A Passive Target Recognition Method Based on LED Lighting for Industrial Internet of Things

Xiao-Xiao Du, Yu Mu, Zi-Wei Ye, and Yi-Jun Zhu

***Abstract*—We consider an industrial internet of things environment, where involves multiple production factors, such as automatic guided vehicles (AGVs), people, container, etc. A deep learning model is presented for multi-target recognition, where the training data is shadow image formed by the nonuniform illumination of LED lighting source. Three shadow models of typical shapes are constructed to describe the shadows at different positions. The performance of the optimal VGG-16-based Faster-RCNN model is analyzed in view of the recognition accuracy and speed, and it is proved that recognizing three, four, and five types of objects, the mean average precision is 93%, 94.8%, and 92.6%, respectively. To enhance the generalization performance, the optimal Faster-RCNN is combined with the motion state of objects and the corresponding threshold. Simulation results show that the proposed deep learning model obtains significant performance gains to reduce missed and false detection.**

***Index Terms*—Visible light, deep learning, supervised learning, industrial internet of things, moving objects.**

I. Introduction

THE industrial Internet of Things (IIoT) is an extension application of the traditional internet of things in the industrial field, involving all links and main parts of industrial manufacturing and internet communication technology. To improve the flexibility of production factors and reduce spending on infrastructure, it is urgent to replace traditional or optical cable communication with wireless networks [1], [2]. Light-emitting diodes (LEDs) have become a compelling device in wireless technology because of constant cost reductions and their excellent performance. Compared with radio frequency (RF) technology, visible light has significant advantages, such as rich spectrum resources, inherent security capability, high immunity to electromagnetic interference, and license-free operation [3], [4]. As a ubiquitous lighting equipment, LED can be used as a new sensor in IIOT.

The IIoT requires real-time monitoring of devices, product tracking, and worker identification. The existing researches of visible light technology focus on transmitter and receiver. The laser serving as a transmitter focuses on three-dimensional imaging lidar is widely used in target reconstruction [5], [6]. However, the detection accuracy is closely related to the number of emitters due to its extremely narrow radiation range. In the receiver terminal, optical camera is susceptible to complex backgrounds, illumination variations and shadow of object [7], [8]. Besides, the frame rate of the camera is within 30 Hz. The captured image will be blurred or distorted, or even lose target when the object moves at high speed. Instead, the visible light exhibits obvious advantages in target recognition, such as wide radiation range, high communication rate and security, no additional equipment [9]. In recent years, researches on visible light mainly focus on active location, which requires photodiode (PD) to be attached to the target. Passive target recognition calls for free from any sensors (or tags) on the target [10]. However, there are few researches about passive detection of multiple targets. In [11], the proposed StarLight system requires a larger number of LED panels to track the human skeleton posture. Li-Tect [12] as another latest approach detects the object by distinguishing the blockage state of the line-of-sight (LOS) signal between each sensor pair. In the reflection-based technologies such as [13]–[15], the performance is affected by the signal-to-noise ratio (SNR) and the number of sensors. Nevertheless, these approaches only detect a single target at a time, and require plenty of LEDs and PDs, leading to more complex calculations. Therefore, it is critical to develop a passive multiple targets recognition technology based on LED lighting.

In this paper, we propose a passive target recognition method, which uses the shadow formed by the blocked lights between the LED and PD. The measured light strength value is regarded as a pixel in the light intensity distribution map, thus the target recognition can be considered from the perspective of image. As an efficient approach in image process, deep learning has exerted great impact on various research fields, and has been applied in visible light communications (VLC). In [16], a two-dimensional convolution neural network (CNN) structure is proposed for data decoding in the rolling shutter based on optical camera communication system. In [17], the author proposed a visible light position (VLP) system, using repeated unit cells and machine learning algorithms to achieve position error within 4-5 cm. In [18], the VLP system is further improved, the SFDP method in machine learning is proposed to achieve an average positioning error about 2 cm. Considering the shadow is related to shape and position, we use CNN to extract features of objects. Among the modern approaches, the Faster-RCNN approach exhibits

Manuscript received May 19, 2021; revised July 1, 2021; accepted July 15, 2021. Date of publication July 21, 2021; date of current version August 18, 2021. The work was supported in part by the National Key Research and Development Project under Grant 2018YFB1801903, in part by the National Natural Science Foundation of China (NSFC) under Grant 61901524. *(Corresponding author: Yi-Jun Zhu.)*

The authors are with the Information Engineering University, Zhengzhou 450000, China (e-mail: xxgcdxx@163.com; muyuvlc123@gmail.com; yzwleaf@outlook.com; yijunzhu@zzu.edu.cn).

Digital Object Identifier 10.1109/JPHOT.2021.3098672

competitive performance [19]. Several researches investigate that the Faster-RCNN approach is capable for various real-time object detection [20]–[22]. However, two challenges of passive target recognition method based on LED lighting are summarized as follows:

- **Non-uniform effect of LED lighting source based on Lambertian model.** Since the Lambertian radiation model of the light source obeys the two-dimensional Gaussian distribution in space, the light intensity received by the receiver is various with different positions.
- **Recognition of similar targets.** Objects of different shapes and sizes may have similar shadows. The shadows of the targets present a certain randomness at different positions.

