Article

Intelligent Detection of Rebar Size and Position Using Improved DeeplabV3+

Wei Chen [1], Xianglin Fu [1],*, Wanqing Chen [2] and Zijun Peng [1]

[1] College of Civil Engineering, Changsha University of Science and Technology, Changsha 410114, China; chenwei85chen@163.com (W.C.)

[2] Wuhan Yucheng Jiufang Construction Co., Wuhan 430050, China

* Correspondence: flyxl@stu.csust.edu.cn; Tel.: +86-19376852298

Abstract: For the development of reinforced concrete structures and infrastructure construction, traditional rebar checking and acceptance methods have shortcomings in terms of efficiency. The use of digital image processing technology cannot easily identify a rebar configuration with complex and diverse backgrounds. To solve this problem, an inspection method combining deep learning and digital image processing techniques is proposed using an improved DeeplabV3+ model to identify reinforcing bars, with the identification results subjected to digital image processing operations to obtain the size information of the reinforcing bar. The proposed method was validated through a field test. The results of the experiment indicated that the proposed model is more accurate than other models, with a mean Intersection over Union (mIoU), precision, recall, and F1 score reaching 94.62%, 97.42%, 96.95%, and 97.18%, respectively. Moreover, the accuracy of the dimension estimations for the test reinforcements met the engineering acceptance standards.

Keywords: intelligent detection; rebar size measurement; DeeplabV3+ model; attention mechanism

1. Introduction

During the construction of reinforced concrete structures and infrastructures, the acceptance of reinforcements before concealment is one of the key tasks for quality control in cast-in-place reinforced concrete structures. The traditional quality inspection methods involve the inspectors climbing the buildings and using tape measures and dial calipers to measure the size and position of the reinforcements. These approaches have disadvantages, such as high human resource and time consumption, and low work efficiency. Meanwhile, intelligent recognition methods based on traditional image processing face many challenges. Their image feature extraction relies mainly on manually designed feature extractors, requiring specialized image segmentation algorithms tailored to specific scenarios and tasks. These methods are unable to simultaneously address different scenarios, and detection accuracy is significantly affected by factors such as the background and lighting. Therefore, exploring a more efficient, intelligent, and versatile method for rebar detection on construction sites holds research significance and practical value.

Recently, various algorithms based on convolutional neural networks (CNNs) [1] have begun to be applied to the construction industry, such as target detection and semantic segmentation. Target detection is mainly divided into one-stage and two-stage detection models. The former has shown outstanding performance in terms of detection speed, which includes the Yolo series [2–8] and SSD [9], while the latter focuses on the accuracy of detection, which includes Faster R-CNN [10] and Mask R-CNN [11]. Semantic segmentation is mainly divided into segmentation methods based on region classification and pixel classification, the former methods including multi-scale patch aggregation (MPA) [12] and simultaneous detection and segmentation (SDS) [13], and the latter methods consisting of the Deeplab series [14–16] and SegNet [17].

Many attempts have been made to apply deep learning algorithms in the construction industry. Guo et al. [18]. used an improved Faster R-CNN algorithm to extract features of helmets using VGG16 as a feature extraction network, which can detect whether helmets were worn or not, with high accuracy. Chen et al. [19] used the lightweight network PP-LCNet as the backbone network of the YOLOv4 model, and used deep separable convolution to significantly improve the detection speed while ensuring the accuracy of helmet detection. Liu et al. [20] proposed a dense-end face detection algorithm based on YOLOV5, which reduced repetitive gradient information by introducing cross-stage connections and incorporating an attention mechanism to improve the feature extraction capability of the network. In terms of structural health monitoring, Park et al. [21] designed a real-time crack detection system using the YOLOV3-tiny model as well as laser sensors. Ruan et al. [22] used the DeeplabV3+ model combined with unmanned aerial vehicle inspection to identify concrete shedding and exposed reinforcements on concrete bridge surfaces. Ahmed et al. [23] used deep learning methods to process bridge GPR data for the rapid detection and localization of bridge deck reinforcements. Regarding rebar detection, Zheng et al. [24] proposed a multi-scale steel bar detection network RebarNet based on the YOLOv5 embedded attention mechanism, which can effectively reduce missed and false detections in real-time detection-based steel bar counting detection tasks. Shin et al. [25] developed an automated system based on convolutional neural network (CNN) computer vision technology for estimating the size of steel bars in bundled packaging and counting the quantity. Yan et al. [26] added bottom-up path and attention mechanisms to the Mask R-CNN model to obtain a more accurate rebar target detection. However, their subsequent treatment of rebars still relied on traditional image processing techniques, limiting its applicability to relatively narrow scenarios. An et al. [27] improved the Harris corner detection algorithm and combined it with a laser rangefinder to propose an image-based intelligent ranging system for measuring steel bar spacing. However, this method requires the manual setting of relevant parameters for different scenarios. While deep learning techniques have been extensively adopted in the construction industry, there remains a research gap in applying this technology to the detection of rebar engineering dimensions. Currently, there is still a lack of a fast, convenient, highly efficient, and widely applicable intelligent method for detecting rebar dimensions.

To solve the current application needs of intelligent inspection and to achieve the detection of a concealed reinforcement's configuration quickly and conveniently, this study innovatively combines smartphones with advanced deep learning technology to propose a new non-contact method for detecting the dimensions of rebars, without the need to set specific feature extractors based on specific scenes, making it more widely applicable, and there is very limited existing research in this area. It utilizes smartphones as capture devices to capture photos of rebars and then employs the DeeplabV3+ [28] model for automatic rebar configuration recognition. Compared to traditional manual measurement methods, the approach in this paper is more efficient, convenient, and rapid. Furthermore, this study applies advanced deep learning technology to the engineering domain, which is conducive to advancing the intelligence in the field of engineering. However, due to the original design intent of the DeeplabV3+ model not being focused on segmenting rebar objects, it faces the following issues in practical applications for rebar segmentation tasks:

(1) The segmentation of rebar edges is different, resulting in missed detections in local areas of rebars;

(2) The identification of the rebar intersections and discontinuous segmentation is incomplete;

(3) Due to the effects of the background and lighting, there are some instances where the background is mistakenly checked as rebar.

