IOP Conference Series: Earth and
Environmental Science

Driver Cell-phone Use Detection Based on CornerNet-Lite Network

To cite this article: Anqing He *et al* 2021 *IOP Conf. Ser.: Earth Environ. Sci.* **632** 042004

Driver Cell-phone Use Detection Based on CornerNet-Lite Network

Anqing He[1], Guohua Chen[1], Wei Zheng[1], Zhenhao Ni[1], Qingqing Zhang[2] and Zhongjie Zhu[2, *]

[1]Zhejiang CRRC electric vehicle Co. Ltd. Wuxiang Town, Yinzhou District, Ningbo, China
[2]College of Information and Intelligence Engineering, Zhejiang Wanli University, Ningbo, China

*Corresponding author e-mail: zhongjiezhu@zwu.edu.cn

Abstract. Driver requires a high level of concentration when driving at high speed. Cell-phone use while driving can cause serious problems. To solve this issue, this paper proposes a proposal scheme about driver cell-phone use detection based on CornerNet-Lite Network. The scheme can eliminate the traffic accident risk caused by cell-phone use through detecting driver cell-phone use. The scheme includes two stages: model training and simulation test. Firstly, the data set of cell-phone use was established. Secondly, the data set was preprocessed by the preset processing method. Finally, the CornerNet-Lite network architecture was optimized to reduce the training time and improve the detection accuracy and real-time detection. Through a large number of experiments, the results showed that the scheme had a good detection effect, with the accuracy of 86.2% and with 30fps. Under the noise interference of simulated cab environment, it still had a high robustness.

Keywords: Cell-phone use, target detection, CornerNet-Lite.

1. Introduction

Public transportation safety is a social topic that we have been paying more attention to [1]. Distracted driving is a common misbehavior, which is one of the important factors leading to traffic accidents [2]. Cell-phone use is a form of distracted driving, this dangerous behavior that can cause loss of life and goods. Therefore, it is important to detect cell-phone use while driving dynamically and issue warnings against such behavior [3].

For the target detection of cell-phone use, the main problem is to solve hand and mobile phone detection and recognition. Target detection technology is a hot topic in modern image processing, which has a good research basis and application scenarios [4, 5]. At present, the target detection technologies are mainly divided into two categories: On the one hand, the traditional methods based on artificial features [6-9]. On the other hand, the methods based on deep learning neural network [10-13]. In the traditional method, various artificial features were designed according to the target features. Then, the target was detected and identified by regression. Viola et al. proposed a fast target detection method was proposed which used the enhanced cascade of simple Haar features [14]. Dalal et al. the feature of directional gradient histogram (HOG) was selected and the support vector machine was used as the classifier to detect the target [15]. With the progress and development of science and technology, deep learning methods play an important role in target detection. First of all, Alex et al. AlexNet was proposed to establish the dominant position of neural network in computer vision [16, 17]. Then, Ross et al. proposed R-CNN series network, which realized target detection through this scheme and greatly improved the speed and accuracy of target detection [18]. Later, VGGNet was proposed by Simonyan, which showed that the network was easier to project and was close to our life [19-22]. Then, He et al. put forward ResNet network which realized the deeper network structure and promoted the development of target detection. Joseph et al. a series of YOLO series networks was proposed, this deep network had been in unprecedented position in computer vision [23-25]. These deep learning networks mentioned can directly predict the confidence of bounding boxes and their categories. However, the method mentioned above is not good enough in the real-time detection of mobile phones in this paper, and the detection accuracy is not enough.

In order to solve the above mentioned problems, the existing driver cell-phone detection technology cannot meet the needs of most people, a more effective technology is urgently needed. Therefore, this paper proposes a scheme for detecting driver cell-phone use based on CornerNet-Lite Network [26]. In our scheme, to solve the problem that the detection accuracy is not high, the data set is established manually and a variety of preset processing methods was used in the data set. For the lack of good real-time performance, we adopted the CornerNet-Lite network architecture of anchor-free and optimized its backbone network [27, 28].

