IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received April 29, 2021, accepted May 24, 2021, date of publication May 31, 2021, date of current version June 10, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3085117

An LED Detection and Recognition Method Based on Deep Learning in Vehicle Optical Camera Communication

XU SUN, WENXIAO SHI, QING CHENG, WEI LIU, ZHUO WANG, AND JIADONG ZHANG
College of Communication Engineering, Jilin University, Changchun 130012, China
Corresponding author: Wei Liu (liuvivi512@163.com)

This work was supported by the Project of Science and Technology Development Plan of Jilin Province, China, under Grant 20200401122GX.

ABSTRACT In the Vehicle to Vehicle (V2V) communication based on Optical Camera Communication (OCC), optical signals are transmitted using LED arrays and received employing cameras. In a complex scene, how to accurately detect and recognize LEDs in real time remains a problem. To solve this problem, this paper designs an end-to-end network based on You Only Look Once version 5 (YOLOv5) object detection model, which can precisely detect the LED array position in real time and alleviate motion blur simultaneously. Further, we propose an LED segmentation recognition method, which is beneficial to more reliable LED status recognition. It allows more light sources to be used for communication, which can effectively improve data rate in the vehicle OCC system. The effectiveness of our method is demonstrated by theoretical analysis and experiments in real traffic scenes. Our code is available at *https://github.com/cq100/D2Net*.

INDEX TERMS Optical camera communication (OCC), vehicle to vehicle (V2V), LED recognition, you only look once version 5 (YOLOv5), image motion deblurring.

I. INTRODUCTION

In Intelligent Transport Systems (ITS), vehicle communication is introduced to reduce traffic accidents. Driving information is transmitted among vehicles [1]. For example, the front vehicle sends speed, lane changing information, emergency braking signals, and safety warnings, then the following vehicles receive these information in real time. In recent years, the existing Radio Frequency (RF) has been mature in wireless communication technology. But current RF technology still has many problems, such as interference, limited available bandwidth, etc [2]. However, Optical Camera Communication (OCC) can well make up for the existing gaps. The advantages of OCC are lower cost, broad spectrum, no harmful effects to human health, and the reduction of the non-Line-of-Sight (nLoS) system interference. Therefore, OCC is suitable for the V2V communication field [3]. In the OCC system, Light Emitting Diodes (LEDs) are used as transmitters [4], which are common illuminating devices [5].

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

Each LED can be individually modulated, which can significantly improve communication speed. Different from Free Space Optical (FSO) communication, Visible Light Communication (VLC), and Light Fidelity technology (LiFi), OCC utilizes image sensors as receivers [6]. Image sensors can not only receive LED light intensity information, but also utilize various color and space domain [7].

Despite the numerous benefits of the OCC system, there are challenges in the development of V2V communication. The first challenge is the noise from the sun, street light sources, and other background lights in the images [8]. These noises lead to more difficulties on the accurate extracting the Region of Interest (RoI) in real traffic scenes. Recently, the RoI signaling technique based on the OCC system has been introduced into the PHY IV mode that is part of IEEE 802.15.7-2018 (revision of IEEE 802.15.7-2011 standard [9]) for vehicle applications. In addition, image blur always degrades the LED recognition accuracy [10], which may be caused by vehicle motion. The second challenge is the low data rate due to the bandwidth limitation of the camera [11]. The third challenge is the long delay as most detection and recognition schemes require higher computations [12].

In order to obtain excellent performance of the OCC system, intensive researches have been carried out. In traditional methods, LED detection relies on pixel intensity [13]. However, when the vehicle moves quickly, it is difficult to identify the LED based on the pixel intensity due to the motion blur inferring the adjacent LED pixels. In [14], the multiple exposure overlay coding method was proposed to reduce blur effects. Although it helped to obtain a high signal quality, it used two cameras, which brought very high deployment costs. In [10], when the images are blurry due to a high vehicle speed, the scheme used pixel intensity, optical flow, and data information from previous frames to detect the LED position. But the algorithm cannot accurately recognize the LED status since it did not alleviate motion blur. Moreover, a video detection algorithm was applied to the vehicle OCC system innovatively [15], which achieved a processing time of 69 ms and lower extraction errors. The previous Cam-Shift algorithm with the Kalman filter scheme had a shorter processing time of 42 ms [16]. Selective Capture (SC) method was also introduced for vehicle taillights to reduce the computational burden [7]. It achieved a data rate of 6.912 kbps and a Bit Error Rate (BER) of 10^{-5} at transmission distance of 1.25 m.

In recent years, deep learning has also provided some methods to solve these problems faced by the OCC system. However, few studies considered the accuracy, robustness, data rate, and real-time ability of the OCC system at the same time. In [17], the RoI was detected by using You Only Look Once version 2 (YOLOv2) algorithm, which took the accuracy and real-time performance into account. It provided an acceptable detection accuracy and a high processing frame rate. But the transmission distance of this LED recognition experiment was only 0.5-1 m. The previously proposed Dimmable Spatial 8-Phase Shift Keying (DS8-PSK) decoder adopted neural networks to recover data on the simulated blur images [18]. In [19], the proposed scheme first introduced the Convolutional Neural Networks (CNN) in the VLC field, which presented more than 95% precision of Optical Fringe Codes (OFC) classification. But it required that the RGB-LED and the camera were parallel to each other. Furthermore, the CNN was also used to distinguish logic 0 and logic 1 [20]. It achieved a higher data rate of 1.2 kbit/s than the previously proposed algorithm [21].