Therefore, to address the aforementioned issues, this study made targeted improvements to the network structure of the DeeplabV3+ model. It adopted ResNet50 as the backbone network for the enhanced model and incorporated an attention mechanism within ResNet50 to handle high-semantic-information feature layers. Simultaneously, the

dilation rates of the ASPP module were reduced to allow the model to more accurately capture information from targets of different sizes. The rebar segmentation configuration obtained from the improved model was further processed using image processing techniques, such as edge contour detection, to obtain the diameter and spacing of the rebars. The field test results showed that the improved DeeplabV3+ model could accurately identify the dimensional information of the rebar in line with acceptance criteria.

The overall structure of the study takes the form of six sections, including this introductory section. Section 2 is the methodology section, primarily introducing the foundational model DeeplabV3+ and analyzing its existing issues in detail. It also focuses on providing a comprehensive introduction to the improved DeeplabV3+ model. Section 3 offers a comprehensive explanation of the experimental parameters and evaluation metrics employed in this study. Section 4 presents experimental results and analyses, validating the effectiveness of the improvements through a comparative analysis of the results. Section 5 delves into the aspect of rebar size measurement. The final section, Section 6, provides a conclusive summary of this paper, outlining the limitations of this study and suggesting directions for future research.

2. Methodology

2.1. DeeplabV3+

To achieve high-precision segmentation and recognition of the rebars, a model with excellent performance in the field of semantic segmentation was selected as the basic algorithm of present study. The Deeplab series was developed on the basis of a fully convolutional network (FCN) [29], which combines deep convolutional networks and conditional random fields (CRFs) and uses Atrous convolution to effectively enlarge the field of view. In 2018, DeeplabV3+, the latest model in the Deeplab series introduced by the Google team, achieved an outstanding result of up to an 87.8% mIoU on the PASCAL VOC-2012 dataset, with segmentation results that far outperformed other models. Therefore, DeeplabV3+ was chosen as the base algorithm. The network structure is shown in Figure 1.
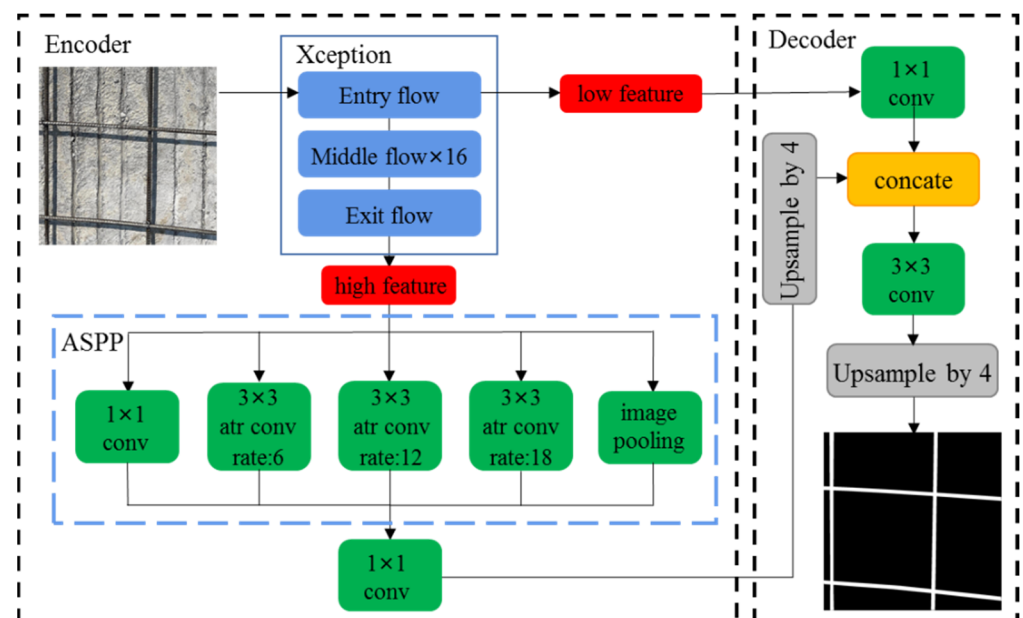

Figure 1. Structure of DeeplabV3+.

The encoder part of the network consists of a backbone and an atrous spatial pyramid pooling (ASPP). A residual mechanism was used to connect separable convolutions, which was divided into three flow steps: entry flow, middle flow, and exit flow. The input image passed the backbone network to extract one high-level feature map and one low-level feature map. The high-semantic-information feature map was passed to the ASPP

module, which was divided into five parallel operations: normal convolution (1×1), three atrous convolution groups with different rates, and image pooling. This approach expanded the convolution field, performed multi-scale feature fusion to reduce the effects of inconsistent input scales, and finally performed a 1×1 convolution operation before passing to the decoder.

In the decoder section, the ASPP-processed feature maps were bilinearly unsampled by a factor of 4 and were then concatenated with the low-semantic-information feature maps output by the backbone network using a 1×1 convolution to the adjust channel. After a tandem operation, the 3×3 convolution kernel and bilinear upsampling by a factor of 4 were applied to obtain the final segmentation map.

While this model has achieved excellent results on the PASCAL VOC-2012 dataset, its original design was not specifically tailored for rebar images, and its performance in rebar image recognition is not ideal. In order to better adapt the model to the task of rebar image segmentation, improvements and optimizations were needed.