The main contributions of this paper are as followed: first, since there is no publicly available data set for the detection of driver cell-phone use, a data set has been built specifically for driver cell-phone use by downloading pictures from the Internet and taking photos manually on the spot. Second, the data set images are optimized, driving vehicle detection is a continuous action belongs to the movement frequency of target detection, the requirements of the data set are higher during the establishment of a sample data set, we added a variety of noise attack to make our data set has better precision and robustness. Third, the CornerNet-Lite network architecture are optimized which adding an hourglass module and reducing the number of layers of the hourglass network to achieve better and faster detection speed for feature extraction in the hourglass module in the backbone network. Through a large number of experiments, the results show that our proposed scheme is encouraged and effective, with the accuracy of 86.2% and 30fps. Under the noise interference of simulated cab environment, it still has a high robustness.

The remainder of this article is arranged as follows: in part 2, the proposed solution is described in detail, including the establishment of the database, the training phase, and the testing phase. Experimental results and analysis are presented in the third part. Finally, we introduce the conclusion of our work in the fourth section.

2. Proposed scheme

The proposal in this paper is CornerNet-Lite network optimization and the combination of image preprocessing and noise interference to simulate the complex scene of cab. Firstly, since there is no publicly available data set for cell-phone use detection, we established the data set of this paper through shooting and online search. Secondly, we expanded the sample set of collected images through a large number of image enhancement techniques and simulated the complex scene of cab with noise attack. Third, CornerNet-Lite network optimization. The backbones in the original CornerNet-Lite network architecture are composed of two hourglass modules. This paper adopts the simplified hourglass moudle to optimize the number of network layers, and three simplified hourglass module are used to replace the original backbones. The proposed scheme block diagram is shown in figure 1.

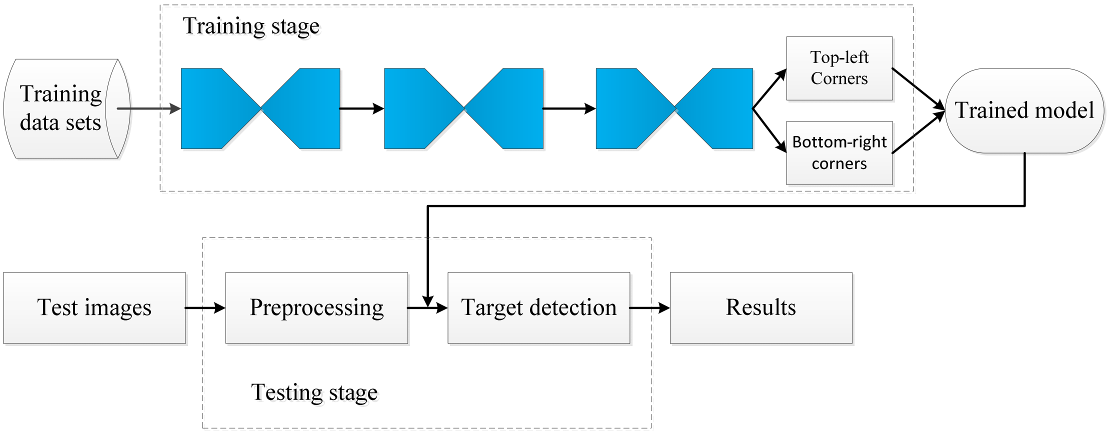


Figure 1. Diagram of the proposed scheme.

2.1. Establishment of data set

In general, the performance of neural network depends on the number of samples. Since there is no public database of cell-phone use, the experiment collected 545 relevant pictures, randomly selected 400 pictures as the training set, and the remaining 145 pictures as the test set. Aiming at the problem of data set sample limitation, this paper uses image inversion, scaling and other image preprocessing methods to expand the sample so as to achieve better detection effect. At the same time, these images cover different lighting, shooting Angle, resolution, detection background and other conditions, which meet the sample diversity and targeted needs and are of great significance for the improvement of algorithm detection robustness.

These images include different lighting conditions, vehicle shooting angles, resolution, road environment and road conditions thus meeting the requirements of sample diversity and ensuring the purposeful optimization of the data set. The data set must be optimized to improve the robustness of the correlation algorithm. If the data set takes no account of lighting and contains only images captured under a single lighting condition, a trained model may cause false detection or mission when evaluating poorly lit images or videos. For the same reason, images involving different road environments and resolutions are also considered. Optimizing the data set is valuable for achieving better detection results with fewer training samples. The following figure shows several sample images from the dataset.

