Research Article

A Deep Neural Network Based on Circular Representation for Target Detection

Cong Lin,[1] **Zhoujian Chen,**[1] **Yiquan Huang,**[1] **Haoyu Jiang,**[1] **Wencai Du** ⓘ**,**[2]
and Qiong Chen ⓘ[3]

[1]*College of Electronics and Information Engineering, Guangdong Ocean University, Zhanjiang /524025, China*
[2]*Institute of Data Engineering and Sciences, University of Saint Joseph, Macao, China*
[3]*Department of Earth System Science, Ministry of Education Key Laboratory for Earth System Modeling, Institute for Global Change Studies, Tsinghua University, Beijing /100084, China*

Correspondence should be addressed to Wencai Du; george.du@usj.edu.mo and Qiong Chen; qiongchen@mail.tsinghua.edu.cn

Convolutional neural network (CNN) model based on deep learning has excellent performance for target detection. However, the detection effect is poor when the object is circular or tubular because most of the existing object detection methods are based on the traditional rectangular box to detect and recognize objects. To solve the problem, we propose the circular representation structure and RepVGG module on the basis of CenterNet and expand the network prediction structure, thus proposing a high-precision and high-efficiency lightweight circular object detection method RebarDet. Specifically, circular tubular type objects will be optimized by replacing the traditional rectangular box with a circular box. Second, we improve the resolution of the network feature map and the upper limit of the number of objects detected in a single detect to achieve the expansion of the network prediction structure, optimized for the dense phenomenon that often occurs in circular tubular objects. Finally, the multibranch topology of RepVGG is introduced to sum the feature information extracted by different convolution modules, which improves the ability of the convolution module to extract information. We conducted extensive experiments on rebar datasets and used AB-Score as a new evaluation method to evaluate RebarDet. The experimental results show that RebarDet can achieve a detection accuracy of up to 0.8114 and a model inference speed of 6.9 fps while maintaining a moderate amount of parameters, which is superior to other mainstream object detection models and verifies the effectiveness of our proposed method. At the same time, RebarDet's high precision detection of round tubular objects facilitates enterprise intelligent manufacturing processes.

1. Introduction

The detection and recognition of circular tubular objects are always a basic problem and difficult problem in image processing. Circular tubular objects are common in the production practice of traditional manufacturing, so the intelligent detection of circular tubular objects is also an important part of the process of enterprise intelligence. Many traditional feature-based images processing methods had been proposed for detection of circular tubular objects, such as Genetic algorithms [1, 2], Gradient pair vectors [3], and Hough transform filters [4–6]. Such methods have low

detection accuracy and strongly rely on "hand-crafted" features from feature engineering, which have great limitations.

In recent years, data-driven deep learning technology has shown superior performance in the field of artificial intelligence, and convolutional neural network is one of the most popular deep learning structures [7, 8]. In the field of computer vision, detection methods based on convolutional neural networks have made great breakthroughs in many fields, such as image classification [9], target detection [10, 11], semantic segmentation [12], instance segmentation [13], and gesture recognition [14]. In 2012, Krizhevsky et al. constructed an 8-layer convolutional neural network

AlexNet [15]. AlexNet used ReLU as the activation function of CNN for the first time, successfully solving the problem of gradient diffusion of Sigmoid when the network was deep and applied Dropout in the network to randomly ignore some neurons to avoid model overfitting. Since then, more and more neural network models have been proposed. The VGG (Visual Geometry Group) [16] network designed by Simonyan and Zisserman mainly improved network performance by increasing the depth of the network. The convolutional layer of the network uses a 3×3 small-size convolution kernel. Compared with a large-size convolutional layer, a small-size convolutional layer has fewer parameters and can increase the nonlinearity of the mapping function. With the deepening of network layer, the VGG network reaches a performance bottleneck at the 16th layer and then tends to be saturated. Increasing the depth of the neural network like VGG can improve the performance of the network to a certain extent, but this approach has two bottlenecks. On the one hand, the deeper the network structure needs to learn more parameters, making the network easy to overfit. On the other hand, the network with more layers requires more computing resources. The GoogLeNet [17] network model developed and designed by Szegedy et al. used a novel inception structure as the basic module for cascading, and the network reaches a depth of 22 layers. Inception uses 3 convolution kernels of different sizes to extract feature information of different scales from the previous input layer. The 1×1 convolution kernel is used to reduce the data dimension of the previous layer and the amount of convolution calculations for the subsequent 3×3 and 5×5 convolutional layers to greatly reduce network parameters while increasing the depth of the network and make full use of computing resources to improve the computational efficiency of the algorithm. Although methods such as ReLU and batch normalization can solve the gradient disappearance or explosion of deep neural networks to a certain extent, the problem of gradient disappearance or explosion is still very serious when training a very deep network. In 2015, He et al. proposed a 152-layer residual networks (ResNet) [18], which added shortcut connections when constructing the network, so that the output of subsequent layers was not the input mapping in the traditional neural network, but the input mapping and the superposition of the input, which solves the problem of the disappearance or explosion of the gradient of the deep neural network training. This type of detection method based on deep learning technology has high accuracy, strong model robustness, and strong transferability, but it also has the disadvantage of slow inference speed due to the large model. At the same time, most of these detection methods are based on the traditional rectangular box to detect and recognize objects, so they are not optimized for circular or tubular objects, resulting in poor recognition and detection of such objects.

The detection and recognition of circular tubular objects are a basic problem in the field of image processing. Zelniker et al. [19] used a maximum likelihood estimation method based on convolutional neural network to better estimate the circular parameters of the center and radius of the circle in digital images as a method of circular object detection.

