Research on Vehicle Appearance Damage Recognition Based on Deep Learning

Qianqian Zhu1,*,a, Wei Hu2,b, Yingnan Liu1,c and Zihao Zhao1,d

1 Automotive Data of China Co.,Ltd., China Automotive Technology & Research Center Co.,Ltd., Tianjin, China
2 Ministry of Industry and Information Technology Equipment Industry Development Center, Beijing, China

*,azhuqianqian@catarc.ac.cn; bhwei@eidc.gov.cn; cluyingnan@catarc.ac.cn; dzhaozihuo@catarc.ac.cn

Abstract. Economic development has promoted the booming of the auto industry. With the increase of the number of cars, car insurance has become the largest type of insurance in the insurance industry with more than half of the market share. After the emergence of traditional vehicles, professional loss assessment personnel need to go to the scene to investigate the accident and complete the loss assessment. In recent years, With the rapid development of science and technology, the insurance industry has been changing from artificial and information to automation and intelligence. This paper presents a vehicle appearance damage recognition algorithm based on deep learning and its model evaluation method, which can accurately judge the vehicle damage in the image. The research shows that the Mask R-CNN model based on KL-loss performs well in vehicle damage detection and has good robustness; at the same time, the accuracy of the evaluation model results is greatly improved by replacing the traditional IOU calculation accuracy method with the component position.

1. Introduction

Internet insurance has begun to take shape in China. With its huge premium scale and standardized products, car insurance has become the driving force for insurance companies to gain customers. In the car insurance business, although the amount involved in small claim cases of non-personal injury is not high, the cases are frequent, which not only occupy a large amount of claim costs, but also have a great impact on customer satisfaction. In recent years, image processing technology based on computer vision has developed rapidly. At the same time, the massive amount of claims data accumulated by insurance companies urgently needs to be revitalized and the value of the data is brought into play. Based on computer vision technology, damage identification of vehicle images can be realized, and rapid compensation can be made based on basic data of insurance companies, which can greatly reduce the cost of claim settlement, shorten the claim settlement period and improve customer satisfaction.

In recent years, deep learning technology [1] represented by CNN has made great achievements in object detection, image recognition, semantic segmentation and other fields [2]. Object detection algorithms based on deep learning are widely used in industry. These models based on CNN can be roughly divided into two categories: one stage method directly predicts the location of bounding box according to CNN forward calculation, such as SSD [3] and Yolo [4]; two stage method generates...
region proposal first, and then classifies proposal boxes, such as Faster R-CNN [5] and Mask R-CNN [6].

In this paper, a method of vehicle appearance damage identification based on computer vision and its evaluation scheme are proposed. Based on the network structure of Mask R-CNN, a model is built to realize the identification of four main types of damage: scrape, deformation, cracking and rupture.

2. Model Establishment

2.1. Damage Recognition Model

2.1.1. Introduction of Mask R-CNN. Mask R-CNN is a network architecture based on Faster R-CNN, which mainly achieves semantic segmentation of object individuals. The model includes the backbone network for feature extraction, the RPN generated by candidate regions, ROIAAlign and the head network. In this paper, Residual Network (ResNet [7]) and Feature Pyramid (FPN) are used to extract image features, and then Area Suggest Network (RPN) is used to extract ROIs. Finally, the parallel FCN layer is used to achieve the segmentation of each ROI. In this paper, Mask R-CNN is used as the basic framework, and a damage identification method is established according to the specific target object. The network structure is shown in Figure 1.

![Figure 1. Schematic diagram of mask R-CNN network structure](image)

2.1.2. Introduction of KL-Loss. Large-scale object detection datasets such as ImageNet and MS-Coco define the basic ground truth bounding box as clearly as possible. However, there are inherent uncertainties in the ground truth bounding box in the self-built vehicle damage data set in some cases, which makes it more difficult to learn the regression function of the bounding box. As shown in Figure 2.

![Figure 2. Images labeled by different people](image)

Currently, the best object detection models (such as Faster R-CNN and Mask R-CNN) rely on the bounding box to locate the object. However, the traditional bounding box regression loss (i.e. smoothing L1 loss) does not consider the fuzziness of the basic ground truth bounding box. In addition, it is common to assume that the bounding box regression is accurate when the classification score is high, but this is not always the case.

