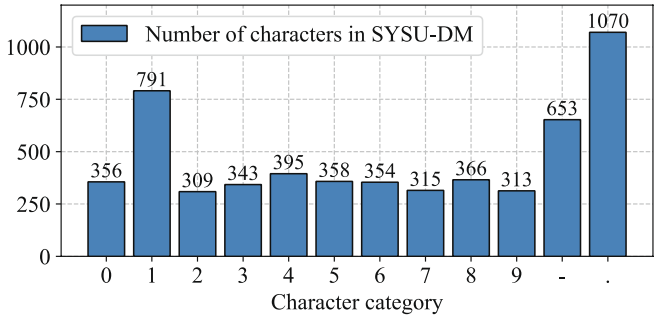

Fig. 8. Histogram of the number of characters in the SYSU-DM dataset.

Table 2
The number of expansions of the three training datasets and the amount of data before and after the expansion.

Dataset	Number before augmentation	synthetic times per image	Number after augmentation
SYSU-DM	753	20	15,813
Water meter	353	30	10,943
Gas meter	102	100	10,302

research. However, most digital instrument reading datasets are private [12] and not yet publicly available. Therefore, we took and collected images from scratch to form the SYSU-DM dataset. In addition, search engines were fully utilized to obtain some publicly available and suitable open-source dataset images to expand the dataset. Ultimately, three datasets were adopted for this study, namely SYSU-DM, Water meter, and Gas meter. These datasets include commonly used mechanical and electronic digital meters to ensure data diversity and reliability.

4.2. SYSU-DM dataset

Considering that the low-cost and flexible Arduino platform is sufficient to meet the needs of data collection and rapid development, it has been used to produce the SYSU-DM dataset which contains 1255 images of digital multimeter readings. The main steps to create the dataset are as follows. First, the Arduino control program was written using the Arduino IDE development tool to simulate the analog signal output of its PWM pin. Then, the pin output was measured using various stops of the DL8490 multimeter with different ranges of DC/AC voltage and DC/AC current. Next, the meter was photographed at different shooting angles and different shooting distances using the smart terminal device. Further, the captured images were labeled with corner points and characters using labeling software and auxiliary scripts such as Labelme and LabelImg. Finally, we randomly divided 753 training images, 251 validation images and 251 test images according to the ratio of 3:1:1.

In the SYSU-DM dataset, the meter reading display area is a seven-segment digital tube. The readings have the numbers 0 ~ 9, negative sign "-" and decimal point "." There are 12 types of characters, and the statistical results of the number of each type of characters are shown in Fig. 8. Since the SYSU-DM dataset is rich in data styles, this paper selects the SYSU-DM dataset to verify the performance of YOLO-CPDM and YOLO-EERRM.

4.3. Water meter dataset

This dataset was collected by multiple users in different locations and natural environments and contains 1244 images of hot or cold-water meters [52] with exactly one meter counter on each image. The dataset is equipped with an annotation file that records the (x, y) coordinates of each meter counter vertex in order. Since the YOLO-CPDM model was designed to detect the four corner points of the reading area, we filtered 505 images labeled with four corner points and their corresponding labels from this dataset (the rest of the data with more than four labeled points are not considered), and divided 353 images for training and 152 images for validation according to 7:3. In the Water meter dataset, the meter reading display area is a mechanical counter. Since the dataset contains different types of water meters with different structures, the amount of data for each character in each water meter reading is too small to train the YOLO-EERRM model for reading recognition,

Fig. 9. Offline data enhancement results for three datasets.

Table 3
Ablation experimental results of YOLO-CPDM.

Method		Ablation variables				#Params. (M)	Metrics			
YOLO-CPDM	Exp1	–	–	–	–	7.1	81.0	78.9	68.0	5.4
	Exp2	√	–	–	–	7.1	91.1	96.4	87.5	5.4
	Exp3	√	√	–	–	7.1	92.6	98.4	90.5	5.3
	Exp4	√	–	√	–	12.4	88.1	89.2	81.3	5.6
	Exp5	√	√	√	–	12.4	88.7	91.6	82.5	5.7
	Exp6	√	–	–	C3CA	6.5	91.6	96.6	88.2	5.6
	Exp7	√	–	–	C3CBAM	6.4	92.3	97.6	89.6	5.5
	Exp8	√	–	–	C3SE	6.4	91.6	96.6	87.9	5.5
	Exp9	√	–	–	C3ECA	6.4	92.1	97.2	89.9	5.6
	Exp10	√	√	–	C3CA	6.5	93.8	97.0	91.5	5.3
	Exp11	√	√	–	C3CBAM	6.4	**94.1**	**98.6**	**93.4**	**5.3**
	Exp12	√	√	–	C3SE	6.4	93.4	98.0	91.5	5.4
	Exp13	√	√	–	C3ECA	6.4	93.2	97.4	90.8	5.5

$mAP50^{:95}$.

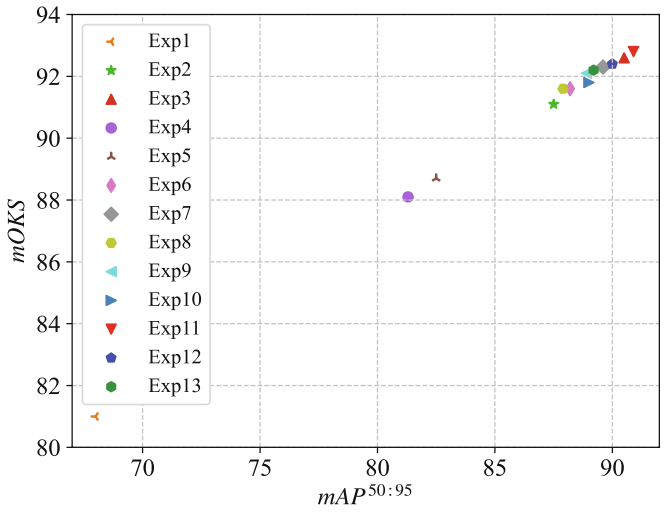

Fig. 10. The scattergram of $mAP50^{:95}$ and mOKS in Exp 1 ∼ Exp 13.

therefore, this paper only uses this dataset to validate the YOLO-CPDM model.

