IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 17, 2019, accepted December 29, 2019, date of publication January 6, 2020, date of current version January 14, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2964055

Vehicle-Damage-Detection Segmentation Algorithm Based on Improved Mask RCNN

QINGHUI ZHANG[1,2], XIANING CHANG[2], AND SHANFENG BIAN[2]

[1]Key Laboratory of Grain Information Processing and Control, Ministry of Education, Henan University of Technology, Zhengzhou 450001, China
[2]College of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China

Corresponding author: Qinghui Zhang (zqh131@163.com)

This work was supported by the National Natural Science Foundation of China under Grant U1404617.

ABSTRACT Traffic congestion due to vehicular accidents seriously affects normal travel, and accurate and effective mitigating measures and methods must be studied. To resolve traffic accident compensation problems quickly, a vehicle-damage-detection segmentation algorithm based on transfer learning and an improved mask regional convolutional neural network (Mask RCNN) is proposed in this paper. The experiment first collects car damage pictures for preprocessing and uses Labelme to make data set labels, which are divided into training sets and test sets. The residual network (ResNet) is optimized, and feature extraction is performed in combination with Feature Pyramid Network (FPN). Then, the proportion and threshold of the Anchor in the region proposal network (RPN) are adjusted. The spatial information of the feature map is preserved by bilinear interpolation in ROIAlign, and different weights are introduced in the loss function for different-scale targets. Finally, the results of self-made dedicated dataset training and testing show that the improved Mask RCNN has better Average Precision (AP) value, detection accuracy and masking accuracy, and improves the efficiency of solving traffic accident compensation problems.

INDEX TERMS Mask RCNN, vehicle-damage-detection, loss function, detection accuracy.

I. INTRODUCTION

Object detection is one of the main research contents of computer vision. It is to determine the category and location information of the object of interest in the image on the instance level. Currently the most popular target detection algorithms include RCNN[1], Fast RCNN[2], Faster RCNN[3] and SSD[4]. However, these frameworks require a large amount of training data, which cannot achieve end-to-end detection. The positioning ability of the detection frame is limited, and when the feature is extracted, as the number of convolution layers increases, gradient disappearance or gradient explosion often occurs. For these drawbacks, He Kaiming et al. proposed a residual network (ResNet) [5] [25], which helps the model to converge by using the residual module, accelerates the training of the neural network, and combines with the target detection model Mask RCNN[6] [26] [27] to realize object detection and segmentation, greatly improving the accuracy of the model detection. Mask RCNN is the first deep learning model that combines both target detection and segmentation in one network [7]. It can achieve challenging instance segmentation tasks, which can not only accurately segment individuals in different categories, but also label each pixel in the image to distinguish different individuals in the same category [8].

Most current instance segmentation algorithms are based on candidate regions. Pinheiro *et al.* [9] proposed a DeepMask segmentation model, which outputs prediction candidate masks through the instances appearing in the input image to segment each instance object, but the accuracy of boundary segmentation is low [10]; Li *et al.* [11] proposed the first end-to-end instance segmentation framework, full convolutional instance segmentation (FCIS). By improving the position-sensitive score map, FCIS predicts both the bounding box and instance segmentation, but it can only roughly detect the boundary of each instance object when processing overlapping object instances [12]; He *et al.* [6] proposed the Mask RCNN framework, which is an algorithm with relatively fine instance segmentation results among existing segmentation algorithms [13].

Compared with the traditional target detection method, the target detection model Mask RCNN not only has a great improvement in detection accuracy, but also has great advantages in the field of small target detection. It is widely used in

The associate editor coordinating the review of this manuscript and approving it for publication was Amr Tolba.

agriculture [14], construction [15] , Medical image segmentation [16] and other fields. Lin *et al.* [17] used Mask RCNN to classify rice planthoppers, and realized the effective and rapid identification of rice planthoppers and non-rice planthoppers, achieving an average recognition accuracy of 0.923. Wang *et al.* [18] used Mask RCNN to ship-target detection, which shows that Mask RCNN has better performance in solving the problem of closely aligned targets and multi-scale targets. Shi *et al.* [19] used Mask RCNN to the existing home-service-robot platform to obtain category information, location information, and item-mask information of the target, and obtained an 85% mAP value. Li *et al.* [20] proposed a building target detection algorithm based on Mask RCNN. In remote sensing images of different scenes, the detection of building targets can achieve an accuracy of 94.6%. The application field of Mask RCNN algorithm is very wide, but no one has used it in the field of automobile damage detection.