2.2. Improved DeeplabV3+

Based on rebar feature information and segmentation detail information, DeeplabV3+ was improved, as shown in Figure 2.
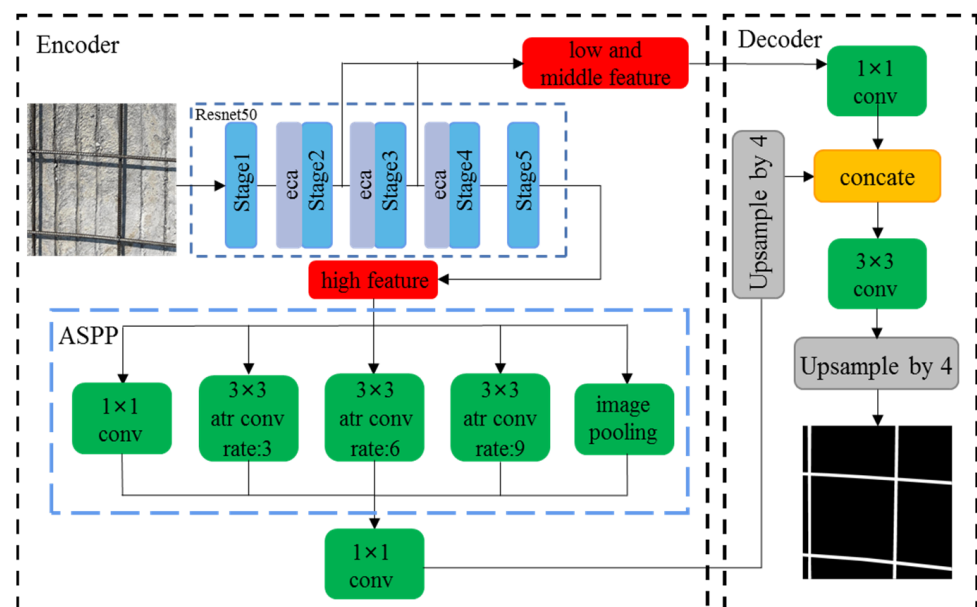

Figure 2. Structure of improved DeeplabV3+.

(1) The detection of rebars was the target of present study. To reduce the complexity of original DeeplabV3+ model, ResNet50 was selected as the backbone extraction network;

(2) According to the feature information and distribution pattern of the rebar dataset, an efficient attention module was added to the backbone network to optimize the feature extraction pattern of the network, as well as to deepen the sensitivity of the network to identify rebar. Thus, the redundant operations of the network to extract non-object features could be avoided;

(3) To solve the problem of incomplete edge information and loss of detailed information in the segmentation effect of the original DeeplabV3+ model, the convolutional dilation rate and convolutional density of the cavity convolution in the ASPP module were changed from 6, 12, and 18 to 3, 6, and 9, so there was no cavity loss when performing scale fusion.

2.2.1. Improvement of Backbone

The deep residual network was proposed by He et al. [30]. In general, the number of neural network layers can be increased to make the network's prediction better, but in practice, it has been found that deepening the number of a neural network layer creates the problem of gradient disappearance or gradient explosion. This can cause the network to saturate or decline in accuracy on the training set, resulting in a degradation in model performance. The advent of deep residual networks has enabled these problems to be solved. A residual block was added to a deep residual network, as shown in Figure 3, where the network obtains an expected value of $H(x)$ when the input is x. Using the characteristics of residual learning, the residual $F(x) = H(x) - x$ is defined, so the original expected value becomes $F(x) + x$, making the network easier to optimize.
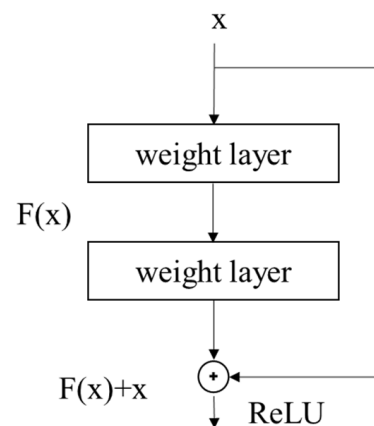

Figure 3. Schematic diagram of the residual blocks.

The original DeeplabV3+ model has a complex and diverse dataset. Thus, the adopted backbone feature extraction network required a more complex feature extraction model, such as the structure of the original backbone network Xception, which uses a residual mechanism to connect the depthwise separable convolutions. This network has a complex structure with a number of parameters, which is not suitable for the detection task of rebar. To address this problem, ResNet50, which also has a residual mechanism, was chosen as the part of the backbone extraction network for the DeeplabV3+ model. More simple direct connections were added to the network, which had the advantages of having fewer parameters, being easier to train, and faster convergence in comparison with Xception, to fully take into account the feature scale matching problem and to improve the model inference speed.

2.2.2. Efficient Channel Attention Module

The efficient channel attention module (ECA) [31] is an adaptation of the SENet [32], which uses one-dimensional convolution to replace the fully connected layer, to solve the problem that the dimensionality reduction operation adversely affects the correlation between the learning channels of the network while sharing information using local cross-channel to reduce the complexity of the network. The mechanism can be implemented as follows: the feature dimensions [H, W, C] of the input feature layer are pooled using the average to obtain the aggregated features [1, 1, C], and then the correlation between channels is obtained by performing a one-dimensional convolution of size k to generate the channel weights w. Finally, the weights are multiplied by the original input feature layer to form a new feature layer. The network structure of ECA is shown in Figure 4.
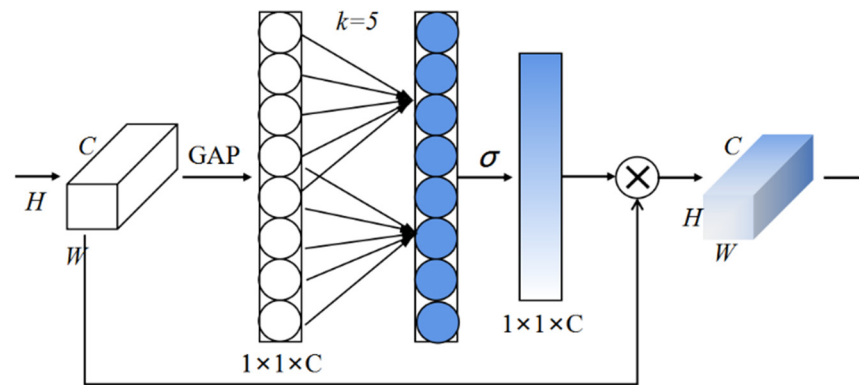
Figure 4. Efficient channel attention module.

The size k of the one-dimensional convolution kernel in Figure 4 was determined adaptively using the number of channels, C, as follows:

$$k = |t|_{odd} = \left| \frac{log_2 C + b}{\gamma} \right|_{odd},\qquad(1)$$

where k denotes the convolution kernel size, C denotes the number of channels, $|t|_{odd}$ denotes the nearest *odd* number to t, and γ and b were set to fixed values of 2 and 1, respectively. After determining k, the channel weights w can be expressed as

$$w = \sigma[C1D_k(\mathrm{y})],\qquad(2)$$

where C1D denotes a one-dimensional convolution, σ is the sigmoid activation function, and y is the channel aggregation feature.