Ayala-Ramirez et al. [20] propose a circle detection method based on genetic algorithm (GA), which can detect subpixel circles in composite images, but this method has poor performance in dealing with small circle targets. Yang et al. propose a circular object detection algorithm based on convolutional neural network in CircleNet [21], which is used to identify and detect the glomeruli of spherical biomedical objects. CircleNet improves the network's detection accuracy of circular tubular objects to a certain extent by adding a circular detection head to the network, but this model also has the problems of high model complexity and low detection efficiency. Because circular tubular objects often appear in a large number of clusters, another difficulty in detecting and identifying circular tubular objects is dense scene detection. Recognition and detection of objects in dense scenes are a difficult problem in the field of object detection. There are roughly two reasons. One is that highly overlapping instances are likely to have very similar features, making it difficult for the detector to generate distinguishable prediction results for each instance; the other is that there is a serious overlap between instances, and the prediction results may be incorrectly suppressed by NMS. For these reasons, mainstream algorithms in the object detection field, such as R-CNN [22–24], YOLO [25–27], and SSD [28], have poor detection effects in dense scenes.

In order to solve the above problems, based on the network architecture of object detection algorithm CenterNet [29], combining with the characteristics of circular tubular type image and the particularity of circular tubular object detection, we design a high-precision and high-efficiency lightweight circular object detection method RebarDet. First, a circular representation structure is introduced on the basis of the CenterNet network architecture, and a circular box representation is used to replace the traditional rectangular box representation. The circular box representation is optimized for the spherical shape of the circular tubular target and has better rotation consistency than the rectangular box representation. At the same time, the parameter amount of detection representation is reduced from 4 to 3 to improve the detection effect of the network while reducing the model parameters. Then, we expand the prediction structure of the CenterNet network. The specific method is to increase the resolution of the network feature map and the upper limit of the number of single detection objects in the network and optimize for the dense phenomenon of circular tubular object detection. In addition, by combining the advantages of RepVGG [30] block in information fusion and information extraction, we introduce RepVGG block into the CenterNet network structure, which reduces the amount of model calculation and improves the feature extraction capability of the network. Combining the above improvements, we obtain our work RebarDet, a high-precision and high-efficiency lightweight circular object detection method. Based on the rebar dataset provided by Glodon, extensive experiments have been done to verify the performance of RebarDet in the rebar detection environment. The experimental results show that our work has brought about a significant improvement in detection performance. We also compare with the current mainstream target detection network on the three

performance indicators of AB-Score, inference speed, and parameter quantity. The experimental results show that the accuracy and inference speed of our proposed RebarDet exceed the existing mainstream object detectors when the parameters are moderate.

In general, the main contributions of this work are as follows:

(1) The circular representation structure is introduced on the basis of the CenterNet network structure, and the spherical shape of the circular tubular object is optimized to improve the detection effect of the network while reducing the model parameters

(2) Improve the resolution of the network feature map and the upper limit of the number of single detection objects in the network, expand the prediction structure of the CenterNet network, and optimize the dense phenomenon of circular tubular object detection

(3) By combining the advantages of RepVGG block in information fusion and information extraction, the RepVGG block is introduced into the CenterNet network structure, which reduces the amount of model calculation and improves the feature extraction ability of the network

(4) The proposed work RebarDet under the condition of moderate parameter quantity, the accuracy and inference speed are both higher than the existing mainstream object detectors, and it reaches 0.8114 AB-Score and 6.9 fps on the rebar dataset provided by Glodon, significantly better than mainstream object detection models

2. Methods

This work follows the general process of key point detection. Suppose the input image is $I \in R^{W \times H \times 3}$, where W and H are the width and height of the image, respectively. In the prediction phase, the network generates a heat map of key points as $\widehat{Y} \in [0, 1]^{W/R \times H/R \times C}$, where R is the step size corresponding to the original image, and C corresponds to the number of detection points in the object detection. For example, in the COCO [31] object detection task, the value of C is 80, representing 80 categories. In this way, $\widehat{Y}_{x,y,c} = 1$ represents a predicted value of a detected object, indicating that for category c, an object of this category is detected at the current coordinate (x, y), and $\widehat{Y}_{x,y,c} = 0$ indicates that there is no object of category c at the current (x, y) coordinate. For a certain C category in each label map, calculate the real key point $p \in R^2$ in it for training, the calculation method of the center point is

$$ p = \left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2} \right), \tag{1} $$

and the coordinate after downsampling is set to $\tilde{p} = [p/R]$, where R is the downsampling factor 4, so the final calculated center point is the center point corresponding to the low resolution. Next, use $Y \in [0, 1]^{W/R \times H/R \times C}$ to mark the image, and use a Gaussian kernel:

$$ Y_{xyc} = \exp\left(-\frac{(x - \tilde{p}_x)^2 + \left(y - \tilde{p}_y\right)^2}{2\sigma_p^2} \right), \tag{2} $$

to distribute the key points on the feature map in the form of markers in the downsampled image, where σ_p is a standard deviation related to the object size w and h. In the whole process, if two Gaussian distributions of a certain class overlap, the one with the larger element is directly selected.