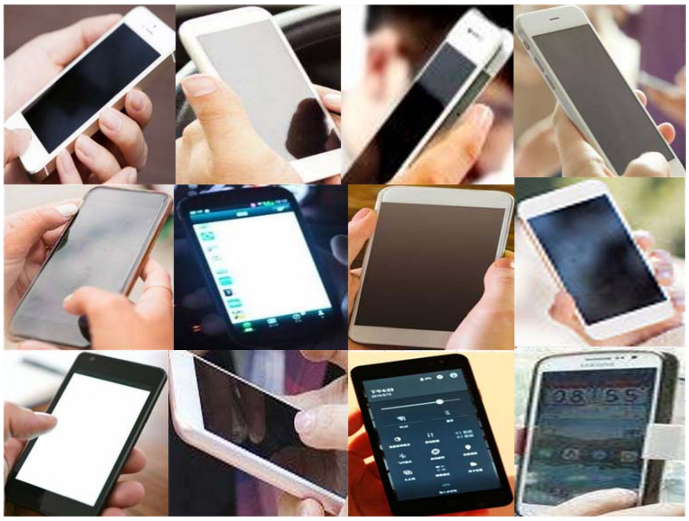


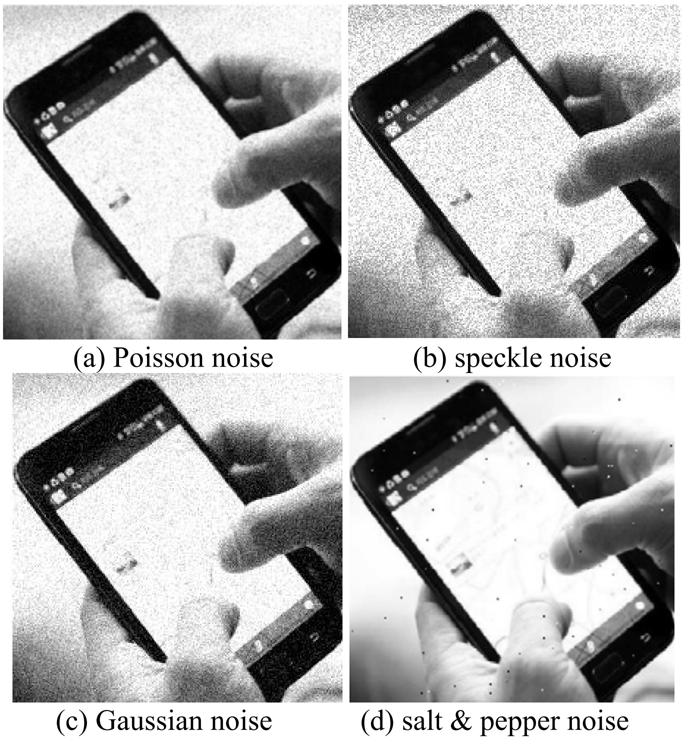
Figure 2. Sample data set of some cell-phones.

2.2. Establishment of data set

This paper also takes some advantageous methods when train set and test set were established. Because the data set contains three types of large object, medium object and small object. Cell-phone use in this paper belongs to a small object. In general, we observed that there were fewer sample data sets of small targets in our self-built data set, which potentially made the target detection model pay more attention to the detection of large targets. Therefore, the small target feature extraction ability will be improved through the optimization of the Cornernet network architecture in this paper. In this background, if not mark small target will have an impact the results of training in some resolution is lower and even to the naked eye looks vague smaller target also on the mark, the purpose is to let all training process model can be fully extracted the characteristics of the target and enhance the robustness of training model. Therefore, when it comes to clear images, it is possible to accurately label specific objects and try to use a fitting label bounding boxes to frame the object when labeling so that the model can more accurately locate the coordinates of the objects in the picture, which is more conducive to the extraction of object features.