The paper uses Mask RCNN algorithm to detect and segment automobile damaged areas in traffic accidents. It has very important research value and has broad application scenarios in the field of transportation. Due to the complexity of car damage detection and segmentation, there are problems such as lower detection segmentation accuracy and slower detection speed. This paper improves the model's network structure by reducing the number of layers in the residual network, and adjusting the internal structure to strengthen the regularization of the model, enhance the generalization ability, and then adjust the parameters of the anchor box and the loss loss function to improve the accuracy of car damage detection and segmentation. In this paper, the improved Mask RCNN is applied to the field of automobile damage detection, and a model based on it proposed for detecting and segmenting the damaged area of a vehicle in an accident. Photos can be taken from both sides of the accident and uploaded for assessment. Insurance companies can also use this model to process claims quickly.

II. CAR-DAMAGE-DETECTION ALGORITHM FRAMEWORK

The vehicle-damage-detection and segmentation system based on the Mask RCNN model designed in this paper is shown in Figure 1.

It can be seen from the figure that an image of the damaged part of the car is selected and collected according to the demand, and the data are marked by the LabelMe annotation tool to make a dataset in the.json format, which is divided into a training set and a test set. The data are sent to the Mask RCNN for feature extraction and classification prediction and segmentation masking, and the car-damage-detection result is output.

A. MASK RCNN ALGORITHM

Mask RCNN is an instance segmentation framework extended by Faster RCNN. It is divided into two stages: the first stage scans the image and generates the proposal, and the second classifies the proposal and generates the bounding box and mask. The network structure block diagram of the Mask RCNN algorithm is shown in Figure 2.

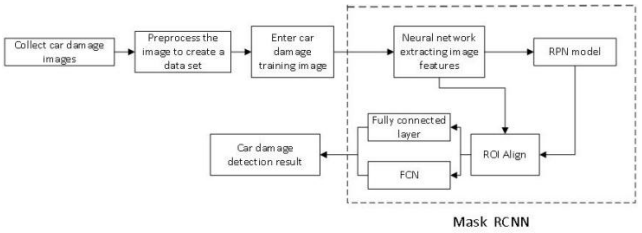

FIGURE 1. Car-damage-detection segmentation system framework.

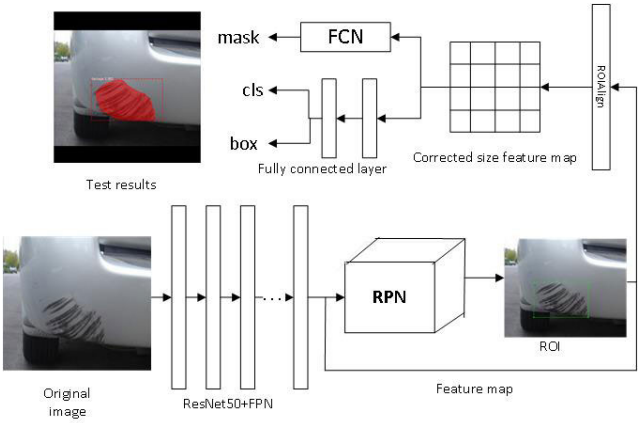

FIGURE 2. Mask RCNN network framework model.

The algorithm flow is the following.

(1) Input the image to be processed into a pre-trained ResNet50+FPN network model to extract features and obtain corresponding feature maps.

(2) This feature map obtains a large number of candidate frames (i.e., the region of interest, or ROI) through RPN, and then uses the softmax classifier to perform binary classification of foreground and background, using frame regression to obtain more accurate candidate-frame position information, and filtering out part of the ROI by non-maximum suppression.

(3) The feature map and the last remaining ROI are sent to the RoIAlign layer, so that each ROI generates a fixed-size feature map.

(4) Finally, the flow goes through two branches, one branch entering the fully connected layer for object classification and frame regression, and the other entering the full convolution network (FCN) for pixel segmentation.

B. BACKBONE NETWORK STRUCTURE IMPROVEMENT

Generally, the backbone network of Mask RCNN adopts ResNet101; that is, the number of network layers is 101, but too many layers will greatly reduce the rate of the network structure. The car-damage category trained in this paper is relatively simple, and the requirements for the network layer are lower; thus, to further improve the running speed of the algorithm, this paper uses ResNet50.

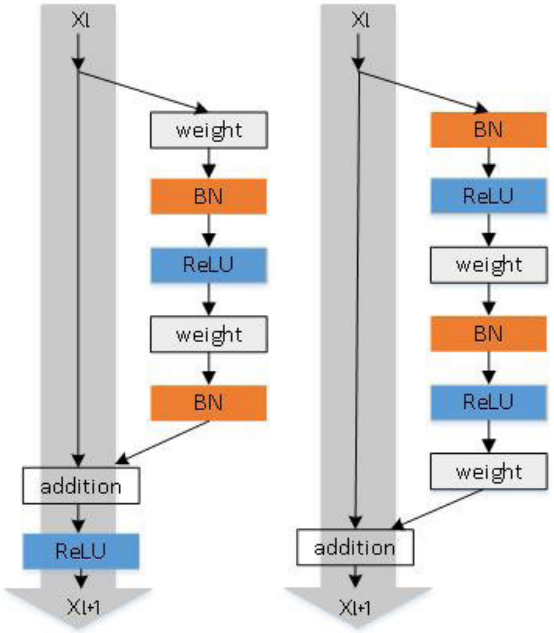

FIGURE 3. ResNet structure and improved ResNetV2 structure.