Although replacing the backbone network could effectively adjust the backbone network parameters, the sensitivity to feature information was not enhanced. To solve this problem, some improvements were made to the backbone network, as shown in Figure 5: the ECA was added to Resnet50 to enhance the sensitivity of the network to feature information and to reduce the unnecessary information processing procedure.
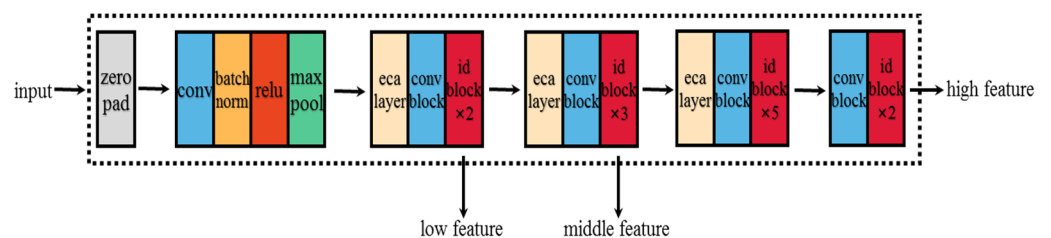

Figure 5. Structure of the improved Resnet50.

2.2.3. Adjusting Atrous Convolution

The ASPP (Atrous Spatial Pyramid Pooling, Figure 6) module is a critical component of the Deeplabv3+ model used to capture multi-scale contextual information on top of the feature maps extracted from the backbone network. Atrous convolution forms the core of the ASPP module. Atrous convolution with a dilation rate of r expands the receptive field of the original $n \times n$ convolution kernel to $N = n + (n-1)(r-1)$ by adding $r-1$ zeros between adjacent filter values in each spatial dimension, which ensures the output feature map size. Figure 7 shows the convolution schematic. The ASPP module uses parallel atrous convolutions with four different atrous rates to segment objects at different scales in combination with image-level features. The global contextual information is fused by applying average pooling on top of the last feature map of the backbone network. The results of each operation along the channel are concatenated, and a 1×1 convolution is performed to obtain the output.
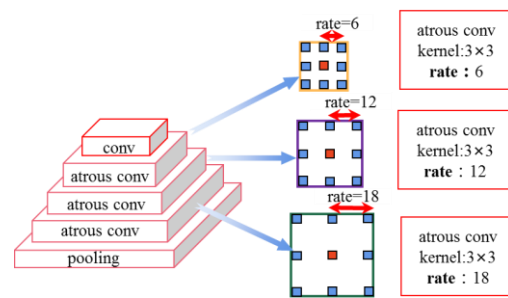
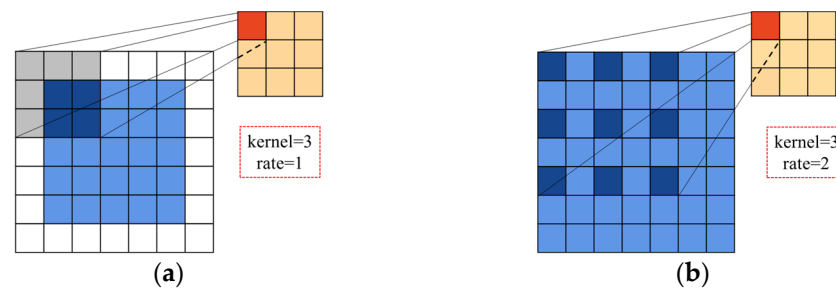**Figure 6.** Schematic of atrous spatial pyramid pooling (ASPP).

Figure 7. (**a**) Normal convolution; (**b**) atrous convolution.

The dilation rates for the parallel atrous convolutions of the ASPP module in the original DeeplabV3+ were 6, 12, and 18. When the backbone network proceeded with feature extraction, the feature map resolution gradually decreased, and the dilation rates of 6, 12, and 18 could not effectively extract features from multi-scale images about rebar if a smaller void rate was not set. This led to a lack of ability to segment small targets. Considering the existence of rebar with multiple diameters, to extract multi-scale image features more effectively and to improve the segmentation capability of targets with different size rebars, the dilation rates of the atrous convolution were adjusted to be 3, 6, and 9.

3. Datasets and Experimental Conditions

3.1. Datasets

The main difficulties in the detection of rebar on a construction site are as follows: the background of the rebar is complex, and there is a possibility that the background color is similar to that of the rebar. Furthermore, pipes, pads, and other tools from utilities, as well as shadows from light, are present with the rebar at construction site. As there is no public rebar dataset, we used smartphones to capture a large number of rebar images at the construction site. To prevent training overfitting and to improve the accuracy of the model, a data enhancement approach was adopted to expand the dataset by processing the images using flipping, exposure and perspective change, and motion blurring to enrich the diversity of the training sample. The final dataset contained 3130 images in total, which was divided into three groups at a ratio of 8:1:1, i.e., 2504 training images, 313 validation images, and 313 test images.

3.2. Experimental Conditions

The data were trained using the improved DeeplabV3+ model. The input image size was 512×512, the batch size was eight, the two training categories were set (i.e., rebar and background), the weight decay index was 0.001, the learning rate was 0.0001, and the number of iterations was set as 100. The specific training environment is shown in Table 1.

Table 1. Experimental environment.

Configuration	Parameter
CPU	Intel Xeon E5-2686 v4
GPU	NVIDIA GeForce RTX 3080 TI
Development environment	Keras 2.3.1, TensorFlow 2.6, CUDA 11.2, cuDNN 8.0
Operating system	Ubuntu 18.04

3.3. Evaluation Index

Semantic segmentation was performed as a pixel-level classification. To evaluate the performance of rebar segmentation algorithm intuitively and quantitatively, the mean Intersection over Union (mIoU), precision, recall and F1 score were used as evaluation metrics, which are defined with respect to the pixel evaluation category. The corresponding formulas are as follows:

$$Precision = \frac{TP}{FP + TP}, \tag{3}$$

$$Recall = \frac{TP}{FN + TP}, \tag{4}$$

$$mIoU = \frac{1}{k}\sum_{i=0}^{k}\frac{TP}{FN + FP + TP}, \tag{5}$$

$$F1\_score = \frac{2TP}{2TP + FP + FN} \tag{6}$$

where *TP* denotes the number of pixels of segmented rebar, *FP* denotes the number of pixels of segmented background, *FN* denotes the number of pixels of unsegmented rebar, and *k* represents the number of categories.