In this paper, a new bounding box regression Loss KL-loss [8] proposed by the Institute of Megvii is applied to simultaneously learn the uncertainty of bounding box regression and positioning, so as to make the object frame clearer and improve the positioning accuracy.
Traditional Mask R-CNN predicts the offset of rectangular frame coordinates relative to ground truth, while KL-loss uses direct prediction diagonal coordinates. The goal of the network is to estimate the location confidence while estimating the location. For simplicity, the coordinates are assumed to be independent of each other and the univariate Gaussian distribution is used. The calculation structure is shown in Figure 3.

![Figure 3. Schematic diagram of KL-loss calculation structure](image)

The goal of object localization is to minimize the KL distance between the predicted distribution and the basic ground truth distribution on the sample. The proposed KL-loss not only considers the correctness of a single prediction box, but also minimizes the KL divergence for N samples. The loss function is as follows:

$$\hat{\Theta} = \arg \min_{\Theta} \frac{1}{N} \sum D_{KL}(P_{\hat{\Theta}}(x) \parallel P_{\Theta}(x))$$  \hspace{1cm} (1)

$$L_{reg} = D_{KL}(P_{\Theta}(x) \parallel P_{\Theta}(x))$$

$$= \int P_{\Theta}(x) \log P_{\Theta}(x) dx - \int P_{\Theta}(x) \log P_{\Theta}(x) dx$$  \hspace{1cm} (2)

$$= \frac{(x_g - x_e)^2}{2\sigma^2} + \log(\sigma^2) + \frac{\log(2\pi)}{2} - H(P_D(x))$$

Where, $x_g$ is the basic ground truth bounding box position, $x_e$ is the estimated bounding box position, $D_{KL}$ is the KL distance, $\sigma$ is the standard deviation, $P_D$ is the basic ground truth $\delta$ function, $P_{\Theta}$ is the predicted Gaussian distribution, and $\Theta$ is a set of parameters that can be learned.

This new loss can greatly improve the positioning accuracy of multiple architectures with little additional computing cost. The location variance learned can also help to fuse adjacent bounding boxes during non-maximum suppression (NMS) to further improve the location effect.

2.2. Model Evaluation

Since humans are experts in object detection, we can know whether the detection results are correct or not. But for the model, we need to judge the correctness of the detection object through quantitative indicators. In this paper, IOU (intersection over union) and component-based assessment methods are proposed to study the accuracy of damage identification.

2.2.1. Evaluation Method Based on IOU. In traditional IOU-based evaluation indicators, the intersection of the detection result and Ground Truth is compared to their union, which is IOU. For each class, the area where the prediction box and the ground truth overlap is the intersection, and the total area across is the union. As shown in Figure 4.
Figure 4. Schematic diagram of IOU calculation

When the IOU is greater than a certain threshold (generally set to 0.5), the output of the model is judged to be correct. However, this method is greatly limited by the location and size of the rectangular box. Inaccurate rectangular box labeling will bring great challenges to this kind of evaluation method. Especially for the self built damage data set, as shown in the figure below, the left is the ground truth, and the right is the model recognition result, but because the IOU < 0.5, the result is judged as wrong, but the actual model recognition result is still correct, which will seriously underestimate the ability of the model. As shown in Figure 5.

![Deformation](image1)

(a)Model Prediction Results  (b)Ground Truth

Figure 5. The result of deformation is wrong (IOU<0.5)

2.2.2 Evaluation Method Based on Component-assisted. Because of the great randomness and uncertainty of the vehicle damage scope, it is difficult to define the obvious scope of a vehicle damage. The IOU evaluation method based on location coordinates will seriously underestimate the ability of the model.

In order to accurately evaluate the effect of damage model detection and give more abundant semantic information output, this paper also adds vehicle component instance segmentation. Firstly, two independent Mask R-CNN networks were used for component segmentation and damage detection respectively. Secondly, the results of the two are fused, and the output of the composite results on the damaged components is used to replace the traditional coordinate position output. That is, if the model identifies a certain type of damage in a certain component, and the component also has such damage in the ground truth, the output of the model is determined to be correct. The calculation structure is shown in Figure 6.