The main contributions of this paper are outlined as follows:

- A passive target recognition method for the industrial internet of things is presented, where a deep learning model is used to identify the target. The transmitter takes an LED as the light source, and multiple photodiodes as the receivers to obtain the light intensity distribution maps.
- The target recognition method is considered in view of the shadow formed by the blocked lights. Based on the Lambertian radiation model, the shadow models of typical shapes are constructed, which describes the changes of shadows in different positions. Therefore, the light intensity distribution map with shadow can be used as the basis of target recognition.
- For the simulated light intensity distribution maps with shadows, the optimal Faster-RCNN model with high recognition accuracy and fast speed for multi-target is developed. To reduce the false and missed detection, target recognition and movement are combined to enhance the generalization performance.

This paper is organized as follows: Section II describes the VLC Lambertian radiation model and shadow models of different objects. Section III presents a Faster-RCNN framework for shadow detection. Section IV provides the simulation results of shadow, analysis the detection and generalization performance of proposed VGG-16-based Faster-RCNN. Finally, conclusions are drawn in Section V.

II. VLC Channel and Shadow Model

The scenario of IIoT is depicted in Fig. 1, which employs an LED as transmitter and many PDs (blue round holes) as receivers. There are multiple objects including AGVs, people, containers and other production factors. Different shapes of shadows are generated in different positions. These shadows can be observed on the received light intensity distribution maps.

A. Lambertian Radiation Model

In this section, we introduce the VLC channel model based on the Lambertian model. The LOS and non-line-of-sight (NLOS) are two propagation paths between LED and PD. The total received power is given as:

$$P_{rx} = P_{tx} H_{LOS}(0) + \int P_{tx} H_{ref}(0) \tag{1}$$

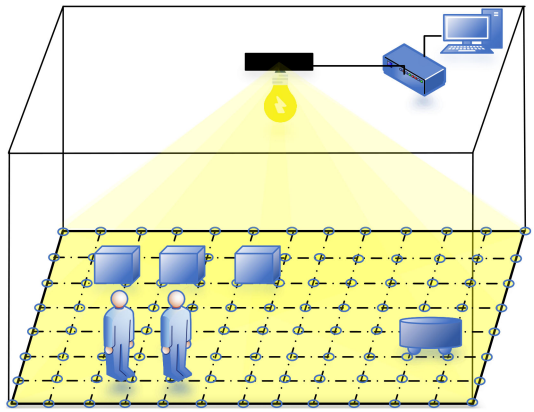

Fig. 1. The scenario diagram of system model.

where P_{rx} and P_{tx} is the total received power and transmit power by LED and PD, respectively. $H_{LOS}(0)$ and $H_{ref}(0)$ are the channel DC gain on the directed path and the first reflection, respectively. The first reflection influence is weak enough to be neglected in IIOT [23]. The channel DC gain of i-th LED and j-th PD on the directed path is given as [24]:

$$H_{LOS}^{ij}(0) = \begin{cases} \frac{\mu A_r (m_0+1)}{2\pi (D_{ij})^2} \cos^{m_0}(\phi) \cos(\psi)\, T_s(\psi) g_s(\psi), & 0 \le \psi \le \Psi_C, \\ 0, & \psi > \Psi_C. \end{cases} \tag{2}$$

where ϕ and ψ represent the angle of irradiance and incidence respectively. Ψ_C is the field of view of a receiver. The parameter $m_0 = \frac{-\ln(2)}{\ln(cos(\Phi_{1/2}))}$ denotes the order of Lambertian model, where $\Phi_{1/2}$ is the semi-angle at half power of transmitter. Additionally, μ is the responsivity of PD. D_{ij} is the distance between i-th LED and j-th PD. A_r denotes the effective receiving area of the PD, $T_s(\psi)$ is the gain of optical filter. $g_s(\psi)$ is the optical concentrator.

B. Shadow Model for Different Objects

We propose a deep learning framework to detect, evaluate and classify the multiple objects in IIoT. In simulation, the ambient light is ignored. When an object presents in the area, a shadow will be formed. We assume the received light intensity under the shadow equals zero. This paper intends to use the shadow area of the light intensity distribution maps to detect multiple objects. In this model, cubes, cuboids and cylinders represent AGVs, containers and people, respectively. In addition, spherical and imitating-person model are introduced, and the effect is explored. The framework takes the maps with shadow as the input and evaluates the category and location of the shadows as the output. Fig. 2 demonstrates the recognition method procedure.

This paper firstly considers the detection of a sphere. We assume there is only one LED located at (x_t, y_t, z_t) and a sphere located at (x_o, y_o, z_o), the PDs located at $(x_p, y_p, 0)$ on the floor.

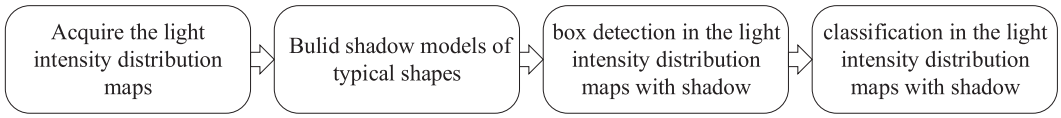

Fig. 2. The procedure of the passive target recognition method.

