Research on Detection Algorithm of Wheel Position based on CenterNet

Yanyang Liu*, Jun Yan
AIPARK, Zhangjiakou city, Hebei province, China

*Corresponding author: liuyanyang@aipark.com

Abstract. In order to accurately determine the stopping state of a vehicle, a wheel location detection method combining enhanced images with CenterNet network was proposed. Image enhancement network was used to process the detected images to improve the contrast and sharpness of images. Then, CenterNet network was used to detect the wheel position in the enhanced image. The algorithm proposed in this paper can efficiently and accurately detect the vehicle wheel location, thus determining a variety of vehicle stopping states, so as to avoid the scene complexity limit, camera Angle and other factors that lead to the wrong judgment of stopping states, and greatly improve the efficiency of roadside parking management.

Keywords: image enhancement, target detection, wheel location detection, deep learning, centernet, (key words)

1. Introduction

In recent years, the surveillance technology based on high level video is more popular. This method collects images and videos of the vehicle and parking space information through cameras, and then analyzes and processes the information through computer vision technology, so as to achieve the monitoring and management of the parking state on the side of the road. Roadside parking management is a typical problem in urban traffic construction, tension and parking difficult problem for urban parking space resources, through the unity of the parking management and dynamic real-time monitoring, can make full use of space resources, realize effective integration of space resources, to solve the problem of drivers find parking difficult, provides the intelligent parking services and integration. Roadside parking, however, tend to have a parking space is not closed, the environment complex, thus to add to the difficulty of parking management, especially how to accurately determine parking parking Spaces within the state, is a difficult problem to be solved, but the existing parking state decision algorithm are susceptible to scene complexity limit, camera Angle, the influence of such factors as all cannot meet the current high camera Settings state judgement of parking demand.

In May 2020, Xingyu Zhou proposed the visual target Detection algorithm [1] LAVD (Large Angle Vehicle Detection) for parking management in complex scenarios based on deep learning. This algorithm is based on the detection of target Center point by Center Net algorithm, and the Anchor point idea of YOLO algorithm is added to solve the problem of missing overlapping targets inherent in the Anchorage-free algorithm. In January 2020, Guoan Zhu et al. proposed a mathematical analysis method [2] to determine whether there was any violation of the traffic line after the vehicle was
stopped. They mainly used probabilistic Hough transform to identify the traffic line area, and based on the established line analysis model. In December 2019, Zengchao Zhang et al. proposed a parking space vehicle detection method [3] based on the combination of geomagnetic sensor and Ultra Wideband (UWB) technology, which can improve the accuracy by about 6.0% compared with the single geomagnetic detection method. In May 2019, Zhan Li proposed a non-standard parking behavior identification method [4] based on surveillance video, and comprehensively utilized GIS technology, computer vision and other technologies to process surveillance video to realize the accurate identification of non-standard parking behaviors. In March 2019, Zhuang Shao proposed a vehicle stopping pressure line detection method [5] based on an improved convolutive neural network, and the improved Faster R-CNN model greatly improved the detection accuracy and speed of the target vehicle. In August 2017, Wei Jiang conducted in-depth research on the algorithm [6] of wheel detection and tracking based on video images, proposed a simple and efficient arc extraction method of wheel contour, and adopted the Hough transform method to conduct wheel detection. This algorithm can accurately detect and track the moving wheels in the practical application environment in real time. In May 2013, Yan Meng et al. proposed a parking space state detection system [7] based on sampling points. This system has good real-time performance and accuracy in embedded devices and can achieve the goal of system miniaturization.

This paper give full consideration to the existing target detection algorithm and the complexity of the actual application scenario, the deficiency of the existing vehicle detection algorithm now overlap between bounding box is more, the box features can not accurately represent the vehicle and large Angle complex circumstance, such as vehicles, keep out in this paper, starting from how to accurately state of a variety of vehicle parking, is proposed in this paper, the algorithm is based on the wheel the location to determine the state of a variety of parking vehicles, make full use of the Center of the Anchor - free type Net algorithm, the advantages of the focus on the characteristics of target Center, testing park in any parking space in the wheels of the vehicle with site location, It provides necessary preconditions for judging the stopping state of the vehicle. Based on the detected wheel position, various stopping states of vehicles can be efficiently and accurately judged, so as to avoid the occurrence of errors in stopping state judgment caused by scene complexity limit, camera Angle and other factors, and further, greatly improve the efficiency of roadside parking management.

2. Detect network

2.1. CenterNet network structure

Objects as Points proposed CenterNet network [8] for the first time. The core idea of this network is to regard the target as a point, the center of the target bounding box. The target detection problem has been transformed into an evaluation problem, and other target properties such as size, 3D position, direction and attitude all carry out parameter regression with the estimated center as the benchmark. The position of the target is determined by the estimation of the upper left corner and the lower right corner of the target.

