Berger Scientfic Press Inc.
Innovation & Technology Advances
Volume 1 Issue 1, June, 2023
www.bergersci.com
DOI: https://doi.org/10.61187/ita.v1i1.1

Rebar Cross-section Detection Based on Improved YOLOv5s Algorithm

Yanmei Zheng[1], Guanghui Zhou[2], Bibo Lu[3,*]

[1,2,3]College of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan, China
[1]*zhengym02@hpu.edu.cn*, [2]*1074050095@qq.com*, [3]*1074870971@qq.com*
[*]*Correspondence Author*

Abstract: ***Rebar is an indispensable material in the construction industry and the counting of rebar is a very important part of the production, transportation, and use process. Due to the problems of dense cross-sections and mutual adhesion of cross-sections boundaries for bundles of rebar, most of the rebar is still counted manually. To address the above problems, this paper proposes an improved YOLOv5s-based rebar cross-section detection method, which aims to solve the problems of missed and false detection in dense small object detection. The K-means algorithm is used to re-cluster the anchor box size; the ECANet (Efficient Channel Attention Net) module is integrated with the backbone network to adjust the attention weight of the feature map and improve the feature extraction ability of the network; To address the problem that small object with a scale smaller than 8×8 in the network cannot be detected, a new detection layer scale is redesigned by fusing the high-level semantic information with the low-level features to reduce the network's miss detection rate for small object. Experiments were conducted on a homemade rebar cross-section dataset with mAP0.5 of 94.3%, which is 6.2% improvement compared with the baseline model YOLOv5s. The improved network was able to better identify the cross-sectional features of the rebar, thus significantly reducing the leakage rate in rebar counting detection.***

Keywords: YOLOv5, Dense small objects, ECANet module, Small object detection layer**.**

1. Introduction

Rebar has an irreplaceable role in construction, manufacturing, and road building. The final process in the rebar production process is the separation of rebar in certain quantities and bundling it into individual units. For commercial management reasons, the quantity of rebar must be counted and accounted for before it leaves the factory before it arrives at the market before it leaves the market, and when it arrives at the building construction site. Rebar counting, as a daily task in most construction sites, is mostly done manually, and this method is highly prone to errors, thus it is urgent to research object detection method for rebar counting, which also plays an important role in reducing workers' work intensity and improving the efficiency and accuracy of rebar counting.

The key to the visual bundle counting technique depends on the ability to detect or segment the rebar cross-sections with the correct method. Xinman Z et al [1] proposed an online rebar counting and automatic separation system by setting the region of interest, segmenting with template matching, and mutation threshold by preprocessing and edge detection of rebar images. Wu Y et al [2] used concave point matching segmentation, K-level fault tolerant counting, and multiple segmentation to accurately count rebar with diameters between 8 mm and 20 mm. Zhao J et al [3] used an improved Otsu algorithm, Sobel operator, and edge clustering algorithm to detect the number of rebar in a set stable environment.

With the development of convolutional neural networks in recent years, many researchers have proposed different object detection algorithms. For example, two-stage R-CNN [4] and Faster R-CNN [5], and single-stage algorithms such as SSD [6], RetinaNet [7], CenterNet [8], YOLOv3 [9], YOL Ov4 [10], and YOLOv5 [11].

Wang H et al [12] combined Canny edge detection, circular Hough transform, and CNN model to calculate the number of rebars, and the highest detection accuracy of this model was 95.99%. Zhu Y et al [13] proposed a CNN-based fusion model for rebar localization and segmentation. Li Y et al [14] introduced a feature pyramid, complete cross-joint IoU in YOLOv3 loss and focal loss to propose a YOLOv3-based rebar counting method in real-time.

None of the above methods achieves the balance between real-time and accuracy in the rebar cross-section detection task, so this paper proposes a rebar cross-section detection algorithm based on improved YOLOv5s. The anchor box is computed using K-means algorithm re-clustering; the ECANet module is added to the backbone part of the YOLOv5 network; Fusion of high-level features and low-level features to add a new detection layer to cope with the detection of small object. The improved method in this paper was tested on a self-constructed dataset for multiple comparisons, and the improved average accuracy reached 94.3%, which is higher than most of the current mainstream object detection models.