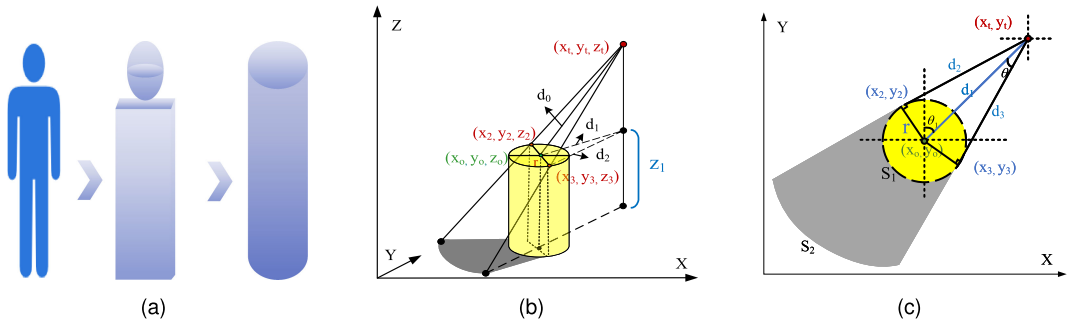

Fig. 3. (a) The basic 3-D human model. (b) 3-D schematic diagram of a cylinder. (c) 2-D plane schematic diagram of a cylinder.

The approximate shadow area (x, y, z) is given by:

$$\begin{cases} \frac{x-x_0}{x_p-x_0} = \frac{y-y_0}{y_p-y_0} = \frac{z-z_0}{z_p-z_0} & (3.1) \\ (x-x_o)^2 + (y-y_o)^2 + (z-z_o)^2 = r^2 & (3.2) \end{cases} \tag{3}$$

Next, we consider the detection of cubes as AGVs. The detection of cube shadow is a great challenge. To ensure accuracy and reduce computational complexity, we select the three sides nearest to light source to obtain the shadow region. Given a vertex $a_0 = (x_o, y_o, z_o)$, a = length, b = width and c = height of a cube, we can calculate the distance from the light source to the center of the six surfaces. The formula of six surfaces in a cube is as follows:

$$\begin{cases} y - y_0 = 0 & (4.1) \\ x - x_0 = 0 & (4.2) \\ z = 0 & (4.3) \\ y - (y_0 + b) = 0 & (4.4) \\ x - (x_0 + a) = 0 & (4.5) \\ z - c = 0 & (4.6) \end{cases} \tag{4}$$

Combining three of the above formulas with formula (3.1), if there is a real solution within the cube, it can be regarded as a shadow region.

Similar to the scenarios outlined above, we consider a 3-D human model as shown in Fig. 3(a). Two approximated models of the actual human body are proposed. One is a combination of a sphere and a cube, the other is a single cylinder. In the sphere-and-cube model, the shadow region can be acquired by formulas (3)-(4). In the cylinder model (the top center located at (x_o, y_o, z_o)), the gray area in Fig. 3(b) represents the shadow region. The shadow can also be analyzed on the XOY plane as shown in Fig. 3(c).

The gray area is represented by S_1 and S_2, which are the boundary of the cylinder and shadow respectively. The tangent point from the light source to the cylinder are (x_2, y_2, z_2) and (x_3, y_3, z_3), which are given as:

$$\begin{cases} x_2 = x_o - r\cos(\theta + \theta_1) \\ y_2 = y_o + r\sin(\theta + \theta_1) \end{cases} \tag{5}$$

$$\begin{cases} x_3 = x_1 + r\cos(\theta_1 - \theta) \\ y_3 = y_1 - r\sin(\theta_1 - \theta) \end{cases} \tag{6}$$

where $\theta_1 = \frac{\pi}{2} - \arctan(\frac{y_t - y_o}{x_t - x_o})$, $\theta = \arcsin(\frac{r}{d_1})$. The shadow area$(S)$ is given as:

$$ld_1 = \sqrt{(y_t - y_o)^2 + (x_t - x_o)^2}$$

$$d_2 = \sqrt{(y_t - y_2)^2 + (x_t - x_2)^2} \tag{7}$$

$$S = S_2 - S_1 = \left(\frac{\mathrm{d}_2 z_t}{z_t - z_o}\right)^2 \theta - rd_1\cos\theta - \frac{\pi + 2\theta}{2} r^2 \tag{8}$$

we can get the shadow of people at any position in the scenario, which can be used for subsequent simulation.

To recognize the shadow of multiple objects, we take the supervised learning approach, in which an optimal Faster-RCNN model is used to recognize and locate shadows on the light intensity maps. Details on the Faster-RCNN are provided in Section III.

III. A Faster-RCNN Framework for Shadow Detection and Localization

The Faster-RCNN network is proposed to handle the classification and localization tasks of the shadows. The characteristic of Faster-RCNN is that the region proposals is generated by the deep learning network. The region proposals task and the feature extraction task share the convolutional layers, which greatly accelerates the network training speed. The specific framework is shown in Fig. 4.