2.1. Overview. Figure 1 shows the overall network structure of RebarDet. We introduce the circular representation structure and RepVGG module into the network and expand the network prediction structure. Compared with the usual object detection backbone networks such as VGG and ResNet, the Hourglass [32] is better used as the feature extraction network of the model. Because compared to other backbone networks, Hourglass is more conducive to key point detection. The Hourglass network structure includes convolutional layers, deconvolutional layers, fully connected layers, etc., repeatedly using top-down and bottom-up methods, and continuously encoding and decoding to infer the location of detection points. The overall network structure is stacked with submodules and subnetworks. It is constructed in such a way that Hourglass has a high degree of flexibility while having a complex structure and has an excellent performance in describing complex features. The repeated encoding and decoding operations of the network make the network has stronger presentation ability and can better mix global and local information. Compared with other object detection networks, the advantage of using the Hourglass network is that the feature points of the object may appear in different layers of the network, and the final feature map of Hourglass can better detect all the key points of the object. We introduced the RepVGG module at the beginning of the network. After the input image enters the network, it first passes through two RepVGG modules for feature extraction. The multibranch topology of RepVGG will sum up the information extracted by different convolution modules, improve the ability of the shallow network to extract image features, reduce the loss of image details, and be more conducive to the subsequent deep network to extract high-level semantic information of the image. Then, the image passes through the residual module of the network and is downsampled, where RebarDet reduces the original two residual modules to one residual module, and the number of downsampling is also reduced from 2 to 1, which increases the resolution of subsequent feature maps and is more conducive to the detection of dense objects by the network. Then, the image enters the Hourglass backbone network for feature extraction, and the extracted key point feature information is input to the final circular detection head. As can be seen from Figure 1, the network has three detection heads: heat map head, local offset head, and circle
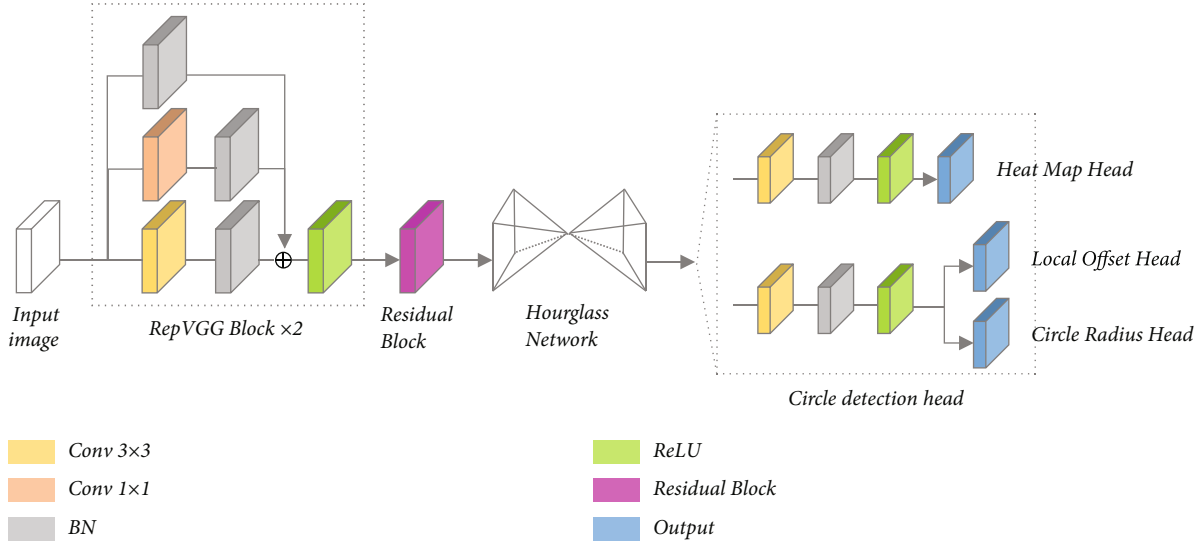
FIGURE 1: The overall network structure of RebarDet. We introduce the RepVGG module and circular detection head on the basis of CenterNet and expanded the network prediction structure.

radius head, which, respectively, detect the object category, center point coordinates, and radius to complete the recognition and detection of circular objects.

2.2. Circle Representation. Existing detectors in the field of object detection, such as Faster R-CNN, YOLO, SSD, etc., generally use rectangular boxes parallel to the horizontal axis to represent the object of network prediction and recognition. Comprehensive considerations are given to the best fitting effect of rectangular boxes for most objects. However, the fitting effect of rectangular boxes is poor in some specific scenes, such as the recognition task of circular tubular objects. As shown in Figure 2(a), the detection effect of using a rectangular box to fit the rebar object is not very good, and the number of missed detections is 5. The reason for the poor performance of rectangular box in circular tubular object recognition task may be that the cross-section of circular tubular object is round, and the distribution of circular tubular object is mostly dense, so it is difficult to fit rectangular box. RebarDet introduces a circular representation structure, using a circular box instead of a traditional rectangular box, to optimize for the spherical shape of a circular tubular object. Specifically, we build a circular detection head to enable the convolutional neural network to predict the regression of the center point (px, py) and the radius r of the circular box and finally obtain the circular representation of the recognition object. As shown in Figure 1, the network uses heat map head ($H \times W \times C$), local offset head ($H \times W \times 2$), and circle radius head ($H \times W \times 1$) to complete the identification and detection of circular objects. Figure 2 (b) is the detection effect of the rebar object after applying the circular box. It can be seen that the detection effect of the circular box on the rebar object is better than that of the rectangular box, which can achieve the industrial accuracy of zero missed detection. Figure 3 is the experimental comparison of the rotation consistency of the rectangular box and the circular box. After the original image is rotated

by 90 degrees for the rectangular box, the number of missed objects increases significantly, while the circular box only has 2 missed objects after the original image is rotated. The above experimental results show that the circular box has better rotation consistency than the rectangular box.