CenterNet is similar to the one-stage method based on Anchor, and its center point can be regarded as the Anchor point without size. Its important difference lies in: (1) The anchor point assigned by CenterNet is only placed on the position without size, and there is no need to manually set the threshold value to distinguish foreground and background; (2) There is only one positive anchor point for each target, so NMS need not be used later. The key point is obtained by the local peak value on the feature map; (3) CenterNet uses a larger resolution feature map output (1/4 original image) than traditional target detection, so multi-scale feature maps such as FPN are not required.
Figure 1. CenterNet network structure diagram

The backbone network uses the following four kinds: ResNet-18 [9], ResNet-101 [10], DLA-34 [11], Hourglass-104 [12]. In the experiment, Deformable convolutional layer is used to optimize ResNet and DLA-34, hourglass-104 network remains unchanged.

2.2. Loss function

The loss function [8] consists of three parts: classification loss, center offset loss, and size loss.

It is assumed that the target of the input image \( I \in \mathbb{R}^{W \times H \times 3} \) is to generate a keypoint heatmap \( Y \in \mathbb{R}^{W \times H \times C} \), and the value of \( Y \) is [0, 1]. If it is 1, it is the detected keypoint, and if it is 0, it is the background. Where \( R \) is the output strid, that is, the size scaling scale, in the experiment \( R \) is 4; \( C \) is the total number of categories. Different full convolutional encoding and decoding networks are used to predict in the experiment.

For the key point \( P \) of the real value of each category \( C \), calculate its low-resolution equation \( \tilde{p} \), and conduct Gaussian processing for \( Y \):

\[
Y_{x,y,c} = \exp\left(-\frac{(x - \tilde{p}_x)^2 + (y - \tilde{p}_y)^2}{2\sigma_p^2}\right)
\]  

(1)

Use Focal Loss to reduce the penalty of pixel-level logistic regression:

(1) Classification of loss [13]:

\[
L_c = \frac{1}{N} \sum_{w=1}^{W} \sum_{h=1}^{H} \left\{ \begin{align*}
(1 - \hat{Y}_{w,h})^\alpha \log \hat{Y}_{w,h} & \text{ if } Y_{w,h} = 1 \\
(1 - \hat{Y}_{w,h})^\alpha (\hat{Y}_{w,h})^\beta & \text{ otherwise }
\end{align*} \right.
\]

(2)

Where \( \alpha \), \( \beta \) is the hyperparameter of Focal loss, which is set to 2 and 4 in the experiment; \( N \) is the number of key points in an image.

(2) Center migration loss:

As the image is sampled under convolution, the key points of Ground Truth will be deviated. In this paper, the prediction of local offset is added for each key point (the same predictive value is used for all categories), \( \hat{O}_{x,y} \in \mathbb{R}^{W \times H \times 1} \), This offset is trained using L1 loss, only at the \( p \) position of the key point, other positions are ignored.

\[
L_{off} = \frac{1}{N} \sum_{p} |\hat{O}_{p} - (\frac{p}{R} - \tilde{p})|.
\]  

(3)
(3) Scale loss:

\((x^k, y^k, x'_k, y'_k)\) is the bounding box of the target \(K\), so its central position is \(p_k = \left(\frac{x^k + x'_k}{2}, \frac{y^k + y'_k}{2}\right)\).

Therefore, the size \(S_k = (x^k - x'_k, y^k - y'_k)\) of the target can be estimated, and L1 Loss is added at the central point:

\[L_{\text{size}} = \frac{1}{N} \sum_{k=1}^{N} |\hat{S}_k - S_k|,\]  

(4)

Where the scale is not normalized, and the original pixel coordinates are directly used. In order to adjust the impact of the loss, it is multiplied by a coefficient. The target loss function [8] of the whole training is:

\[L_{\text{det}} = L_k + \lambda_{\text{size}} L_{\text{size}} + \lambda_{\text{off}} L_{\text{off}}\]  

(5)

Where \(\lambda_{\text{size}} = 0.1\), \(\lambda_{\text{off}} = 1\), the entire network predicts that it will output C+4 values at each location (i.e., \(x, y\), size \(w, h\) of the key point class \(C\), offset) all outputs share a full convoluted backbone.

3. Wheel location detection network

3.1. The network structure

The algorithm proposed in this paper is to enhance the original image containing the vehicle, and then input the enhanced image to the wheel landing location detection Network CenterNet, which is a new wheel landing location detection method. First, use treat detection images for image enhancement, image enhancement network [14] to improve the image contrast and resolution, CenterNet use "convolutional coding decoding of backbone network, CenterNet before all of the samples are deformable convolution, make network receptive field becomes more precise, used on sample is transposed convolution, can better restore image semantic information and location information. Fig. 2 is the overall flow chart of the method in this paper.