2. Related Work

2.1 YOLOv5 Introduction

In 2020 Ultralytics released the YOLOv5 algorithm, which consists of four main components: input, backbone, neck, and output. The input part uses image enhancement techniques to improve the detection effect. The backbone network consists of a convolutional module, a bottleneck layer, and a spatial pyramid pooling. The neck part uses a combination of top-down and bottom-up fusion for the better fusion of multi-scale features, then passes the feature information into the detection layer, eliminates redundant box after non-maximum suppression and other operations, and finally

outputs the prediction category with the highest confidence score, while returning information about the bounding box coordinates of the detection object. YOLOv5 is a one-stage detection algorithm with accurate and fast detection and has an excellent performance in many detection YOLOv5 is divided into s, m, l, and x versions according to the depth and width of the network, and YOLOv5s is selected as the benchmark network structure for research and analysis in view of the light weight of the model.

2.2 Difficulties of Rebar Cross-section Detecting

The cross-sections of bundled rebar are a kind of dense small object, which has the characteristics of a single object category, more concentrated object distribution, and the low distinction between each other during the detection. Detection will encounter the following difficulties:

(1) The cross-sectional size of rebar varies: the cross-sectional diameter of rebar varies greatly depending on the specification, and the diameter of rebar tested in this paper is between 18 and 28 mm.

(2) The number of bars per bundle varies greatly, ranging from 70 to 180 depending on the type of bars.

(3) The bundling of rebar bars causes serious adhesion between cross-sections, resulting in depressions between bars, which makes it difficult to extract object features and makes detection more difficult.

3. YOLOv5 Improvements and Innovations

3.1 K-means Clustering Algorithm

The YOLOv5 algorithm adds a priori information about the object box to the network prediction information. Before training, the network is "informed" of the range of information about the object box, such as the range of box center coordinates and width and height. The network is then trained to learn from this range of values to obtain more accurate anchor box detection information, which largely increases the stability and convergence speed of the network.

In the YOLOv5 algorithm, the idea of auto-learning bounding box anchors is used to calculate the appropriate anchor box size. Since the detection object size of the self-built dataset differs significantly from the COCO dataset [15], the anchor box size is recalculated using the K-Means algorithm, and three new anchor boxes for detecting small object with obscure boundaries are added in this paper. The addition of three small object anchor boxes can reduce the missed detection of small object to some extent. The clustering center is calculated by the K-means clustering algorithm, and iterations are repeated until the clustering center no longer changes, and the width and height of the three prior anchor boxes are finally clustered well. The specific function T is shown in Equation (1):

$$T = \min \sum_{i=1}^{n} \sum_{j=1}^{k} [1 - \frac{box_i \cap cen_j}{box_i \cup cen_j}] \quad (1)$$

In the above equation, box_i is the area of the ith labeled box in the object, cen_j is the area of the jth clustering center, n is the number of detected objects, and k is the number of clustering centers.

The indicator of well clustering is expressed by the best possible recall (BPR), which is calculated as shown in Equation (2):

$$P = \frac{1}{n} \sum_{i=1}^{n} \max(\frac{box_i \cap cen_1}{box_i \cup cen_1}, \frac{box_i \cap cen_2}{box_i \cup cen_2}, ..., \frac{box_i \cap cen_k}{box_i \cup cen_k}) \quad (2)$$

In this dataset, when k=12, BPR=0.999, which is in better performance, the final generated anchor box is assigned according to the different detection layer scale sizes, as shown in Table 1.

Table 1: Anchor box allocation table

Feature map	20×20	40×40	80×80	160×160
Receptive Field	Larger	Large	Middle	Small
Anchor box	(16,15) (16,17) (16,18)	(12,12) (15,14) (14,15)	(9,9) (9,10) (11,9)	(7,7) (7,9) (9,7)

3.2 Adding ECANet Module

The dark environment in which the rebar is located and the small and dense object make it difficult for the network to extract the feature information of the rebar cross-section. At the same time, when the image background is convolved in the network, the iterative accumulation will produce a large amount of redundancy, which makes the positive and negative samples unbalanced and floods part of the object information, thus leading to low detection accuracy. To address the above problems, this paper adds the ECANet (Efficient Channel Attention Net) module [16] to the backbone network.