2.3. Expansion of Prediction Structure. The CenterNet proposed by Zhou et al. have a shortcoming in actual training, that is, if the center points of multiple objects of the same category in the image overlap when the network is sampled, CenterNet is also powerless for this situation, because there is only one center point, so these two objects can only be trained as one object. For the detection of circular tubular objects, the distribution of circular tubular objects is dense and compact in most cases, so the overlapping phenomenon of object centers is more serious. To alleviate this phenomenon, we expand the prediction structure of the network in RebarDet. Specifically, we increase the resolution of the feature map before the Hourglass subnetwork, so that the network can detect the center point of the object at a larger feature map level, which can reduce the overlap of the object center point. As shown in Figure 4(a), the size of the feature map has been increased from the original 128×128 to 256×256. Figure 4(b) is a simplified schematic diagram of the object detection effect after the feature map is enlarged. It can be seen that compared to the small feature map, the large feature map can divide the object to be detected more dispersed, improve the detection ability of the network in dense scenes, and reduce the risk of missed detection. In addition, we expand the upper limit of the number of objects that the network can detect at a time from 128 to 256, which makes the network perform better in dense multitarget recognition scenarios.

2.4. RepVGG Block. The architecture of RepVGG is very simple and effective, which is equivalent to adding identity and residual branches to the block of the VGG network.
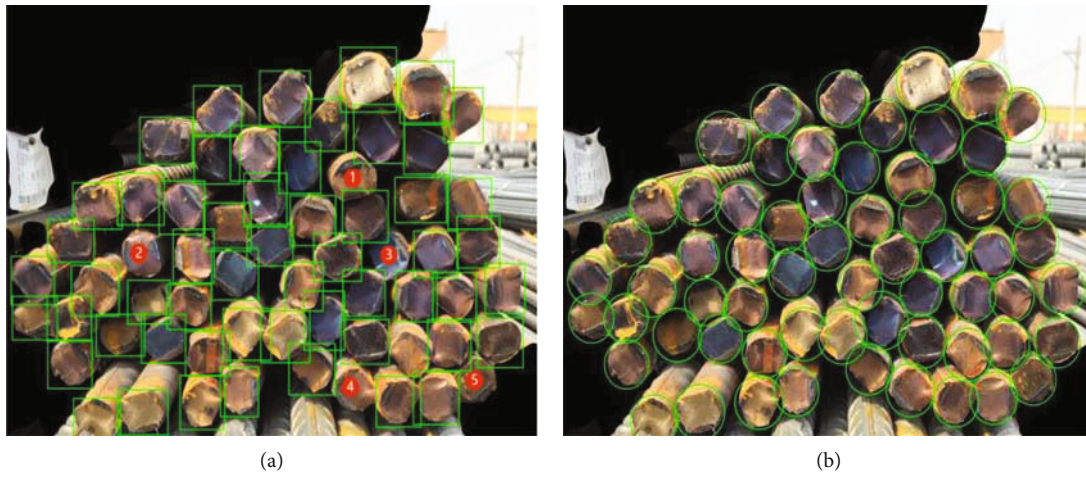
FIGURE 2: The actual effect of RebarDet under the background of rebar detection, (a) is the detection effect of the rectangular box with the number of missed detection being 5, and (b) is the detection effect of the circular box, achieving zero missed detection.
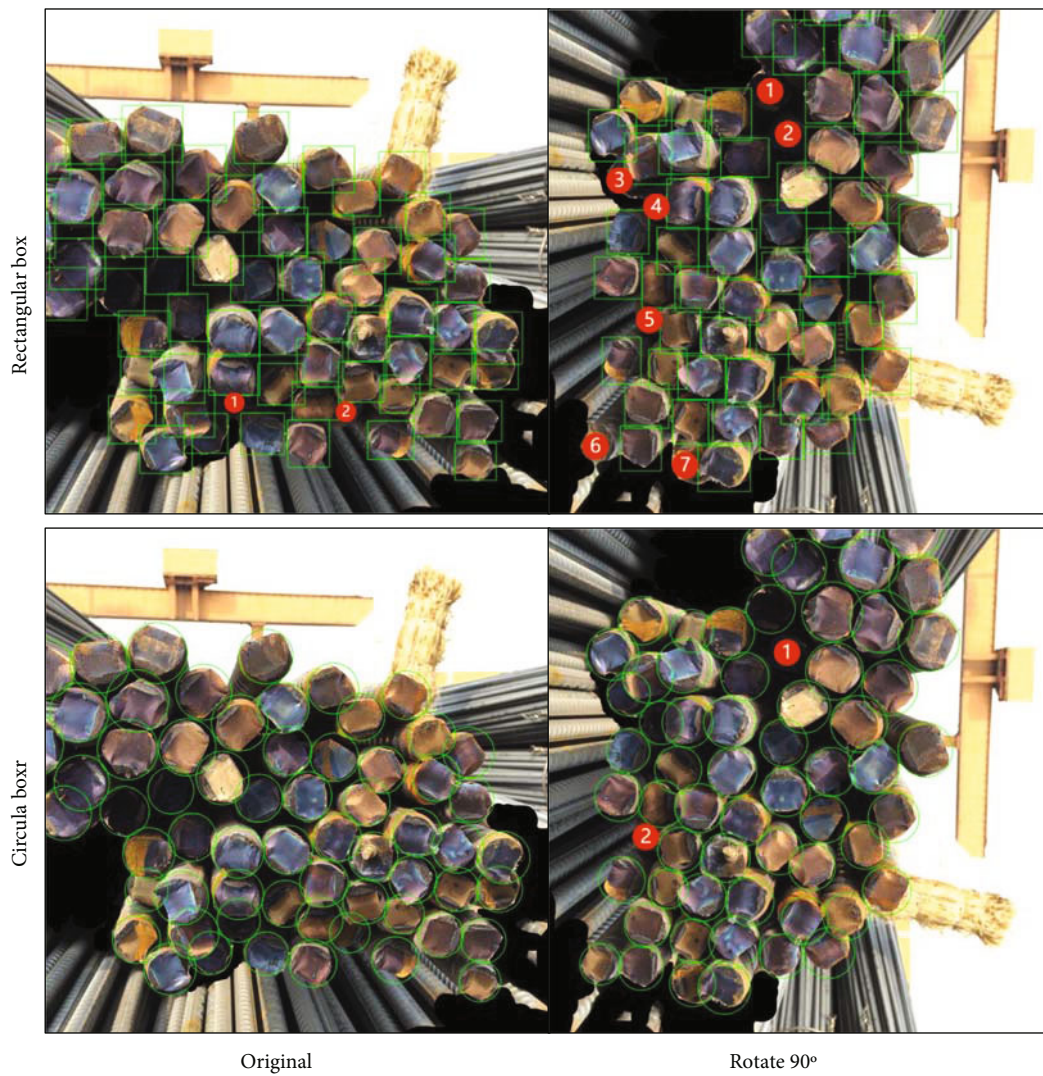