4. Results and Analysis

An independent comparative analysis on the test results was conducted in terms of three aspects: backbone network, adding attention mechanism, and atrous convolutional adjustment, to validate the effectiveness of the improved model.

4.1. Performance Comparison of Improved Backbone

Using the original DeeplabV3+ model, the detection of the rebar often resulted in missed and false results. The main reason is that the features of the targets were not sufficiently utilized by the backbone network, and the detection scale did not match well with the target scale. Based on the distribution characteristics of steel rebars, to reduce the computational time, Resnet50 was chosen as the backbone network. In such a network, the tail pooling layer and the fully connected layer were cut and an attention mechanism was introduced before the different convolutional stages to improve the ability to extract feature information. To verify the effectiveness of the improved backbone network, ablation experiments were conducted with Xception, Mobilenetv2, Resnet50, and Resnet101. The experimental results are shown in Table 2.

Table 2. Comparison experiment of the backbone network.

Method	Backbone	Size	Param (M)	mIoU (%)	Precision (%)	Recall (%)	F1_Score (%)
original DeeplabV3+	Xception		42.1	89.66	97.20	91.91	94.48
	Mobilenetv2		2.7	84.12	94.88	89.08	91.89
	Resnet50	512 × 512	26.9	88.58	96.16	90.43	93.21
	Resnet101		45.9	90.11	94.42	93.91	94.16
	Ours		27.4	92.98	97.32	95.29	96.29

With the same input data, the improved backbone network resulted in a 3.32% improvement over Xception, 4.4% over Resnet50, 8.86% over Mobilenetv2, and 2.87% over Resnet101 in terms of the mIoU. For the precision and recall, the detection performance of the improved backbone network was also improved compared to those of other backbone networks. Although the precision was only slightly improved by 0.12%, the recall was improved by 3.38% in comparison with Xception. The selection of Resnet50 and the combination of the attention mechanism could effectively reduce the number of network parameters and improve the detection performance.

4.2. Ablation Experiment of Improved Module

To demonstrate the effectiveness of each improvement module, ablation experiments were conducted. The network was trained and evaluated using the corresponding datasets. The training parameters and loss functions of each module were consistent. The comparison of the experimental results is shown in Table 3.

Table 3. Ablation experiments of improved module.

Serial Number	IB *	AT *	AC *	Param (M)	mIoU (%)	Precision (%)	Recall (%)	F1_Score (%)
1				42.1	89.66	97.20	91.91	94.48
2	√			26.9	88.58	96.16	90.43	93.21
3	√	√		27.4	92.98	97.32	95.29	96.29
4	√	√	√	27.4	94.62	97.42	96.95	97.18

IB *: improved backbone network; AT *: addition of an efficient channel attention module; AC *: adjustment of atrous convolution.

After replacing the backbone network, the mIoU value decreased from 89.66% to 88.53%, and the precision and recall values decreased slightly, mainly because the replaced simpler backbone network reduced the complexity of the model and resulted in a significant reduction in the number of parameters. When an efficient channel attention module was introduced to improve the model's sensitivity to rebar information, The mIoU value increased from 88.53% to 92.98%, the precision value increased from 96.16% to 97.32%, the recall value increased from 90.43% to 95.29%, and the F1 score value increased from 93.21% to 96.29%. Finally, when the atrous convolution was adjusted the mIoU value reached 94.62%, which was 4.96% higher than that of the original model, and the precision, recall and F1 score also had some improvement, with the values reaching 97.42%, 96.95%, and 97.18%, respectively, while the number of parameters decreased to 27.4 M. The experimental results showed that the improved DeeplabV3+ model could effectively improve the detection efficiency of the rebar images.

4.3. Comparative Results of Different Models

Other advanced models, including FCN, PSPNet [33], U-Net [34], U-Net++ [35], SegNet, and R2U-Net [36] were chosen for a comparison. Table 4 presents the diverse performance of these eight models across four evaluation metric categories, along with the average inference time for each model on a single image. Our proposed method was outstanding in terms of each metric category. To illustrate the superiority of the proposed method more intuitively, four images were randomly selected from the test set, and the detection results of the above methods for rebar image segmentation were compared, as shown in Figure 8. It was found that our proposed model showed better segmentation results than the other methods when the images contained shadows, occlusion, and rebar intersection at the rebar edges.

Table 4. Evaluation of different models for rebar detection.