The attention mechanism obtains more critical information by focusing on the important regions of the input object. The basic idea is to score the semantic information for each part on the input image and then weight the semantic features according to the score as a way to highlight the impact of important features on downstream models or modules. In the ECANet module, the global features of the image are aggregated using a global average pool (GAP) without channel dimensionality reduction, learned directly using 1D convolution, and then the weight coefficients of the channel are fixed between 0 and 1 by computing the Sigmoid function in a single pass. Finally, a weighted multiplication between these coefficients and the input features is performed to increase the "interest" of important features and to suppress the focus on less important features. In contrast to the SE module [17], ECANet uses 1D convolution to replace fully connected layers to avoid significantly increasing the parameters of the model and thus improving its nonlinear capability. The negative impact of dimensionality degradation on the model performance is avoided and a local cross-channel information interaction strategy is used to improve the performance of the ECANet model. Figure 1 shows the structure of the SE module and Figure 2 shows the structure of the ECANet module. Based on the above analysis, this paper embeds an efficient channel attention module into

YOLOv5 to enhance the model learning capability and thus improve the detection accuracy of rebar cross-section.

To make the network pay more attention to the "interested" object features, ECANet is added to the convolutional layer module of the network, which can effectively extract the information of the object features and reduce the missed detection of small object of rebar cross-section.

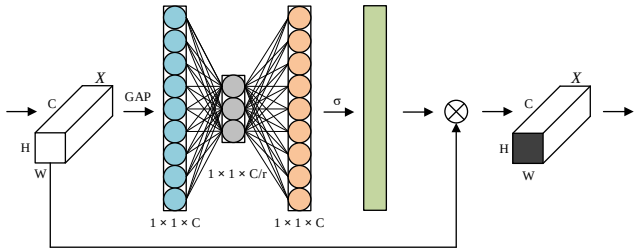


Figure 1: SE structure

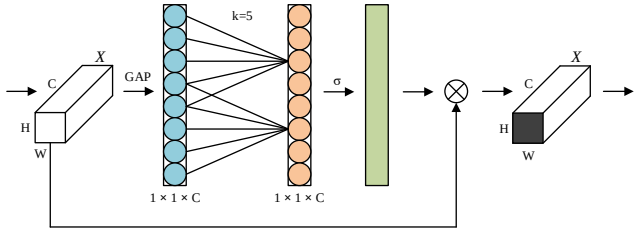


Figure 2: ECANet structure

In Figures 1 and 2, $X \in R^{W \times H \times C}$ is the output of the convolution module. W, H and C are the width, height, and channel dimensions of the feature map, respectively. Kernel size k of the 1D convolution indicates the coverage of cross-channel interactions, which indicates how many adjacent channels are involved in the attentional prediction of the channel. The kernel size k is proportional to the channel dimension C and can be determined adaptively by a nonlinear mapping of the channel dimensions. r is the reduction rate, σ is the Sigmoid activation function and ⊗ denotes the element product.

The backbone network is mainly for feature extraction of the object. In order to increase the focus of the network on the object features, ECANet module is added to the convolutional layer module of the network before the end of the feature extraction process, i.e., the end of the backbone, and this operation can effectively extract the information of the object features and reduce the missed detection of small object.

3.3 Addition of Small Object Detection Layer

The rebar cross-sectional diameters vary in size and have a wide range of specifications, and the cross-sectional diameters of the rebar collected in this paper range from 18-28 mm. For this multi-scale detection problem, the network processes the feature maps by down-sampling and uses multi-scale feature fusion to extract features of different sizes from different layers to do prediction. However, for rebar cross-sections with small object, the object region information is lost after multiple convolutions in the network, and using the original scale of the network to detect such object is very likely to cause missed detection, so a new detection layer is added in this paper to cope with the detection of small object and further reduce the missed detection rate of the model for small object[18].

The YOLOv5 algorithm has three detection layers, corresponding to three sets of an anchor box at the same time. The detection layer corresponding to the eighth layer of the backbone network has a size of 20×20 and can be used to detect object above the 32×32 scale; the detection layer corresponding to the sixth layer of the backbone network has a size of 40×40 and can detect object above the 16×16 scale; the detection layer corresponding to the fourth layer of the backbone network has a size of 80×80, which corresponds to the smallest perceptual field at this point and is used to detect object above the 8×8 scale. However, small object such as rebar cross-sections, cannot be detected by the network when the object scale is smaller than 8×8. To address this problem, after the 18th layer of the Neck, the feature map is continued to be up-sampled, and the up-sampling scales the smaller feature map to the same size as the larger feature map to satisfy the condition of feature fusion. Meanwhile, in layer 21, the acquired feature map of size 160×160 is spliced and fused with the layer 2 feature map of the backbone network, and in layer 22 of the network, a new detection layer is added. The improved network has four detection layers, and by deepening the network, the feature information of the object can be better extracted. This approach, by transferring down the higher-level semantic features to complement the lower-level semantics, allows for high-resolution, strongly semantic features, thus facilitating the detection of small object[19]. The improved overall YOLOv5 model structure is shown in Figure 3, and the red dashed box part in the figure shows the newly added detection layer structure.