Rectangular box

Circula boxr

Original   Rotate 90°

FIGURE 3: Experimental comparison of the rotation consistency between the rectangular box and the circular box.

Feature map

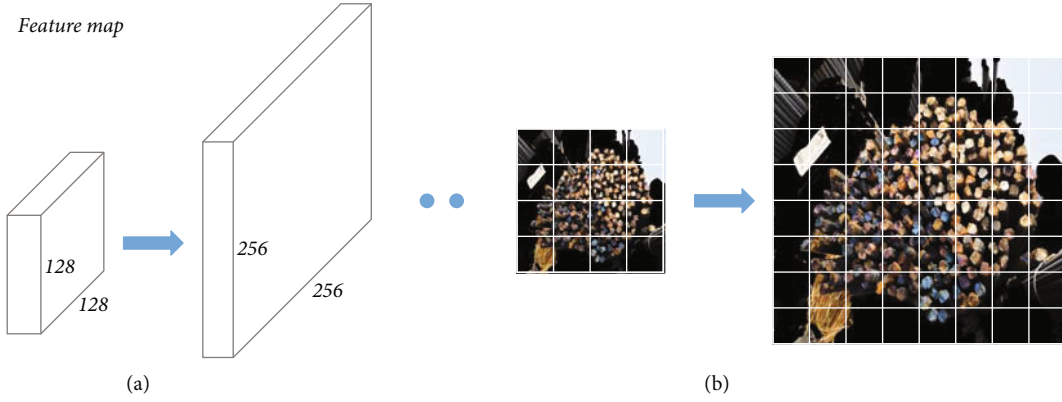(a)   (b)

FIGURE 4: (a) is a schematic diagram of the enlarged network feature map, and (b) is a schematic diagram of the effect. For the purpose of simplification and easy observation, in (b) the schematic diagram of the effect is simplified from 128- >256 to 4- >8.

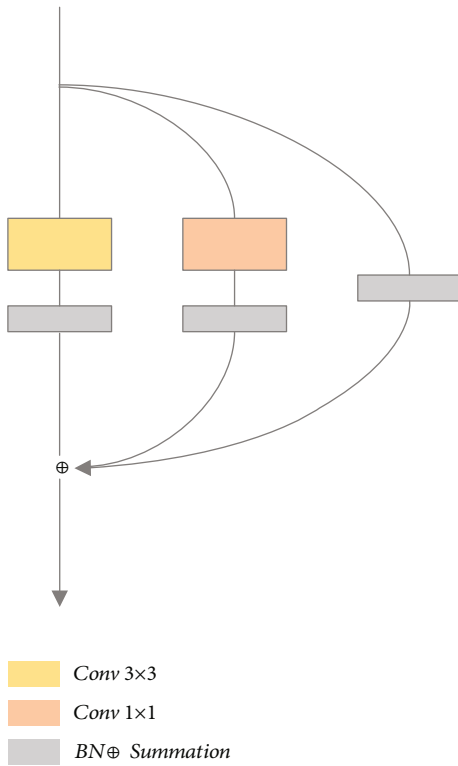

Conv 3×3

Conv 1×1

BN ⊕ Summation

FIGURE 5: Schematic diagram of the RepVGG block structure, a 1 × 1 convolution (additional BN layer) and a BN layer are connected in parallel on both sides of the 3 × 3 convolution (additional BN layer), then, the sum operation is performed.

TABLE 1: The score table of part B of AB-score. If the predicted number is completely correct, the score is 1. If the predicted number differs by 1, the score is 0.5. If the predicted number differs by 2, the score is 0.1. In other cases, the score is 0.

Number deviation range	Score (task B)
0	1
±1	0.5
±2	0.1
Other	0

The entire network is only composed of 3×3 convolution, BN layer, and ReLU modules. Figure 5 is a schematic diagram of the structure of RepVGG block. In the training phase, RepVGG has a multibranch topology in which 1×1 convolution (additional BN layer) and BN layer are connected in parallel on both sides of 3×3 convolution (additional BN layer). This structure sums up the feature information extracted by different convolution modules, which can improve the information extraction ability of a single convolution module. Considering that the shallow convolution of the network is responsible for extracting low-level semantic features, and the information richness of this part of the low-level semantic features directly determines the effectiveness of the high-level semantic features of the subsequent convolutional layer, so we set the first two layers of the backbone network as RepVGG block. The feature of multiple modules in parallel is utilized to improve the information extraction ability of the network's shallow convolution and improve the feature extraction ability of the whole network while reducing the amount of model calculation [33]. Regarding the number of RepVGG blocks and the number of channels, we discuss in Section 3.3.3 and obtain experimental results.