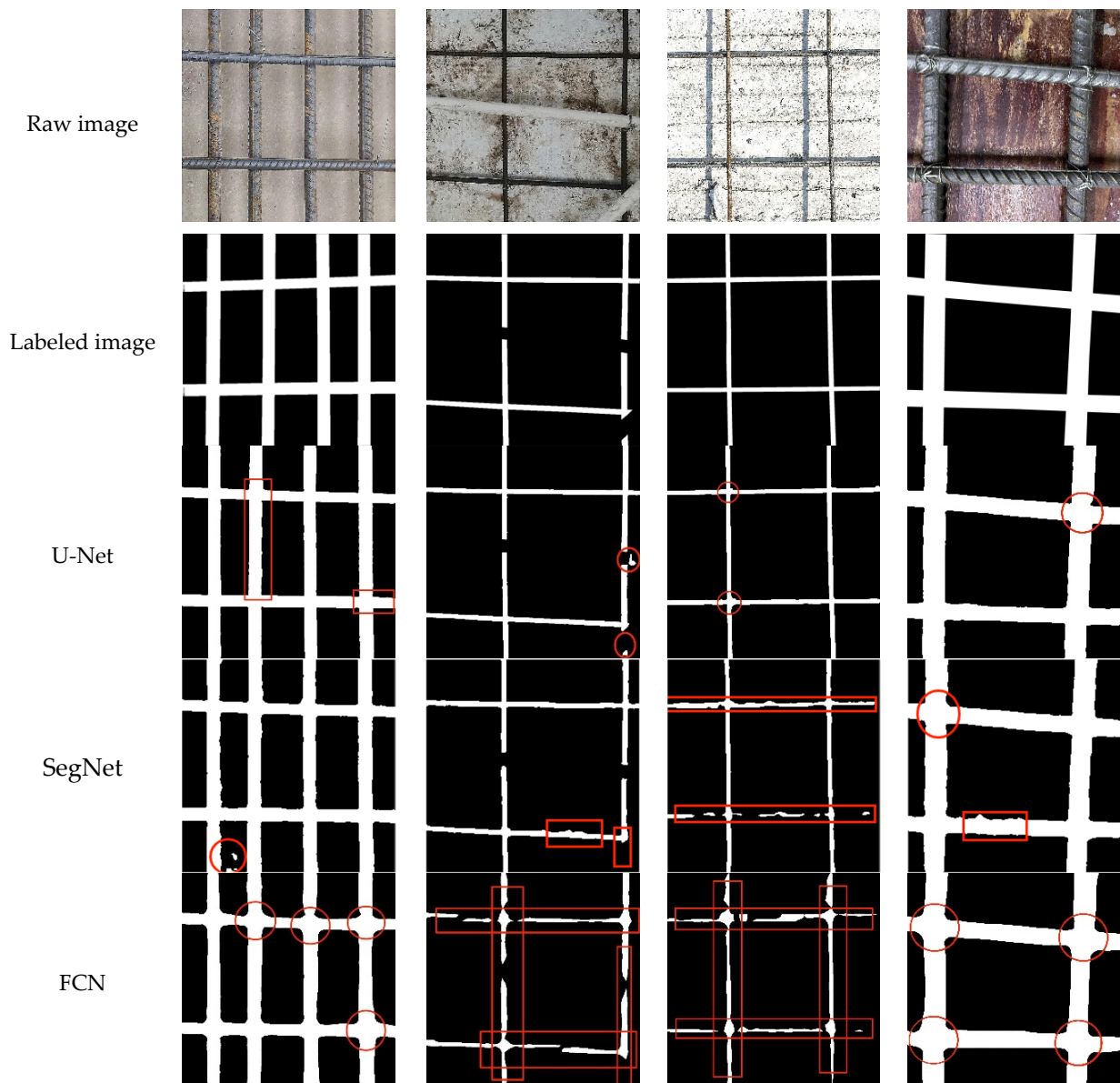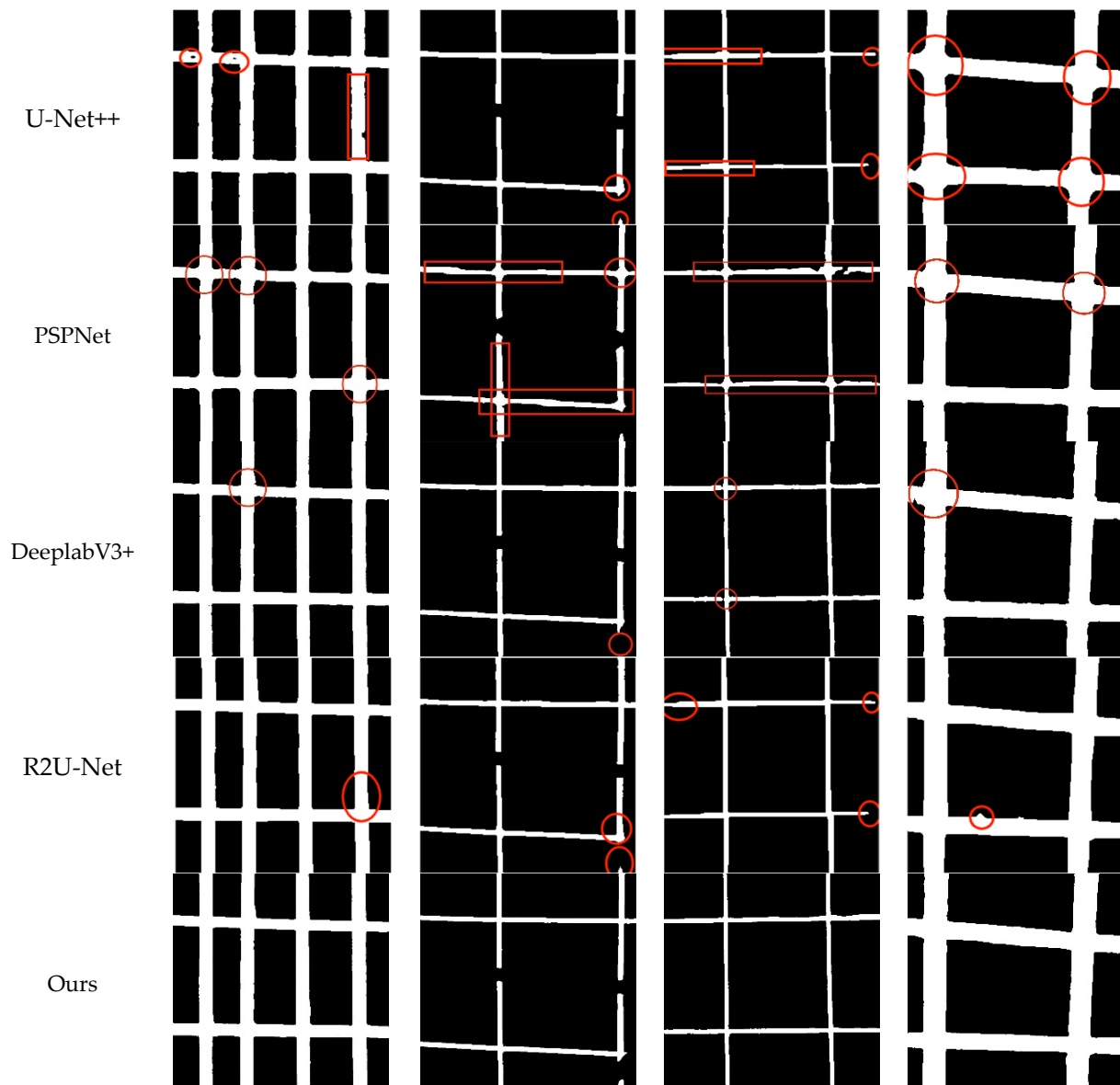| Model	mIoU (%)	Precision (%)	Recall (%)	F1_Score (%)	Time (s/Item)
U-Net	92.88	96.42	95.98	96.20	2.37
SegNet	86.72	90.87	88.52	89.68	2.14
FCN	82.30	88.43	85.19	86.78	2.72
U-Net++	92.81	96.56	95.67	96.11	2.11
PSPNet	81.97	91.81	87.42	89.56	2.25
Deeplab v3+	89.66	97.20	91.91	94.48	1.86
R2U-Net	92.98	96.68	95.48	96.08	1.42
Ours	94.62	97.42	96.95	97.18	1.21

Figure 8. *Cont.*

Figure 8. Visualization of segmentation results of each model (the red circles and boxes in the figures represent inaccuracies in the segmentation results).