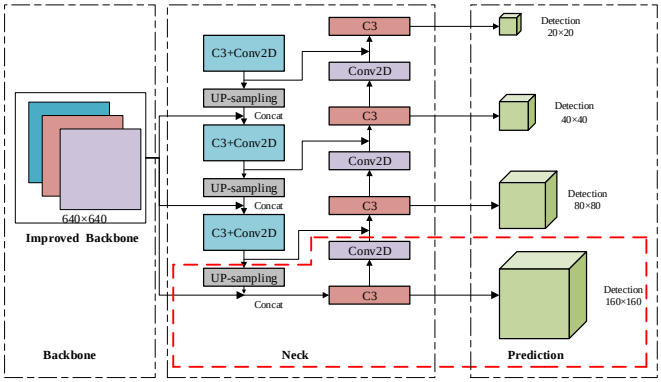


Figure 3: Overall network structure

4. Experiments

4.1 Experimental Configuration

The experimental environment is Ubuntu 20.04, IntelXeon(R) Silver 4214R CPU @ 2.40GHz x48 processor, 128GB RAM, NVIDIA Corporation TU102GL [Quadro RTX 8000] graphics card, CUDA11.4.2 accelerated GPU, and Python version 3.8.12, Python version is 3.8.12, and the deep learning framework is Pytorch 1.9.1.

In this paper, we use YOLOv5s as the benchmark network, trained under the same parameters with an initial learning rate of 0.01, a decay rate of 0.001, a momentum parameter of 0.937, a batch size of 8, and an epochs setting of 200, using stochastic gradient descent (SGD) [20] was used as the optimization function to train the model.

4.2 Experimental Dataset

In this paper, we built our own image dataset of the rebar cross-sections, and the data was provided by Anyang Iron and Steel Group. According to the needs of the production process, the acquisition equipment is a depth camera, and the images to be processed are those obtained after mapping the depth point cloud data to speed up the processing speed. The dataset contains a total of 458 images of rebar cross-sections of 28mm and 22mm, each containing 70~180 rebar cross-sections object, totaling 74260 rebar cross-sections. Figure 4 is a statistical plot of the label size of the dataset.

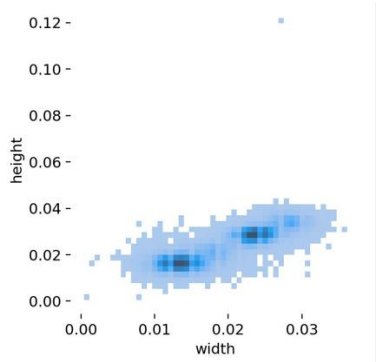
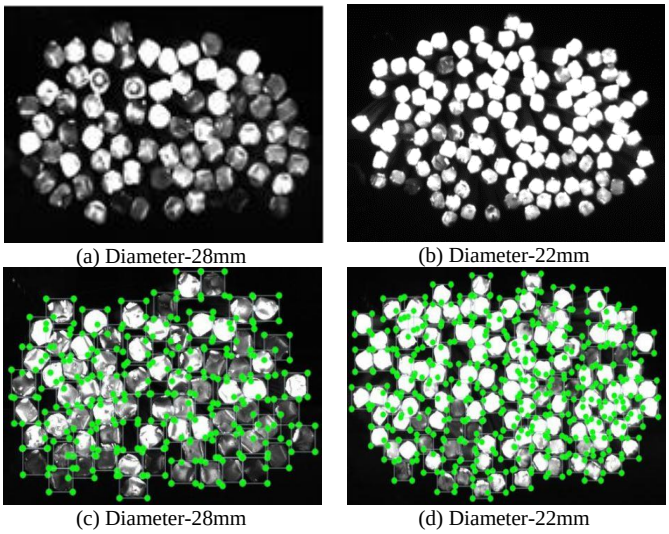


Figure 4: Dataset label size statistics

In Figure 4, the horizontal coordinate width is the ratio of object width to image width, and the vertical coordinate height is the ratio of object height to image height. the larger the value of width and height, the larger the percentage of the object in the image. As can be seen from Figure 4, the data in this paper are mainly small and medium-sized object data, which fits the research problem of small object detection in this paper. In addition, to ensure the generalization of the model, the data also contains some data of other size scales.

