Vehicle LED detection and segmentation recognition based on deep learning for optical camera communication

CHENG Qing, MA Haitao*, and SUN Xu

Department of Communications Engineering, Jilin University, Changchun 130012, China

(Received 18 February 2022; Revised 28 April 2022)

In the vehicle to vehicle (V2V) communication based on optical camera communication (OCC) system, how to achieve high reliability and low latency communication is still a problem. In this paper, we propose a lightweight light-emitting diode (LED) detection algorithm based on deep learning to detect the vehicle LED position at different communication distances, which can improve LED detection accuracy and inference speed. In addition, we design an LED segmentation recognition algorithm to reduce the bit error rate (*BER*) of the vehicle OCC system. The experimental results demonstrate the effectiveness of the proposed algorithms in real traffic scenes.

Document code: A **Article ID:** 1673-1905(2022)08-0508-5

DOI https://doi.org/10.1007/s11801-022-2026-5

In the intelligent transportation system (ITS), vehicle to vehicle (V2V) communication exchanges driving information between vehicles, which can reduce traffic accidents and improve road safety[1]. The current V2V communication generally uses radio frequency (RF) communication technology. Although this technology tends to mature, it is still faced with the shortage of spectrum resources and the sensitivity of electromagnetic radiation[2].

Optical camera communication (OCC) has the advantages of rich spectrum, no radio frequency interference and high confidentiality, which can effectively supplement the defects of radio frequency (RF) communication[3,4]. The OCC is a potential communication technology in the V2V scenes[5]. The vehicle OCC system uses the light-emitting diode (LED) as transmitter and the camera as receiver, which can utilize the light intensity, the LED color and spatial domain information[6].

Although the vehicle OCC system has numerous benefits, it is easily affected by environment factors, such as street lights and LED billboards[7]. The captured image often contains many interference signals, which makes the vehicle LED detection more difficult. At present, the main solutions include two aspects. On the one hand, the LED detection method based on traditional image processing is generally simple, but its robustness is poor, which affects the reliability of the vehicle OCC system in complex traffic scenes[8]. On the other hand, the LED detection method based on deep learning has strong learning ability and generalization ability, but its computational complexity is usually high, which increases the latency of the vehicle OCC system[9].

In 2019, LIU et al[10] improved the particle filter and Cam-Shift algorithm to fast determine the vehicle LED position in the captured image, but the error was still high in real traffic scenes. In 2020, PHAM et al[11] employed the you only look once version2 (YOLOv2) to detect the vehicle LED, which achieved the detection accuracy of 60% and the inference speed of around 40 fps on GPU NVIDIA GTX 1080Ti. In 2021, SUN et al[12] proposed a D2Net algorithm based on deep learning to detect the LED array and remove the vehicle motion blur in real traffic scenes. The algorithm had a detection accuracy of 99.81%, and its inference speed was only 39 fps.

In this paper, we propose an LED detection algorithm to detect the vehicle LED position at different communication distances. This algorithm adopts a lightweight backbone network, an efficient detection neck and a multi-scale detection head to improve LED detection accuracy and inference speed. We optimize the LED segmentation recognition algorithm[12] to identify the vehicle LED states. The algorithm can dynamically expand the predicted box obtained by the LED detection algorithm, which is more flexible for real traffic scenes, and can reduce the vehicle OCC system bit error rate (*BER*).

Fig.1 is the function diagram of the vehicle OCC system. In the transmitter, the on-off keying (OOK) modulation is applied to the data signals. The LED driver controls the light and dark flickering of the vehicle LED to send these data signals. In the receiver, the camera is

* E-mail: maht@jlu.edu.cn

employed to capture images with the vehicle LED. We use the proposed LED detection algorithm and the optimized LED segmentation recognition algorithm to obtain the state of each LED.