The Faster-RCNN network mainly contains four modules: the feature extractor, region proposal network (RPN), region

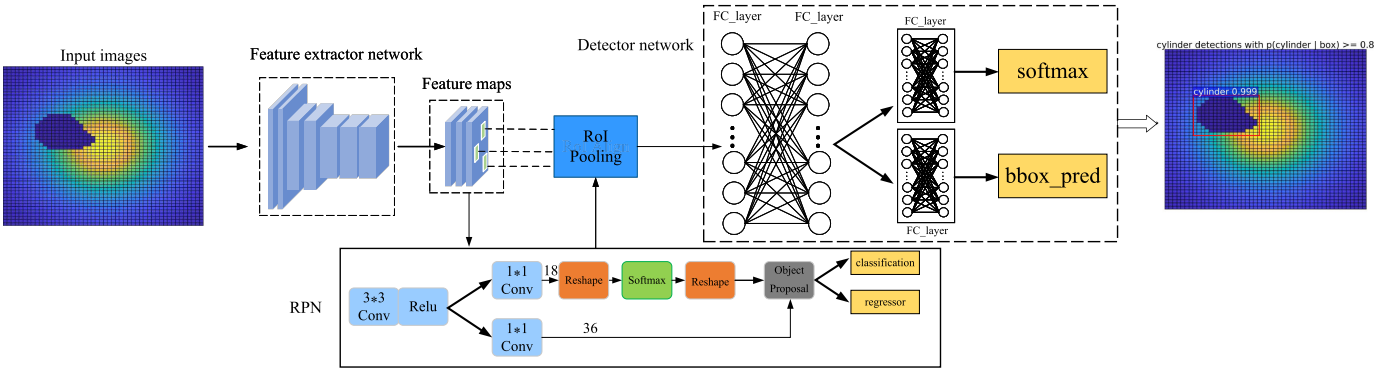

Fig. 4. The Faster-RCNN architecture for shadow detection.

of interest (RoI) pooling, and detector network. The feature extractor module extracts features according to the format of input dataset. Where the dataset is the light intensity maps with shadows. The RPN module contains a convolution layer with kernel size 3×3 and two convolution layers with kernel size 1×1. The convolution layer of 3×3 is used to fuse the information of feature maps, while the two 1×1 convolution layers are used to predict the classification score and anchor offset respectively. The anchors with higher foreground score are selected by the NMS operation and fed into the pooling module as a region proposal. The RoIs obtained from the RPN may have different sizes, which depend on the choice of anchors and the output of the regression. The RoI module performs a typical 7×7 max-pooling operation on each proposal to obtain the same size of feature maps. In the detector network, the fully connected layer and the softmax classifier are used to evaluate the category of the proposals, and the boundary regression algorithm is used to predict position. In the training phase, if the intersection over union (IoU) of any RoI with the ground-truth is greater than detector positive overlap, the RoI will be judged as a positive target. If the IoU is less than the detector positive overlap but greater than the detector negative overlap, the RoI will be a negative target.

In this paper, several experiments have been performed to compare the detection performance of different feature extractors. Typically, the feature extractor is a CNN, and its depth depends on the complexity of extracted features. There are many standard feature extractor networks in the field of computer vision, such as VGG-16, VGG-19, Resnet-50, Resnet-101 [25]. The networks of VGG-16 and VGG-19 employ the same size of convolution kernel and maximum pooling. The latter has more convolution layers to verify the impact of network structure on performance. Resnet framework designs a residual module to train deeper networks. The output of any layer in the network can be used as the input of a subsequent layer, which solves the accuracy reduction caused by the increase of layers in deep learning. Resnet-50 and Resnet-101 represent the number of hidden layers. These experiments are simulated by the indoor visible light radiation model and the existing sever. Considering the limited number of images, this paper adopts the pre-trained model as one of the common transfer learning in the Faster-RCNN network. The performances of different pre-trained models and CNN structures are analyzed. An optimal Faster-RCNN model in terms of detection accuracy and speed is constructed for multiple targets in IIoT.

TABLE I
SIMULATION PARAMETERS

Parameters	Value
Transmitted optical power of LED P_t	$25w$
LED position (x_t, y_t, z_t)	$(2.5m, 2.5m, 3m)$
Semi-angle at half power $\Phi_{1/2}$	$60°$
Physical area of PD A_r	$0.1cm^2$
FOV (Field of View) of PD Ψ_C	$90°$
Responsivity of PD μ	$0.4A/W$
Optical filter gain T_s	1.0
Optical concentrator gain g_s	2.25
human body measurement	
Width	$15cm$
Length	$40cm$
Height	$175cm$
Radius (r)	$12cm$
Sphere	$25cm$
Cube	$(50cm, 50cm, 50cm)$
Cuboid	$(100cm, 50cm, 50cm)$
Cylinder	
Radius (r)	$12cm$
Height	$175cm$

IV. SIMULATION AND ANALYSIS

A. Simulation of Shadow

In this section, we consider the indoor environment (5 m,5m,3 m) with an LED in the center of ceiling. The simulation results based on the lambertian radiation model can be summarized as follows: (1) the PDs will receive stronger light intensity when the PDs closer to the LED. While the PDs located at the boundary receive weak light intensity, it is related to the half-power angle of the LED. The half-power angle exceeds $60°$, the light intensity changes slightly. Thus, the $60°$ is used in this paper. (2) The resolution of the shadow is related to the layout of PD. When the layout is sparse, the shadow will be rough and difficult to be identified, so the 50×50 grid of PD is adopted in this paper. Other simulation parameters are shown in Table I.