In this work, we put forward an LED detection and recognition method based on deep learning, which can simultaneously consider the accuracy, robustness, data rate, and real-time performance of the vehicle OCC system in real traffic scenes. Our contributions in this paper are presented as follows:

- We design a network based on YOLOv5 model, named D2Net, which is an end-to-end network that can accurately detect the LED array and achieve image motion deblurring at the same time. The network not only lays a foundation for the LED recognition, but also enhances the robustness of vehicle OCC system.
- We propose an LED segmentation recognition method that combines machine learning and traditional image processing. The method can achieve multiple light sources communication, which can improve data rate.
- We compare other state-of-the-art methods to probe the effectiveness of our scheme. Related experiments and network training parameters will also be provided. Furthermore, our code is available at *https://github.com/cq100/D2Net*.

The rest of this paper is arranged as follows: In Section II, we present the proposed vehicle OCC system architecture. In Section III, we illustrate the proposed D2Net. In Section IV, we illustrate the proposed LED segmentation recognition method. In Section V, we conduct experiments and analyze the effectiveness of the proposed method. Finally, in Section VI, we summarize the work of this paper.

II. PROPOSED VEHICLE OCC SYSTEM ARCHITECTURE

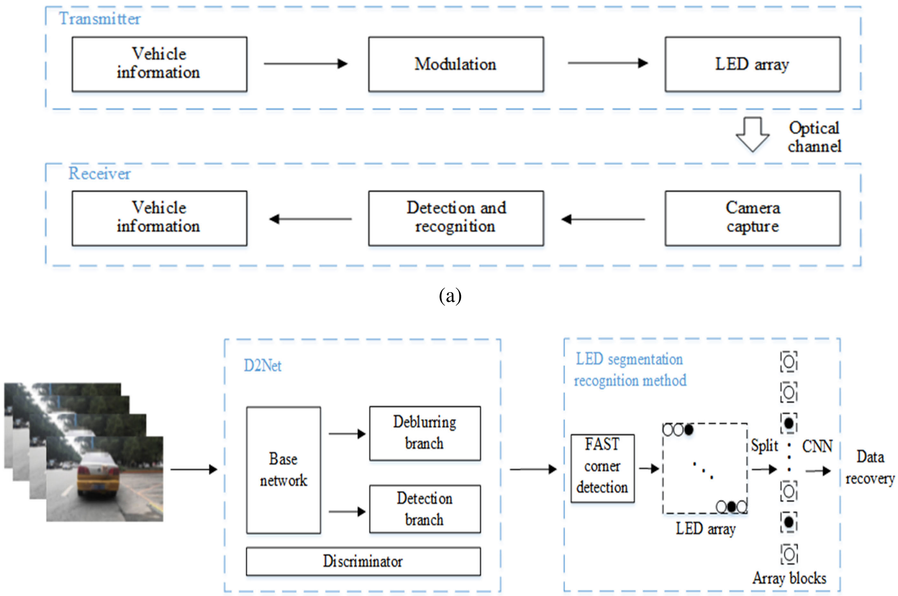
Fig. 1(a) shows the vehicle OCC system reference architecture. In the transmitter, the driving information of the front vehicle is modulated. Various data are transmitted through different LED statuses. The LED array is powered by the LED driver and controlled by the Micro Control Unit (MCU). In the receiver, the camera installed in the following vehicle can continuously capture images. By using detection and recognition methods, the LED position and status are determined in the images. Finally, we are able to obtain the information sent by the front vehicle.

In real traffic scenes, the LED array may be affected by motion blur. Thus, the key is image detection and deblurring for LED status recognition precisely. Fig. 1(b) shows the proposed LED detection and recognition architecture. We initially detect the LED array position and alleviate motion blur at the same time by using the proposed D2Net. Then, we use the LED segmentation recognition method that employs Features from Accelerated Segment Test (FAST) corner detection algorithm [22] and the LED segmentation method to identify the LED status through a CNN model. Finally, the data are recovered.

On-Off Keying (OOK) is a common modulation scheme, which is used for transmitter modulation in our experiment. There are only two statuses of the LED: 1 represents On; 0 represents Off.

III. PROPOSED D2NET

How to detect LED array accurately and in real time from a complex environment is still a challenge. YOLOv5 is a suitable and efficient model for object detection because it can be operated at a fast inference speed and guarantee the detection accuracy simultaneously. There are several models that can be selected to satisfy different requirements, such as YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x, etc. So, it is a good choice to use the YOLOv5 model to detect LED arrays in real traffic scenes. However, the captured images may be blurred when the vehicle moves, which greatly affects the communication quality. Most earlier deblurring methods



(a)

(b)

FIGURE 1. The vehicle OCC system. (a) Reference architecture (b) The proposed LED detection and recognition architecture.

cannot well capture the complex blur variations, because motion blur typically has unknown and spatially varying blur kernels in real scenes. The DeblurGAN-v2 is a single image motion deblurring network, which utilizes Generative Adversarial Network (GAN) to boost deblurring effectiveness, quality, and flexibility [23]. It can be applied to image deblurring in real traffic scenes. However, when YOLOv5 model and the DeblurGAN-v2 are cascaded, it leads to high computations and long delay. Thus, inspired by the DeblurGAN-v2, we propose a new end-to-end network D2Net based on YOLOv5 model, which can achieve LED array detection and image motion deblurring simultaneously.