![Figure 2. The network structure](image)

The wheel states include three states: the wheel location is visible in the image, the wheel location is not visible in the image, and the wheel location is not in the image. If the state of the wheel to wheel the site visible in the image, in view of the video image is generated for each each wheel of the vehicle in the respective first heat map, among them, the first heat of the graph is 128 pixels by 128 pixels two-dimensional heat map, and the method respectively corresponding to the wheel of the first heat to the value of the position, the first heat to each position the value of the said probability of each position for the wheels with location if any location in the wheel center on site, is the location value is close to 1, the farther from the center point, if the position is, the smaller the value of the position.

(1) If the state of the wheel is that the wheel setting point is not visible in the image or the wheel setting point is not in the image, then the value of the wheel corresponding to the position in the first thermal diagram is a predetermined value of 0.
(2) If the state of the wheel is the place where the wheel is hitting, which can be seen in the image, the value of the position corresponding to the wheel in the first thermal diagram is close to 1. The value of the position in the first thermal diagram corresponding to the wheel is calculated by the following formula:

$$H'(x', y') = \exp\left(\frac{-(x' - x)^2 + (y' - y)^2}{2\sigma^2}\right), (c = 1, 2, 3, 4)$$ (6)

Where, $H'(x', y')$ is the vehicle of the first hot wheels to $H$ the pixel values, $i$ is the label of each wheel, $\exp(x) = e^x$ is the exponential function based on $e$, $\sigma$ is the standard deviation of the Gaussian function. In this paper, $\sigma = 2$, $x'$ is the horizontal axis coordinate value of the first wheel landing site, $y'$ is the vertical axis coordinate value of the second wheel landing site.

After generating the first thermal diagram of each wheel of each vehicle, the wheel landing location detection model based on convolutional neural network is trained. First of all, in this paper, resnet-18 is adopted as the backbone network, and 3 upper sampling layers are added after the last feature layer in ResNet-18. Specific training for the will of the image frame is set to 512 pixels by 512 pixels, input ResNet-18 of the size of the output characteristics of the last layer of 16 pixels x 16 pixels, after sampling on three layers, the final output size of 128 pixels by 128 pixels, spread via the ResNet-18 before the neural network output of each image in the second heat to each wheel location value $\hat{H}'$; 

The loss function $E_{loss}$ of the value of the position in the first thermal diagram of each wheel of each vehicle and the value of the position in the second thermal diagram of each wheel of each vehicle is obtained, i.e:

$$E_{loss} = \frac{1}{n} \sum_{i=1}^{n} \sum_{c=1}^{4} ||H'_c - \hat{H}'_c||^2$$ (7)

Where $n$ is the number of training image frames, $c$ is the label of each wheel, $i$ is a positive integer, $H'_c$ is the first thermal diagram of the $c$ th wheel generated from the first vehicle image, and $\hat{H}'_c$ is the second thermal diagram of the $c$ th wheel generated from the second vehicle sample; Through gradient descent algorithm, the predetermined residual network is adjusted by the principle of loss function $E_{loss}$, so that the predetermined residual network is optimized, and the optimal wheel landing location detection model based on convolutional neural network is obtained. Where, the value of each position in the second thermal diagram represents the probability that each position is a wheel setting point. If any position is in the center of the wheel setting point, the value of this position is close to 1; if the position is farther from the center point, the value of this position is smaller. The schematic diagram of wheel setting point detection network is shown in Fig. 3.

![Diagram](image)

**Figure 3.** Schematic diagram of wheel location detection network

4. **Experimental environment and data set**

In order to verify the effectiveness of the wheel location detection network proposed in this paper, AIPARK shot and constructed about 110,000 sample data sets in Haidian District, Beijing from June to November, 2019. The data sets covered a variety of images of actual scenes, such as video images of scenes during the day, night, rain, snow and so on. The coordinate information and wheel state of each vehicle in the image in the video image are manually marked. A sample annotation is shown in Fig.4.
The hardware configuration of the PC used in the experiment in this paper is as follows: CPU core-i7, GeForce RTX2080, and 16 GB of memory. The environment configuration is: TensorFlow1.10.0, Python 3.5, Linux Ubuntu 18.04.

5. Comparison of experimental results

![Figure 4. Example figures in dataset](image)

![Figure 5. Results of wheel landing location detection algorithm](image)

To test and verify the effectiveness of the proposed approach, evaluating the wheel with the site test network performance, this method respectively and the Hourglass - 104 network and DLA-34 the
wheels of the site testing method, no matter in the day light in good conditions or poor lighting conditions, the proposed method can effectively detect all wheels with locations, and under the condition of low exposure of the night, and DLA-34 Hourglass-104 network in the different levels of residual and the phenomenon of false detection. Fig.5 shows an example of the detection effect of the algorithm presented in this paper. Due to the introduction of an image enhancement network before segmentation, the method presented in this paper is still able to effectively detect the location of wheels in a variety of complex environments.