4.4. Gas meter dataset

The dataset Gas-Meter-Counter [53], consists of 102 training images and 51 testing images. All images were acquired by different operators using different typology of devices. The images in the dataset were collected under conditions of occlusion, different lighting conditions, blurring and perspective distortion. Under these conditions, the detection and segmentation problem become very difficult. In the Gas meter dataset, the meter reading display area is a mechanical counter. Similarly, since the amount of data for each character in the dataset with different types of structured gas tables is too small to train the YOLO-EERRM model for reading recognition. So, this paper only uses this dataset to validate the YOLO-CPDM model.

5. Offline data enhancement for YOLO-CPDM

Since the digital meter reading data set is captured in a realistic limited scene, but the smart meter reading system needs to be applicable to different conditions, such as different angles, orientations, positions, scaling, and luminance. Therefore, to mitigate the impact of noise and uncertainties, and to enhance the robustness of the model, additional data is synthesized through data augmentation to improve the performance of the model in real-world applications.

In addition, capturing and labeling the digital meter reading dataset is labor intensive, and even then, it is difficult to obtain large datasets; therefore, the original dataset needs to be augmented by offline data enhancement method. The offline data augmentation includes luminance enhancement, random blurring, panning, rotation, flipping, perspective transformation, and scaling. The data enhancement script fully automatically transforms the original image and the corresponding labels to fit the input/output of the YOLO-CPDM model, which eliminates the need to manually label the enhanced image and can theoretically acquire an infinite number of synthetic data. In addition, the script can automatically determine whether the four corner points of the enhanced reading area are outside the image, and thus can discard and recompose the image. Specifically, the probability that a picture is brightness enhanced and randomly blurred is 0.3, the probability that it is panned, rotated, flipped, perspective transformed and scaled is 0.7, and the image is transformed at least once, and the RGB value of the background fill is set to (114, 114, 114). We generate multiple synthetic images for each image in the training set, so that the training set can be expanded to a multiple of the original one, and the amount of expanded data is shown in Table 2.

Fig. 9 shows a sample of offline data augmentation for each of the three datasets, including luminance enhancement, random blurring, panning, rotation, flipping, perspective transformation, and scaling. Synthesizing additional data using data augmentation techniques can greatly improve the generalization ability of the model. We label each of the four corner points clockwise in the reading area and draw the serial numbers of the corner points.

5.1. Experimental results of the YOLO-CPDM

5.1.1. Ablation study

After training the YOLO-CPDM model on the augmented training set, we test it on the test set. The test set contains images with different illumination conditions, different shooting tilt angles and different shooting distances. The hyperparameters of the experiment are set to batch = 8, epoch = 300, and the stochastic gradient descent (SGD) optimizer is used with a momentum value of 0.937, a weight decay value of 0.0005, an initial learning rate $lr_0 = 0.01$, a final learning rate $lr_f = 0.0001$, and a warmup and cosine learning rate decay strategy. The variables of the ablation experiment include data enhancement, dynamic loss, increasing detection head and attention mechanism.

For the speed test, we recorded the total latency for input image preprocessing, model prediction and non-maximum suppression postprocessing. For accuracy tests, we used the commonly keypoint detection evaluation metric, called object keypoint similarity (OKS) [54], i.e.,

$$\mathrm{OKS}_m = \frac{\sum_i \exp(-d_{mi}^2/2s_m^2\sigma_i^2)\delta(v_{mi} > 0)}{\sum_i \delta(v_{mi} > 0)} \tag{30}$$

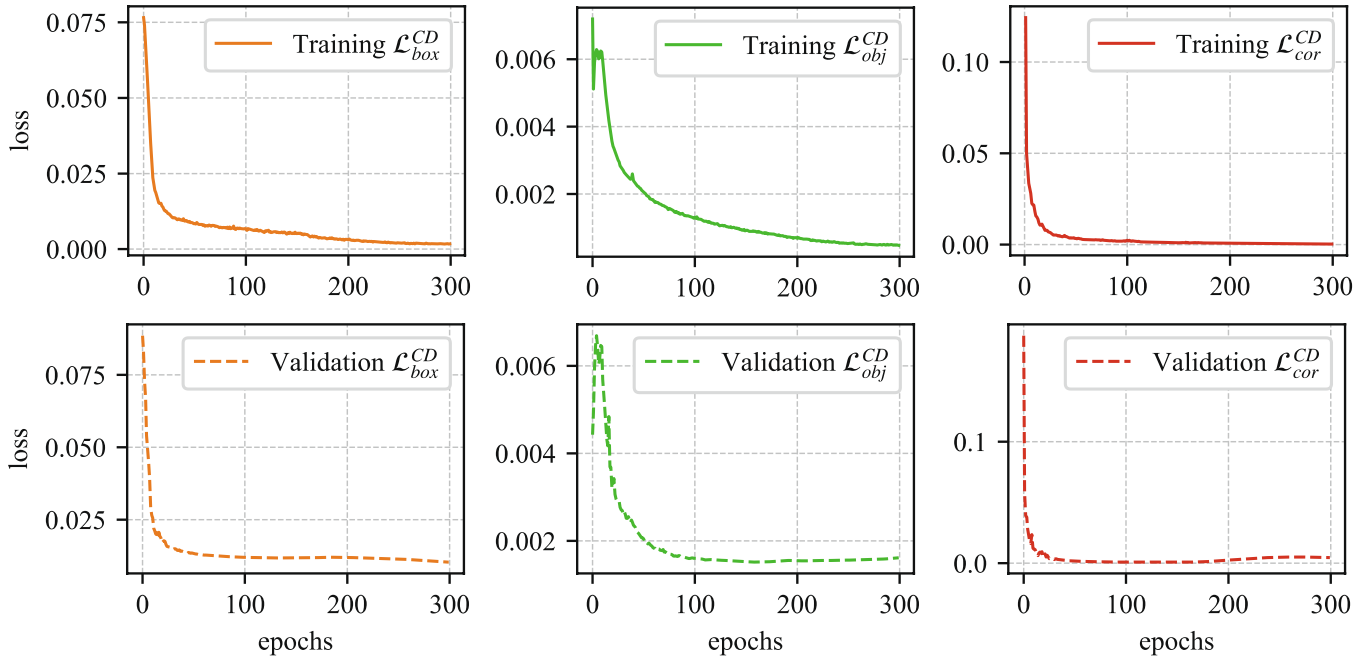

Fig. 11. YOLO-CPDM-based model training and reading area detection loss curve with Exp 11.

