Open Access Library Journal
2019, Volume 6, e5784
ISSN Online: 2333-9721
ISSN Print: 2333-9705

Quantity Detection of Steel Bars Based on Deep Learning

Han Yang, Chenghua Fu

College of Mechanical Engineering of Sichuan University of Science & Engineering, Zigong, China
Email: 1594737268@qq.com

How to cite this paper: Yang, H. and Fu, C.H. (2019) Quantity Detection of Steel Bars Based on Deep Learning. *Open Access Library Journal*, **6**: e5784.
https://doi.org/10.4236/oalib.1105784

Received: September 12, 2019
Accepted: October 7, 2019
Published: October 10, 2019

Abstract

In the actual production environment, the number of steel bars in the construction site is mainly counted manually. For the special task of steel bar detection, a detection and counting method based on depth learning is proposed. The method is applied to the actual production environment instead of the traditional time-consuming and labor-consuming manual counting method. By comparing the traditional detection algorithm with the one-stage and two-stage detection in depth learning. After the algorithm and considering the efficiency of the model, the improved detection algorithm is proposed to adapt to the special task of steel bar detection. In the final evaluation index, the improved one-stage detection algorithm is superior to the improved detection algorithm in the special task of steel bar detection, showing the improvement of performance, and compared with the single-stage detection algorithm. The law has also been improved to a certain extent.

Subject Areas

Mechanical Engineering

Keywords

Deep Learning, Object Detection, Steel Bar Detection

1. Introduction

Target detection is a basic task in computer vision. Its purpose is to detect the type of object and the position of the object in a video, and the object detection is applied to image retrieval, pedestrian detection, tracking, etc. In the sub-task of computer vision, with the craze of deep learning, a large number of products using deep learning technology have gradually entered the field of vision, and are widely used in intelligent security, such as face recognition gates, vehicle identification systems, etc. Therefore, integrating deep learning technology into industrial production, improving the efficiency of traditional industrial production, and reducing the repetitive labor force, and liberating people from the repetitive workforce are a very worthwhile research and application. Therefore, based on the target detection, the target detection algorithm is applied to a specific task, such as the detection of the number of steel bars, which are a major work of this paper.

2. Background Introduction

2.1. Traditional Detection Method

The traditional target detection algorithm consists of the extraction of candidate regions of the target and the classification of the candidate regions. Therefore, the traditional target detection algorithm differs by the difference between the two parts, wherein for the target candidate region extraction part, It is divided into sliding window based detection such as DPM [1], and detection based on texture feature extraction, such as Selective Search [2], and the classification of the extracted candidate regions, such as AdaBoost [3], SVM [4], Decision Tree [5], Random Forest [6], etc. structure.

The traditional detection method can meet the requirements of the application to a certain extent, but with the improvement of the detection accuracy and the real-time detection requirements in the industry, the traditional detection algorithm usually has high time complexity and the characteristics of the manual design are low. Hierarchical, cannot express a large number of multi-category goals, not very robust. Therefore, the industry needs a more efficient detection algorithm to replace the traditional detection algorithm.

2.2. Development of Deep Learning

After 2012, Alexnet [7] and GoogleNet [8] greatly improved the convolutional neural network. After the effect of the Imagenet [9] competition, many researchers applied convolutional neural networks to all directions of traditional computer vision problems and made great breakthroughs. Therefore, convolutional neural networks have become very effective tools for solving computer vision problems.

The algorithms based on meaningful learning target detection can be divided into two main categories: two-stage detection and single-stage detection. The former refers to a detection algorithm that requires a regional proposal network like the Faster RCNN [10]. Such algorithms can achieve high accuracy, but at a slower speed. Although speed can be achieved by reducing the number of proposals or decreasing the resolution of the input image, there is no qualitative improvement in speed. The latter refers to a detection algorithm similar to YOLO [11] and SSD [12] that does not require regional proposal and direct regression. Such algorithms are fast, but the accuracy is not so good as the former. However, RetinaNet [13], which combines accuracy and speed, proposes that the focus loss implementation achieves the same level of accuracy in a two-stage detection algorithm. Therefore, RetinaNet is utilized to detect the number of bars.