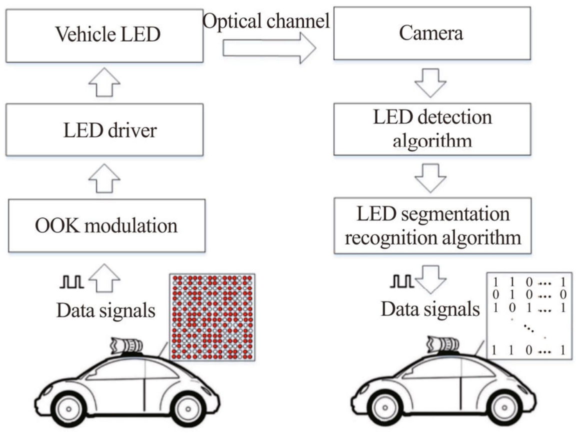

Fig.1 Function diagram of the vehicle OCC system

The input image is sampled to a fixed size of 320×320×3 by using bilinear interpolation. The proposed LED detection algorithm detects the vehicle LED position in the input image. The structure of this algorithm is shown in Fig.2. It mainly consists of backbone network, LED detection neck, and multi-scale LED detection head.

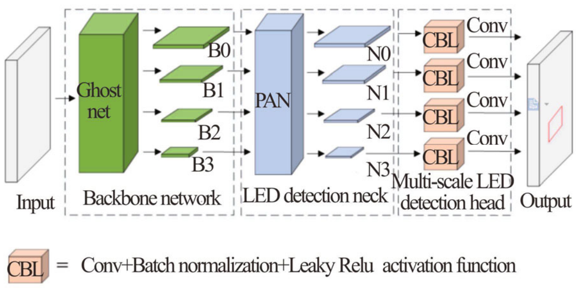

Fig.2 Structure of the proposed LED detection algorithm

The lightweight GhostNet[13] is employed as backbone network to extract the basic features of the input image. It outputs four feature maps B0: 80×80×24, B1: 40×40×40, B2: 20×20×112 and B3: 10×10×160. The core of the network is the Ghost module whose structure is shown in Fig.3. Firstly, the feature maps generated by convolution are linearly transformed. Then, these transformed feature maps are concatenated with the feature maps obtained by the identity map. The Ghost module can effectively reduce feature redundancy and enhance computational efficiency.

The shallow feature maps of the backbone network usually contain local information, such as texture, contour and shape. The deep feature maps usually contain global semantic information. We use the path aggregation network (PANet)[14] as the LED detection neck to fuse these feature maps. Fig.4 is the PANet structure. Firstly, the B0, B1, B2 and B3 are fused from the bottom to the top by convolution, upsampling and concatenating operations. Then, the feature maps N0: 80×80×24, N1: 40×40×40, N2: 20×20×112 and N3: 10×10×160 are obtained by top-down fusion of the feature maps P0, P1, P2 and P3.

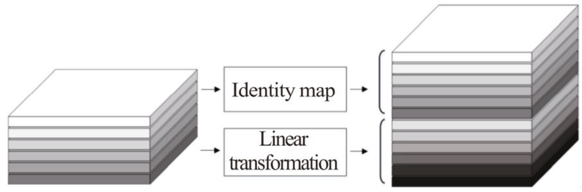

Fig.3 Structure of the Ghost module

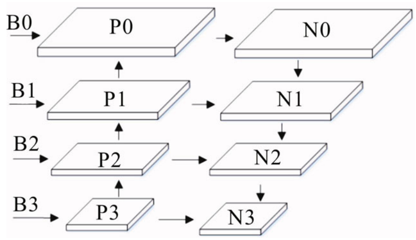

Fig.4 PANet structure

We utilize multiple feature maps of different scales to detect the vehicle LED at various communication distances. The multi-scale LED detection head is similar to the YOLOv4-tiny[15] detection head. Unlike it, an output feature map Out0: 80×80×18 is added to accurately detect extra-small vehicle LED. The other feature maps are Out1: 40×40×18 for the small vehicle LED prediction, Out2: 20×20×18 for the medium vehicle LED prediction, and Out3: 10×10×18 for the large vehicle LED prediction. Specifically, the N0, N1, N2 and N3 are converted to many candidate anchor boxes based on grids, which are distributed on four feature planes. These anchor boxes contain the center point coordinates, width and height, confidence score and category of the predicted box.