For the situation that the data sample size is insufficient, the images obtained after enhancing the data using horizontal flip, on-the-fly crop, and translation are 1374, and the total sample has 222780 rebar cross-sections object, which are labeled using LabelImg software, and the data are divided into a training set, validation set, and test set according to the ratio of 8:1:1. The two specifications and labeling examples of rebar section images are shown in Figure 5.



(a) Diameter-28mm (b) Diameter-22mm

(c) Diameter-28mm (d) Diameter-22mm

Figure 5: Dataset presentation and annotation

4.3 Evaluation Indicators

Precision (P), Recall (R), mean Average Precision (mAP@.5), frames per second, FPS), and other objective evaluation indicators to verify the performance of the model. mAP reflects the comprehensive detection capability of the algorithm for different categories of object. mAP@.5 represents the average detection accuracy of all object categories when the threshold of IoU is 0.5. The higher the threshold of IoU, the higher the requirement for the regression ability of the model. FPS refers to the number of images that can be detected per second, which is used to measure the real-time performance of model detection. The calculation of the above indexes is shown in Formula (3) ~ (5):

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$Pression = \frac{TP}{TP + FP} \tag{4}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AceP(i) \tag{5}$$

In the above equation, TP is the number of correct rebar detection, FN is the number of missed rebar detection, FP is the number of false rebar detection, N is the number of categories, and $AceP(i)$ is the average accuracy of the ith category.

4.4 Ablation Test

In order to verify the effectiveness of K (K-means clustering algorithm), E (ECANet) and S (small object detection layer) improvement points in improving the model detection effect, the ablation experiment was conducted based on YOLOv5s, using the same experimental equipment and data for testing, and P (Precision), R (Recall), mAP@.5 are selected as the experimental evaluation indicators. "-" means that the improvement strategy was not adopted, and "+" means that the improvement strategy was used, the experimental results of each model are shown in Table 2.

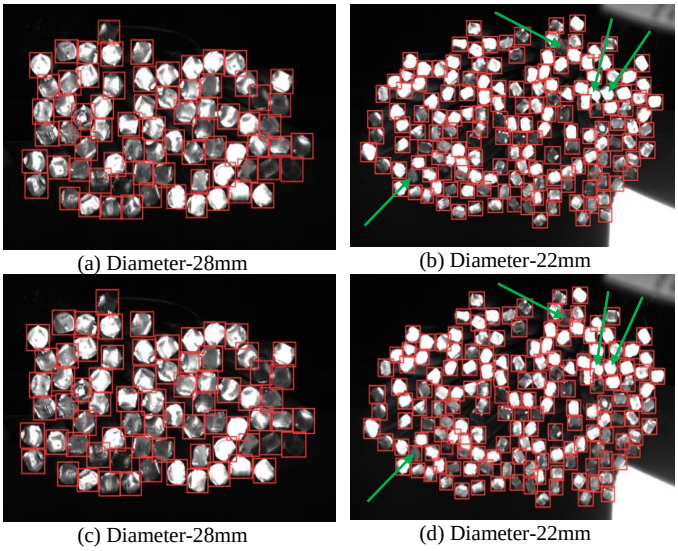
Table 2: Ablation experiments

Model	K	E	S	P(%)	R(%)	mAP@.5(%)
YOLOv5s	-	-	-	95.3	78.6	88.1
Improvement1	+	-	-	95.5	79.0	88.9
Improvement2	+	+	-	95.8	79.7	89.1
Improvement3	+	+	+	95.9	82.5	94.3

From Table 2, we can see that: from the experimental results of Improvement1, after re-clustering the size of the anchor box, there is a small improvement in all indicators; from the experimental results of Improvement2, the attention semantic compensation module brings a small performance improvement in all indicators compared with the original YOLOv5s algorithm, in which mAP@.5 is improved by 1%; comparing the experimental results of Improvement2 and Improvement3, we can see that by comparing the experimental results of Improvement2 and Improvement3, it can be seen that by increasing the detection layer and fusing the high-resolution information with the deep feature semantic information, all the indexes of the algorithm are also improved, among which the mAP@.5 reaches 94.3%, which

is 5.2% higher than that of Improvement2, and greatly reduces the missed detection rate of small object. From the four sets of ablation experiments, it can be seen that the combined effect of calculating the anchor box size by K-means clustering, adding the ECANet module, and adding new detection layers can effectively improve the effect of rebar cross-section detection.