2.3. Status of Calculation of the Number of Steel Bars

At the site of the construction site, for the steel trucks entering the field, the inspection personnel need to manually carry out the on-site manual reinforcement of the steel bars on the vehicle. After confirming the quantity, the steel truck can complete the loading and unloading. Currently, the manual counting method is adopted, as showing in Figure 1.

In the above manual counting process, manpower is consumed, and the speed is slow. Generally, it takes half an hour for the steel bar of the car, and it takes several hours for one entry to take the inventory. In the face of such cumbersome and repetitive work, it is proposed for the above work to reduce the repeatability and the labor of a lot of manpower based on the detection of the number of steel bars for deep learning.

3. Rebar Quantity Detection Method

3.1. Feature Network

With the development of deep learning, the feature extraction network has also undergone great changes. From the early days of AlexNet [7], VGG-Net [14], GoogleNet [8], and today's ResNet, the network learning effect is getting better and better.

In the ResNet network, as showed in Figure 2, in addition to the normal convolutional layer output, there is a branch that directly connects the input to the output. Output and the convoluted output are arithmetically added to obtain the final output. The expression is $H(x) = F(x) + x$, x is the input, $F(x)$ is the output of the convolution branch, and $H(x)$ is the output of the entire structure. It can be shown that if all parameters in the $F(x)$ branch are 0. $H(x)$ is an identity map. The residual structure artificially creates an identity map, which allows the entire structure to converge toward the direction of the identity map.

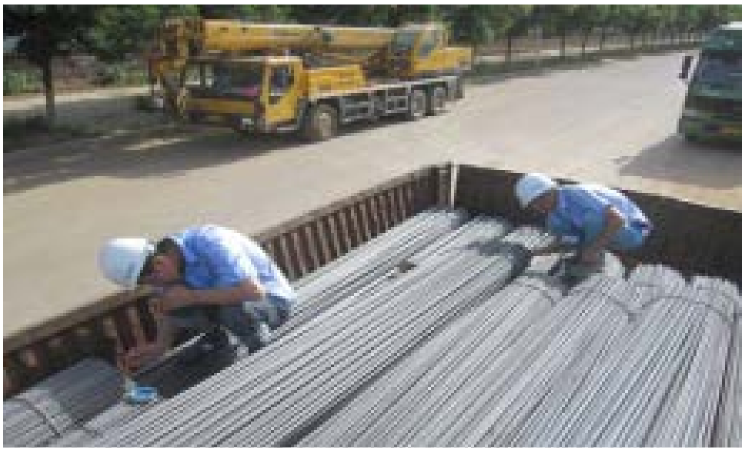

Figure 1. Manual counting of steel bars.

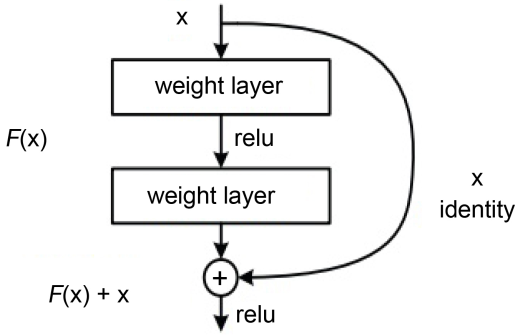

Figure 2. Residual link structure diagram.

3.2. Characteristic Pyramid Structure

Figure 3 is a block diagram of the network. A bottom-up line, a top-down line, connected horizontally. The enlarged area in the figure is the horizontal connection. Here, the main function of the 1 × 1 convolution kernel is to reduce the number of convolution kernels, that is, to decrease the number of feature maps and not to change the size of the feature map.

Bottom-up is actually the forward process of the network. In the forward process, the size of the feature map changes after passing through some layers, but does not change when passing through other layers, and the layer that does not change the size of the feature map is classified into one stage, so each extracted feature It is the last layer output of each stage, so that it can form a feature pyramid.

The top-down process is performed using up sampling, while the horizontal connection combines the result of the up sampling with the same size feature map generated from the bottom up. After the fusion, a 3 × 3 convolution kernels is used to convolve each fusion result in order to eliminate the aliasing effect of the up sampling. It is assumed that the generated feature map results are P2, P3, P4, P5, and the convolution results C2, C3, C4, and C5 from the bottom up are one-to-one correspondence.