Because the size of the car damage in the images will be different, only a single convolutional neural network cannot extract all the image attributes well. Therefore, the backbone structure of ResNet50 and a FPN feature pyramid network is used in this paper. FPN [21] uses a top-down hierarchy with lateral connections, from single-scale input to building a network feature pyramid, which solves the multi-scale problem of extracting target objects in images. This structure has strong robustness and adaptability, and requires fewer parameters.

To further improve the detection accuracy, the backbone network structure is improved and the order of each layer is adjusted, as shown in Figure 3. The right-hand part of the diagram in each figure is called the ''residual'' branch and the left-hand part the ''identity'' branch. The value of the ''identity'' branch cannot be easily changed. Keep the input and output consistent, otherwise it will affect the information transmission and hinder the loss of loss. Adjust the order of the layers on the ''residual'' branch, the improved ResNet structure has two advantages. First, back-propagation basically meets the requirements, and information transmission is unimpeded. Second, the BN layer acts as a pre-activation, and the concept of ''pre'' is relative to the weight (conv) layer. This can enhance the regularization of the model, and the generalization performance is better.

C. RPN MODEL IMPROVEMENT

In this paper, the Feature Pyramid Networks structure is adopted, and the images are made into different sizes to generate features corresponding to different sizes. The shallow features can distinguish simple large targets and the deep features can distinguish small targets. The different-size feature maps generated by the FPN are input into the RPN [22],

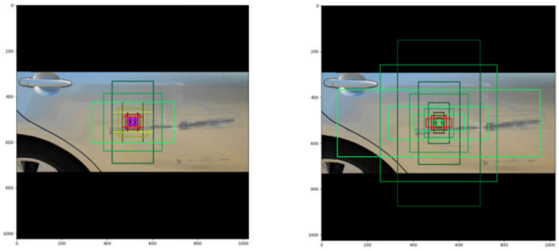

FIGURE 4. ROI generated by the original RPN and the improved RPN.

and then the RPN can extract the RoI features from different levels of the feature pyramid according to the size of the target object. Thereby, the simple network structure changes, without substantially increasing the calculation amount, greatly improving the detection performance of small objects and achieving excellent improvement in accuracy and speed.

RPN is equivalent to a sliding-window-based classless target detector. It is based on the structure of a convolutional neural network. The sliding frame scan produces anchor frame anchors. A suggested area can generate a large number of anchors of different sizes and aspect ratios, and they overlap to cover as many images as possible; the size of the suggested area and the desired area overlap (IOU) will directly affect the classification effect. To be able to adapt to more damaged car areas, the algorithm adjusts the scaling scale of the ''anchor point'' to $\{32 \times 32, 64 \times 64, 128 \times 128, 256 \times 256, 512 \times 512\}$, and the aspect ratio of the anchor point is changed to {1:2,1:1,3:1}, as shown in Figure 4. The so-called IoU is the coverage of the predicted box and the real box, the value of which is equal to the intersection of the two boxes divided by the union of the two boxes. In this paper, the value of IoU is set to 0.8; that is, when the overlap ratio of the area corresponding to the anchor frame and the real target area is greater than 0.8, it is the foreground; when the overlap rate is less than 0.2, it is the background; between the two values, it is discarded. This reduces the amount of computation underlying the model, saves time, and the improved RPN produces less ROI, which, in turn, increases the efficiency of the model.

D. RoIAlign MODEL

In the Mask RCNN network structure, the mask branch must determine whether a given pixel is part of the target, and the accuracy must be at the pixel level. After the original image is heavily convolved and pooled, the size of the image has changed. When the pixel-level segmentation is directly performed, the image target object cannot be accurately positioned, so the Mask RCNN is improved on the basis of Faster RCNN, and the RoI Pooling layer is changed into the interest-region alignment layer (RoIAlign). The bi-linear interpolation [23] method preserves the spatial information on the feature map, which largely solves the error caused by the two quantizations of the feature map in the RoI Pooling layer, and solves the problem of regional mismatch of the image object. Pixel-level detection segmentation can thus be achieved.