A. THE D2NET ARCHITECTURE

The D2Net architecture based on YOLOv5l is shown in Fig. 2 (s represents the convolution stride, and s is 1 by default). It consists of the following four parts: base network, detection branch, deblurring branch, and discriminator. The base network mainly employs the focus module, the Spatial Pyramid Pooling (SPP) module [24], and the Cross Stage Partial connections (CSP) module [25]. The focus module can reduce spatial information loss when using downsampling whereas the SPP module is to increase the receptive field. The CSP module solves duplicate gradient problems in large-scale base networks, so the parameters and Floating-point Operations per second (FLOPs) can be reduced. The detection branch uses the Path Aggregation Network (PANet) [26] to output three feature maps of different shapes to predict the LED array location in the captured images. These feature maps enable our network to detect LED arrays of different sizes at different transmission distances. The deblurring branch uses the Convolutional Block Attention Module (CBAM) attention mechanism [27] and GAN technology. The attention mechanism can focus on important features and suppress unnecessary ones for image deblurring. The discriminator is divided into a global discriminator and a local discriminator. It provides a reasonable gradient descent direction for the deblurring branch, which is conducive to removing possible blur in the images.

B. LOSS FUNCTION

Our loss function is mainly composed of the following two parts: detection loss and deblurring loss.

In the LED array detection task, we deploy the same class loss L_{class} and confidence loss L_{conf} as YOLOv5. The difference is localization loss L_{local} that we use Distance Intersection over Union (DIoU) loss L_{DIoU} [28], which converges much faster in the training stage by minimizing the normalized distance between the center points of the predicted box B and the ground truth box B_{gt}. The calculation formulas are presented as follows:

$$L_Y = L_{class} + L_{conf} + L_{local} \tag{1}$$

$$L_{DIoU} = 1 - IoU + \frac{\rho^2\left(b, b_{gt}\right)}{c^2} \tag{2}$$

where L_Y is the overall loss of the detection task, Intersection over Union (IoU) is the ratio of the intersection and union of the two boxes, b and b_{gt} are the center points of B and B_{gt}, $\rho(\cdot)$ is the Euclidean distance, and c is the diagonal length of the smallest enclosing box containing the two boxes.

In the image deblurring task, we employ the loss function that is similar to the DeblurGAN-v2, which contains generator (deblurring branch) loss L_G and discriminator