3.3. Box Setting Improvement and Loss Function Improvement

The network model draws on the idea of Region Proposal Network. After the convolution of 3 × 3, there will be padding in the convolution process. Therefore, the size of the convolution is not changed, and the obtained feature map is obtained. Each point of the set is set as the anchor point, centered on the anchor point, artificially set different sizes, different proportions of the anchor frame, because it is the detection of the steel bar, and the shape of the steel bar is mostly circular, therefore, this article is in the original The anchor frame has been modified to make it meet the special task of steel bar inspection. The specific modification is as follows.

The original three aspect ratios were modified to be {1:1.5, 1:1, 1.5:1}, three different scales $\left\{2^0, 2^{\frac{1}{3}}, 2^{\frac{2}{3}}\right\}$, and a total of nine boxes to satisfy the detection of

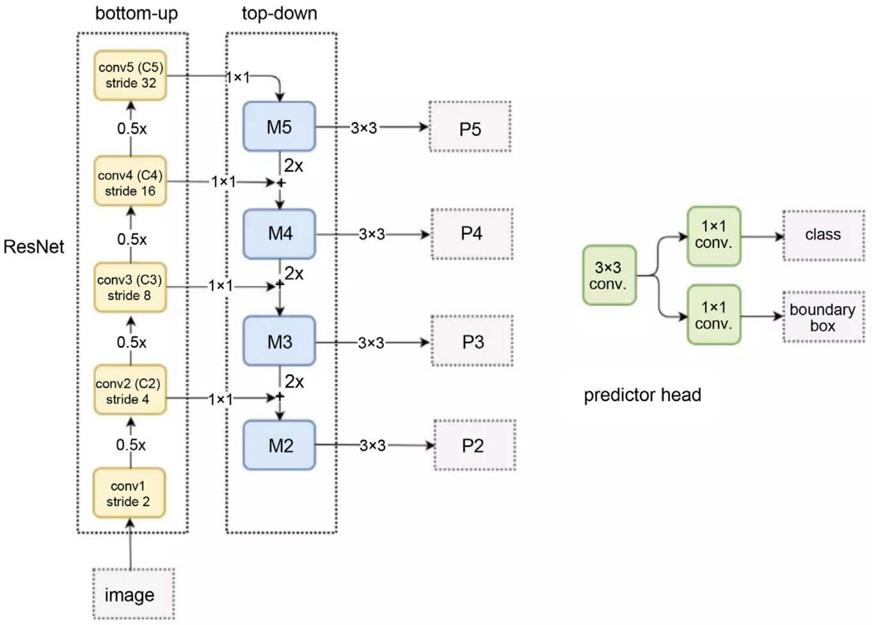

Figure 3. Network structure diagram.

steel bars. As shown in the predictor head section of Figure 2, the resulting feature map goes through two different branches, the classification branch (used to clearly identify each frame being foreground and background), and the regression branch (used to regress the detected foreground). That is, the coordinates of the frame of the reinforcing bar are detected, so there are different loss functions for different branches.

3.3.1. Classification Loss

In the classification task, the cross entropy loss function is commonly used. Suppose there are M samples, the classification target has N categories, y represents the label, and the probability that the i-th sample is predicted to be the N-th class is $p_{i,n}$, and the defined CE is:

$$CE = \frac{1}{M}\sum_{i=1}^{M}\sum_{n=1}^{N} -y_{i,n}\log\left(p_{i,n}\right) \tag{1}$$

In order to solve the positive and negative category imbalance, the parameter α is introduced to control the contribution weight to the total loss of the classification. The new loss function is defined as following:

$$CE_{\alpha} = \frac{1}{M}\sum_{i=1}^{M}\sum_{n=1}^{N} -\alpha y_{i,n}\log\left(p_{i,n}\right) \tag{2}$$

Since for the application of steel bar detection, only the category of steel bars is detected for the detected frame, the trade-off factor for the simple and difficult samples in the original loss is not increased, and the final loss function is as shown in Equation (2).