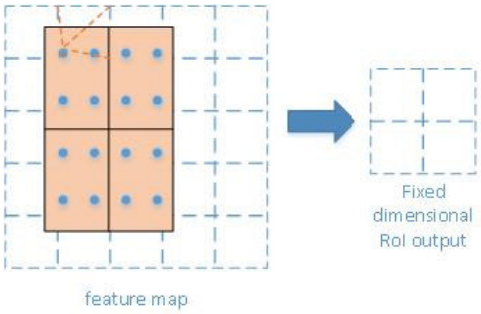

FIGURE 5. RoIAlign schematic.

The interest-area alignment layer RoIAlign differs from the ROI pooling in that it eliminates the quantization operation and does not quantize the ROI boundary and the unit, but uses bi-linear interpolation to calculate the exact position of the sample points in each unit, retaining its decimal, and then uses the maximum pooling or average pooling operation to output the last fixed-size RoI. As shown in Fig. 5, the blue dotted line is the 5×5 feature map after convolution, the solid line is the feature small block corresponding to the ROI in the feature map, and RoIAlign maintains the floating-point number boundary, without quantization processing. First, the feature small block is divided into 2×2 units (each unit boundary is not quantized) and then divided into four small blocks in each unit; the center point is taken as four coordinate positions, as shown by the blue dot in the figure. Then, the values of the four positions are calculated by bi-linear interpolation, and finally the maximum pooling or average pooling operation performed to obtain the feature map of 2×2 size.

E. IMPROVEMENT OF LOSS FUNCTION

The multitasking loss function of Mask RCNN is

$$\mathrm{L} = L_{cls} + L_{box} + L_{Mask} \tag{1}$$

The above equation is the same as the loss function in the Faster RCNN model, which represents the classification error and detection error, respectively. The mask branch and the class prediction branch are decoupled, and a binary mask is independently predicted for each category, without relying on the prediction results of the classification branch. The loss function in Faster RCNN:

$$\mathrm{L}(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \tag{2}$$

In the above formula, i is the index of the anchor box in the mini-batch;N_{cls} and N_{reg} indicate the number of classification layers and regression layers respectively;P_i represents the predicted probability value of anchor i being an object;P_i^* is 0 if the anchor box is negative, and is 1 if the anchor box is positive;t_i indicates 4 parameterized coordinates of the prediction candidate box;t_i^* refers to 4 parameterized coordinates of the true value region;L_{cls} and L_{reg} represent classification loss and regression loss, respectively. λ Represents the balance coefficient, which is used to control the proportion of the two loss functions.

In the Faster RCNN,a hyperparameter $\lambda = 10$ control balance is introduced between the classification loss and regression loss, and the large-scale target and small-scale target share this one parameter.

The error function of the Class prediction branch in the Mask RCNN can be calculated by the following formula:

$$L(p, u, t^u, v) = L_{cls}(p, u) + \lambda[u \geq 1] L_{loc}(t^u, v) \tag{3}$$

where p is the predicted class, u is GT class, t^u is the predicted bounding box for class u, v is GT bounding box.

If the hyperparameter $\lambda = 10$ is still introduced in the Mask RCNN, it will cause a phenomenon. The high-level semantic information is introduced on the underlying feature. The small-scale target has obvious rise points and the large-scale target is not obvious. On the high-level feature, more underlying feature information is introduced or maintained. The large-scale target has obvious rise points and the small scale is not obvious. The frame of the large target is actually more accurate, but the position drift is more serious, so the underlying information that is good for positioning is needed, which contributes to the improvement of the large target on the map indicator. The possible position of the small target is more accurate. However, the judgment of semantic information is relatively weak, so high-level semantic information is needed to assist the discrimination, which contributes to the improvement of the small target in the map index. To summarize, the focus is on optimizing the location information for large targets. For small targets, the focus is on optimizing category prediction. That is, for different scale targets, different weights should be introduced in the loss function to improve the detection accuracy of the detection branches.

III. EXPERIMENTAL RESULTS AND ANALYSIS

To reduce the number of steps in making dataset labels and to improving the detection accuracy of car-damage images, transfer learning and Mask RCNN are used in this paper to process and detect images showing damage.

A. TRANSFER LEARNING

Deep learning requires a significant amount of data, but in most cases it is difficult to find enough training data for a specific problem within a certain range. To solve this problem, a solution is proposed, namely to use transfer learning [24].

Transfer learning includes a source domain and a target domain, defined as

$$D(s) = \{x, P(x)\}, D(t) = \{x, P(x)\} \tag{4}$$

where $D(s)$ is the source domain, $D(t)$ the target domain, x the feature space, and P(X) the marginal probability distribution,$X = \{x_1, K, x_n\} \in x$.