3. Experiments

3.1. Rebar Dataset.
To verify the effectiveness of RebarDet on circular tubular object detection, we evaluate RebarDet in the context of rebar detection. All experiments in this paper are based on the rebar dataset released by Glodon on the DataFountain platform in 2019, which is also the only open-source rebar dataset so far. The dataset contains a total of 250 cross-sectional images of rebars and provides rectangular box annotations for object detection task training. Among them, 24,442 rebar samples in 200 rebar images are used as training data, and 6,499 rebar samples in 50 rebar images are used as test data. Eventually, we formed a cohort with 200 training and 50 testing images.

3.2. Score Metrics.
In the rebar quantity counting competition published by Glodon on the DataFountain platform, the competition uses $F1$-score as the scoring algorithm.
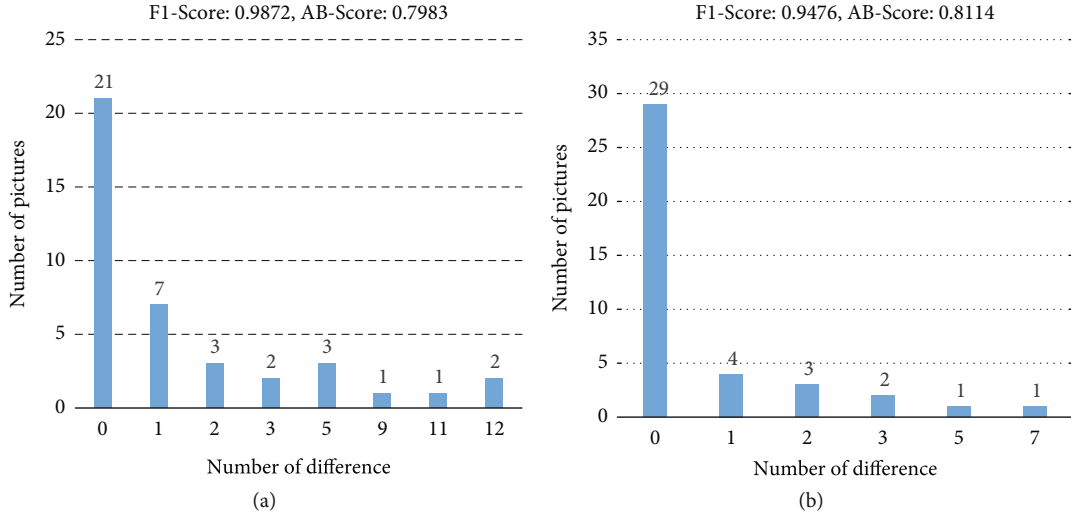
FIGURE 6: Statistics of the test results of the trained RebarDet model in 40 rebar pictures, the abscissa is the difference between the predicted results and the real results, and the ordinate is the number of predicted pictures. (a) is the statistical chart of the test results of the RebarDet model with F1-score of 0.9872 and AB-score of 0.7983, (b) is the statistical chart of the test results of the RebarDet model with F1-score of 0.9476 and AB-score of 0.8114.

TABLE 2: RebarDet ablation experiment results table.

Component	RebarDet			
Circle representation	√	√	√	√
Expansion of prediction structure				
Max objects	√	√	√	
Feature map resolution	√	√		
RepVGG block	√			
AB-score	0.8114	0.7751	0.7568	0.6984 0.4631

However, F1-score is actually a measure of classification problems, and it is not necessarily suitable for counting problems such as rebar counting. In our research, it is found that the model with high F1-score is not effective in the actual rebar counting detection on the construction site. So in this paper, we use the scoring algorithm AB-score of a counting competition on iFLYTEK OPEN PLATFORM to evaluate RebarDet. AB-score consists of two parts, A and B. Part A is the mAP value with IoU equal to 0.5. For part B, as shown in Table 1, different scores are assigned to different counting results obtained from a single picture. If the predicted number is completely correct, the score is 1. If the predicted number differs by 1, the score is 0.5. If the difference is 2, the score is 0.1. Otherwise, the score is 0. The final score in part B is the mean of all the image scores. The final score of AB-score is as follows:

$$AB\text{-}Score = A * 0.4 + B * 0.6. \quad (3)$$

Figure 6 is a comparison chart of the detection effect of the trained circular tubular object detection model on the test set. The abscissa is the difference between the predicted

result and the real result, and the ordinate is the number of predicted pictures. The F1-score and AB-score of the model in Figure 6(a) are 0.9872 and 0.7983, and the F1-score and AB-score of the model in Figure 6(b) are 0.9476 and 0.8114. It can be seen from the figure that the number of pictures that are completely predicted correctly by the latter is more than that of the former, and the error in a smaller range is also better than that of the former. This shows that the model with high F1-score is not as effective as the model with high AB-score in the actual test, indicating that AB-score is more suitable for the detection task of circular tubular objects.

3.3. Loss Function. The loss function of network training follows CenterNet and consists of three parts: L_k, L_{radius}, and L_{off}. The predicted heat map is optimized by pixel regression loss L_k with focal loss:

$$L_k = \frac{-1}{N} \sum_{xyc} \begin{cases} \left(1 - \widehat{Y}_{xyc}\right)^{\alpha} \log\left(\widehat{Y}_{xyc}\right) & \text{if } Y_{xyc} = 1, \\ \left(1 - \widehat{Y}_{xyc}\right)^{\beta} \left(\widehat{Y}_{xyc}\right)^{\alpha} \log\left(1 - \widehat{Y}_{xyc}\right) & \text{otherwise,} \end{cases}$$
$$(4)$$

where α and β are the hyperparameters of the focal loss, and N is the number of keypoints in the image I. The radius of the circular box is optimized with L_{radius}:

$$L_{radius} = \frac{1}{N} \sum_{k=1}^{N} \left| \widehat{R}_{pk} - r_k \right|, \quad (5)$$

where r_k is the true radius of each circular object k. Center-Net performs an $R = 4$ downsampling operation on the image in the network. When the feature map is remapped to the original image, it will cause an accuracy error. Therefore, an additional local offset: $\widehat{O} \in R^{W/R \times H/R \times 2}$ is used for

TABLE 3: The experimental result table of the number of RepVGG Blocks and the number of channels. Using two RepVGG blocks and the number of channels of 128, the experimental accuracy is the highest.