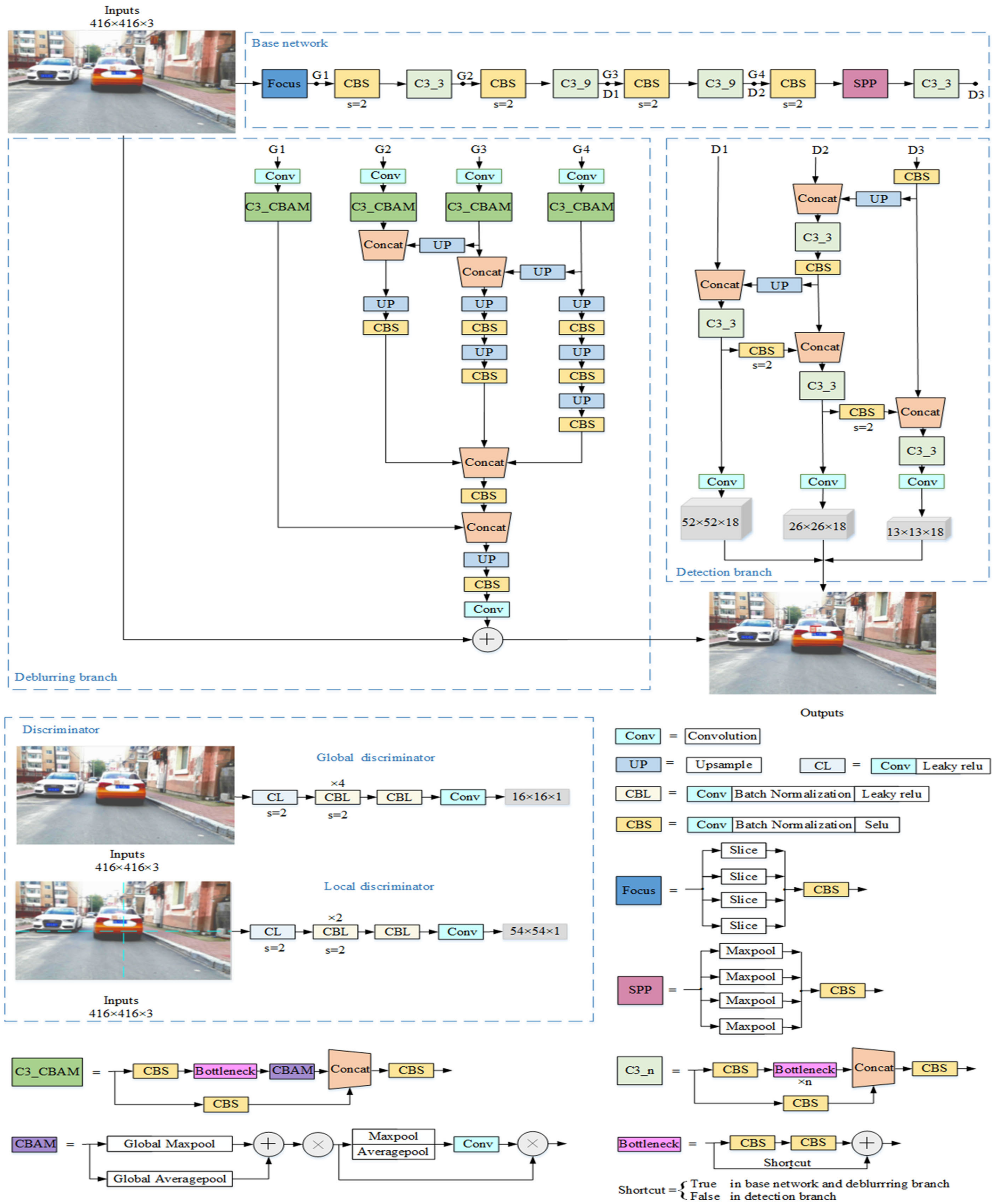


FIGURE 2. The D2Net architecture based on YOLOv5l.

loss L_D as follows:

$$L_G = 0.5 * L_p + 0.005 * L_x + 0.01 * L_{adv} \tag{3}$$

$$L_p = mean(MSE(x, z)) \tag{4}$$

$$L_x = mean(MSE(x', z')) \tag{5}$$

$$L_{adv} = E\left[(D(x) - E(D(G(z))) + 1)^2\right] + \frac{1}{2} * E[(D(G(z)) - E(D(x)) - 1)^2] \tag{6}$$

$$L_D = E\left[(D(x) - E(D(G(z))) - 1)^2\right] + \frac{1}{2} * E[(D(G(z)) - E(D(x)) + 1)^2] \tag{7}$$

where MSE is mean-square-error loss, L_p is pixel-space loss, x and z are the clean image and the D2Net output image, L_x represents perceptual loss, x' and z' denote the feature vectors of x and z through a pre-training VGG16 network (the first 15 layers), L_{adv} denotes adversarial loss, and G is a generator. When the global adversarial loss is calculated, D is the global discriminator; When the local adversarial loss is calculated, D is the local discriminator.

In general, multitask end-to-end network is difficult to train because the loss functions of different tasks have different convergence speeds, which causes an unbalanced training process. Thus, we employ the gradient normalization (GradNorm) algorithm [29], which can automatically balance multitask training by adjusting the weights of different tasks. The total loss function $L(t)$ of the D2Net is calculated as follows:

$$L(t) = \omega_1(t) * L_Y + \omega_2(t) * L_G \tag{8}$$

where $\omega_1(t)$ and $\omega_2(t)$ are the weights, which are updated by calculating the gradient loss $L_{grad}(t; \omega_i(t))$ as follows:

$$L_{grad}\,(t; \omega_i(t)) = \sum_i \left| G_w^i(t) - E_{task}\left[G_w^i(t)\right] * [\gamma_i(t)]^\alpha \right|_1 \tag{9}$$

where $G_w^i(t)$ is the gradient norm, w is the subset of the full network weights, $E_{task}[G_w^i(t)]$ is the average gradient norm, $\gamma_i(t)$ is the relative inverse training rate for each task, which is used to balance gradients, and α is a hyperparameter, which can tune training rate.

IV. PROPOSED LED SEGMENTATION RECOGNITION METHOD

In the vehicle OCC system, data rate is one of the key indicators. The data rate is obtained by multiplying three factors [11]: frame rate, the number of symbols per frame, and the number of bits per symbol. Increasing any one of the three factors will improve data rate. Some ways use a high frame rate camera that is extremely expensive. Other schemes used more complex modulation methods [30]. In fact, the use of multiple light sources is the most economical way. However, using a large number of LEDs for communication will greatly increase the system complexity, and it is difficult to ensure LED recognition accuracy. For example, a 3×3 LED array was used in [31], which had an acceptable computation. If a 16×16 LED array is used, it will result heavy computations. The reason is that there are still 2^{256} combinations even with simple OOK modulation method. Whether using traditional image processing method or deep learning, it is difficult to obtain reliable data in real time.

Therefore, we propose an LED segmentation recognition method to reliably identify the LED status through a simple CNN model. The overall structure of LED segmentation recognition method is shown as Fig. 3. By using FAST corner detection algorithm, we obtain the precise LED array. It is split into 256 array blocks to reduce the computational burden. Then, we only need to utilize a simple CNN model to identify two statuses of an array block, which utilizes the generalization ability of neural networks to reduce the impact of noise. Finally, we employ batch to recover all data of a LED array at the same time.

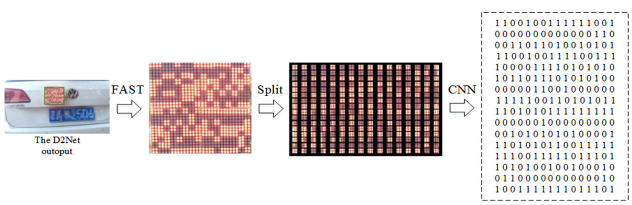


FIGURE 3. **The overall structure of LED segmentation recognition method.**

A. FAST CORNER DETECTION

In order to further reliably identify the LED status in real time, we adopt FAST corner detection algorithm to obtain more precise LED array position. When the pixel value difference between a pixel and three quarters of the pixels in the circular neighborhood is greater than a predetermined threshold, this pixel is considered as a candidate feature point. The advantage of this algorithm is that it has a faster detection speed and guarantees LED array detection accuracy.