5. Rebar Size Measurements

After an accurate rebar segmentation configuration was obtained using the improved DeeplabV3+ model, Canny operator edge detection was applied to process the segmentation configuration to accurately locate the rebar edge information. Then, the edge contour lines were determined for rebar dimensional measurements. As shown in Figure 9, the blue lines show the result of the line fitting. The green box was the inner contour of the rebar grid, and the purple line shows the rebar spacing. To convert the image size into an actual one, rebar with a known diameter was placed on the photograph or the rebar size specification was obtained from the design drawing.

A variety of different diameters of rebar was selected including 8, 10, 12, 14, 16, and 20 mm. For each diameter of rebar, eight isometric positions were selected to measure the diameter, and the average value of the diameter was obtained. Both the original and the improved DeeplabV3 + model were used to identify the rebars and obtain their diameter. Table 5 shows the results of the diameter detection before and after the improvement

of DeeplabV3+. According to the requirements of the rebar diameter measurements, a diameter error should be within 0.8 mm.
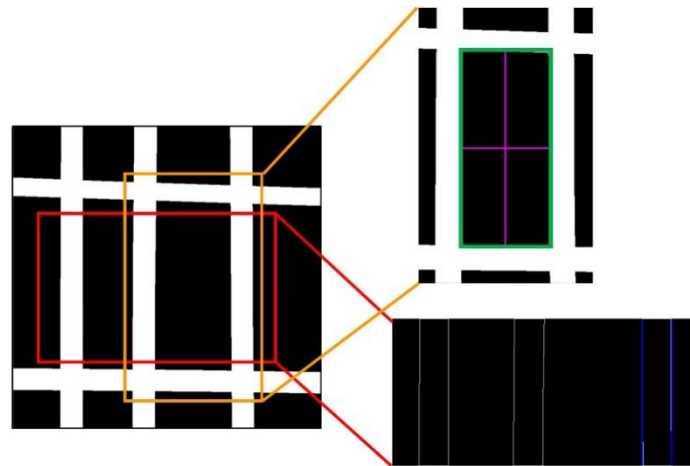

Figure 9. Schematic diagram of the diameter and spacing of rebar.

Table 5. Comparison of diameter detection results using original and improved DeeplabV3+.

Type	Real Diameter /mm	Model	Test Results /mm	Error /mm	Qualified or Not
8	8.32	DeeplabV3+	9.21	+0.89	No
		Ours	7.61	−0.71	Yes
12	11.86	DeeplabV3+	13.12	+1.26	No
		Ours	12.44	+0.58	Yes
14	14.16	DeeplabV3+	15.05	+0.89	No
		Ours	14.55	+0.39	Yes
16	15.94	DeeplabV3+	17.17	+1.23	No
		Ours	15.48	−0.66	Yes
18	18.02	DeeplabV3+	18.93	+0.91	No
		Ours	18.44	+0.42	Yes
20	19.92	DeeplabV3+	21.04	+1.12	No
		Ours	20.68	+0.76	Yes

From Table 5, the measurement error using original DeeplabV3+ for the above six different types of steel bars was greater than 0.8 mm, which did not meet the diameter acceptance requirements. Meanwhile, the absolute values of the diameter detection errors using our improved DeeplabV3+ for six types of steel bars were 0.71, 0.58, 0.39, 0.66, 0.42, and 0.76 mm, all of which were less than 0.8 mm and met the requirements of the rebar diameter acceptance.

Six rebar grids were randomly selected for spacing measurements using a ruler. For each grid, five isometric positions were selected and averaged for the spacing measurement. The spacings of the selected six locations were detected using the original and the improved DeeplabV3+ model to record the measured values and errors. According to the requirements of spacing measurements, a spacing error should be within 10 mm. The spacing detection results using the original and the improved DeeplabV3+ model are shown in Table 6. The results showed that the minimum value of the spacing detection error of the model before the improvement was 14.25 mm, which did not meet the spacing acceptance requirement, while the maximum value of the error after the improvement of the model was 7.27 mm, which was less than 10 mm and met the spacing acceptance requirement. Therefore, the detection accuracy of our proposed improved DeeplabV3+ model can meet the requirements of rebar detection acceptance.

Table 6. Comparison of spacing detection results using original and improved DeeplabV3+.

Number	Real Spacing /mm	Model	Test Results /mm	Error /mm	Qualified or Not
1	203.5	DeeplabV3+	184.4	−19.1	No
		Ours	196.2	−7.3	Yes
2	202.4	DeeplabV3+	186.2	−16.2	No
		Ours	199.6	−2.8	Yes
3	201.9	DeeplabV3+	181.4	−15.5	No
		Ours	199.3	−5.6	Yes
4	201.6	DeeplabV3+	183.1	−18.5	No
		Ours	195.0	−6.6	Yes
5	206.8	DeeplabV3+	184.4	−12.4	No
		Ours	199.4	−7.4	Yes
6	202.3	DeeplabV3+	188.1	−14.2	No
		Ours	197.2	−5.1	Yes

6. Concluding Remarks

To address the challenge of rebar size detection in construction quality acceptance, an improved DeeplabV3+ model combined with image processing techniques was proposed for the fast identification and measurement of rebar. ResNet50 was used as the backbone network instead of Xception for multi-scale feature fusion. The dilation rate and density of the atrous convolution in the ASPP was adjusted, and an attention mechanism was added to obtain a rebar segmentation configuration with high accuracy. The recognized segmentation configuration was then used for image processing, such as edge detection, to obtain an accurate measurement of the rebar's size and spacing.