Table 4
Comparison results of YOLO-CPDM with different methods.

Method	Size (H & W)	#Params. (M)	SYSU-DM		Water meter		Gas meter		Lat. (ms)
HRNet-W32 [54]	512	28.5	82.8	70.8	51.1	24.5	50.9	25.2	79.4
	640	28.5	87.0	78.8	54.7	30.4	55.0	32.8	112.4
HRNet-W48 [54]	512	63.6	81.7	68.5	47.0	22.2	46.8	22.1	106.4
	640	63.6	85.9	76.9	52.7	28.9	53.7	28.8	169.5
HigherHRNet-W32 [56]	512	28.6	89.8	85.2	60.5	40.0	53.5	28.5	80.6
	640	28.6	91.2	87.1	62.7	44.4	56.3	34.9	126.6
HigherHRNet-W48 [56]	512	63.8	91.1	86.9	61.9	42.0	54.0	33.5	117.6
	640	63.8	92.6	90.6	64.0	46.3	57.9	38.2	172.4
YOLO-CPDM (Ours)	512	6.4	93.1	91.5	65.4	48.6	57.7	37.9	**4.5**
	640	6.4	**94.1**	**93.4**	**67.6**	**50.4**	**60.2**	**39.3**	5.3

mAP50:95.
mAP50:95.

where m denotes the id of the target, i denotes the id of the keypoint, d_{mi} denotes the Euclidean distance between the predicted and true positions of the ith keypoint of target m, s_m is the pixel area of target m. σ_i is the normalization factor, it is used to describe the difficulty of labeling each keypoint, and is set to 0.03 for all 4 vertices of the reading area. v_{mi} denotes the visibility of the keypoint of the target, if the keypoint is visible or occluded, then $v_{mi} > 0$; if the keypoint is outside the picture range, $v_{mi} = 0$. We filter the dataset so that all the corner points of the reading area are within the range of the picture (i.e., $v_{mi} > 0$).

Further, the above OKS can be simplified as:

$$\mathrm{OKS}_m = \sum_i \exp(-d_{mi}^2/0.0018s_m^2) \tag{31}$$

According to (31), the mean OKS (mOKS) can be calculated for all targets in the test dataset. In order to improve the performance index of the model, the following dynamic loss function can be established, namely:

$$s_{L_1}(x,y) = \begin{cases} \dfrac{0.5\times(x-y)^2}{\beta}, & \text{if } |x-y| < \beta \\ |x-y| - 0.5\times\beta, & \text{otherwise} \end{cases} \tag{32}$$

where $x \in \mathbb{R}$ is the predicted value, $y \in \mathbb{R}$ is the true value, and β is the set dynamic loss update variable, specified as:

$$\beta = \begin{cases} 1, 0 \leqslant e_p < 100 \\ 0.75, 100 \leqslant e_p < 150 \\ 0.5, 150 < e_p \end{cases} \tag{33}$$

Based on OKS, we report the mAP50:95 and mAP75 of the standard precision and recall scores, as shown in Table 3. From Exp 1 and Exp 2, it can be seen that the data enhancement method can synthesize sample data under different angles, orientations, positions, scaling ratios and luminance conditions, which makes the model generalization ability significantly improved. From Exp 2 and Exp 3, it can be seen that the established loss function s_{L_1} can improve the performance index of the model. In fact, with the convergence of iterations, the total loss value $\mathscr{L}^{CD}$ keeps decreasing and the model tends to be stable, when decreasing the β parameter of the loss function can make the loss value increase and the gradient of model back propagation increases to make it further trained.

Add Head in Exp 4 and Exp 5 refers to the adoption of the yolov5s6 [55] model, which has four detection heads. As seen from Exp 2 and Exp 4, the performance of the model decreases with more detection heads.