It can be seen from the above formula that transfer learning is used to transfer the model parameters already trained in the source domain to new models in the target domain to help the new model training. Considering that most of the

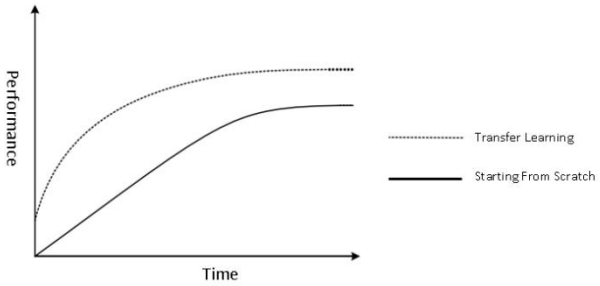

FIGURE 6. Comparison between transfer learning and starting from scratch.

image data have similar basic features, such as color and mode, in this paper pre-training is first done on large coco datasets, and then the trained weight files are migrated to the dedicated datasets collected in this article for training, fine-tuning the network parameters. This allows the convolutional neural networks to achieve good results on small datasets, thereby alleviating the problem of insufficient data sources.

Ideally, a comparison of successful Transfer Learning with Starting From Scratch is shown in Figure 6.

It can be seen that using migration learning can bring three advantages. First, the initial performance of the model is higher. Second, the rate of performance improvement of the model is greater during the training process. Third, the final performance of the trained model is better.

B. BUILDING A DATASET

The main research object of this paper is a picture of a vehicle that is scratched. The experiment collected 2,000 damaged vehicle images (1600 training sets and 400 test sets) from online downloads and daily photographs.

The main steps in acquiring a dedicated dataset for detecting vehicle scratches in complex environments include the following two parts.

1) Image collection: Images of damaged vehicles at different angles and of different sizes in different scenes are photographed and downloaded from the Internet. Because the downloaded images vary in size, and the sample of Mask RCNN must be normalized to a uniform size, a script is used to normalize the images to 1024 × 1024 pixels, and the insufficient portions are filled with 0.
2) Image processing: The captured images are marked using the marking tool Labelme and divided into training sets and test sets. The specific steps in this process are the following.

First, a folder ''datasets'' is created, and then two subfolders, ''train'' and ''val'', are created for storing training samples and test samples. The images in each folder correspond to a.json annotation information file with the same name. The labeling interface is shown in Figure 7.

C. EXPERIMENTAL ENVIRONMENT

See Table 1.

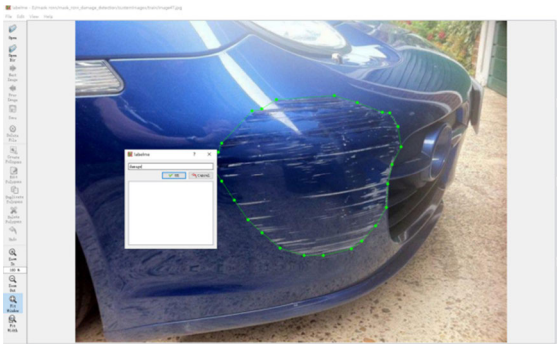

FIGURE 7. Labelme-marked car-damage image.

TABLE 1. Experimental environment information table.

Attribute name	Attribute value
TensorFlow version	1.14.0
Keras version	2.2.5
RAM	31.3G
Processor	Inter(R) Core(TM)i7-6700K CPU @4.00GHz x 8
Graphics	GeForce GTX 1080/PCIe/SSE2
Operating system version	Ubuntu16.04，64bit

TABLE 2. Experimental part parameter table.

Parameter	Value
LEARNING_RATE	0.001
LEARNING_MOMENTUM	0.9
WEIGHT_DECAY	0.0001
DETECTION_MIN_CONFIDENCE	0.8
STEPS_PRE_EPOCH	100
NUM_CLASSES	2
MASK_POOL_SIZE	14
POOL_SIZE	7
VALIDATION_STEPS	50

D. PARAMETER SETTINGS

See Table 2.

E. EVALUATION INDEX

The evaluation index of the experimental results consists of two aspects: detection performance and segmentation performance. In this experiment, the P-R curve and the AP value were used to evaluate the performance of the target detection, and the mean intersection over union (MIoU) and the running speed were used to evaluate the image segmentation performance.

$$P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN} \tag{5}$$

where TP is the correct number of samples correctly classified. FP is the number of negative samples of a positive sample that is incorrectly marked. FN is the number of positive samples that are incorrectly marked as negative samples.

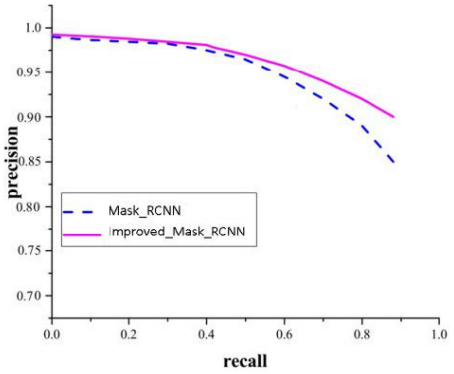

FIGURE 8. P-R curve.

