Urban road users detection and velocity estimation from top-view fish-eye imagery under low light conditions

Masoomeh Shireen Ansarnia, Etienne Tisserand, Alain Tremeau, Patrick Schweitzer

Urban road users detection and velocity estimation from top-view fish-eye imagery under low light conditions

Masoomeh Shireen Ansarnia
Institut Jean Lamour
Université de Lorraine-CNRS
Nancy, France
masoomeh.ansarnia@univ-lorraine.fr

Etienne Tisserand
Institut Jean Lamour
Université de Lorraine-CNRS
Nancy, France
etienne.tisserand@univ-lorraine.fr

Alain Tremeau
ECLATEC
Maxeville, France
alain.tremeau@eclatec.com

Patrick Schweitzer
Institut Jean Lamour
Université de Lorraine-CNRS
Nancy, France
patrick.schweitzer@univ-lorraine.fr

***Abstract*—In this study we present a technique which performs pedestrians and vehicles detection through the use of off-the-shelf object detectors alongside existing optical flow based velocity estimation techniques. Road users detection and apparent movement quantification are carried out respectively by YOLOv4 and FlowNet2.0. The speed of the users is then estimated from the average optical flow in each bounding box. Experimental results show that our proposed methods can effectively detect road users and estimate their velocity and movement direction under low light and low contrast urban video scenes. The videos are made with a low cost fish-eye RGB camera placed in a vertical position at a height of less than 10 m. The estimated speed by taking into account the deformations generated by the wide-angle lens, is in very good agreement with reality.**

***Index Terms*—Object detection, Optical flow estimation, Orthophotography, Low-illumination, Deep learning**

I. Introduction

With massive increases in the urban population, cities face various urban planning challenges. In recent years, artificial intelligence (AI) has started to manifest itself at an unprecedented pace. With its sophisticated capabilities, cities are leveraging the benefits of technological advancements and implementing the latest AI technologies [1]. Many researchers have used pedestrian or vehicle detection to provide services such as alerts or emergencies in case of danger. Artificial neural networks, support vector machines and genetic algorithms are among the most used AI techniques for accident prediction or unsafe driving pattern analysis during the last ten years [2]. Furthermore, concerning pedestrian detection, low illumination is a major problem in urban and rural environments and different geographic places. It has several inconveniences, especially in the field of security where the obtained night images often lose valuable information due to low brightness, low contrast and high noise. Road user detection is a critical problem in computer vision with a significant impact on safety in urban autonomous driving. Within a fully autonomous driving environment, driver-less vehicles have to communicate and share perceived data with their neighboring vehicles for safer navigation [3]. Also accurately estimating the speed of road vehicles is becoming increasingly important as the enforcement of appropriate speed limits is considered one of the most effective means to increase road safety. We believe that it would be extremely useful to perceive and share these important data from any city infrastructure since it provides another perspective that autonomous vehicles do not necessarily have.

In this work, we study the feasibility to detect and estimate the speed of road users at night and in bad weather conditions with off-the-shelf object detection and optical flow estimation models. We use a combination of YOLOv4 and FlowNet2.0 that allows to detect the presence of users, to analyze their movement and to estimate their real speed in a low light and low contrast video sequence.

II. related work

Object detection is one of the most important challenges in computer vision, due to its variety of usages. This field has vastly advanced in recent years [4], especially with the state of the art CNN-based models such as scaled-YOLOv4 [5], YOLOR [6], and YOLOX [7]. However, for object detection tasks, existing works under low illumination are scarce. In [8] authors trained state-of-the-art object detection models on the illumination-enhanced images. However, they found that the detection performance of models trained on the enhanced image is not better than the performance of models trained directly on original dark image. The experiments imply that

most current image enhancement algorithms, though achieving visually pleasing results, could not enhance low-illumination computer vision tasks, therefore we ignore image enhancement techniques. Many opt for the usage of thermographic camera (FIR) [9]. A study [10] has found that at nighttime, FIR features get the best result compared to the RGB camera by a considerable margin of 50% in average miss rate, although the light level or the illuminance at nighttime is not specified. They also mentioned that concatenating the FIR and RGB features, produces just a slight increase in the average miss rate. Nonetheless, these cameras are relatively more expensive than RGB cameras. Likewise, improving the image by using a DSLR (digital single-lens reflex) camera is not cost-effective and not always adaptable for outdoor use, i.e., in bad weather conditions. Therefore, researchers are also focused on algorithmic solutions for low-cost RGB cameras. Although the above CNN-based models can have an indirect relation with apparent movement estimation, neither of them can achieve speed vector estimation alone.

The speed of an object is usually estimated by two main methods: telemetry (e.g. radar and laser) or image processing methods [11]. Radar is the most widely used method in urban areas but it has limitations such as: (1) can only measure the speed of vehicles, (2) can only measure one object at a time, (3) measuring area is limited. As for the image processing method, there are multiple solutions from background subtraction to optical flow estimation. Optical flow estimation is a computer vision problem of finding pixel-wise motions between consecutive images. RAFT [12], PWC-Net [13] and FlowNet2.0 [14] are amongst the state of the art in this field. These models provide the field of the velocity vector over the entire image and they are more accurate and less noisy than dense flow analytical calculation.