In our experiment, the four corner lights of the LED array keep On, which is conducive to corner detection. The process of accurately detecting the LED array using FAST corner detection algorithm is shown in Fig. 4. The predicted box of the D2Net is expanded by 1.2 times (the confidence score in the figure is 0.99.), which makes the LED array completely contained. Then, the FAST algorithm is performed on the expanded area to obtain the accurate LED array position. If the LED array or the camera is tilted, LED recognition accuracy will be seriously affected. So we use perspective transformation [32] to rectify the LED array, because it can project an arbitrary quadrilateral into a regular rectangle.

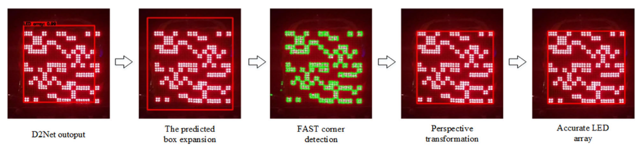


FIGURE 4. **The process of accurately detecting the LED array using FAST corner detection algorithm.**

B. LED SEGMENTATION METHOD

Fig. 5 shows the segmentation array blocks for an LED array. After comparing the LED array of an image with the known real data from the transmitter, we label the categories of each array block in order to train CNN model. It consists of several convolutional layers and polling layers. Fig. 6 presents the labeling results of array blocks. The idea of array block segmentation can simplify a complex task into several subtasks, which greatly improves LED recognition accuracy. The larger number of LEDs, the greater the advantage of this LED segmentation recognition method.

V. EXPERIMENT AND ANALYSIS

A. EXPERIMENT ENVIRONMENT

To verify the performance of our LED detection and recognition method, the experiment dataset is collected in real traffic scenes. We record several videos in the daytime and nighttime. Then, we generate 6000 blurry and clean image pairs by averaging consecutive short-exposure frames [33]. We obtain 30000 image pairs by using data augmentation. These image pairs are divided into a training set consisting of 20000 image pairs for the D2Net, a test set consisting of 5000 image pairs for the D2Net, and a test set consisting of 5000 image pairs for overall system performance evaluation. In the experiment, we employ the OOK modulation scheme and a 32×32 LED array where 4 LEDs send the same logic

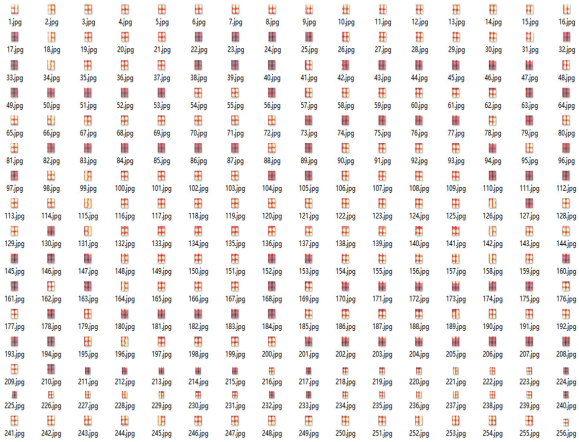


FIGURE 5. The segmentation array blocks for an LED array.

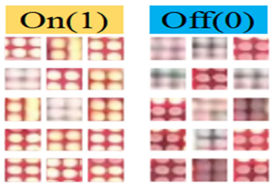


FIGURE 6. The labeling results of array blocks.

0/1. The training parameters related to the D2Net are shown in Table. 1. Our network uses tensorflow framework, and the training equipment is GPU Tesla P100. For reference, our experimental parameters are shown in Table. 2.

B. EXPERIMENT ANALYSIS

To evaluate the D2Net performance quantitatively, we use *Average IoU*, Precision (*P*), Recall rate (*R*), Average Precision (*AP*), Peak Signal-to-Noise Ratio (*PSNR*), Structural Similarity (*SSIM*), and Frames Per Second (FPS). The calculation formulas are given as follows:

$$IoU = \frac{B \cap B_{gt}}{B \cup B_{gt}} \tag{10}$$

$$Average\,IoU = \frac{\sum_{i=1}^{N} IoU_i}{N} \tag{11}$$

where N represents the total number of the objects (LED arrays) in the images.

$$P = \frac{TP}{TP + FP} \tag{12}$$

$$R = \frac{TP}{TP + FN} \tag{13}$$

where *TP* is the number of the LED array correctly detected, *FP* is the number of the non-LED array treated as the LED array, and *FN* is the number of the LED array treated as the non-LED array. The *AP* considers precision and recall rate, which can be represented by the area under the P-R curve.

TABLE 1. The training parameters related to the D2Net.