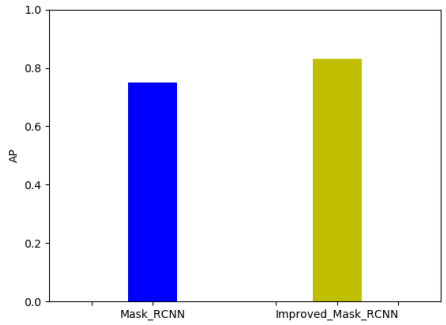

FIGURE 9. AP values of the two algorithms.

P is the accuracy rate and R is the recall rate. A P-R graph is made based on the prediction results of the test set in the network model, and then the average accuracy of the model is obtained from the area under the P-R. The larger the AP value, the better the detection performance.

$$MIoU = \frac{1}{k+1} \sum_{i=1}^{k} \frac{p_{ii}}{\Sigma_{j=0}^{k} p_{ij} + \Sigma_{j=0}^{k} p_{ji} - p_{ii}} \tag{6}$$

where k is the total number of output classes in the model, p_{ij} represents the number of pixels that belong to category i but have been misjudged as category j. p_{ii} indicates the number of pixels correctly classified, while p_{ij} and p_{ji} represent pixels that are misclassified.

F. EXPERIMENTAL RESULTS AND ANALYSIS

In order to study the detection performance of the improved algorithm on the car damage data set, it is compared with the advanced detection algorithm Mask RCNN algorithm. Figure 8 shows the P-R curve obtained using two algorithms. Then, the area under the P-R curve is obtained by integration, and the average accuracy of the two algorithms for car damage detection, that is, the AP value, is obtained, and the result is shown in Figure 9.

It can be seen from Fig.9 and Fig.10 that the improved Mask RCNN algorithm has a significant improvement in detection performance by comparing the Mask RCNN algorithm. As can be seen from Figure 10, the Mask value of the Mask RCNN is 0.75, and the AP value of the improved

FIGURE 10. Vehicle damage detection result based on Mask RCNN algorithm.

TABLE 3. Comparison of test results accuracy and time (fps denotes frame per second).

Algorithm	Detection accuracy (%)	Mask accuracy (MIoU) (%)	Running speed (fps)
Mask RCNN	94.53	81.25	4.26
Improved Mask RCNN	96.68	83.14	4.78

detection algorithm is 0.83, which is 0.08 higher than the advanced target detection algorithm Mask RCNN.

As can be seen from Table 3, compared with the Mask RCNN, the improved Mask RCNN improves the detection accuracy by 2.15%, the mask accuracy by 1.89%, and the running speed by 0.52fps. It can be seen that the improved algorithm not only improves the accuracy, but also speeds up the detection speed, has better performance advantages,

TABLE 4. Statistical Table of Automobile Damage Detection Results Based on Mask RCNN.

Detect picture	a	b	c	d	e	f
lab environment	Normal light	Weak light	Close distance	Multiple damage	Strong exposure	Minor injury
Target quantity detected	1	1	1	3	0	1
Detection accuracy	0.983	0.911	0.906	0.977 0.960 0.977	0	0.968

TABLE 5. Statistical Table of Automobile Damage Detection Results Based on Improved Mask RCNN.

Detect picture	a	b	c	d	e	f
Lab environment	Normal light	Weak light	Close distance	Multiple damage	Strong exposure	Minor injury
Target quantity detected	1	1	1	4	1	2
Detection accuracy	0.995	0.952	0.986	0.972 0.974 0.977 0.905	0.947	0.992 0.933

and has higher applicability in the damaged area of the automobile.

To verify the accuracy and reliability of the improved Mask RCNN for automobile damage detection, experiments were conducted on images under the following conditions: normal illumination, weak illumination, close distance, multiple damage, strong exposure, and insignificant damage. These conditions map to images (a)–(f), respectively, in the original Mask RCNN algorithm and the improved Mask RCNN for testing, and the test results are shown in Table 4 and Table 5. The rectangular box indicates the detected target position, the number on the rectangular frame the probability of belonging to the damaged area of the car, and the binary mask the approximate outline of the damaged area of the car.

Statistics on the car damage detection results of the Mask RCNN algorithm above are shown in the following table:

The results of the car damage detection of the improved Mask RCNN algorithm in the above figure are counted as shown in the following table:

Comparing the Figure 10, Figure 11 and Table 4, Table 5, it is shown that the improved Mask RCNN exhibits improvements in missed detection and low accuracy. The improved algorithm thus shows strong robustness and adaptability for vehicle-damage detection. It can be further seen from the comparison of experimental results that it is difficult to detect the damaged area of the vehicle with high exposure using the original Mask RCNN. Areas in which the damage is not obvious are also difficult to detect, but the improved Mask RCNN has a good performance improvement in this area.