Image augmentation technology produces a series of random changes to the original images to generate similar but different samples, thereby expanding the size of data set and also reducing the dependence of model on certain attributes, thereby increasing the generality of the model. In this paper, the methods of image augmentation include panning, zooming, horizontal flipping, color transformation and so on. Each augmented image is obtained by lots of random combination transformations of the edge images.

In this paper, in order to measure the robustness of the model, Poisson noise, speckle noise, Gaussian noise and Salt and Pepper noise are used to simulate the attack on the test sample set so as to test the robustness and universality of our improved CornerNet-Lite network in the complex driving environment. The test set picture of part of the simulated noise attack is shown in figure 3.

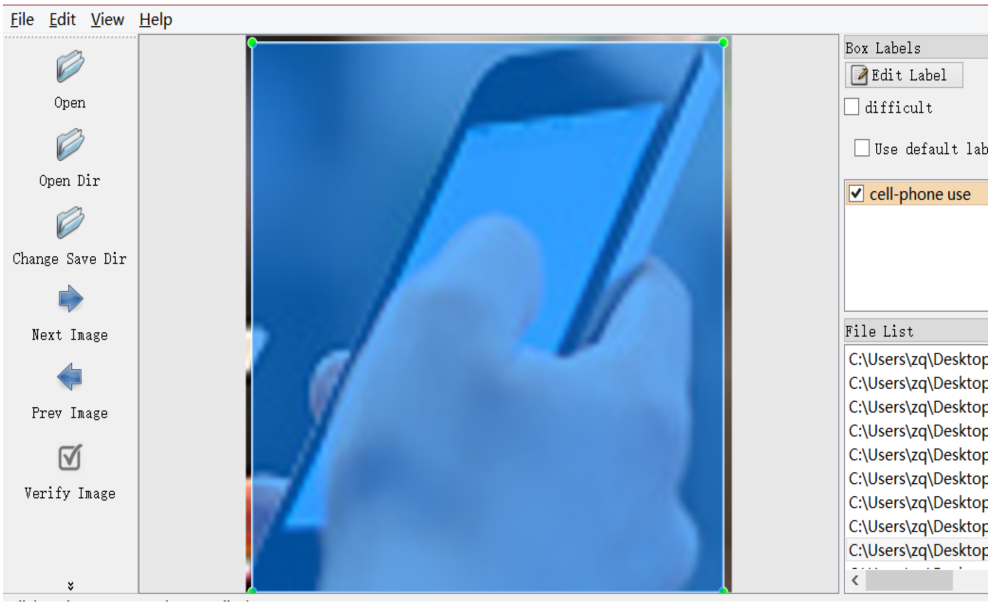


(a) Poisson noise (b) speckle noise

(c) Gaussian noise (d) salt & pepper noise

Figure 3. Different types of noise simulate attack pictures.

In this paper, cell-phone use is set as the detection target. After the annotation is completed, the category of the target and the XML file of the target object will be obtained, as shown in figure 4.



□ difficult

□ Use default lab

☑ cell-phone use

C:\Users\zq\Desktop
C:\Users\zq\Desktop
C:\Users\zq\Desktop
C:\Users\zq\Desktop
C:\Users\zq\Desktop
C:\Users\zq\Desktop
C:\Users\zq\Desktop
C:\Users\zq\Desktop
C:\Users\zq\Desktop

Figure 4. Labeling process.

2.3. *Optimization of CornerNet-Lite network*

The backbone of the CornerNet-Lite network is the Hourglass module. The module in the original CornerNet-Lite network is characterized by two Hourglass. In this paper, cell-phone use detection is a smaller target, therefore more and more accurate image information should be extracted. An Hourglass module is added in the improved CornerNet-Lite network. The time of object detection depends on the depth layer of the network. The more depth layers there are, the longer the training time of the model will be and the greater the calculation amount will be. The number of network layers of the Hourglass module is optimized to simplify the backbone network and its depth is changed to 54 layers.