In comparison with the original DeeplabV3+ model, our proposed model improved the mIoU, precision, recall, and F1 score by 4.96%, 0.22%, 5.04%, and 2.70%, respectively, and the number of parameters was reduced from 42.1 M to 27.4 M. Moreover, in comparison with other algorithms, our proposed model had higher detection accuracy and shorter inference time for a single image. Furthermore, the diameter and spacing detection accuracy of steel rebars reached the standard of acceptance; thus, it can be effectively applied for the intelligent detection of steel rebars.

However, our research method also has certain limitations. The images in this research dataset were collected under clear or dry weather conditions, without considering the interference caused by rainy weather, which would result in wet backgrounds on the template that could lead to reflections, as well as variations in lighting at different times. When applied in situations of excessive brightness or darkness, as well as backgrounds with wet reflections, this method led to significant errors in segmenting rebar images. In future studies, we plan to expand the dataset by capturing steel images in different weather conditions and at various times. We will also use data augmentation algorithms to enhance image features, mitigate the impact of weather and lighting factors, and improve the model's robustness and generalization capabilities.

Author Contributions: Conceptualization, W.C. (Wei Chen); data curation, W.C. (Wei Chen), X.F. and Z.P.; formal analysis, W.C. (Wei Chen), X.F. and W.C. (Wanqing Chen); funding acquisition, W.C. (Wei Chen); methodology, W.C. (Wei Chen), X.F. and W.C. (Wanqing Chen); project administration, W.C. (Wei Chen); resources, W.C. (Wei Chen); software, W.C. (Wei Chen) and X.F.; supervision, W.C. (Wanqing Chen); validation, W.C. (Wei Chen) and X.F.; writing—original draft, W.C. (Wei Chen), X.F. and Z.P.; writing—review and editing, W.C. (Wei Chen), X.F. and W.C. (Wanqing Chen). All authors have read and agreed to the published version of the manuscript.

References

1. Saon, G.; Picheny, M. Recent advances in conversational speech recognition using convolutional and recurrent neural networks. *IBM J. Res. Dev.* **2017**, *61*, 1:1–1:10. [CrossRef]
2. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. 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, 26 June–1 July 2016; pp. 779–788.
3. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
4. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
5. Bochkovskiy, A.; Wang, C.; Liao, H.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2018**, arXiv:2004.10934.
6. Zhu, X.; Lyu, S.; Wang, X. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, Virtual, 11–17 October 2021.
7. Li, C.; Li, L.; Jiang, H. YOLOv6: A single-stage object detection framework for industrial applications. *arXiv* **2022**, arXiv:2209.02976.
8. Wang, C.; Bochkovskiy, A.; Liao, H.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023.
9. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.; Berg, A.C. SSD: Single shot multibox detector. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2016; pp. 21–37.
10. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. In Proceedings of the Advances in Neural Information Processing Systems 28 (NIPS 2015), Montreal, QC, Canada, 7–12 December 2015; Volume 28.
11. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
12. Liu, S.; Qi, X.; Shi, J.; Zhang, H.; Jia, J. Multi-scale patch aggregation (mpa) for simultaneous detection and segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 3141–3149.
13. Hariharan, B.; Arbeláez, P.; Girshick, R.; Malik, J. Simultaneous detection and segmentation. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2014; pp. 297–312.
14. Chen, L.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv* **2014**, arXiv:1412.7062.
15. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [CrossRef] [PubMed]
16. Chen, L.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587.
17. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *39*, 2481–2495. [CrossRef] [PubMed]
18. Guo, S.; Li, D.; Wang, Z.; Zhou, X. Safety helmet detection method based on Faster R-CNN. In *International Conference on Artificial Intelligence and Security*; Springer: Singapore, 2020; pp. 423–434.
19. Chen, J.; Deng, S.; Wang, P.; Huang, X.; Liu, Y. Lightweight Helmet Detection Algorithm Using an Improved YOLOv4. *Sensors* **2023**, *23*, 1256. [CrossRef] [PubMed]
20. Liu, H.; Xu, K. Densely End Face Detection Network for Counting Bundled Steel Bars Based on YoloV5. In *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*; Springer: Cham, Switzerland, 2021; pp. 293–303.
21. Park, S.E.; Eem, S.-H.; Jeon, H. Concrete crack detection and quantification using deep learning and structured light. *Constr. Build. Mater.* **2020**, *252*, 119096. [CrossRef]

22. Ruan, X.; Wang, B.; Wu, J. Identification of SpaUed Concrete and Exposed Reinforcement in Reinforced Concrete Bridge Based on Deep Learning. *World Bridges* **2020**, *48*, 88–92.
23. Ahmed, H.; Le, C.P.; La, H.M. Pixel-level classification for bridge deck rebar detection and localization using multi-stage deep encoder-decoder network. *Dev. Built Environ.* **2023**, *14*, 100132. [CrossRef]
24. Zheng, Y.; Zhou, G.; Lu, B. A Multi-Scale Rebar Detection Network with an Embedded Attention Mechanism. *Appl. Sci.* **2023**, *13*, 8233. [CrossRef]
25. Shin, Y.; Heo, S.; Han, S.; Kim, J.; Na, S. An Image-Based Steel Rebar Size Estimation and Counting Method Using a Convolutional Neural Network Combined with Homography. *Buildings* **2021**, *11*, 463. [CrossRef]
26. Yan, T.; Ma, X.; Rao, Y.; Du, Y. Rebar size detection algorithm for intelligent construction supervision based on improved Mask R-CNN. *Comput. Eng.* **2021**, *47*, 274–281.
27. An, M.; Kang, D. The distance measurement based on corner detection for rebar spacing in engineering images. *J. Supercomput.* **2022**, *78*, 12380–12393. [CrossRef]
28. Chen, L.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
29. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
30. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
31. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 14–19 June 2020; pp. 13–19.
32. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
33. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
34. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2015; pp. 234–241.
35. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Proceedings of the 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, 20 September 2018*; Springer: Cham, Switzerland; pp. 3–11.
36. Alom, M.Z.; Hasan, M.; Yakopcic, C. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. *arXiv* **2018**, arXiv:1802.06955.