Number of blocks Number of channels	128	256
RepVGG block × 1	0.7525	0.7437
RepVGG block × 2	0.8114	0.7758
RepVGG block × 3	0.7825	0.7638

TABLE 4: Comparison table between RebarDet and other mainstream object detection models in AB-score, inference speed, and parameter quantity.

Method	Backbone	Input size	AB-score	Inference speed (fps)	Params (M)
SSD	VGG-16	512×512	0.7233	3.5	33.51
RetinaNet [34]	ResNet-101	512×512	0.7536	3.7	55.37
Faster-RCNN	ResNet-101	512×512	0.7745	3.5	60.18
ATSS [35]	ResNet-101	512×512	0.7812	3.5	50.91
YOLOv5	yolov5x	512×512	0.7983	7.1	40.97
RebarDet (ours)	Hourglass-104	512×512	0.8114	6.9	35.24

each center point to compensate. All center points of class c share the same offset prediction, and this offset value is trained by $L1$ loss:

$$L_{\text{off}} = \frac{1}{N} \sum_p \left| \widehat{O}_{\tilde{p}} - \left(\frac{p}{R} - \tilde{p} \right) \right|. \qquad (6)$$

Finally, the overall objective is

$$L_{\text{det}} = L_k + \lambda_{\text{radius}} L_{\text{radius}} + \lambda_{\text{off}} L_{\text{off}}, \qquad (7)$$

we set $\lambda_{\text{radius}} = 0.1$ and $\lambda_{\text{off}} = 1$ referring from CenterNet.

3.4. Ablation Study. To verify the effectiveness of our proposed RebarDet model on detection of circular tubular objects, we conduct ablation experiments on the rebar dataset provided by Glodon. In addition, we conduct comparative experiments on mainstream object detection models and compare the performance of various parameters of the models to prove the superiority of the RebarDet model. Specifically, we resize all training set images to 512×512 resolution and keep the relevant hyperparameters of model training consistent to ensure a fair comparison. All models are evaluated using AB-score, see Section 3.2 for details of AB-score.

3.5. Circle Representation. In order to verify the effectiveness of the circular representation structure, we use the original CenterNet model and our improved CenterNet model with the circular representation structure for comparative experiments. The experimental results are shown in the fifth and sixth columns of Table 2. We obtained an AB-score of 0.6984 after adding the improved circular representation structure to the basic CenterNet model. Compared with the basic CenterNet model, adding the circular representation structure can bring an improvement of 0.2353 AB-score to the detector, which proves that by adding a circular representation structure to the network can improve the performance of the detector. We think there are several reasons why the circular representation structure can bring such a big improvement. First, in the detection of circular tubular objects, the fitting effect of the circular box is better than that of the traditional rectangular box. Second, the circular detection head is used to detect and represent the object, which reduces the amount of parameters for detection and representation. Third, the circular box has better rotation consistency than the traditional rectangular box.

3.6. Expansion of Prediction Structure. Max objects represent the upper limit of the number of objects detected by the network at a time. As shown in the third and fourth columns of Table 2, max objects and feature map resolution obtain AB-scores of 0.7568 and 0.7751, respectively. The entire expansion of prediction structure brings an improvement of 0.0767 AB-score to the model, which proves the effectiveness of our designed expansion of prediction structure. We believe that increasing the size of the feature map can divide the objects to be detected more scattered, which greatly improves the detection ability of the network in dense scenes. The increase of max objects improves the loadability of network detection of dense objects to a certain extent.

3.7. RepVGG Block. In order to verify the effectiveness of RepVGG Block, we discuss how many RepVGG blocks are applied in the network and how many channels are adopted and conduct corresponding comparative experiments and obtain the optimal number and number of channels for applying RepVGG block. The experimental results are as

shown in Table 3. Applying a number of RepVGG blocks of 2 and a number of channels of 128, the model achieves the best accuracy. As shown in the second and third columns of Table 2, applying RepVGG block brings an improvement of 0.0363 AB-score to the model, which proves the effectiveness of the multibranch topology of RepVGG on the model. We think that the multibranch topology of RepVGG can improve the extraction ability of shallow convolutions at the beginning of the network, which parallels 1×1 convolution (additional BN layer) and BN layer on both sides of 3×3 convolution (additional BN layer) and BN layer to sum up the feature information extracted by different convolution modules, thus improving the information extraction ability of a single convolution module. The richness of low-level semantic information extracted by this part of the convolutional module has a great influence on the high-level semantic features of the subsequent convolutional layers. Therefore, improving the extraction ability of this part of the convolutional module can improve the performance of the entire network.

3.8. Network Performance. In order to prove the superiority of the model, we use three performance indicators of AB-score, inference speed, and parameter quantity to conduct comparative experiments on mainstream object detection models. The experimental results are shown in Table 4. In terms of detection accuracy, RebarDet obtains the highest AB-score, 0.8114. While obtaining the highest detection accuracy, RebarDet also maintains a high model inference speed of 6.9 fps, which is only slightly lower than the model inference speed of YOLOv5 of 7.1 fps. On the model parameters, the SSD model using VGG-16 as the backbone network has the least amount of parameters, but the AB-score and inference speed of the SSD model are not satisfactory. While the RebarDet model based on Hourglass-104 maintains a moderate amount of parameters, its detection accuracy is the highest and the model inference speed is also better than other mainstream object detection models.