Training parameters	Value
Input image size	416 × 416
Initial learning rate	1e-3
End learning rate	1e-6
Optimizer	Adam
Epoch	200
Bach size	8
IoU threshold	0.5

The calculation formula is given as follows:

$$AP = \int_0^1 P(R)dR \tag{14}$$

To evaluate the deblurring performance of the proposed network, we employ *PSNR* and *SSIM* as follows:

$$PSNR = 10 * \lg\left(\frac{(2^n - 1)^2}{MSE(x, z)}\right) \tag{15}$$

$$SSIM = \frac{(2 * \mu_x * \mu_z + c_1) * (2 * \sigma_{xz} + c_2)}{\left(\mu_x^2 + \mu_z^2 + c_1\right)\left(\sigma_x^2 + \sigma_z^2 + c_2\right)} \tag{16}$$

where n is the number of binary bits representing each pixel, μ_x, μ_z are the average pixel values of the two images, σ_{xz} is the pixel covariance of the two images, σ_x^2, σ_z^2 are the pixel variances of the two images, and c_1, c_2 are constants.

Table. 3 shows the performance of the D2Net and some state-of-the-art methods on the model test set. Obviously, the inference speed of YOLOv3-tiny is the fastest, but the *Average IoU* and *AP* values are lowest. YOLOv5x has the best detection accuracy, but the inference time is longer than other detection models. The *Average IoU* and *AP* values of YOLOv5l are similar to YOLOv4 [34], but the inference speed is faster than YOLOv4. This result demonstrates that YOLOv5l can balance speed and accuracy well. In addition, the evaluation metrics of YOLOv5l outperform Single Shot MultiBox Detector (SSD) [35] and YOLOv3 [36]. So we choose YOLOv5l as the basis of the D2Net for detecting LED array, which is reasonable. Compared with the above models, the accuracy and inference speed of the D2Net are acceptable, and it can improve the image quality. Although the DeblurGAN-v2 has better *PSNR* and *SSIM*, it has the longer inference time than the D2Net.

We obtain the following experimental results by using the D2Net: Fig. 7(a) and Fig. 7(b) are the test results in the daytime; Fig. 7(c) and Fig. 7(d) are the test results in the nighttime. Obviously, the D2Net can effectively detect LED array and reduce blur effects. Meanwhile, the original images are still clean through the D2Net.

To evaluate the system performance, we design comparative experiments. Table. 4 shows the average BER performance relative to different LED detection and recognition methods on the system test set. Obviously, our LED segmentation recognition method has better robustness than the average and the center gray thresholds. When the images are clean, the BER of our scheme is slightly higher than the combination of YOLOv5x and LED segmentation recognition

TABLE 2. Experimental parameters.

Experimental parameters		value
LED array size		15 cm
LED illuminance		350 lux (at 10 cm)
In the daytime	Ambient light	2940 lux
	Exposure time	1/1000 s
	ISO	50
In the nighttime	Ambient light	14 lux
	Exposure time	1/320 s
	ISO	1600
Camera frame rate		240 fps
Camera resolution		1280 × 720
Lens focal length		27 mm

TABLE 3. The performance of the D2Net and some state-of-the-art methods on the model test set.

Model	P	R	AP (%)	Average IOU	PSNR	SSIM	Model size (MB)	FPS
SSD300 (VGG16)	0.801	0.859	82.33	0.690	–	–	92	58
YOLOv3	0.865	0.917	89.78	0.756	–	–	235	50
YOLOv3-tiny	0.639	0.765	71.28	0.598	–	–	33	235
YOLOv4	0.997	1.000	99.91	0.854	–	–	245	42
YOLOv4-tiny [37]	0.708	0.836	85.34	0.609	–	–	23	217
YOLOv5s	0.868	0.922	89.81	0.763	–	–	27	112
YOLOv5m	0.953	0.965	96.52	0.817	–	–	84	79
YOLOv5l	0.997	0.998	99.83	0.842	–	–	192	55
YOLOv5x	0.998	1.000	99.99	0.912	–	–	367	35
DeblurGAN-v2 (Inception)	–	–	–	–	32.57	0.819	233	21
D2Net	0.996	0.998	99.81	0.833	31.70	0.810	206	39

FIGURE 7. Test results using the D2Net. (a) (b) in the daytime. (c) (d) in the nighttime.

method. When the images are blurred, our BER performance is greater than other approaches. At the same time, it also has a high processing frame rate of 36 FPS to ensure real-time communication.

The BER performance of our scheme at different communication distances is presented in Fig. 8. When the distance varies from 1 to 3 m, the BER is 0, indicating that our scheme performs well on these clean images. The BER increases slightly when the distance is 3 to 4 m, which always meets the Forward Error Correction (FEC) requirements (BER < 3.8×10^{-3}). The transmission distance is longer than 0.5-1 m in [17]. We can achieve error free transmission within 2 m on the blur images. Note that the BER increases severely when the distance exceeds 2 m. The reason is that our LED array

TABLE 4. The average BER performance relative to different LED detection and recognition methods on the system test set.