3.3.2. Return Loss

In the regression branch, the value predicted by the network is the deviation between the prediction frame and the setting frame. Since the shape of the reinforcing bar is relatively fixed, the basic frame size is set to 8, and the frame is applied to 9 different sizes and sizes. The loss function is L1 loss, assuming r is the deviation between the prediction area and the anchor frame, and r^* is the deviation between the label and the anchor frame (x, y, w, h), (x^*, y^*, w^*, h^*) respectively indicate the prediction area, the anchor frame area, and the horizontal and vertical coordinates of the area marked by the label, the width and height, then (r_x, r_y, r_w, r_h) indicates the deviation between the prediction area and the anchor frame area, and $(r_x^*, r_y^*, r_w^*, r_h^*)$ indicates the deviation of the anchor frame area from the label.

$$\begin{gathered} r_x = \frac{x - x_a}{w_a} \\ r_y = \frac{y - y_a}{h_a} \\ r_w = \log \frac{w}{w_a} \\ r_h = \log \frac{h}{h_a} \end{gathered} \tag{3}$$

$(r_x^*, r_y^*, r_w^*, r_h^*)$ The same is true, the final loss of $\text{smooth}_{\text{L1}}$ is as follows:

$$\text{smooth}_{\text{L1}}\left(r - r^*\right) = \begin{cases} 0.5\left(r - r^*\right)^2 & \text{if } |x| < 1 \\ \left|r - r^*\right| - 0.5 & \text{other} \end{cases} \tag{4}$$

4. Experiment

4.1. Data Set

The data in this paper contain a total of 250 training sets, as shown in **Figure 4**, and 200 test sets, all taken from the actual scene, and the collected data set is

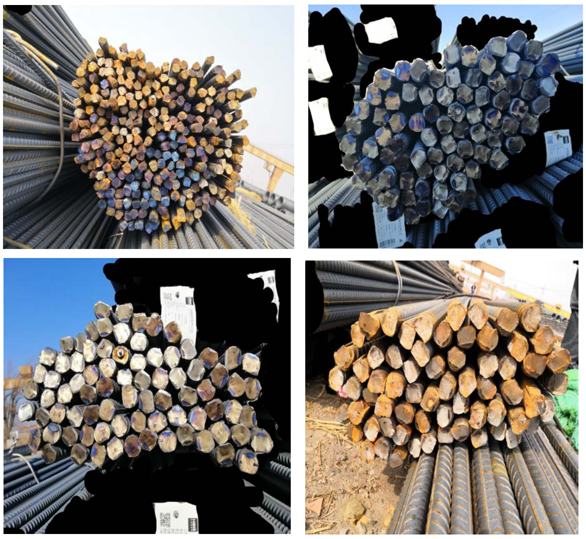

Figure 4. Partial data set.

labeled as the format of the VOC data set for training. Among them, training and verification sets are divided into 8:2 ratios in the training set.

4.2. Training Process

Since only the reinforcing bars are detected, in the classification and regression layers, the corresponding output layers are respectively modified, and in the classification layer, only the reinforcing bars are divided, and in the regression layer, only the single-category coordinate frame of the reinforcing bars is returned. A box with an anchor frame and a label that are greater than 0.8 is marked as a positive sample, and a frame smaller than 0.3 is a negative sample. The learning rate for is initialized to 0.002, the attenuation rate is 0.9, and it is attenuated every 20 epochs, and the total training epoch is 200 times.

During training, data is horizontally and vertically flipped, and brightness is increased by three data enhancement techniques.

4.3. Judging Criteria

The trained model is evaluated on the test set according to the following F1-Score calculation method:

$$R(\text{Recall rate}) = \frac{\text{Detect the correct amount}}{\text{Correct quantity} + \text{Missed quantity}} \tag{5}$$

$$P(\text{Preciseness}) = \frac{\text{Detect the correct amount}}{\text{Correct quantity} + \text{Number of errors}} \tag{6}$$

$$\text{F1} = \frac{2PR}{P+R} \tag{7}$$

As in Equations (5), (6), and (7), the final F1 score on the test set is used as the model criterion.

4.4. Experimental Result

Using YOLOV1, SSD, the improved algorithm of this paper is evaluated on the test set. The final experimental results are shown in Table 1. The test results are shown in Figure 5. It can be seen that the F1 score on the test set. The improved algorithm applied in this paper is higher than the comparison detection algorithm.

This article uses Python language, pytorch's deep learning framework, and the experimental equipment is i7 8700 RTX2070.

Table 1. Comparison of detection effects of different algorithm models on test sets.