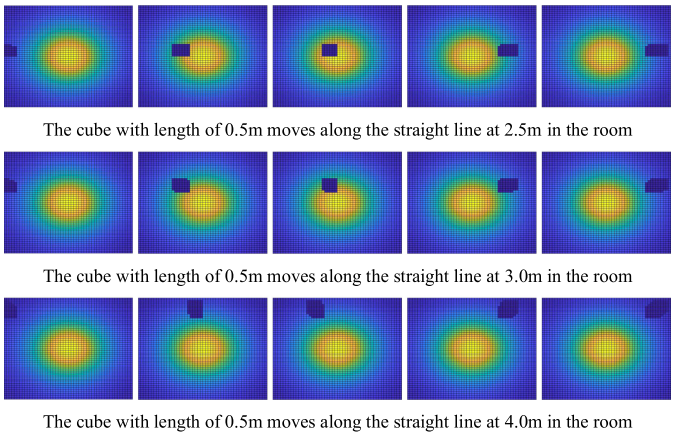

Fig. 5. The changes of shadow at different positions.

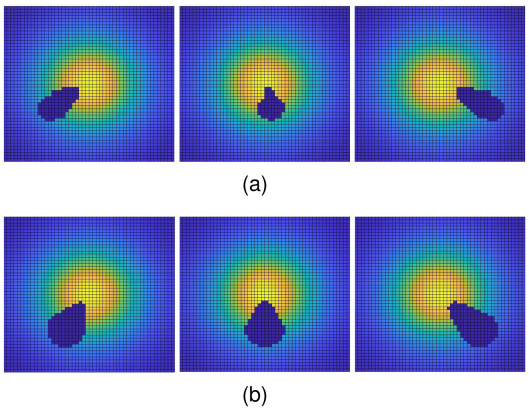

Fig. 6. (a) The shadows of the sphere-and-cube model. (b) The shadows of the cylinder model.

Fig. 5 shows the changes of shadow when the object straight move at several typical positions, such as 2.5 m, 3.0 m and 4.0 m. Apparently, the proposed shadow model is almost consistent with the realistic environment. The simulation results of the two approximate people models are shown in Fig. 6. The shadows obtained by the two models are extremely similar. Unlike the sphere and cube, the shadows of people have a obvious changes, increasing the recognition difficulty.

The above simulation results indicate that the shadows of sphere and cube are insensitive to the changes of position, but the shadows of people are greatly affected by the position. It is difficult to recognize multiple targets with a certain feature. Deep learning can extract a number of features from massive data, thus we adopt the Faster-RCNN to recognize multiple targets.

B. The Detection Effect of Faster-RCNN Model

In this section, we propose an optimal Faster-RCNN model by evaluating the classification and regression task with test data. The loss function is a weighted sum of the binary cross-entropy loss term for classification and a robust loss term for regression. It is used to estimate the difference between the predict value and the real value. The smaller of the loss function value, the better the robustness of the model. The total loss function of Faster-RCNN contains the loss of RPN and the loss of detector network in the training. The RPN is to pre-check which location of feature maps contains objects by a classifier, and adjusts the corresponding location by a regressor. The classifier and regressor are trained according to the formula (9) to generate region proposal for the detector network where detect object categories and regions.

$$\begin{aligned} L(p_i, t_i) &= \frac{1}{N_{cls}} \sum\nolimits_i L_{cls}(p_i, p_i^*) \\ &\quad + \lambda \frac{1}{N_{reg}} \sum\nolimits_i p_i^* L_{reg}(t_i, t_i^*) \\ L_{cls}(p_i, p_i^*) &= -log|p_i p_i^* + (1 - p_i)(1 - p_i^*) \\ L_{reg}(t_i, t_i^*) &= \sum\nolimits_{i \in \{x,y,w,h\}} smooth_{L1}(t_i - t_i^*) \\ Smooth_{L1}(x) &= \begin{cases} 0.5x^2 & if\ |x| < 1 \\ |x| - 0.5 & otherwise \end{cases} \end{aligned} \tag{9}$$

where, i is the index of an anchor in a mini-batch. p_i is the probability that an anchor is predicted as a object. If p_i is positive, the probability of ground-truth label $p_i^* = 1$. Conversely, $p_i^* = 0$ is negative. $t_i = (t_x, t_y, t_w, t_h)$ represents the center coordinates, width and height of the predicted bounding box (or the adjusted anchor), and t_i^* is the ground-truth box corresponding to positive anchor. L_{cls} is the log form of the classification loss, L_{reg} is the regression loss. $N_{cls} = 256$ and N_{reg} is the number of anchor, which are two normalized terms to balance both loss functions by combining the parameter λ ($\lambda = 10$).

The classification task is to recognize multiple objects. The regression task is to evaluate the difference between estimated locations and manually annotated locations. The average precision (AP) is widely employed to evaluate both tasks. The AP score is calculated as the mean precision over the 11 recall values (default values) given a predetermined IoU threshold. The MAP is the average precision of all classification. The formula (10) are as follows:

$$\begin{aligned} Precision &= \frac{\mathrm{TP}}{TP + FP} \\ Recall &= \frac{\mathrm{TP}}{TP + FN} \\ AP &= \frac{1}{11} \sum_{Recall_\alpha} Precision(Recall_i), \\ &\qquad \alpha = 1, 2, 3 \ldots, 11 \\ MAP &= \frac{\sum_{\alpha=1}^{\mathrm{k}} \mathrm{AP}(\alpha)}{k} \end{aligned} \tag{10}$$

where, TP is the number of true positives, i.e., the number of objects correctly detected, while FP is the number of false positives. FN is the number of a false negative, which presents the undetected number. k is the number of classifications, and $AP(\alpha)$ is the average precision of category α.