To more intuitively demonstrate the performance of the method in this paper for the rebar cross-section detection task, the trained model is tested on the test set (confidence parameter conf=0.45, non-maximum suppression threshold IOU=0.6), and the detection results of the YOLOv5s algorithm and the method in this paper are compared, and the test results are shown in Figure 6. The rebar specifications in Figure 6 (a) and Figure 6 (c) are 28 mm; the rebar specifications in Figure 6 (b) and Figure 6 (d) are 22 mm. In Figure 6, (a) and (b) are the detection results on the original YOLOv5s algorithm, and (c) and (d) are the detection results of the method in this paper. As can be seen from the overall comparison graph, the detection performance of both the original algorithm and this our method is good when the rebar specification is 28mm, and all object can be detected correctly; when the rebar specification is 22mm, the original YOLOv5s algorithm has the problem of missed object, which are identified with green arrows in Figure 6(b), and Figure 6(d) is the detection effect graph of this ours method, both can correctly detect the missed object of the original algorithm The missed object are identified by green arrows in Figure 6(b). In summary, it can be seen that the improved detection algorithm is better than the YOLOv5s algorithm for the detection of rebar cross-sections in general.



(a) Diameter-28mm (b) Diameter-22mm

(c) Diameter-28mm (d) Diameter-22mm

Figure 6: Results before and after algorithm improvement

4.5 Comparison Experiment

To further verify the overall performance of the improved algorithm, eight sets of comparison experiments were conducted in this paper, and the precision (Precision), average precision (mAP@.5), frames per second (FPS), and model weights were used as evaluation indexes, and the experimental results are shown in Table 3, in which Faster R-CNN is a two-stage detection count, and SSD, RetinaNet, CenterNet, CornerNet, and YOLO series are one-stage detection algorithms.

Table 3: Comparison experiments

Model	P(%)	mAP@.5(%)	Size(MB)	FPS(frame)
Faster R-CNN	87.8	91.2	182.1	25
SSD	66.7	86.9	90.6	32
RetinaNet	65.8	72.5	118	18
CenterNet	56.6	73.9	154	27
YOLOv3	95.7	87.5	19.0	36
YOLOv4	76.6	92.0	488.4	34
YOLOv5s	95.3	88.1	14.4	52
Ours	93.8	94.3	16.7	40

From Table 3, we can see that: in terms of accuracy, the P of this our methods are all better than other object detection algorithms; in terms of average accuracy, the mAP@.5 of the method in this paper is 94.3%, which is 3.1%, 7.4%, 21.8%, 20.4%, 6.8%, 2.3%, and 6.2% higher compared to Faster R-CNN, SSD, RetinaNet, CenterNet, YOLOv3, YOLOv4 and YOLOv5s, respectively; in terms of model weight size, the model weights obtained from the training of the method in this paper are 16.7 MB, which is 1/11, 1/5, 1/7, 1/9, 7/8, 1/29 of the model weights of Faster R-CNN, SSD, RetinaNet, CenterNet, YOLOv3, and YOLOv4, respectively, although compared with the previous model of YOLOv5s weight files slightly increased, the mAP@.5 improved by 6.2%; in terms of frames per second transmission, the FPS of this our method is 40 frames, which is slightly inferior to YOLOv5s in terms of detection rate, but still satisfies the detection in real-time. In summary, the method in this paper achieves a balance of detection accuracy and detection speed, and the overall performance of the algorithm is better than other object detection algorithms, which can meet the basic requirements of rebar cross-section detection.