The prediction module starts with a modified remnant where CornerNet-Lite replaces the first convolution module with a corner pooling module. The modified residuals are followed by a convolution module. CornerNet-Lite has multiple branches for predicting Heatmaps, Embeddings, and Offsets. The first part of the module is the modified version of the residual module. In this modified residual module, CornerNet-Lite replaces the first 3×3 convolution module with a corner pooling module. The residual module first processes the characteristics through the network of two 3×3 convolution modules with 128 channels and then applies a corner pooling layer. After the residual module, CornerNet-Lite inputs the pooled feature into a 3×3 convolutional layer with 256 channels and then adds the reverse residual module. The modified residue is followed by a 3×3 convolution module with 256 channels and 3 convolution layers of 256 channels to generate Heat maps, Embeddings, and Offsets.

For each corner, there is a positive position of ground-truth and all other positions are negative. Instead of penalizing negative positions equally during training, CornerNet-Lite penalizes negative positions within the positive position radius. The reason why is that if a pair of false corner point detectors are close to their respective ground-truth positions, it can still produce a boundary box that fully overlaps with ground-truth. CornerNet-Lite determines the size of the object by ensuring that the bounding box generated by a pair of points within the radius is greater than or equal to the IOU of ground-truth. We set IOU to 0.7 in experiments. For a given radius, the reduction in punishment is given by the non-normalized 2D Gaussian with its center in the positive position, and it is 1/3 of the radius. Pcxy is the score of the predicted c position (x, y) in the heat map, while Ycxy is the ground-truth heatmaps enhanced by non-standardized Gaussian. CornerNet-Lite has a variant of focal loss.

$$F\det = \frac{-1}{n}\sum_{c=0}^{A}\sum_{x=1}^{B}\sum_{y=0}^{C}\begin{cases}(1-Pcxy)^{\alpha}\log(Pcxy) & \text{if } Ycxy=1\\ (1-Xcxy)\sigma^{\beta}(Pcxy)^{\alpha}\log(1-Pcxy) & \text{if } Ycxy\neq 1\end{cases} \tag{1}$$

Where n is the number of targets in the image, α and β are the super parameters that control the contribution of each point. Gaussian convex points encoded in Ycxy (1−Ycxy) were used to reduce the punishment around the ground-truth.

Many networks involve a lower sampling layer to collect global information and reduce memory usage. When their full convolution is applied to the image so the size of the output is usually smaller than the image. Therefore, the position (m, n) in the image is mapped to the position in the Heatmaps where i is the down sampling factor. When CornerNet-Lite remolts the position in the Heatmaps to the input image and some precision can be lost, which can greatly affect the IOU between the small bounding box and the ground-truth. To solve this problem, CornerNet-Lite predicts the position offset to slightly adjust the corner position, and then remakes them to the input resolution.

$$\Delta o = \{\frac{Mo}{i} - [\frac{Mo}{i}], \frac{No}{i} - [\frac{No}{i}]\} \tag{2}$$

Where is the offset, M_O and N_O are the coordinates of *m* and *n* of Angle *O*. In particular, CornerNet-Lite predicts that the top-left corner of all categories shares offsets and the bottom-right corner shares another offsets. For training, CornerNet-Lite applies a smooth L_1 loss to the ground-truth corner position.

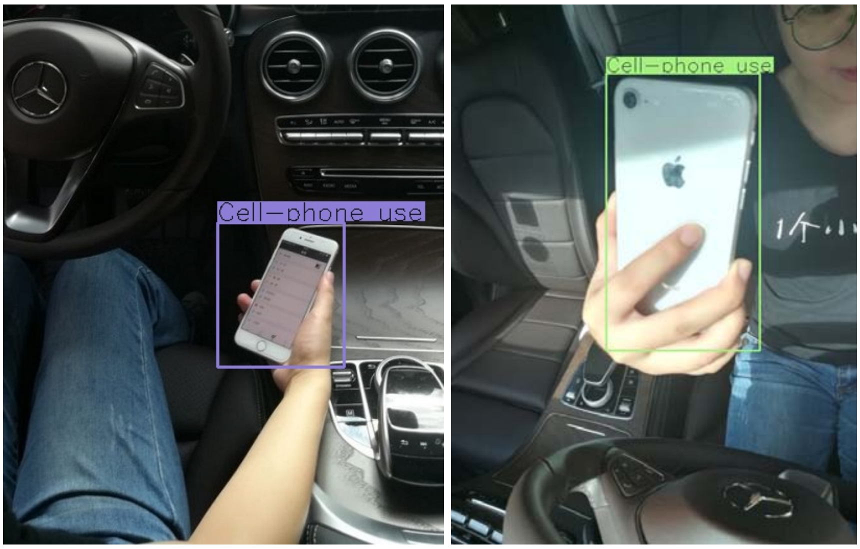
$$Foff = \frac{1}{n}\sum_{o=1}^{n} smoothL1Loss(Oo, \hat{O}o) \tag{3}$$