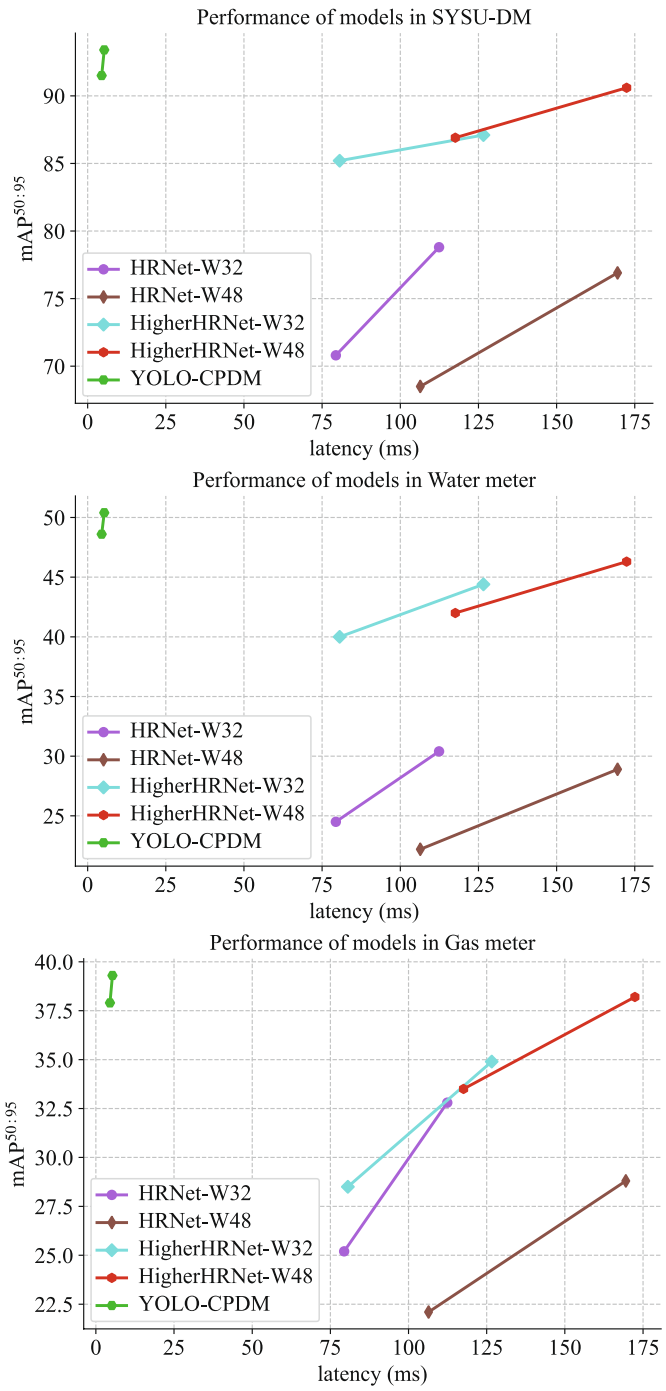

Fig. 12. mAP50:95 vs latency (ms) in three datasets.

The reason is that yolov5s6 is pretrained on high-resolution images, it is unsuitable for images with relatively low resolution. In addition, higher resolution requires more training resources and real-time performance will deteriorate. Exp 6 ∼ Exp 9 compare the effect of adding different attention mechanisms, adding attention mechanisms can make the model focus on information that is more important to the current task and reduce the attention to other information, which makes the mOKS, mAP scores of the model further improved. From Exp 10 to Exp 13, the combination of offline data augmentation, dynamic loss, and attention mechanism can lead to greater performance improvement. Compared with other attention modules, the C3CBAM attention module in Exp 11 improves the performance the most, with mOKS improving by 13.1 points (16.2 %), mAP75 improving by 19.7 points (25.0 %), and mAP50:95 improving by 25.4 points (37.4 %) compared to Exp 1.

To more clearly demonstrate the performance differences of the ablation experiments, we plotted the scatter plots of mAP50:95 and mOKS for Exp 1 ∼ Exp 13. As shown in Fig. 10, the proximity of the points in the figure to the upper right corner indicates a higher degree of accuracy for the method, while those further away from this region exhibit lower accuracy. Therefore, it shows that Exp 11 has the highest accuracy among all ablation experiments.

The loss curves of $\mathscr{L}_{box}^{CD}$, $\mathscr{L}_{obj}^{CD}$, and $\mathscr{L}_{cor}^{CD}$ during its training and validation are shown in Fig. 11. Since YOLO-CPDM only needs to detect one category, the reading region, and no other detection categories, there is no $\mathscr{L}_{cls}^{CD}$. As shown in Fig. 11, the losses of training and validation are decreasing, and the total loss reaches the lowest at 160 epochs, indicating that the model has converged normally and achieved good performance.

5.1.2. Comparison study

We compare the detection accuracy and inference delay of YOLO-CPDM with the current SOTA key-point detection model [54,56] on three datasets (i.e., SYSU-DM, Water meter, and Gas meter), and the experimental results of the comparison are shown in Table 4.

The results of inference delay and mAP50:95 scores for each group of comparison experiments are shown in Fig. 12. The larger the ordinate of the point in the figure, the higher the positioning accuracy of this method. In addition, the smaller the abscissa of the point, the better the real-time performance of this method.

From the comparison experiments, it can be seen that the YOLO-CPDM model has the highest mOKS and mAP50:95 scores on the three datasets. mOKS scores reach 94.1 and mAP50:95 scores reach 93.4 on the SYSU-DM dataset, and the inference latency is only 5.3 ms, which is much lower than the state-of-the-art (SOTA) key-points detection model named HigherHRNet-W48 [56], which performed second only to YOLO-CPDM. Compared to HigherHRNet-W48, YOLO-CPDM improved mOKS by 1.5 points (1.6 %) and mAP50:95 by 2.8 points (3.1 %) on the SYSU-DM dataset, mOKS by 3.6 points (5.6 %) and mAP50:95 by 4.1 points (8.9 %) on the Water meter dataset, and mOKS by 2.3 points (4.0 %) and mAP50:95 by 1.1 points (2.9 %) on the Gas meter dataset. In addition, the YOLO-CPDM-512 model is slightly less accurate than the YOLO-CPDM-640 model due to the reduced resolution of the input image, but it is 38.3 times faster than HigherHRNet-W48, and YOLO-CPDM-640 is 32.5 times faster than HigherHRNet-W48. The experimental results show that YOLO-CPDM has high accuracy and real-time performance and can be easily embedded into mobile devices.

The detection results of YOLO-CPDM on the three data sets are shown in Fig. 13, which shows that the model can accurately detect the top-left, top-right, bottom-right, and bottom-left corner points of the digital meter reading area in sequence. All corner points and sequential markers plotted in the figure are generated by the YOLO-CPDM detection program YOLO-CPDM. Experimental Results of the YOLO-EERRM.

5.2. Ablation study

Before validating the YOLO-EERRM model, we cleaned the dataset by removing the blurred images that could not be recognized by human eyes, and finally, 702, 228 and 226 images were used for training, validation and testing, respectively. The hyperparameters in the experiments are set to batch = 8, epoch = 300, and optimized by SGD optimizer with warmup and cosine learning rate decay strategies, whose initial learning rate lr_0 is set to 0.01 and the size of the input image is 640 × 640. We compare the models with different final learning rates and different attention mechanisms, and the results of the ablation experiments are shown in Table 5.

As seen from the above table, the mAP50 and F1-score scores of all groups of experiments are close to 100, and the inference delay is only

Fig. 13. Detection results of YOLO-CPDM on three datasets.

Table 5
Ablation study for the YOLO-EERRM.