Schemes	BER		FPS
	clean	blur	
YOLOv3-tiny + Average threshold	0.649	0.928	202
YOLOv3-tiny + Center threshold	0.635	0.927	202
YOLOv3-tiny + Segmentation recognition	0.466	0.919	143
YOLOv5x + Average threshold	0.536	0.924	34
YOLOv5x + Center threshold	0.526	0.923	34
YOLOv5x + Segmentation recognition	0.268	0.912	32
D2Net + Average threshold	0.538	0.629	38
D2Net + Center threshold	0.527	0.614	38
Our scheme (D2Net + Segmentation recognition)	0.272	0.439	36

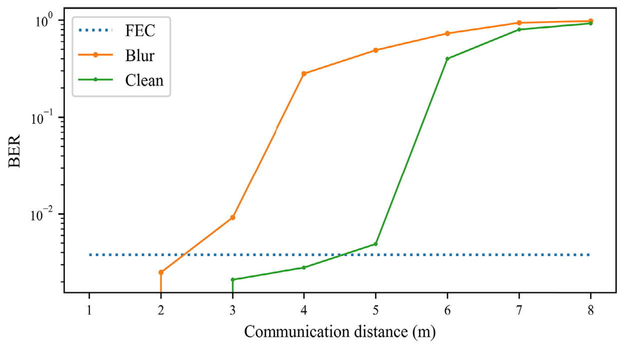


FIGURE 8. The BER performance of our scheme at different communication distances.

area becomes smaller when the distance gets longer, which makes the deblurring effect less obvious. Besides, the proposed scheme can be operated at data rate of 9.216 kbit/s, which is higher than 6.912 kbps in [7].

VI. CONCLUSION

This paper proposes an LED detection and recognition method based on deep learning to determine the position and status of LED in the vehicle OCC system. We design a new network D2Net based on YOLOv5 object detection model, which is an end-to-end network that can precisely detect LED arrays and reduce motion blur at the same time. Moreover, we propose an LED segmentation recognition method, which can be beneficial when multiple light sources are used. The proposed scheme always meets the FEC requirements at the transmission distance of 2 m on the blur images, and achieves a processing frame rate of 36 FPS. Our scheme provides an approach for deep learning applications in the future vehicle OCC system.

REFERENCES

[1] R. M. Mare, C. L. Marte, and C. E. Cugnasca, "Visible light communication applied to intelligent transport systems: An overview," *IEEE Latin Amer. Trans.*, vol. 14, no. 7, pp. 3199–3207, Jul. 2016.

[2] A.-M. Căilean and M. Dimian, "Current challenges for visible light communications usage in vehicle applications: A survey," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2681–2703, May 2017.

[3] N. Saeed, S. Guo, K.-H. Park, T. Y. Al-Naffouri, and M.-S. Alouini, "Optical camera communications: Survey, use cases, challenges, and future trends," *Phys. Commun.*, vol. 37, Dec. 2019, Art. no. 100900.

[4] T.-H. Do and M. Yoo, "Visible light communication-based vehicle-to-vehicle tracking using CMOS camera," *IEEE Access*, vol. 7, pp. 7218–7227, 2019.

[5] F. A. Dahri, F. A. Umrani, A. Baqai, and H. B. Mangrio, "Design and implementation of LED–LED indoor visible light communication system," *Phys. Commun.*, vol. 38, Feb. 2020, Art. no. 100981.

[6] T. Nguyen, A. Islam, T. Yamazato, and Y. M. Jang, "Technical issues on IEEE 802.15.7m image sensor communication standardization," *IEEE Commun. Mag.*, vol. 56, no. 2, pp. 213–218, Feb. 2018.

[7] S. Teli and Y.-H. Chung, "Selective capture based high-speed optical vehicular signaling system," *Signal Process., Image Commun.*, vol. 68, pp. 241–248, Oct. 2018.

[8] T. Yamazato, M. Kinoshita, S. Arai, E. Souke, T. Yendo, T. Fujii, K. Kamakura, and H. Okada, "Vehicle motion and pixel illumination modeling for image sensor based visible light communication," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 9, pp. 1793–1805, Sep. 2015.

[9] *IEEE Standard for Local and Metropolitan Area Networks—Part 15.7: Short-Range Wireless Optical Communication Using Visible Light*, Standard 802.15.7-2011, IEEE Standards Association, Piscataway, NZ, USA, 2011, pp. 1–309.

[10] P. Huynh, T.-H. Do, and M. Yoo, "A probability-based algorithm using image sensors to track the LED in a vehicle visible light communication system," *Sensors*, vol. 17, no. 2, p. 347, Feb. 2017.

[11] L. Liu, R. Deng, and L.-K. Chen, "47-kbit/s RGB-LED-based optical camera communication based on 2D-CNN and XOR-based data loss compensation," *Opt. Exp.*, vol. 27, no. 23, pp. 33840–33846, 2019.

[12] M. D. Thieu, T. L. Pham, T. Nguyen, and Y. M. Jang, "Optical-RoI-signaling for vehicular communications," *IEEE Access*, vol. 7, pp. 69873–69891, 2019.

[13] H. C. N. Premachandra, T. Yendo, M. P. Tehrani, T. Yamazato, H. Okada, T. Fujii, and M. Tanimoto, "High-speed-camera image processing based LED traffic light detection for road-to-vehicle visible light communication," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2010, pp. 793–798.

[14] T.-H. Do and M. Yoo, "Multiple exposure coding for short and long dual transmission in vehicle optical camera communication," *IEEE Access*, vol. 7, pp. 35148–35161, 2019.

[15] Z. Liu, W. Guan, and S. Wen, "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," *IEEE Photon. J.*, vol. 11, no. 6, pp. 1–20, Dec. 2019.

[16] M. Huang, W. Guan, Z. Fan, Z. Chen, J. Li, and B. Chen, "Improved target signal source tracking and extraction method based on outdoor visible light communication using a cam-shift algorithm and Kalman filter," *Sensors*, vol. 18, no. 12, p. 4173, Nov. 2018.