The loss function L of the proposed LED detection algorithm is expressed as

$$L = L_{\mathrm{class}} + L_{\mathrm{conf}} + L_{\mathrm{local}}, \tag{1}$$

where L_{class} represents the classification loss that adopts cross entropy loss function, which is used to quantify the error between the predicted class and the ground truth class, L_{conf} represents the confidence loss that utilizes cross entropy loss function, which is used to quantify whether there is the vehicle LED in the anchor boxes, and L_{local} represents the regression loss that employs the distance intersection over union (DIoU) loss function[16], which is used to quantify the position error between the predicted box and the ground truth box.

As the communication distance increases, the deviation between the predicted box and the ground truth box increases, which seriously affects the vehicle LED detection accuracy. The LED segmentation recognition algorithm[12] expands the predicted box by 1.2 times to contain the vehicle LED, but its poor flexibility introduces noise. Therefore, we optimize the algorithm by dynamically expanding the predicted box, and the formulas are shown as

$$u_{\min} = x_{\min} - \beta(x_{\max} - x_{\min}), \tag{2}$$

$$u_{\max} = x_{\max} + \beta(x_{\max} - x_{\min}), \tag{3}$$

$$v_{\min} = y_{\min} - \beta(y_{\max} - y_{\min}), \tag{4}$$

$$v_{\max} = y_{\max} + \beta(y_{\max} - y_{\min}), \tag{5}$$

$$\beta = 0.2 \times \frac{\frac{x_{\max} - x_{\min}}{w} \times \frac{y_{\max} - y_{\min}}{h}}{S_k} + 1, \tag{6}$$

where $x_{\min}$, $x_{\max}$, $y_{\min}$ and $y_{\max}$ are the minimum and maximum values of the abscissa x and ordinate y of the predicted box, respectively, $u_{\min}$, $u_{\max}$, $v_{\min}$ and $v_{\max}$ are the minimum and maximum values of the abscissa x and ordinate y of the expanded box, respectively, w is the image width, h is the image height, S_k is the average area of the ground truth box on the unit plane at 1 m (S_k=0.015), and β is the expansion factor that can be adjusted automatically.

The process of dynamically expanding the predicted box is shown in Fig.5.

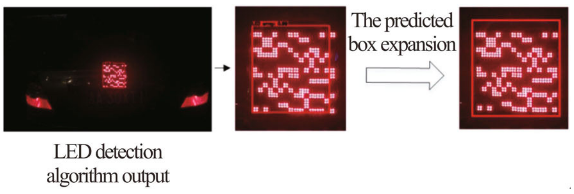

Fig.5 Process of the predicted box expansion

After that, the features from accelerated segment test (FAST) corner detection algorithm on the expanded box is used to detect four bright corners. The pixel coordinates of four corners can be obtained. We use these pixel coordinates and the nearest neighbor coordinates to calculate the perspective transformation matrix. The vehicle LED position can be rectified through perspective transformation. The rectified vehicle LED is divided into 16 blocks equally that are fed to the convolutional neural network (CNN) recognition model in the form of batch. The diagram of LED state recognition is shown in Fig.6.

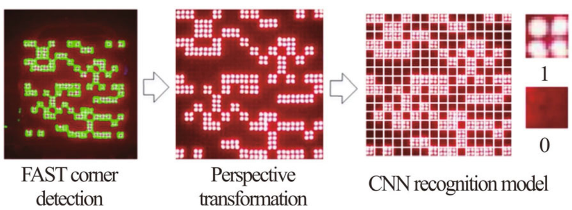

Fig.6 Diagram of LED state recognition

Our dataset is collected from 1 m to 8 m in real traffic scenes, which is composed of total 5 000 images. We use 3 500 images for the LED detection algorithm training, and 1 500 images for testing and the vehicle OCC system testing. The data augmentation methods are used during training stage, such as flipping, cutting and translation. Training parameters of the LED detection algorithm are shown in Tab.1.