Method		Ablation variables		#Params. (M)	Metrics			
		lr_f	Attention Mechanism		mAP50	mAP50$^{:95}$	F1-score	Lat.(ms)
YOLO-EERRM	Exp1	0.002	–	7.1	99.8	91.2	99.9	4.2
	Exp2	0.0002	–	7.1	99.9	91.4	99.9	4.2
	Exp3	0.002	C3CA	6.5	99.9	91.6	99.9	3.9
	Exp4	0.0002	C3CA	6.5	99.9	91.8	99.9	3.9
	Exp5	0.002	C3SE	6.4	99.9	91.3	99.9	3.8
	Exp6	0.0002	C3SE	6.4	99.9	91.8	99.9	3.8
	Exp7	0.002	C3ECA	6.4	99.9	91.5	99.9	3.8
	Exp8	0.0002	C3ECA	6.4	100	92.2	100	3.8
	Exp9	0.002	C3CBAM	6.4	99.9	91.9	99.9	3.8
	Exp10	0.0002	C3CBAM	6.4	**100**	**92.4**	**100**	**3.8**

Table 6
AP^{75} scores for each type of character on the SYSU-DM dataset for ablation study.

Class	AP^{75}									
	Exp1	Exp2	Exp3	Exp4	Exp5	Exp6	Exp7	Exp8	Exp9	Exp10
0	100	100	100	100	100	100	100	100	100	100
1	100	100	100	100	100	100	100	100	100	100
2	100	100	100	100	100	100	100	100	100	100
3	100	100	100	100	100	100	100	100	100	100
4	100	100	99.3	100	100	100	99.3	100	100	100
5	100	100	100	100	100	100	100	100	100	100
6	100	100	100	100	100	100	100	100	100	100
7	100	100	100	100	100	100	100	100	100	100
8	100	100	100	100	100	100	100	100	100	100
9	100	100	100	100	100	100	100	100	100	100
–	100	100	100	100	100	100	100	99.6	100	100
–	96.1	97.4	96.7	97.6	97.1	98.2	99.0	98.9	98.2	99.4

about 4 ms, which indicates that the YOLO-EERRM model has very good accuracy and real-time performance. In addition, as seen by Exp1 vs. Exp 2, Exp 3 vs. Exp 4, Exp 5 vs. Exp 6, Exp 7 vs. Exp 8, and Exp 9 vs. Exp 10, the final learning rate $lr_f = 0.0002$ performs better compared to $lr_f = 0.002$. The mAP50, mAP50$^{:95}$, and F1-score are all improved, indicating that reducing the final learning rate enables the model to converge better to the global optimum. Further, it can be seen from Exp 1, Exp 2 and the rest of the experiments that the introduction of the attention mechanism can improve the accuracy of the model in detecting the real target with the same final learning rate, thus improving the performance of the model. As can be seen from Exp10, YOLO-EERRM with C3CBAM attention mechanism has the optimal performance compared to Exp1, with an improvement of 1.2 points (1.3 %) in mAP50$^{:95}$.

In addition, due to the powerful generalization performance of

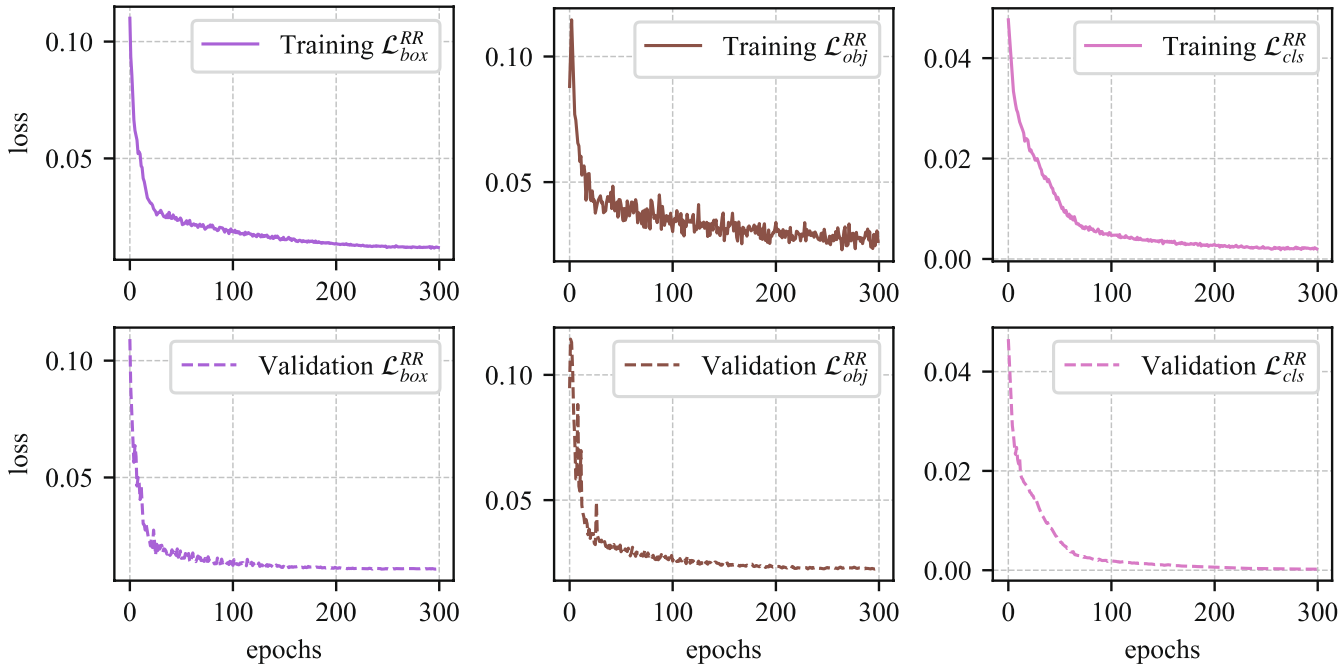

Fig. 14. YOLO-EERRM-based model training and readings recognition loss curve with Exp 10.

Table 7
Comparison results of YOLO-EERRM with different methods. (▲ indicates that this method uses multi-scale training strategy).