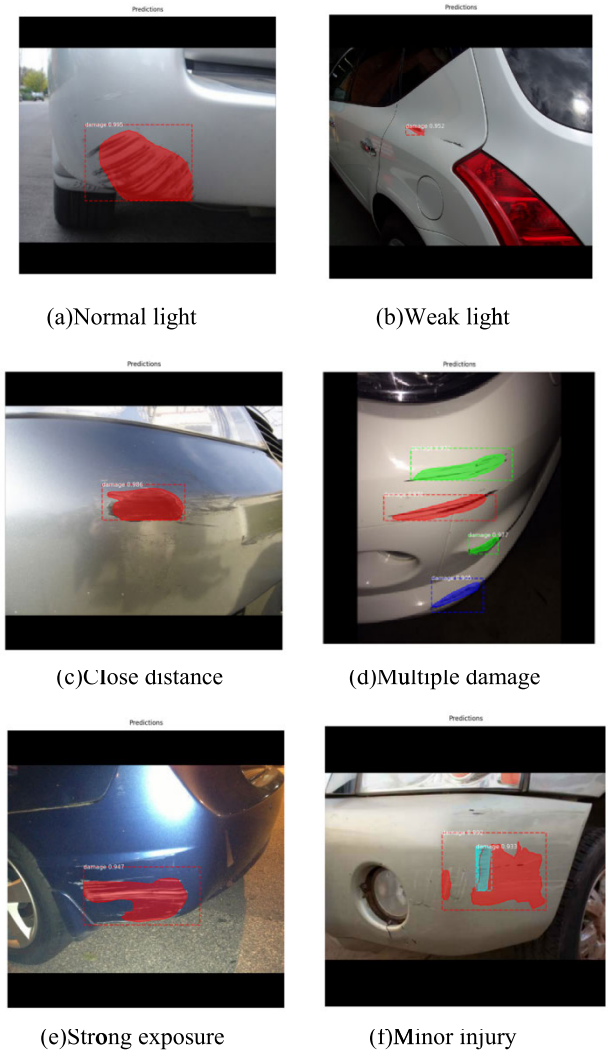

FIGURE 11. Car damage detection results based on the improved Mask RCNN algorithm.

IV. CONCLUSION

In the work described in this paper, a detection algorithm based on deep learning for vehicle-damage detection is used to deal with the compensation problem in traffic accidents. After testing and improvement, the proposed transfer-learning and improved Mask RCNN–based vehicle-damage-detection method is more universal, and can better adapt to various aspects of car-damage images. The algorithm achieved good detection results in different scenarios. Regardless of the strength of the light, the damaged area of multiple cars, or a scene with an overly high exposure, the fitting effect is better and the robustness is strong.

Although the robust Mask RCNN algorithm is adopted in this paper and it improves on the original algorithm and obtained ideal experimental results, some aspects have yet to be studied. For example, the detection accuracy is very high, but the mask instance segmentation cannot be completely correct, and some areas in which the damage is not obvious cannot be segmented. In future work, data expansion can be

carried out to increase the size of the dataset, collect more car damage images under different weather conditions and different levels of illumination, enhance the data, improve the edge-contour enhancement of images, and make the masking of the damaged areas of the car more accurate.