3. Experimental results and analysis

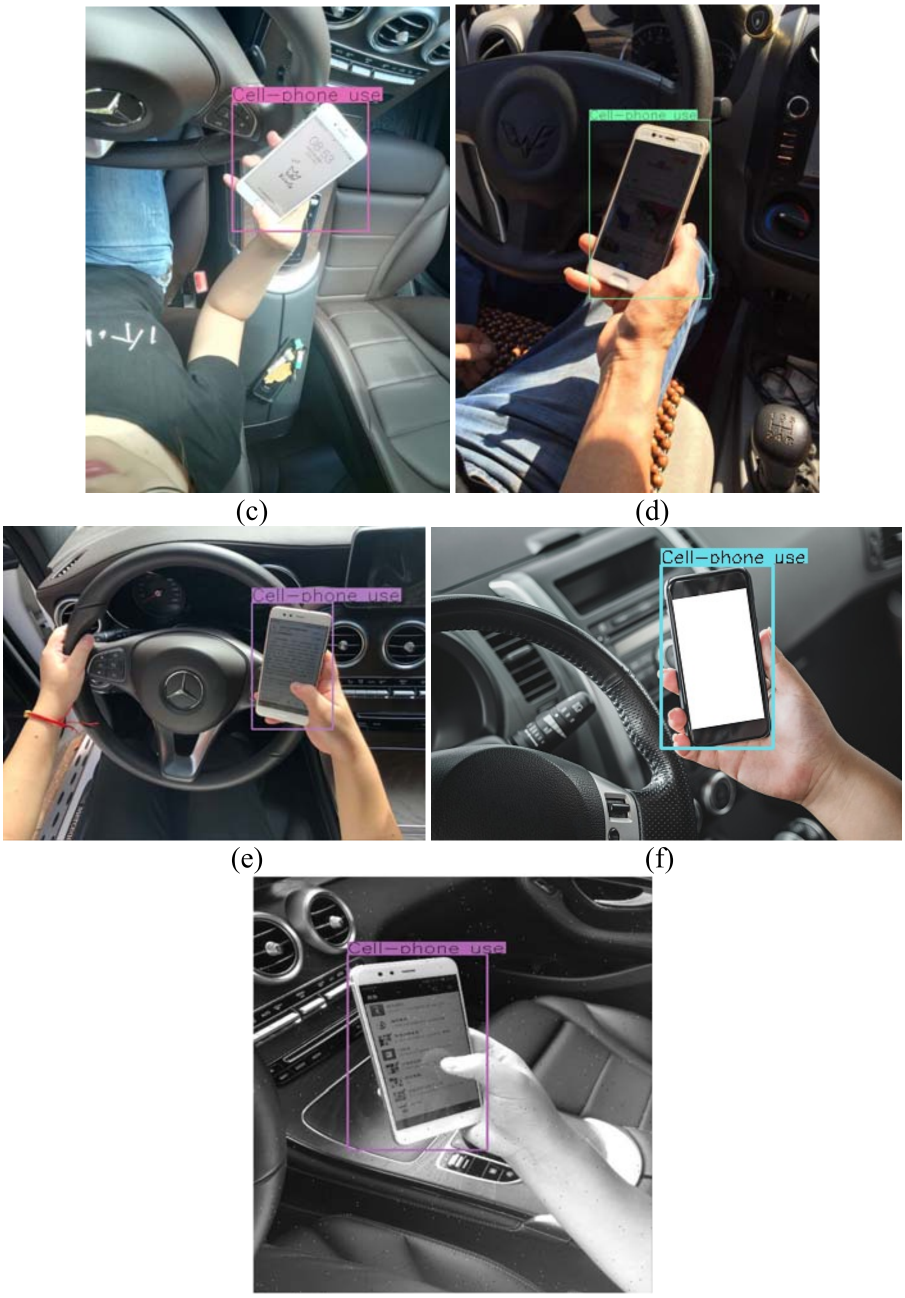
Intel i9-9920x processor, RTX2080, and cuda10.2 and cudnn7.4 under Linux operating system.