Method	Backbone	Size ($H \times W$)	Recognition accuracy (%)	#Params. (M)	Lat. (ms)
CRNN [32]	The same as [32]	32 × 100	96.9	7.8	5.8
Faster R-CNN [29]	ResNet50 [58]	315 × 960	98.2	28.4	25.1
Fcos [30]		320 × 992	96.9	32.1	33.6
CenterNet [31]		512 × 512	98.2	32.7	24.5
▲ YOLOv3 [26]	DarkNet53 [26]	448 × 448	97.8	61.6	19.1
EfficientDet-D0 [57]	EfficientNet [59]	512 × 512	96.0	3.8	22.7
EfficientDet-D1 [57]		640 × 640	97.8	6.6	29.3
EfficientDet-D2 [57]		768 × 768	99.1	8.0	34.8
▲ YOLO-EERRM-448	YOLOv5-C3CBAM-backbone	448 × 448	**100**	6.4	**3.1**
▲ YOLO-EERRM-640		640 × 640	**100**	6.4	3.8

YOLO-EERRM, the AP^{50} scores of all kinds of test results are basically 100, so we use the more stringent AP^{75} index to test the performance of the model, and the AP^{75} scores of all kinds are shown in Table 6.

As can be seen from the table, the AP^{75} scores of all types of experiments in each group are close to 100, which verifies the reliability of YOLO-EERRM in reading recognition, although there are still some individual not yet full scores, but this does not affect us to get the correct reading recognition results, because we only need to obtain their approximate positions to get the correct reading recognition results. Therefore, in practical applications, we can get the correct readings by adjusting the confidence threshold and IoU threshold as well as the post-processing method of eliminating the overlapping boxes of similar positions.

Exp10 has the highest accuracy and speed in all ablation experiments, and its loss curves of $\mathcal{L}_{box}^{RR}$, $\mathcal{L}_{obj}^{RR}$, and $\mathcal{L}_{cls}^{RR}$ during training and validation are shown in Fig. 14. As can be seen from the figure, each loss of training and validation is decreasing, indicating that the model has converged normally and achieved good performance. Finally, the model with the highest score of $mAP50^{:95}$ in the validation set is taken as the optimal model.

5.3. Comparison study

We compare the accuracy and inference latency of YOLO-EERRM with the current mainstream models (i.e., OCR model [32], two-stage model [29], single-stage model [26,57], and anchor free model [30,31]) for reading recognition on the SYSU-DM dataset. Here, the CRNN [32] model is able to output the complete end-to-end reading results, and the other methods are all object detection models, and the complete reading recognition results can be obtained by sorting the x-axis coordinates of the character detection box in increasing order, and their comparative experimental results are shown in Table 7.

In the comparison experiments, the backbone of YOLOv3 was built using the darknet53 [26] network, and the backbone of CRNN was built according to Ref. [32]. For a fairer comparison experiment, the backbone of Faster R-CNN, Fcos, and CenterNet used ResNet50 [58]. the backbone of EfficientDet used EfficientNet [59], and the backbone of EfficientDet D1 ∼ EfficientDet D3 with increasing number of model parameters and input image size, the inference speed gradually becomes slower, but the accuracy also gradually improves. Both YOLO-EERRM-448 and YOLO-EERRM-640 achieve 100 % accuracy on the SYSU-DM dataset. In addition, the inference speed delay of YOLO-EERRM-448 is only 3.1 ms, proving its excellent operation speed and recognition accuracy.

5.4. Performance of digital meter recognition method

To obtain complete readings identification results, we integrated YOLO-CPDM, perspective transformation and YOLO-EERRM into an

Fig. 15. Partial detection results of the SYSU-DM dataset.

Table 8
Accuracy and average latency of readings recognition.

Dataset	Correct amount / Total amount	Recognition accuracy (%)	Lat.(ms)
SYSU-DM	225 / 226	99.6	8.6

end-to-end pipeline for digital meter reading recognition. Fig. 15 shows the reading recognition results of the SYSU-DM test images. YOLO-CPDM can precisely locate the corner points of the reading area and mark the reading area in the detection result with the numbers $0 \sim 3$. By perspective transformation of the four corner points in the upper left, upper right, lower right and lower left corners, YOLO-EERRM can accurately identify each character in the reading area, and the final identification results can be combined according to the ordering of the x-coordinate of the center of the character envelope box.

The test performance of the system is shown in Table 8, which comprehensively records the reading recognition accuracy and time consumption of the system on the SYSU-DM test images. Statistically, the complete digital meter reading recognition method has 99.6 % accuracy and 8.6 ms inference latency. The experimental results prove that the proposed digital meter reading recognition system can not only adapt to complex scenes with arbitrary shooting angles and distances, but also has high accuracy and real-time performance.

6. Conclusion

For the intelligent scenario needs such as data monitoring and billing system, this paper proposes an autonomous recognition method for digital meter readings in natural scenes, which solves the recognition problem of multi-angle and multi-scale digital meter readings in complex environments. A robust and effective YOLO-CPDM is proposed by reconstructing the detection head, and its localization accuracy is improved by embedding an attention mechanism module, dynamic loss function, and offline enhancement techniques. Next, a perspective transformation is applied to geometrically correct the corner point coordinates of the distorted reading area. Further, YOLO-EERRM is used to identify the characters in the reading area. Finally, all processes are integrated into an end-to-end pipeline for complete readings identification results.

Experiments have been conducted on three datasets containing mechanical and electronic digital meters to evaluate the practicability and robustness of the proposed method. The experimental results show that the method has certain advantages, reflecting in the following aspects:

1. The proposed YOLO-CPDM can solve the problem of image distortion caused by tilted shooting angles and multi-variable scales. As shown in Table 4, YOLO-CPDM achieves the highest accuracy and lowest inference latency on three datasets compared with the existing methods.
2. The proposed YOLO-EERRM can solve the problem of inaccurate recognition due to errors in character segmentation of two-step methods. As shown in Table 7, YOLO-EERRM exhibits optimal accuracy and real-time performance.
3. The proposed integrated pipeline can solve the problem of real-time detection and accurate readings in complex environments. The integrated pipeline has 99.6 % accuracy and only 8.6 ms inference latency on the SYSU-DM test images.