Model name	F1 score
YOLOv1	0.853
SSD	0.947
RetinaNet	0.956
Improved RetinaNet	0.974

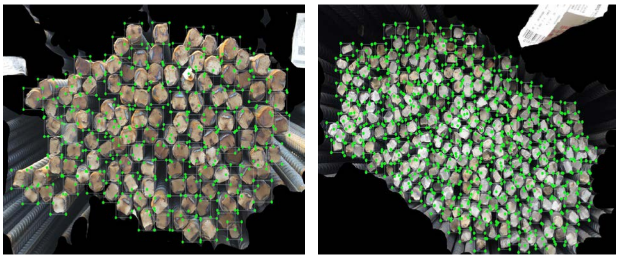

Figure 5. Test results.

5. Conclusion

In this paper, deep learning detection technology is applied to the detection of the actual number of engineering steel bars, and the traditional detection algorithm in the literature is compared with the deep learning detection algorithm. Based on the original RetinaNet algorithm, the modification of the anchor frame size and the modification of the loss function make it more suitable for the special field of steel quantity detection. The test performance demonstrated by it shows that it can be applied to the actual engineering field, reducing the manpower and material resources consumed in the single link of the quantity of steel bars. At present, the detection of the number of reinforcing bars is run on a server with a GPU, and the network model is slightly larger. The next step is to change the network model to a more lightweight network so that it can be applied to the mobile terminal.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Felzenszwalb, P.F., Girshick, R.B., Ramanan, D. and McAllester, D. (2010) Object Detection with Discriminatively Trained Part-Based Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **32**, 1627-1645. https://doi.org/10.1109/TPAMI.2009.167

[2] Uijlings, J.R.R., Van De Sande, K.E.A., Smeulders, A.W.M. and Gevers, T. (2013) Selective Search for Object Recognition. *International Journal of Computer Vision*, **104**, 54-171. https://doi.org/10.1007/s11263-013-0620-5

[3] Hu, W. and Maybank, S. (2008) AdaBoost-Based Algorithm for Network Intrusion Detection. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, **38**, 577-583. https://doi.org/10.1109/TSMCB.2007.914695

[4] Niu, X.X. and Suen, C.Y. (2012) A Novel Hybrid CNN-SVM Classifier for Recognizing Handwritten Digits. *Pattern Recognition*, **45**, 1318-1325. https://doi.org/10.1016/j.patcog.2011.09.021

[5] Pal, M. (2005) Random Forest Classifier for Remote Sensing Classification. *International Journal of Remote Sensing*, **26**, 217-222.

[6] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing*

Systems, Harrahs and Harveys, Lake Tahoe, NV, 1097-1105.

[7] Szegedy, C., Liu, W., Jia, Y., *et al.* (2015) Going Deeper with Convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, 7-12 June 2015, 1-9. https://doi.org/10.1109/CVPR.2015.7298594

[8] Russakovsky, O., Deng, J., Su, H., *et al.* (2015) ImageNet Large-Scale Visual Recognition Challenge. *International Journal of Computer Vision*, **115**, 211-252. https://doi.org/10.1007/s11263-015-0816-y

[9] Ren, S., He, K. and Girshick, R. (2015) Faster R-CNN: Towards Real-Time Object Detection with Regional Proposal Networks. *Advances in Neural Information Processing Systems*, Montreal, 7-12 December 2015, 91-99.

[10] Redmon, J., Divvala, S. and Girshick, R. (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, 27-30 June 2016, 779-788. https://doi.org/10.1109/CVPR.2016.91

[11] Liu, W., Anguelov, D., Erhan, D., *et al.* (2016) SSD: Single Shot MultiBox Detector. In: Leibe, B., Matas, J., Sebe, N. and Welling, M., Eds., *Computer Vision-ECCV* 2016. *Lecture Notes in Computer Science*, Springer, Cham, October 8-16 2016, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2

[12] Lin, T.Y., Goyal, P. and Girshick, R. (2017) Focal Loss for Dense Object Detection. 2017 *IEEE International Conference on Computer Vision*, Venice, Italy, 22-29 October 2017, 2980-2988. https://doi.org/10.1109/ICCV.2017.324

[13] Simonyan, K. and Zisserman, A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition. ArXiv reprint arXiv, 1409-1556.

[14] Fang, L.P., He, H.J. and Zhou, G.M. (2018) Research on Target Detection Algorithm. *Computer Engineering and Applications*, **13**, 7-24.