3.1. Visualization of the experimental results

A part of experimental results for driver cell-phone use is shown below.



(a) (b)



(c) (d)

(e) (f)

(g)

Figure 5. Results of some experiments

Table 1. Comparison of detection effect of CornerNet-Lite network before and after improvement

Type of network	Recall	Precision	Test images	time
Original CornerNet-Lite	0.753	81.4%	145	30ms
Improved CornerNet-Lite	0.838	86.2%	145	30ms

It was found that the original CornerNet-Lite network had poor detection effect under the same data test set. The Improved CornerNet-Lite network was improved in both recall and precision. It is proved that the improved scheme is effective. It has high timeliness and has a great improvement effect in recall rate and precision rate.

Table 2. Comparison of detection effects of different noises

Type of Noise	Recall	Precision	Test images	time
Improved CornerNet-Lite	0.838	86.2%	145	30ms
Noise simulation	0.800	80.2%	145	30ms

According to table 2, Gaussian noise, Pepper and salt noise, Multiplier noise and Poisson noise are respectively used to simulate the attack on the test set images. Under the influence of complex driving environment, the network model in this paper has some influence on the detection effect but still has high accuracy and timeliness.

4. Conclusion

A cell-phone use behavior detection scheme that is based on improved CornerNet-Lite is proposed in this paper. This scheme can effectively eliminate the function of timely warning when the driver is driving the vehicle at high speed. Firstly, data set is established through online collection and artificial field in this paper. secondly, through setting the preprocessing of image preprocessing of data sets, through a variety of noise including Gaussian noise, Poisson noise and Multiplicative noise and Salt and Pepper noise for simulating test set pictures in order to prove the detection model robustness and reliability. Finally, through the network optimization of the backbone network to reduce the hourglass network layer and to reduce the testing time and quantity of the budget. A large number of experiments have proved that this scheme can accurately detect the driver cell-phone use behavior in real time, with an operating time of 30fps and an accuracy rate of 86.2%. At the same time, the design of the end-to-end network structure will be the main research direction in the future.

Acknowledgments

National Natural Science Foundations of China: 61671412. Zhejiang Provincial Natural Science Foundation of China: LY19F010002. Natural Science Foundation of Ningbo, China: 2018A610053. Ningbo Municipal Project of Science and Technology beneficial to People: 2017C50011. Innovation and consulting project from Zhejiang CRRC electric vehicle Co.Ltd. Wuxiang Town, Yinzhou District, Ningbo, China

References

[1] V Žuraulis, Nagurnas S, R Pečeliūnas, et al. “The analysis of drivers' reaction time using cell phone In the case of vehicle stabilization task,” International Journal of Occupational Medicine & Environmental Health, vol. 31, no. 5, pp: 633-648, Oct. 2018.

[2] Mccartt A T, Kidd D G, Teoh E R. “Driver Cellphone and Texting Bans in the United States,” Evidence of Effectiveness. Ann Adv Automot Med, vol.58, pp: 99-114, Dec. 2014.

[3] Rodríguez-Ascariz J M, Boquete L, Cantos J, et al. “Automatic system for detecting driver use of mobile phones,” Transportation Research Part C Emerging Technologies, vol. 19, no. 4, pp: 73-681, Dec. 2011.

[4] Jiang, Huaizu , et al. "Joint salient object detection and existence prediction." Frontiers of Computer ence, pp: 778-788, 2019.

[5] Lesani, Mohsen. Transaction Protocol Verification with Labeled Synchronization Logic. Springer, Cham, 2019.