In addition, this paper has the limitation that capturing and labeling the dataset is labor intensive, and even then, it is difficult to obtain large datasets.

In future work, we intend to research Generative Adversarial Networks (GANs) methods for generating digital meter reading images, addressing the challenges of acquiring large datasets and rectifying imbalanced categories. This will further enhance the detection and recognition performance in extremely difficult samples. Additionally, we aim to undertake industrial application work, including application development and model deployment. For example, using cellular networks such as 4G or 5G to transmit drone telemetry data to the server in real-time for processing and storage.

CRediT authorship contribution statement

Jianqing Peng: Conceptualization, Methodology, Supervision. **Wei Zhou:** Writing – original draft, Software. **Yu Han:** Writing – review & editing. **Mengtang Li:** Writing – review & editing, Data curation. **Wanquan Liu:** Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgment

This work was supported in part by the National Key R&D Program of China under Grant 2022YFB4703103, the National Natural Science Foundation of China under Grant 62103454, and the Shenzhen Science and Technology Program under Grant JCYJ20220530150006014.

References

[1] Y. Liu, J. Liu, Y. Ke, A detection and recognition system of pointer meters in substations based on computer vision, Measurement 152 (2020).
[2] G. Salomon, R. Laroca, D. Menotti, Image-based automatic dial meter reading in unconstrained scenarios, Measurement 204 (2022).
[3] L. Hou, S. Wang, X. Sun, et al., A pointer meter reading recognition method based on YOLOX and semantic segmentation technology, Measurement 218 (2023).
[4] F. Martinelli, F. Mercaldo, A. Santone, Water meter reading for smart grid monitoring, Sensors 23 (1) (2022) 75.
[5] P. Van Aubel, E. Poll, Smart metering in The Netherlands: What, how, and why, Int. J. Elect. Power Energy Syst. 109 (Jul. 2019) 719–725.
[6] J.D. Hmielowski, A.D. Boyd, G. Harvey, J. Joo, The Social Dimensions of Smart Meters in the United States: Demographics, Privacy, and Technology Readiness, Energy Research and Social Science 55 (2019) 189–197.
[7] Q. Hong, et al., Image-based automatic watermeter reading under challenging environments, Sensors 21 (2) (2021) 434.
[8] W. Li, S. Wang, I. Ullah, X. Zhang, J. Duan, Multiple attention-based encoder-decoder networks for gas meter character recognition, Scientific Reports 12 (1) (2022) 10371.
[9] K. Kanagarathinam, K. Sekar, Text detection and recognition in raw image dataset of seven segment digital energy meter display, Energy Reports 5 (2019) 842–852.
[10] M. Imran, H. Anwar, M. Tufail, A. Khan, M. Khan, D. Athiar Ramli, Image-Based Automatic Energy Meter Reading Using Deep Learning, Computers, Materials & Continua 74 (1) (2023) 203–216.
[11] K. Koščević and M. Subašić, "Automatic visual reading of meters using deep learning", In: Proc. Croatian Comput. Vis. Workshop, 2018, pp. 1–6.
[12] R. Laroca, et al., Towards image-based automatic meter reading in unconstrained scenarios: A robust and efficient approach, IEEE Access 9 (2021) 67569–67584.
[13] V. Moysiadis, P. Sarigiannidis, V. Vitsas, A. Khelifi, Smart farming in Europe, Comput. Sci. Rev. 39 (2021).
[14] L. Zhou, L. Zhang, N. Konz, Computer Vision Techniques in Manufacturing, IEEE Trans. Syst. Man Cybern. -Syst. (2022).
[15] A. Esteva, K. Chou, S. Yeung, N. Naik, A. Madani, A. Mottaghi, Y. Liu, E. Topol, J. Dean, R. Socher, Deep learning-enabled medical computer vision, NPJ Digit. Med. 4 (1) (2021) 1–9.
[16] J. Van Brummelen, M. O'Brien, D. Gruyer, H. Najjaran, Autonomous vehicle perception: The technology of today and tomorrow, Transp. Res. c, Emerg. Technol., Apr. 89 (2018) 384–406.
[17] A. Naim, A. Aaroud, K. Akodadi, C. El Hachimi, A fully AI-based system to automate water meter data collection in Morocco country, Array 10 (2021).
[18] A. Anis, M. Khaliluzzaman, M. Yakub, N. Chakraborty, and K. Deb, "Digital electric meter reading recognition based on horizontal and vertical binary pattern," In: Proc. 3rd Int. Conf. Elect. Inf. Commun. Technol. 650 (EICT), Khulna, Bangladesh, Dec. 2017, pp. 1–6.
[19] J. Canny, A computational approach to edge detection, IEEE Transactions on Pattern Analysis and Machine Intelligence vol. PAMI-8 (6) (1986) 679–698.
[20] D. B. P. Quintanilha, R. W. S. Costa, J. O. B. Diniz, J. D. S. de Almeida, G. Braz, A. C. Silva, A. C. de Paiva, E. M. Monteiro, B. R. Froz, L. P. A. Piheiro, and W. Melho, "Automatic consumption reading on electromechanical meters using HoG and SVM", In: Proc. 7th Latin Amer. Conf. Netw. Electron. Media (LACNEM), 2017, pp. 57–61.
[21] L. A. Elrefaei, A. Bajaber, S. Natheir, N. AbuSanab, and M. Bazi, "Automatic electricity meter reading based on image processing", In: Proc. Appl. Elect. Eng. Comput. Technologies Conf., 2015, pp. 1–5.
[22] Q. Bai, L. Zhao, Y. Zhang, and Z. Qi, "Research of automatic recognition of digital meter reading based on intelligent image processing", In :Proc. 2nd Int. Conf. Comput. Eng. Technol., vol. 5, Apr. 2010, pp. 1–5.
[23] A. Krizhevsky et al., "ImageNet classification with deep convolutional neural networks", In: Proc. Int. Conf. Neural Inf. Process. Syst., 2012, pp. 1097–1105.
[24] 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., Las Vegas, NV, USA, 2016, pp. 779–788.
[25] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger", In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 6517–6525.
[26] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, pp. 1–6, arXiv:1804.02767.
[27] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, Computer Vision and Pattern Recognition (2014) 580–587.
[28] R. Girshick, "Fast R-CNN", In: Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Dec. 2015, pp. 1440–1448.
[29] S. Ren et al., "Faster R-CNN: Towards real-time object detection with region proposal networks", In: Proc. Conf. Neural Informat. Process. Syst., 2015, pp. 91–99.
[30] Z. Tian et al., "FCOS: Fully convolutional one-stage object detection", In: Proc. Int. Conf. Comput. Vis., 2019, pp. 9627–9636.
[31] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," 2019, arXiv: 1904.07850. [Online]. Available: http://arxiv.org/abs/1904.07850.
[32] B. Shi, X. Bai, C. Yao, An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (11) (Nov. 2017) 2298–2304.
[33] Y.J. Chong, K. Huat Chua, M. Babrdel, L.C. Hau, L. Wang, Deep Learning and Optical Character Recognition for Digitization of Meter Reading, in: 2022 IEEE 12th Symposium on Computer Applications & Industrial Electronics (ISCAIE), 2022, pp. 7–12.
[34] H. Shuo, Y. Ximing, L. Donghang, L. Shaoli, P. Yu, Digital recognition of electric meter with deep learning, in: 2019 14th IEEE International Conference on Electronic Measurement & Instruments (ICEMI), 2019, pp. 600–607.
[35] W. Zhou, J. Peng, and Y. Han, "Deep Learning-based Intelligent Reading Recognition Method of the Digital Multimeter", In: Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC), 2021, pp. 3272–3277.
[36] J. Zhang, W. Liu, S. Xu, X. Zhang, Key point localization and recurrent neural network based water meter reading recognition, Displays 74 (2022).
[37] C. Wang, L. Guo, C. Wen, X. Yu, J. Huang, Attitude coordination control for spacecraft with disturbances and event-triggered communication, IEEE Transactions on Aerospace and Electronic Systems 57 (1) (Feb. 2021) 586–596.
[38] C. Yang, W. Lu, Y. Xia, Reliability-constrained optimal attitude-vibration control for rigid-flexible coupling satellite using interval dimension-wise analysis, Reliability Engineering and System Safety 237 (2023).
[39] C. Yang, W. Lu, Y. Xia, Uncertain optimal attitude control for space power satellite based on interval Riccati equation with non-probabilistic time-dependent reliability, Aerospace Science and Technology 139 (2023).
[40] N. Sripanuskul, P. Buayai, X.J.I.A. Mao, Generative Data Augmentation for Automatic Meter Reading Using CNNs, IEEE Access 10 (2022) 28471–28486.
[41] R. Zhang, Z. Bahrami, T. Wang, Z. Liu, An adaptive deep learning framework for shipping container code localization and recognition, IEEE Transactions on Instrumentation and Measurement 70 (Aug. 2021) 1–13.
[42] R. Zhang, Z. Bahrami, Z. Liu, A vertical text spotting model for trailer and container codes, IEEE Transactions on Instrumentation and Measurement 70 (2021) 1–13.
[43] W. Wu, H. Liu, J. Li, An automatic meter recognition method for in-orbit application, in: International Conference on Man-Machine-Environment System Engineering, 2020, pp. 403–410.
[44] A. Iqbal, A. Basit, I. Ali, J. Babar, I. Ullah, Automated meter reading detection using inception with single shot multi-box detector, Intelligent Automation & Soft Computing 27 (2) (2021) 299–309.
[45] R. Carvalho, J. Melo, R. Graça, G. Santos, M.J.M. Vasconcelos, Deep learning-powered system for real-time digital meter reading on edge devices, Applied Science 13 (4) (2023) 2315.