5. Conclusion

Rebar is an important material in the field of building construction, and it is an important task to inventory the quantity of reinforcing rebar accurately and quickly instead of manually inventorying it. In this paper, a YOLOv5s-based algorithm for rebar cross-section detection is proposed by combining the K-means clustering algorithm, attention mechanism, and multi-scale detection. The experimental results show that the average accuracy of the proposed method in this paper is substantially improved, while the comprehensive performance is better than other mainstream object detection models, which can improve the detection accuracy on the basis of satisfying the real-time rebar detection, and has a catalytic effect on the application of rebar detection in smart construction sites. Further lightweight of the deep structure of the network will be considered in the future to improve the detection speed on the basis of ensuring the accuracy of the model.

Acknowledgement

This paper is supported by the fund: 2022 Henan Province Key R&D and Promotion Special (Science and Technology Research) Project funding, Fund Number: 222102210131, Project Name: Intelligent analysis of wire harness terminal images based on deep learning.

References

[1] Xinman Z, Mei M, Tingting H, et al. Steel bars counting method based on image and video processing[C]//2017

International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS). IEEE, 2017: 304-309.
[2] Wu Y, Zhou X, Zhang Y. Steel bars counting and splitting method based on machine vision[C]//2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER). IEEE, 2015: 420-425.
[3] Zhao J, Xia X, Wang H, et al. Design of real-time steel bars recognition system based on machine vision[C]//2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC). IEEE, 2016, 1: 505-509.
[4] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.
[5] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 28.
[6] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//European conference on computer vision. Springer, Cham, 2016: 21-37.
[7] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.
[8] Duan K, Bai S, Xie L, et al. Centernet: Keypoint triplets for object detection[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 6569-6578.
[9] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.
[10] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.
[11] Zhu X, Lyu S, Wang X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 2778-2788.
[12] Wang H, Polden J, Jirgens J, et al. Automatic rebar counting using image processing and machine learning[C]//2019 IEEE 9th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER). IEEE, 2019: 900-904.
[13] Zhu Y, Tang C, Liu H, et al. End-face localization and segmentation of steel bar based on convolution neural network[J]. IEEE Access, 2020, 8: 74679-74690.
[14] Li Y, Lu Y, Chen J. A deep learning approach for real-time rebar counting on the construction site based on YOLOv3 detector[J]. Automation in Construction, 2021, 124: 103602.
[15] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context[C]//Computer Vision-ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. Springer International Publishing, 2014: 740-755.
[16] Wang Q, Wu B, Zhu P, et al. Supplementary material for 'ECA-Net: Efficient channel attention for deep convolutional neural networks[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, Seattle, WA, USA. 2020: 13-19.
[17] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.
[18] Liang Z, Shao J, Zhang D, et al. Small object detection using deep feature pyramid networks[C]//Pacific Rim Conference on Multimedia. Springer, Cham, 2018: 554-564.
[19] Deng Z, Sun H, Zhou S, et al. Multi-scale object detection in remote sensing imagery with convolutional neural networks[J]. ISPRS journal of photogrammetry and remote sensing, 2018, 145: 3-22.
[20] Ketkar N. Stochastic gradient descent[M]//Deep learning with Python. Apress, Berkeley, CA, 2017: 113-132.

Author Profile

Yanmei Zheng, female, born in 1984, Han nationality, graduated from the Institute of Computational Mathematics, Chinese Academy of Sciences in 2011, with a Ph.D. in Science. She worked at the Institute of Software, Chinese Academy of Sciences. He has presided over a key research project in Henan Province and participated in 5 projects of the National Natural Science Foundation of China, such as surface projects and major projects. He has published 12 papers in important journals and international conferences at home and abroad, and 5 papers have been indexed by SCI/EI.

Bibo Lu, Male, born in 1978, Ph.D., professor, master's supervisor, visiting scholar of Michigan State University, U.S.A. He graduated from Jilin University in June 2008 and mainly engaged in research on artificial intelligence, image video processing, and analysis. He has been in charge of two National Natural Science Foundation of China (NSFC) projects, three NSFC projects, and more than 10 provincial and ministerial projects, including the Local Cooperation Project of the National Scholarship Council, Henan Provincial Government Decision-making Research Bidding Project, University-Industry Cooperation and Collaborative Education Project of the Ministry of Education, Henan Provincial Education Department Key Science and Technology Research Project, Henan Science and Technology Think Tank Research Project, Henan Provincial Education Science Planning Project, etc. He has also participated in the completion of the Ministry of Education Science and Technology Research Key Project, the New Teacher Fund of the Ministry of Education, the Natural Science Fund of Henan Provincial Education Department, etc.