Tab.1 Training parameters of the LED detection algorithm

Training parameters	Value
Input image size	320×320×3
Bach size	32
Epoch	50
Optimizer	Adam
Learning rate descent method	Cosine annealing scheduler
Initial learning rate	0.001
Decay rate	0.96

In the experiments, we use a 32×32 vehicle LED where four LEDs as a group to send the same signals. The experiment uses the Ubuntu 18.04 operating system, the Python 3.7 programming language, the VScode code editor, and the Torch 1.10 deep learning framework. The equipment in this paper is central processing unit (CPU) Xeon(R) E5-2609, graphics processing unit (GPU) Tesla P100. The experimental parameters are shown in Tab.2.

Tab.2 Experimental parameters

Experimental parameter	Value
LED array size	15 cm
LED illuminance	350 lux (at 10 cm)
ISO	50—1 600
Camera frame rate	30 fps
Camera resolution	1 280 pixel×720 pixel
Lens focal length	27 mm

Tab.3 shows the performance of different LED detection algorithms. The YOLOv5x has the highest detection accuracy. The YOLOv4-tiny has the smallest model size and the shortest inference time. The proposed LED detection algorithm can greatly balance detection accuracy and inference speed. Its average precision (AP) value is only 0.15% lower than the YOLOv5x and much higher than the YOLOv4-tiny. Our algorithm takes up less memory than the D2Net, and it is able to reach 83 fps on the GPU.

The proposed LED detection algorithm can fast and accurately detect the vehicle LED, which is suitable for deployment on the mobile devices. Fig.7(a) and Fig.7(b) are the vehicle LED detection results in the daytime and nighttime, respectively.

Tab.3 Performance of different LED detection algorithms

Algorithm	Model size (MB)	*AP* (%)	Inference speed (fps)
YOLOv4[17]	245	99.67	42
YOLOv4-tiny	**23**	86.27	**217**
YOLOv5s	27	89.38	112
YOLOv5m	84	96.09	79
YOLOv5l	192	99.40	55
YOLOv5x	367	**99.75**	35
D2Net	206	99.34	39
Ours	25	99.60	83

(a) In the daytime

(b) In the nighttime

Fig.7 Vehicle LED detection results

The system testing results are shown in Tab.4. Obviously, the optimized LED segmentation recognition algorithm is able to achieve a lower *BER* than the original algorithm. Our method has a high inference speed of 69 fps to ensure real-time communication.

We measure the relationship between the *BER* and communication distance in Fig.8. As we see, the designed method can achieve error-free transmission within 3 m, and meet the forward error correction (*FEC*) of 3.8×10^{-3} requirements within 4 m. When the distance exceeds 4 m, the system performance begins to decrease. The reason is that the longer the distance, the fewer pixels the vehicle LED occupies, which is difficult for the vehicle LED detection.

Tab.4 System testing results

Methods	*BER*	Inference speed (fps)
D2Net +original segmentation recognition	0.276	36
YOLOv4-tiny+original segmentation recognition	0.323	**148**
YOLOv5x+original segmentation recognition	0.216	33
Proposed LED detection + original segmentation recognition	0.218	69
Ours	**0.196**	69

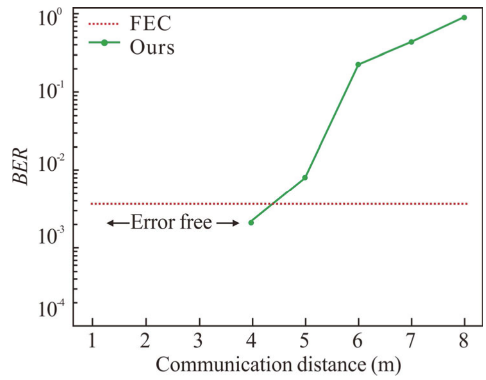

Fig.8 Relationship between the *BER* and communication distance

In this paper, we propose an LED detection algorithm based on deep leaning for the vehicle OCC system, which can improve LED detection accuracy and reduce inference time. Moreover, we optimize the LED segmentation recognition algorithm to enhance the reliability of the vehicle OCC system. The experiment is conducted in real traffic scenes, and the results show the vehicle LED detection accuracy of 99.60% and the inference speed of 83 fps. The vehicle OCC system can achieve error-free transmission within 3 m, and reach an inference speed of 69 fps.