[46] M. Bin, M. Xiangbin, M. Xiaofu, L. Wufeng, H. Linchong, and J. Dean, "Digital recognition based on image device meters", In: Proc. 2nd WRI Global Congr. Intell. Syst., vol. 3, Dec. 2010, pp. 326–330.
[47] L. Gómez, M. Rusinol, and D. Karatzas, "Cutting Sayre's knot: Reading scene text without segmentation. Application to utility Meters", In: Proc. 13th IAPR Int. Workshop Document Anal. Syst. (DAS), 2018, pp. 97–102.
[48] W. Wang et al., "Efficient and accurate arbitrary-shaped text detection with pixel aggregation network", In: Proc. IEEE/CVF Int. Conf. Comput. Vis., 2019, pp. 8439–8448.
[49] S. Woo, J. Park, J.-Young Lee, and I. S. Kweon, "CBAM: Convolutional block attention module", In: Proc. Eur. Conf. Comput. Vis., 2018, pp. 1–19.
[50] J.A. Hartigan, M.A. Wong, Algorithm AS 136: A k-means clustering algorithm, J. Roy. Stat. Soc. C (appl. Stat.) 28 (1) (1979) 100–108.
[51] Z. Zheng, P. Wang, W. Liu, J. Li, Y. Rongguang, and R. Dongwei, "Distance-IoU loss: Faster and better learning for bounding box regression", In: Proc. AAAI Conf. Artif. Intell., 2020.
[52] R. Kucev, Water meter dataset, IEEE Dataport (2019).
[53] A. Nodari and I. Gallo, "A multi-neural network approach to image detection and segmentation of gas meter counter", In: Proc. 12th IAPR Conf. Mach. Vis. Appl. (MVA), 2011, pp. 239–242.
[54] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation", In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Long Beach, CA, USA, 2019, pp. 5693–5703.
[55] D. Maji, S. Nagori, M. Mathew, D. Poddar, YOLO-Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 2637–2646.
[56] B. Cheng, B. Xiao, J. Wang, H. Shi, T. S. Huang, and L. Zhang, "HigherHRNet: Scale-aware representation learning for bottom-up human pose estimation," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020, pp. 5386–5395.
[57] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2020, pp. 10778–10787.
[58] K. He et al., "Deep residual learning for image recognition," in Proc. Conf. Comput. Vis. Pattern Recognit., 2016, pp. 770–778.
[59] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. Int. Conf. Mach. Learn., 2019, pp. 6105–6114.