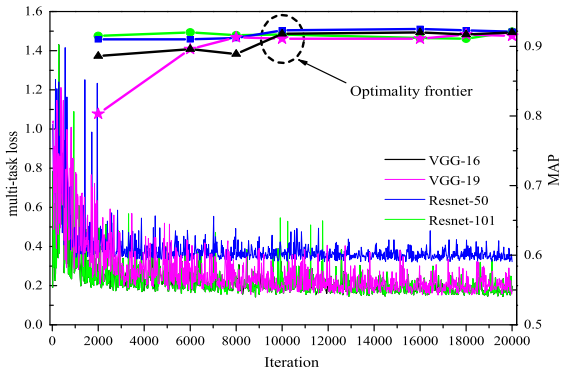

Fig. 7. The total loss values covergence and the MAPs of different feature extractor in the Faster-RCNN network.

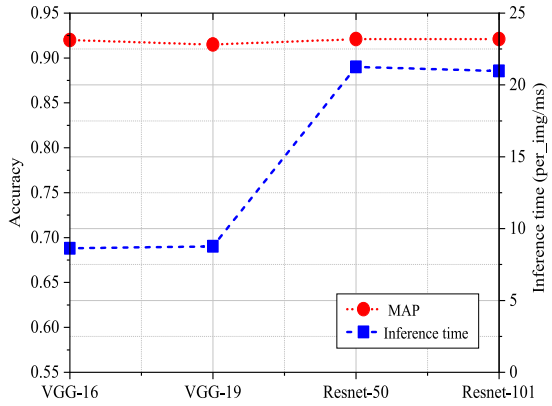

Fig. 8. The effect of the different feature extractor on the accuracy and inference time.

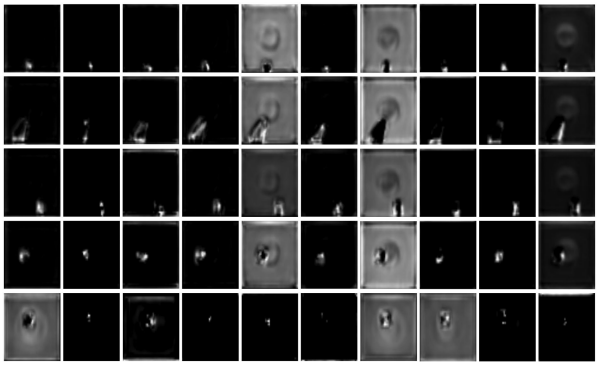

Fig. 9. Some feature maps extracted by CNN-based feature extractor.

We obtain the optimal detection model through the following experiments:

1) Optimality frontier construction to determine the trade-off model. The light intensity distribution maps of sphere, cube, cuboid, cylinder and people are obtained by the shadow model, and trained with the Faster-RCNN. Four feature extractors are trained to acquire the optimal frontier of the Faster-RCNN. Fig. 7 shows the change of the total loss value with the increase of iterative times in the training process. When the iteration reaches 10000 times, the loss values of different extractors are nearly convergent, and the MAPs tend to be consistent. Therefore, the current model is selected to the off-line test, and measured the generalization performance for new objects. Fig. 8 shows the performance of four feature extractors based on the pre-trained model in terms of accuracy and inference time (detect each image). It can be seen that with the increase of the depth and parameters of the feature extractor, the MAPs are basically unchanged. Compared with the general datasets (such as Pascal VOC and COCO datasets), the light intensity map with shadow contains less information. The Faster-RCNN extracts fewer features can achieve a higher recognition rate. However, it may be

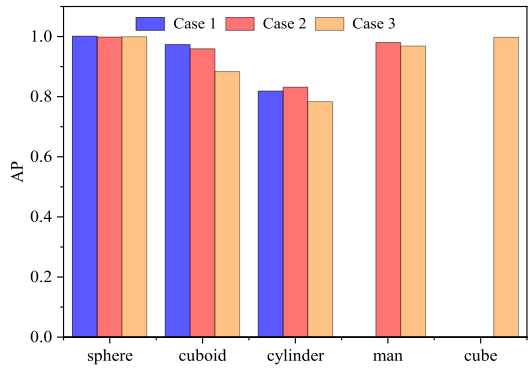

Fig. 10. The APs of objects in three cases. Case 1 contains sphere, cuboid, cylinder, the MAP ups to 93%. Case 2 includes sphere, cuboid, cylinder, and people. The MAP reaches 94.8%. Case 3 adds the cube objects, the MAP is 92.6%.

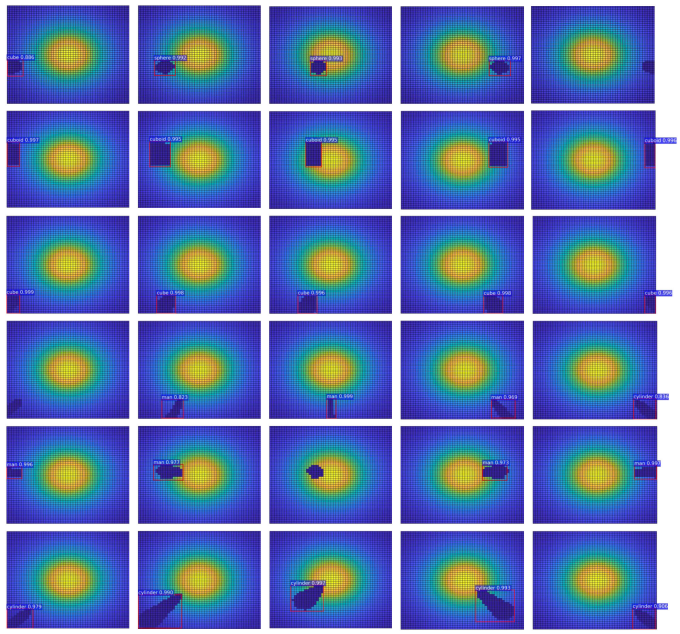

Fig. 11. The detection results for new samples by the optiaml Faster-RCNN. The output box is associated with a category label and a softmax score in $[0, 1]$.