[6] Viola P, Jones M. “Rapid object detection using a boosted cascade of simple features,” CVPR (1), vol. 1, no. 511-518, pp: 3, Dec. 2001.

[7] Dalal N, Triggs B. “Histograms of Oriented Gradients for Human Detection,” 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005.

[8] Liao S, Jain A K, Li S Z. “A Fast and Accurate Unconstrained Face Detector,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 2, pp: 211-223, Jun. 2015.

[9] Felzenszwalb P, McAllester D, Ramanan D. “A discriminatively trained, multiscale, deformable part model,” 2008 IEEE Conference on Computer Vision and Pattern Recognition, pp: 1-8, Jun. 2008.

[10] Yang J, Sidhom S, Chandrasekaran G, et al. “Sensing driver phone use with acoustic ranging through car speakers,” IEEE Transactions on Mobile Computing, vol. 11, no. 9, pp: 1426-1440, Apr. 2012.

[11] Rosen M. Method and system for automated detection of mobile phone usage: U.S. Patent 8,384, 555 [P]. 2013-2-26.

[12] Zeinstra M L, Vanderwall P J. In-Vehicle Electronic Device Usage Blocker: U.S. Patent Application 13/978,540 [P]. 2013-11-7.

[13] Nian-Feng L , He G , Meng Z , et al. “Study on mobile phone call monitoring and positioning and shielding algorithms,” Proceedings 2011 International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE), pp: 1771-1774, Changchun, Dec. 2011.

[14] Li N F, Zhang M, Yang Y J, et al. “Key Technologies and Implementation,” In Study on a Kind of Mobile Phone Signals Monitoring and Shielding System, G.Lee (Ed.): Advances in Intelligent Systems, pp: 321-326, 2012.

[15] Guang-Long, Wang, et al. HOGHS and Zernike Moments Features-Based Motion-Blurred Object Tracking. International Journal of Humanoid Robotics. 2019,2,16.

[16] Lesani, Mohsen. Transaction Protocol Verification with Labeled Synchronization Logic. Springer, Cham, 2019.

[17] Krizhevsky, A.;Sutskever, I.;Hinton,G.E. In proceedings of ImageNet Classification with Deep Convolutional Neural Networks. International Conference on Neural Information Processing Systems, Lake Tahoe, 3-6 December 2012; pp.1097-1105.

[18] Menendez,O.A.;Perez,M.;Cheein,F.A.A. Vision based inspection of transmission lines using unmanned aerial vehicles. In proceedings of the 2016 IEEE International Conference on Multisensory Fusion and Integration for Intelligent Systems (MFI), 2016.

[19] Lu, X.; Li, B.Y.; Yue, Y. X. Grid R-CNN. 2018, 11, 29.

[20] Z.Cai.; N. Vasconcelos. Cascade r-cnn: Delving into high quality object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018; pp. 6154–6162.

[21] B.Cheng.; Y.Wei.; H.Shi.; R.Feris.; J.Xiong.; T.Huang. Revisiting rcnn: On awakening the classification power of faster rcnn. In Proceedings of the European Conference on Computer Vision (ECCV), 2018; pp.453–468.

[22] Girshick, Ross. Fast R-CNN. In proceedings of 2015 IEEE International Conference on Computer Vision (ICCV). 2016.

[23] He, Kaiming. Deep Residual Learning for Image Recognition. In proceedings of IEEE Conference on Computer Vision& Pattern Recognition IEEE Computer Society. 2016.

[24] Redmon, J.; Divvala,S.; Girshick, R. You Only Look Once: Unified, Real-Time Object Detection. 2015.

[25] Redmon, J.; Farhadi, A.YOLO9000: better, faster, stronger. In proceedings of 2017IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI. New York: IEEE, 2017; pp.6517-6525.

[26] Redmon, J.; FarhadI, A. YOLOv3: An incremental improvement. 2018, 04, 08.

[27] Tian, Yunong. Apple detection during different growth stages in orchards using the improved YOLO-V3 model. Computers & Electronics in Agriculture, vol 157, 2019; pp: 417-426.

[28] Zhou, X.Y.; Wang,D.Q.; Krähenbühl. Objects as Points. 2019, 04, 16.