[17] T. L. Pham, M. Shahjalal, V. Bui, and Y. M. Jang, "Deep learning for optical vehicular communication," *IEEE Access*, vol. 8, pp. 102691–102708, 2020.

[18] T. L. Pham, H. Nguyen, T. Nguyen, and Y. M. Jang, "A novel neural network-based method for decoding and detecting of the DS8-PSK scheme in an OCC system," *Appl. Sci.*, vol. 9, no. 11, p. 2242, May 2019.

[19] W. Guan, J. Li, S. Wen, X. Zhang, Y. Ye, J. Zheng, and J. Jiang, "The detection and recognition of RGB-LED-ID based on visible light communication using convolutional neural network," *Appl. Sci.*, vol. 9, no. 7, p. 1400, Apr. 2019.

[20] K.-L. Hsu, C.-W. Chow, Y. Liu, Y.-C. Wu, C.-Y. Hong, X.-L. Liao, K.-H. Lin, and Y.-Y. Chen, "Rolling-shutter-effect camera-based visible light communication using RGB channel separation and an artificial neural network," *Opt. Exp.*, vol. 28, no. 26, pp. 39956–39962, 2020.

[21] Y.-C. Chuang, C.-W. Chow, Y. Liu, C.-H. Yeh, X.-L. Liao, K.-H. Lin, and Y.-Y. Chen, "Using logistic regression classification for mitigating high noise-ratio advisement light-panel in rolling-shutter based visible light communications," *Opt. Exp.*, vol. 27, no. 21, pp. 29924–29929, 2019.

[22] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2006, pp. 430–443.

[23] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, "DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8878–8887.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[25] C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 390–391.

[26] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8759–8768.

[27] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.

[28] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-iou loss: Faster and better learning for bounding box regression," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, 2020, pp. 12993–13000.

[29] Z. Chen, V. Badrinarayanan, C.-Y. Lee, and A. Rabinovich, "GradNorm: Gradient normalization for adaptive loss balancing in deep multitask networks," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 794–803.

[30] C.-W. Chow, R.-J. Shiu, Y.-C. Liu, Y. Liu, and C.-H. Yeh, "Non-flickering 100 m RGB visible light communication transmission based on a CMOS image sensor," *Opt. Exp.*, vol. 26, no. 6, pp. 7079–7084, 2018.

[31] A. Islam, M. T. Hossan, and Y. M. Jang, "Convolutional neural networkscheme–based optical camera communication system for intelligent Internet of vehicles," *Int. J. Distrib. Sensor Netw.*, vol. 14, no. 4, pp. 1–19, 2018.

[32] J. Mezirow, "Perspective transformation," *Adult Educ.*, vol. 28, no. 2, pp. 100–110, 1978.

[33] J. Telleen, A. Sullivan, J. Yee, O. Wang, P. Gunawardane, I. Collins, and J. Davis, "Synthetic shutter speed imaging," *Comput. Graph. Forum*, vol. 26, no. 3, pp. 591–598, Sep. 2007.

[34] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*. [Online]. Available: http://arxiv.org/abs/2004.10934

[35] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 21–37.

[36] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*. [Online]. Available: http://arxiv.org/abs/1804.02767

[37] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Scaled-YOLOv4: Scaling cross stage partial network," 2020, *arXiv:2011.08036*. [Online]. Available: http://arxiv.org/abs/2011.08036

XU SUN received the B.Eng. degree from the School of Electronic Information Engineering, Changchun University of Science and Technology, Changchun, China, in 2019. She is currently pursuing the M.S. degree with the Department of Communications Engineering, Jilin University.

Her research interests include optical camera communications (OCC) and deep learning.

WENXIAO SHI received the B.Eng. degree in communications engineering from the Changchun Institute of Posts and Telecommunications, Changchun, China, in 1983, the M.S. degree in electrical engineering from the Harbin Institute of Technology, Harbin, China, in 1991, and the Ph.D. degree in communication and information systems from Jilin University, Changchun, in 2006.

He is currently a Professor with the Department of Communications Engineering, Jilin University. His research interests include wireless optical communications, wireless resource management, heterogeneous wireless networks, and mobile edge computing.

QING CHENG received the B.Eng. degree from the School of Electronic Information Engineering, Changchun University of Science and Technology, Changchun, China, in 2019. He is currently pursuing the M.S. degree with the Department of Communications Engineering, Jilin University.

His research interests include deep learning and embedded systems.

WEI LIU received the B.Eng., M.S., and Ph.D. degrees from the Department of Communications Engineering, Jilin University, Changchun, China, in 2009, 2012, and 2015, respectively.

She is currently an Assistant Professor with the Department of Communications Engineering, Jilin University. Her research interests include adaptive optics of wireless optical communications and atmospheric channel compensation technology of free space optical communications.

ZHUO WANG received the B.Eng. degree from the Department of Communications Engineering, Jilin University, Changchun, China, in 2016, where she is currently pursuing the Ph.D. degree.

Her research interests include the performance analysis of mixed radio frequency/free space optical relay systems.

JIADONG ZHANG received the B.Eng. degree from the Department of Communications Engineering, Jilin University, Changchun, China, in 2018, where he is currently pursuing the Ph.D. degree.

His research interests include deep reinforcement learning, computation offloading, and resource allocation.