![Deformation](image2)

Figure 6. Schematic diagram based on component-assisted calculation
Because there is an obvious inclusion relationship between the vehicle exterior components and the damage on the exterior components, the damage scope is limited to small enough vehicle components, which has better robustness than the evaluation method based on coordinate position only. The schematic diagram of the case is shown in Figure 7.

Figure 7. The results are all correct

3. Experiment

3.1. Dataset
The data set in this article is derived from historical claims of insurance companies, including panoramic images, close-up images and other forms. The training and verification on the data set can well represent the image effect of practical application. All the data are divided into three data sets: training set, verification set and test set.

In addition, data augmentation is carried out in this experiment. In the field of deep learning, the increase of data often improves the accuracy and generalization ability of the model. Because there is no vehicle damage data set, all the data is self-built by the project. In order to reduce labor costs, data augmentation is the best choice. In addition, in order to provide better damage assessment and richer semantic information, we also annotated the vehicle components and augmented the data.

3.2. Experiment Result
Based on the self-built damage data set, this paper uses KL-loss to carry out experiments on the network model combining ResNet-50+FPN and Mask R-CNN.

In order to evaluate the IOU (IOU>0.5) evaluation method based on location coordinates and the evaluation method based on components, this paper carried out two rounds of experiments, respectively: (1) Model test output based on location coordinates IOU (IOU>0.5); (2) Model test output based on component assistance. Finally, a random inspection of the recognition results by professional damage assessors is added. The experimental results are shown in Table 1.

| Evaluation method | AP  | AP50  | AP75  | AR1  | AR10 | AR100 |
|-------------------|-----|-------|-------|------|------|-------|
| Based on IOU      | 24.3% | 45.7% | 23.4% | 22.5% | 30.1% | 31.2% |

| Evaluation method          | Comprehensive Precision | Comprehensive Recall |
|----------------------------|--------------------------|----------------------|
| Based on Component-assisted| 81.6%                     | 67.4%                |

3.3. Brief Summary
Through the comparison of experimental results and manual secondary confirmation, the component-based evaluation method is closer to the result of artificial damage assessment, which can better reflect
the ability of the model and help to train a better model for vehicle damage identification. The reason for this phenomenon is the difference of the damage itself and the interference of subjective factors when labeling. Different labeling personnel have different definitions of the damage scope, which leads to the influence of the trained model on the identification of the damage scope and the comparison between the identified and the true value.

4. Conclusion

The method of vehicle appearance damage identification based on computer vision proposed in this paper has a qualitative improvement compared with the traditional car insurance claim settlement business. In addition, the research on vehicle appearance damage identification in this paper indicates that the object recognition problem with unclear scope, such as appearance damage, brings challenges to model training and evaluation, and the traditional evaluation methods have limited evaluation of model effect.

Finally, the experiment indicates that our Mask R-CNN model based on KL-loss performs well in vehicle damage recognition, with good robustness and high accuracy. At the same time, the method of replacing the traditional IOU calculation accuracy with the component location greatly improves the effectiveness of the evaluation model results, and has good robustness in the actual experimental evaluation and commercial application. However, because the component identification model can not achieve 100% accuracy, there is still a certain gap between the evaluation method and manual evaluation. In the future, we will continue to improve the performance of the model and carry out commercial application.

References

[1] LeCun, Y., Bengio, Y., Hinton, G. Deep learning. Nature, 2015(521), pp. 436-444.
[2] Zeiler, M. D., Fergus, R. Visualizing and understanding convolutional neural networks. In ECCV, 2014, pp. 818-833.
[3] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]. European conference on computer vision. Springer, Cham, 2016, pp. 21-37.
[4] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. pp. 779-788.
[5] Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In NIPS, 2015, pp. 91-99.
[6] Kaiming He, Georgia Gkioxari, Piotr Dollar, Ross Girshick. Mask R-CNN. In ICCV, 2017, pp. 2980-2988.
[7] He K , Zhang X , Ren S , et al. Deep Residual Learning for Image Recognition[C]. IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2016, pp. 770-778.
[8] He Y, Zhu C, Wang J, et al. Bounding box regression with uncertainty for accurate object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019. pp. 2888-2897.