REFERENCES

[1] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, vol. 13, no. 1, pp. 580–587.
[2] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015. pp. 1440–1448.
[3] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/tpami.2016.2577031.
[4] W. Liu, D. Anguelov, and D. Erhan, "SSD: Single shot multibox detector," in *Proc. IEEE Eur. Conf. Comput. Vision*, Jun. 2016, pp. 21–37.
[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[6] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.
[7] N. Kumar and R. Verma, "A multi-organ nucleus segmentation challenge," *IEEE Trans. Med. Imag.*, vol. 11, no. 1, pp. 34–39, Oct. 2019, doi: 10.1109/TMI.2019.2947628.
[8] A. K. Jaiswal, P. Tiwari, S. Kumar, D. Gupta, A. Khanna, and J. J. Rodrigues, "Identifying pneumonia in chest X-rays: A deep learning approach," *Measurement*, vol. 145, pp. 511–518, Oct. 2019, doi: 10.1016/j.measurement.2019.05.076.
[9] P. Pinheiro and R. Collobert, "Learning to segment object candidates," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 1990–1998.
[10] W. Tang, H.-L. Liu, L. Chen, K. C. Tan, and Y.-M. Cheung, "Fast hypervolume approximation scheme based on a segmentation strategy," *Inf. Sci.*, vol. 509, pp. 320–342, Jan. 2020, doi: 10.1016/j.ins.2019.02.054.
[11] Y. Li, H. Qi, J. Dai, X. Ji, and Y. Wei, "Fully convolutional instance-aware semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4438–4446.
[12] X. Rong, C. Yi, and Y. Tian, "Unambiguous scene text segmentation with referring expression comprehension," *IEEE Trans. Image Process.*, vol. 29, pp. 591–601, Jul. 2019, doi: 10.1109/tip.2019.2930176.
[13] Y. L. Qiao, M. Truman, and S. Sukkarieh, "Cattle segmentation and contour extraction based on mask R-CNN for precision livestock farming," *Comput. Electron. Agricult.*, vol. 165, Oct. 2019, Art. no. 104958, doi: 10.1016/j.compag.2019.104958.
[14] C. Shuhong, Z. Shijun, and Z. Dianfan, "Water quality monitoring method based on feedback self correcting dense connected convolution network," *Neurocomputing*, vol. 349, pp. 301–313, Jul. 2019, doi: 10.1016/j.neucom.2019.03.023.
[15] J. Yang, L. Ji, X. Geng, X. Yang, and Y. Zhao, "Building detection in high spatial resolution remote sensing imagery with the U-rotation detection network," *Int. J. Remote Sens.*, vol. 40, no. 15, pp. 6036–6058, Aug. 2019, doi: 10.1080/01431161.2019.1587200.
[16] E. K. Wang, X. Zhang, L. Pan, C. Cheng, A. Dimitrakopoulou-Strauss, Y. Li, and N. Zhe, "Multi-path dilated residual network for nuclei segmentation and detection," *Cells*, vol. 8, no. 5, p. 499, May 2019, doi: 10.3390/cells8050499.
[17] X. Lin, S. Zhu, and J. Zhang, "Rice planthopper image classification method based on transfer learning and mask R-CNN," *Trans. Chin. Soc. Agricult. Mach.*, vol. 13, no. 4, pp. 181–184, Dec. 2019.
[18] G. Wang and S. Liang, "Ship object detection based on mask RCNN," in *Proc. Radio Eng.*, 2018, pp. 947–952.
[19] J. Shi, Y. Zhou, and Q. Zhang, "Service robot item recognition system based on improved mask RCNN and Kinect," in *Proc. Appl. Res. Comput.*, Jun. 2019, pp. 1–9.
[20] J. Li and W. He, "Building target detection algorithm based on mask RCNN," in *Proc. Sci. Surv. Mapping*, Apr. 2019, pp. 1–13.
[21] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 320–329.
[22] X. Zhang, J. Zou, X. Ming, K. He, and J. Sun, "Efficient and accurate approximations of nonlinear convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1984–1992.
[23] S. Wang and K. Yang, "An image scaling algorithm based on bilinear interpolation with VC++," in *Proc. Techn. Autom. Appl.*, 2008, pp. 168–176.
[24] A. Mathew, J. Mathew, M. Govind, and A. Mooppan, "An improved transfer learning approach for intrusion detection," *Procedia Comput. Sci.*, vol. 115, pp. 251–257, Jan. 2017.
[25] G. Han, J. Su, and C. Zhang, "A method based on multi-convolution layers joint and generative adversarial networks for vehicle detection," in *Proc. KSII Trans. Internet Inf. Syst.*, 2019, pp. 1795–1811.
[26] Y. Yu, K. Zhang, L. Yang, and D. Zhang, "Fruit detection for strawberry harvesting robot in non-structural environment based on mask-RCNN," *Comput. Electron. Agricult.*, vol. 163, Aug. 2019, Art. no. 104846.
[27] Y. Liu, P. Zhang, Q. Song, A. Li, P. Zhang, and Z. Gui, "Automatic segmentation of cervical nuclei based on deep learning and a conditional random field," *IEEE Access*, vol. 6, pp. 53709–53721, 2018.

QINGHUI ZHANG received the B.S.E. degree from the College of Fire Control, Zhengzhou Institute of Anti-Aircraft, in 1996, the M.E. degree in navigation guidance and control from the Ordnance Engineering College, Shijiazhuang, in 2003, and the Ph.D. degree from the Beijing Institute of Technology, Beijing, China, in 2006. He is currently a Professor with the College of Information Science and Engineering, Henan University of Technology. His research interests include artificial intelligence information processing and embedded systems.

XIANING CHANG received the B.S.E. degree from the College of Science, Henan Agricultural University, in 2018. She is currently pursuing the master's degree in computer technology with the College of Information Science and Engineering, Henan University of Technology, China. Her research interests include artificial intelligence information processing, road scene target detection, and deep learning.

SHANFENG BIAN received the B.S.E. degree from the College of Science, Huanghuai University, in 2017. He is currently pursuing the master's degree in signal and information processing with the College of Information Science and Engineering, Henan University of Technology, China. His research interests include intelligent information processing and embedded systems, vehicle detection, and deep learning.