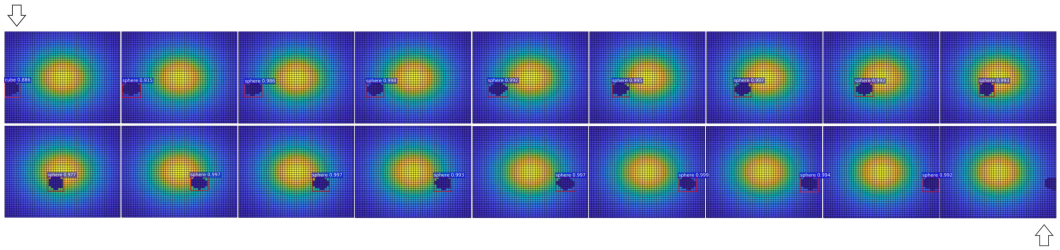

Fig. 12. The detection results of combining recognition and movement of a sphere. The sphere moves at a certain speed and continuously collecting 18 images. More than 70% of the detection results are judged as sphere.

necessary to further verify the performance by collecting more data, especially more complex-shaped objects. Although the accuracy is almost consistent, it is obvious that the inference time of Resnet-50 and Resnet-101 is larger almost 2.5 times than VGG. Therefore, the VGG-16 feature extractor seems to be more suitable for developing the Faster-R-CNN model with great detection accuracy and fast speed.

2) The recognition performance of the different objects. The optimal Faster-RCNN is used to detect the sphere, cube, cuboid and people, and measure the performance. Fig. 9 shows the visualized feature maps, which extracted by the CNN-based feature extractor. The morphological changes of the shadow is consistent with the artificial decision. The multiple targets recognition task is divided into three cases. Case 1 indicates the recognition of sphere, cuboid, cylinder; case 2 is the recognition of sphere, cuboid, cylinder and people; case 3 adds the recognition of cube based on case 2. The APs of the three cases are shown in Fig. 10. In the case 1, the APs of the sphere and cube are superior to the cylinder. The MAP ups to 93%. In the case 2, the people as sphere-and-cube model is added to the test, and the AP reaches up to 98%. We speculate that it may be related to the sphere-and-cube model, which combines cuboid and sphere resulting in a higher recognition accuracy. The APs of sphere, cuboid and cylinder are improved slightly. The MAP reaches 94.8%. In the future, the sphere-and-cube model can be used as the people model. In the case 3, the cube is added to the test, and has a higher recognition accuracy. The MAP decreases to 92.6%. It may be because the increase of categories. To sum up, the optimal Faster-RCNN as a passive target recognition method has great feasibility in the IIoT.
3) The generalization performance of the optimal Faster-RCNN. We test the generalization performance of the Faster-RCNN, which refers to the ability for adapting new samples. The purpose of learning is to learn the rules hidden in the dataset and give an appropriate output for other data. Therefore, the cuboids and cubes in different positions, the sphere with a different radius (30 cm), and people with different heights (160 cm) are tested by the optimal Faster-RCNN model. The confidence threshold is set to 0.8. When the predict confidence is greater than the threshold, the object will be judged as the category. Fig. 11 displays partial visualization results. The first line represents the recognition results of sphere with radius 30 cm. There are false and missed detection, such as the first one, the fifth one. The fourth line is the recognition results of people, in which the first one is missed detection. The optimal Faster-RCNN combined with movement is used to solve the problem of false and missed detection. We assume the object moves at a certain speed and continuously collect 10 or more samples. If 70% of these samples are recognized as the same category, the object will be judged as the category. The results of combining recognition and movement of a sphere are shown in Fig. 12. We continuously collect 18 images, the first of them is false detected, and the last of them is missed detected, which are indicated by the arrow in the figure. However, more than 70% of results are recognized as sphere, so the object is judged as sphere. Therefore, the optimal Faster-RCNN combined with the movement can effectively detect targets and reduce the false and missed detection.

V. Conclusion

In this paper, we have proposed a passive target recognition method based on LED lighting. Compared with lidar-based and vision-based detector, the proposed optimal Faster-RCNN can effectively recognize multiple targets by the shadow. We assume the AGV as a cube, the container as a cuboid, and the people as a cylinder or sphere-and-cube model in IIoT. The simulated light intensity maps with shadow are input into the Faster-RCNN for supervised learning. In the test, we consider three cases: case 1 including sphere, cuboid and cylinder, and the MAP is 93%; case 2 adds the sphere-and-cube model, the MAP is 94.8%; case 3 adds the cube, the MAP is 92.6%. The results reveal that VGG-16-based Faster-RCNN is more sensitive to the changes of cylinder and people, and has high robustness to the sphere and squares. To reduce the false and missed detection, we have developed a Faster-RCNN combined with movement to recognize multiple targets. Simulation results show that the proposed passive recognition method can recognize multiple target with a higher accuracy. Our current and future work include developing this method to a realistic environment and designing a recognition framework related to time.