Statements and Declarations

The authors declare that there are no conflicts of interest related to this article.

References

[1] NGUYEN B L, NGO D T, DAO M N, et al. Scheduling and power control for connectivity enhancement in multi-hop I2V/V2V networks[J]. IEEE transactions on intelligent transportation systems, 2021：1-11.

[2] TUAN N M, PHUONG T V, DO T H, et al. An highly realistic optical camera communication simulation framework for internet of things applications[C]//21st ACIS International Winter Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD-Winter), January

28-30, 2021, Ho Chi Minh City, Vietnam. New York：IEEE, 2021：240-242.

[3] MOHSAN S A H. Optical camera communications：practical constraints, applications, potential challenges, and future directions[J]. Journal of optical technology, 2021, 88(12)：729-741.

[4] HUANG Z, HE J, YU K, et al. Efficient demodulation scheme based on adaptive clock extraction and mapping-sampling for a mobile OCC system[J]. Applied optics, 2021, 60(12)：3308-3313.

[5] THOTA J, ABDULLAH N F, DOUFEXI A, et al. V2V for vehicular safety applications[J]. IEEE transactions on intelligent transportation systems, 2019, 21(6)：2571-2585.

[6] MORENO D, RUFO J, GUERRA V, et al. Optical multispectral camera communications using LED spectral emission variations[J]. IEEE photonics technology letters, 2021, 33(12)：591-594.

[7] ISLAM A, HOSSAN M T, JANG Y M. Convolutional neural network scheme-based optical camera communication system for intelligent Internet of vehicles[J]. International journal of distributed sensor networks, 2018, 14(4)：155014771877015.

[8] SHI J, HE J, JIANG Z, et al. Enabling user mobility for optical camera communication using mobile phone[J]. Optics express, 2018, 26(17)：21762-21767.

[9] HSU K L, CHOW C W, LIU Y, et al. Rolling-shutter-effect camera-based visible light communication using RGB channel separation and an artificial neural network[J]. Optics express, 2020, 28(26)：39956-39962.

[10] LIU Z, GUAN W, WEN S. Improved target signal source tracking and extraction method based on outdoor visible light communication using an improved particle filter algorithm based on Cam-Shift algorithm[J]. IEEE photonics journal, 2019, 11(6)：1-20.

[11] PHAM T L, SHAHJALAL M D, BUI V, et al. Deep learning for optical vehicular communication[J]. IEEE access, 2020, 8：102691-102708.

[12] SUN X, SHI W, CHENG Q, et al. An LED detection and recognition method based on deep learning in vehicle optical camera communication[J]. IEEE access, 2021, 9：80897-80905.

[13] HAN K, WANG Y, TIAN Q, et al. Ghostnet：more fea tures from cheap operations[C]//33st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. Piscataway：IEEE, 2020：1577-1586.

[14] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18-23, 2018, Salt Lake City, UT, USA. Piscataway：IEEE, 2018：8759-8768.

[15] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. Scaled-YOLOv4：scaling cross stage partial network[C]//34st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. Piscataway：IEEE, 2021：13024-13033.

[16] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss：faster and better learning for bounding box regression[C]//34th AAAI Conference on Artificial Intelligence, February 7-12, 2020, New York, USA. Menlo Park：AAAI Press, 2020：12993-13000.

[17] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4：optimal speed and accuracy of object detection[EB/OL]. (2020-04-23) [2022-04-25]. https：//arxiv.org/abs/2004.10934?sid=V22o3y.