In order to evaluate the accuracy of the detection algorithm of wheel connection point, this paper uses the Object KeyPoint Similarity (OKS) [15], namely the similarity of key points.

\[
OKS = \frac{\sum [e^{\frac{-D_i^2}{2\sigma^2}} \delta(v_i > 0)]}{\sum [\delta(v_i > 0)]}
\]

(8)

Where \( s \) is the object scale, \( k_i \) is a keystroke control constant that controls attenuation, \( v_i \) is the visibility marker that represents the actual truth value, the value of 1 means that this key point is visible on the graph.

The calculation formula of AP (average precision) is as follows:

\[
AP @ s = \frac{\sum \delta(OKS_{\mu} > s)}{\sum \mu^{1}}
\]

(9)

The average detection accuracy pairs of each type of algorithm in the experiment are shown in Table 1. As can be seen from Table 1, the three methods all have different degrees of missing detection, which is due to the ambiguity of wheel landing sites under different lighting environments, frequent occlusion of wheel landing sites, and difficulty in accurate extraction of features. However, compared with the other two methods, the detection rate of the method proposed in this paper is significantly higher than the other two methods, and the speed is the fastest, reaching 69 fps.

| DataSet          | Algorithm      | FPS | AP S0 | AP S25 | AP S75 |
|------------------|----------------|-----|-------|--------|--------|
| Test Sets        | Hourglass-104  | 7   | 69.8  | 54.6   | 49.3   |
|                  | DLA-34         | 28  | 65.7  | 51.5   | 46.1   |
|                  | The algorithm used in this article | 69  | 71.8  | 60.6   | 52.3   |

6. Conclusion

In this paper, a wheel landing location detection method combining enhanced image and CenterNet was proposed to solve the problem that the scene complexity limit, camera Angle and other factors lead to the wrong judgment of stopping state. Meanwhile, it has been verified on multiple data sets in practical application scenarios. The experimental results showed that, on the same data set, the detection accuracy and speed of the method proposed in this paper were higher than those of other network detection methods when applied to the wheel landing site detection. The method performed stably in night scenes and had high robustness. In the future research, the method of image enhancement network combined with detection network will be further improved to reduce the number of missed detection and improve the accuracy of detection.

References

[1] Zhou Xingyu. Research and Implementation of Parking management Vision Algorithm based on deep learning complex scenes [D]. Chengdu: University of Electronic Science and Technology, 2020.
[2] Zhu Guoan, Wang Jinxiang. Research on the Detection Method of vehicle pressure line in parking Space based on video [J]. Neijiang Science and Technology. 2020, 1: 84-137.

[3] Zhang Zengchao; Li Qiang et al. Vehicle detection method and Implementation of Parking Space based on geomagnetic sensor and UWB technology [J]. Journal of sensing technology, 2019, 12: 1917-1922.

[4] li exhibition. Nonstandard parking Behavior identification method based on Surveillance video [D]. Nanjing: Nanjing Normal University, 2019.

[5] Shao Zhuang. Vehicle stopping line detection based on improved convolutional neural network [D]. Hefei: Anhui University, 2019.

[6] Jiang Wei. Research on Wheel Detection and Tracking based on video Images [D]. Wuhan: Huazhong University of Science and Technology, 2019.

[7] Meng Yan. Research on parking Space State Detection Technology based on machine vision [D]. Hangzhou: Zhejiang University of Technology, 2013. Xingyi Zhou, Dequan Wang. Objects as Points. In CVPR, 2019.

[8] Xingyi Zhou, Dequan Wang. Objects as Points. In CVPR, 2019.

[9] B. Xiao, H. Wu, and Y. Wei. Simple baselines for human pose estimation and tracking. In ECCV, 2018.

[10] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In CVPR, 2016.

[11] H. Law and J. Deng. CornerNet: Detecting objects as paired keypoints. In ECCV, 2018.

[12] F. Yu, D. Wang, E. Shelhamer, and T. Darrell. Deep layer aggregation. In CVPR, 2018.

[13] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar. Focal loss for dense object detection. ICCV, 2017.

[14] IGNATOV A,KOBYSEV N,TIMOFTET,et al,DSLR-qualityphotos on mobile devices with deep convolutional networks [C] /Proceedings of the 2017 IEEE International Conference on Computer Vision.Washington,DC: IEEE Computer Society, 2017: 3277- 3285.

[15] Khang Nguyen, Nhut T. Huynh, Phat C. Nguyen, Khanh-Duy Nguyen, Nguyen D. Vo, Tam V. Nguyen. Detecting Objects from Space: An Evaluation of Deep-Learning Modern Approaches.Volume 9, Issue 4. 2020.