REFERENCES

[1] C. Lu *et al.*, "Real-time wireless sensor-actuator networks for industrial cyber-physical systems," *Proc. IEEE*, vol. 104, no. 5, pp. 1013–1024, 2016.

[2] A. A. Kumar, S. K. Ovsthus, and L. M. Kristensen, "An industrial perspective on wireless sensor networks—A survey of requirements, protocols, and challenges," *IEEE Commun. Surv. Tut.*, vol. 16, no. 3, pp. 1391–1412, Jul.–Sep. 2014.

[3] H. Haas, L. Yin, Y. Wang, and C. Chen, "What is lifi," *J. Lightw. Technol.*, vol. 34, no. 6, pp. 1533–1544, 2016.

[4] M. A. Khalighi and M. Uysal, "Survey on free space optical communication: A communication theory perspective," *IEEE Commun. Surv. Tut.*, vol. 16, no. 4, pp. 2231–2258, Oct.–Dec. 2014.

[5] A. McCarthy, R. J. Collins, N. J. Krichel, V. Fernández, A. M. Wallace, and G. S. Buller, "Long-range time-of-flight scanning sensor based on high-speed time-correlated single-photon counting," *Appl. Opt.*, vol. 48, no. 32, pp. 6241–6251, 2009.

[6] L. Ye, G. Gu, W. He, H. Dai, J. Lin, and Q. Chen, "Adaptive target profile acquiring method for photon counting 3-D imaging lidar," *IEEE Photon. J.*, vol. 8, no. 6, Dec. 2016, Art. no. 6805510.

[7] H. Hosseinianfar, J. Lian, and M. Brandt-Pearce, "Probabilistic shadowing model for indoor optical wireless communication systems," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, 2019, pp. 936–941.

[8] J. Z. Liang, N. Corso, E. Turner, and A. Zakhor, "Image based localization in indoor environments," in *Proc. 4th Int. Conf. Comput. Geospatial Res. Appl.*, 2013, pp. 70–75.

[9] S. Kumar and S. K. Das, "Target detection and localization methods using compartmental model for internet of things," *IEEE Trans. Mobile Comput.*, vol. 19, no. 9, pp. 2234–2249, Sep. 2020.

[10] M. Youssef, M. Mah, and A. Agrawala, "Challenges: Device-free passive localization for wireless environments," in *Proc. 13th Annu. ACM Int. Conf. Mobile Comput. Netw.*, pp. 222–229, Sep. 2007.

[11] T. Li, Q. Liu, and X. Zhou, "Practical human sensing in the light," *GetMobile: Mobile Comput. Commun.*, vol. 20, no. 4, pp. 28–33, 2017.

[12] E. Alizadeh Jarchlo *et al.*, "Li-Tect: 3-D monitoring and shape detection using visible light sensors," *IEEE Sensors J.*, vol. 19, no. 3, pp. 940–949, Feb. 2019.

[13] N. Faulkner, F. Alam, M. Legg, and S. Demidenko, "Smart wall: Passive visible light positioning with ambient light only," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf.*, pp. 1740–1745, May 2019.

[14] D. Konings, N. Faulkner, F. Alam, E. M. Lai, and S. Demidenko, "Field-light: Device-free indoor human localization using passive visible light positioning and artificial potential fields," *IEEE Sensors J.*, vol. 20, no. 2, pp. 1054–1066, Jan. 2020.

[15] N. Faulkner, F. Alam, M. Legg, and S. Demidenko, "Watchers on the wall: Passive visible light-based positioning and tracking with embedded light-sensors on the wall," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 5, pp. 2522–2532, May 2020.

[16] 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.

[17] Yu-Cheng *et al.*, "Visible light communication and positioning using positioning cells and machine learning algorithms," *Opt. Exp.*, vol. 27, no. 11, pp. 16377–16383, 2019.

[18] Y.-C. Wu *et al.*, "Received-signal-strength (RSS) based 3D visible-light-positioning (VLP) system using kernel ridge regression machine learning algorithm with sigmoid function data preprocessing method," *IEEE Access*, vol. 8, pp. 214269–214281, 2020, doi: 10.1109/ACCESS.2020.3041192.

[19] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[20] C. Li, D. Song, R. Tong, and M. Tang, "Illumination-aware faster R-CNN for robust multispectral pedestrian detection," *Pattern Recogn.*, vol. 85, pp. 161–171, 2019, doi: 10.1016/j.patcog.2018.08.005

[21] M. Pham and S. Lefèvre, "Buried object detection from b-scan ground penetrating radar data using faster-RCNN," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 6804–6807.

[22] X. Sun, P. Wu, and S. Hoi, "Face detection using deep learning: An improved faster rcnn approach," *Neurocomputing*, vol. 299, no. 19, pp. 42–50, 2018.

[23] T. Komine and M. Nakagawa, "Fundamental analysis for visible-light communication system using led lights," *IEEE Trans. Consum. Electron.*, vol. 50, no. 1, pp. 100–107, Feb. 2004.

[24] F. R. Gfeller and U. Bapst, "Wireless in-house data communication via diffuse infrared radiation," *Proc. IEEE*, vol. 67, no. 11, pp. 1474–1486, 1979.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.