4. Conclusion

In this paper, we introduce RebarDet, a high-accuracy and high-efficiency lightweight circular tubular object detection method. By introducing the circle representation structure and RepVGG module on the basis of CenterNet and expanding the network prediction structure, the circle detection performance of the network has been greatly improved. RebarDet is tested on rebar dataset for object detection. The experimental results show that RebarDet's detection accuracy and model inference speed are better than existing mainstream object detectors with moderate parameters.

Data Availability

The datasets used and analysed during the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that there are no conflicts of interest associated with the manuscript.

References

[1] L. M. Schmitt, C. H. Garcia-Capulin, A. Perez-Garcia, and R. E. Sanchez-Yanez, "Theory of genetic algorithms," *Theoretical Computer Science*, vol. 259, no. 1-2, pp. 1–61, 2001.

[2] M. Srinivas and L. M. Patnaik, "Genetic algorithms: a survey," *Computer*, vol. 27, no. 6, pp. 17–26, 1994.

[3] A. A. Rad, K. Faez, and N. Qaragozlou, *Fast circle detection using gradient pair vectors*, Dicta, 2003.

[4] D. J. Kerbyson and T. J. Atherton, *Circle Detection Using Hough Transform Filters*, IET, 1995.

[5] T. D'Orazio, C. Guaragnella, M. Leo, and A. Distante, "A new algorithm for ball recognition using circle Hough transform and neural classifier," *Pattern Recognition*, vol. 37, no. 3, pp. 393–408, 2004.

[6] M. Smereka and I. Duleba, "Circular object detection using a modified Hough transform," *International Journal of Applied Mathematics and Computer Science*, vol. 18, no. 1, pp. 85–91, 2008.

[7] J. Garland, M. Hu, K. Kesha et al., "An overview of artificial intelligence/deep learning," *Pathology*, vol. 53, p. S6, 2021.

[8] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5455–5516, 2020.

[9] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: a comprehensive review," *Neural Computation*, vol. 29, no. 9, pp. 2352–2449, 2017.

[10] L. Jiao, F. Zhang, F. Liu et al., "A survey of deep learning-based object detection," *IEEE Access*, vol. 7, pp. 128837–128868, 2019.

[11] X. Wu, D. Sahoo, and S. C. H. Hoi, "Recent advances in deep learning for object detection," *Neurocomputing*, vol. 396, pp. 39–64, 2020.

[12] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: a survey," *IEEE transactions on pattern analysis and machine intelligence*, 2021.

[13] A. M. Hafiz and G. M. Bhat, "A survey on instance segmentation: state of the art," *International Journal of Multimedia Information Retrieval*, vol. 9, no. 3, pp. 171–189, 2020.

[14] Y. Chen, Y. Tian, and M. He, "Monocular human pose estimation: a survey of deep learning-based methods," *Computer Vision and Image Understanding*, vol. 192, article 102897, 2020.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances*

in *Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.

[16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, https://arxiv.org/abs/1409.1556.

[17] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Boston, MA, USA, 2015.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Las Vegas, NV, USA, 2016.

[19] E. E. Zelniker, I. Vaughan, and L. Clarkson, "Maximum-likelihood estimation of circle parameters via convolution," *IEEE Transactions on Image Processing*, vol. 15, no. 4, pp. 865–876, 2006.

[20] V. Ayala-Ramirez, C. H. Garcia-Capulin, A. Perez-Garcia, and R. E. Sanchez-Yanez, "Circle detection on images using genetic algorithms," *Pattern Recognition Letters*, vol. 27, no. 6, pp. 652–657, 2006.

[21] H. Yang, R. Deng, Y. Lu et al., "CircleNet: anchor-free detection with circle representation," 2020, https://arxiv.org/abs/2006.02474.

[22] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580–587, Columbus, OH, USA, 2014.

[23] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, Santiago, Chile, 2015.

[24] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 28, pp. 91–99, 2015.

[25] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Las Vegas, NV, USA, 2016.

[26] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition.*, Honolulu, HI, USA, 2017.

[27] J. Redmon and A. Farhadi, "Yolov3: an incremental improvement," 2018, https://arxiv.org/abs/1804.02767.

[28] W. Liu, D. Anguelov, D. Erhan et al., *Ssd: Single Shot Multibox Detector*, European Conference on Computer Vision, Springer, Cham, 2016.

[29] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," 2019, https://arxiv.org/abs/1904.07850.

[30] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, and J. Sun, "Repvgg: making vgg-style convnets great again," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13733–13742, 2021.

[31] T.-Y. Lin, M. Maire, S. Belongie et al., "Microsoft coco: common objects in context," in *European conference on computer vision*, Springer, Cham, 2014.

[32] A. Newell, K. Yang, and J. Deng, *Stacked Hourglass Networks for Human Pose Estimation*, European conference on computer vision, Springer, Cham, 2016.

[33] C. Lin, Y. Zheng, X. Xiao, and J. Lin, "CXR-RefineDet: single-shot refinement neural network for chest X-ray radiograph based on multiple lesions detection," *Journal of Healthcare Engineering*, vol. 2022, Article ID 4182191, 11 pages, 2022.

[34] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, Venice, Italy, 2017.

[35] S. Zhang, C. Chi, Y. Yao, Z. Lei, and S. Z. Li, "Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9759–9768, Seattle, WA, USA, 2020.