Previously, researchers have developed networks that jointly perform the functions of apparent motion analysis and object detection. Liu et al. [15] introduce a method to real-time estimate the speed of an object by combining two CNNs: YOLOv2 and FlowNet. This was destined to be applied to robotics and autonomous driving. In [16] the authors propose the video object detection method based on the YOLO-v3 together with the FlowNet 2.0 optical flow extraction network. El-Nouby et al. [17] present a model based on YOLOv2 and FlowNet 2.0 for action detection, such as horse riding, skiing, etc. Another method [18] uses Faster R-CNN and landmark-based scanlines on the road surface for vehicle detection and velocity estimation, for bidirectional movement, e.g., on highways, and does not apply to the objects that can have any given direction. Zhang et al. [19] use YOLOv3 to reduce motion blur, video defocus, and partial occlusion, in optical flow estimated by FlowNet 2.0. Finally, in [20], the authors use YOLOv2 for object detection and FlowNet for optical flow estimation. They also merge their results to estimate the speed of detected objects. On the other hand, they do not provide the direction of the velocity vector. In all of these works, two-stream networks successfully solved the problem of specific objects' speed estimation. Although, the methods above do not include top-down view and do not test on all types of road users (e.g. drivers, bikers, pedestrians); and more importantly does not consider low illumination and bad weather conditions.

In this paper, we provide a method to detect road users and estimate their velocity and apparent direction, based on a combination of YOLOv4 object detection and FowNet2.0 optical flow estimation. The method uses the label of the road user and bounding boxes provided by the object detection network to select the optical flow of the object from the whole flow image provided by FlowNet2.0. We evaluate the performance of this method at night.

III. Materials and Methods

A. Filming and shooting setup

The camera is mounted at a height of 7.6 m and tilted vertically towards the ground. This prevents obstacles in the field of view (FoV). The FoV and the height H of the camera are the only geometrical parameters of the system. Detection of humans in this perspective could not be subject to mass surveillance since facial features are not visible enough to be recognized. Another advantage is that the location of the objects in the horizontal plane is more precise and simple to determine. On the other hand, with standard non-deforming lenses, the monitored scene is small. A wide-angle lens is more optimal since it provides a larger area of interest. Nighttime images were shot under urban lighting (10-30 lux). The specifications of the camera are given in Table I.

The algorithm is primarily developed in python and tested on Google Colaboratory. Table II shows the content of our dataset which is divided into 70% for training, 20% for validation and 10% for test. We will consider the detection of those objects whose bounding box side is higher than 60 pixels (about 5% of the height of the registered frames) mandatory. Table III provides the parameters of training.

TABLE I
SPECIFICATIONS OF THE CAMERA USED FOR TESTS AND TRAINING

Specifications	**ARDUCAM B0261**
Sensor	Sony IMX291 1/2.8"
Sensor size	5.64×3.18 mm
Image Resolution	1920×1080
Frame rate	30 fps
FoV_H	$160^\circ = 2,79\ rd$
Focal length	1.77 mm = 602 pixels

TABLE II
DATASET DETAILS

Categories	**Samples obtained**
Total positive frames	4960
Total Negative frames	3040
Annotated pedestrians	3620
Annotated cars	744

TABLE III
TRAINING PARAMETERS

Batch size	Image resolution	Learning rate
64	416×416	0.001

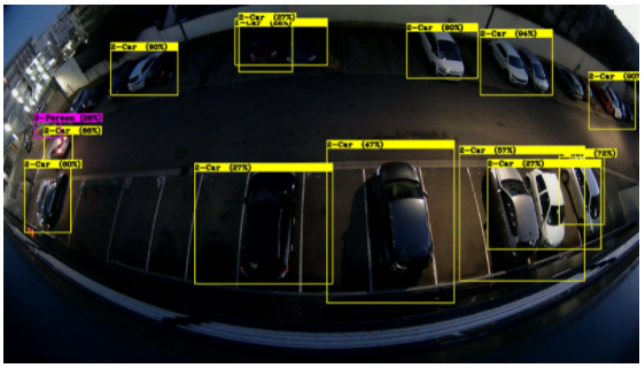

Fig. 1. Example of YOLOv4 detection on an urban parking.

Fig. 2. Example of FlowNet2.0 estimation on an urban parking.

B. Individual model evaluation

- Road User Detection

YOLOv4 object detector was used for road user detection. The training and evaluation of this model have been done on our specific dataset (top view images at nighttime) and the parameters used for training the model are presented in Table III. Also, transfer learning has been done using the pre-trained weights from the COCO dataset for better performance and faster training. Fig. 1 shows an example of detection by YOLOv4.

- Optical Flow Estimation

The FlowNet2.0 model was tested on our video sequences. For a moving car, an output example of this model is shown in Fig. 2. The blue color indicates the movement towards the left, and the intensity of color is related to the apparent speed; hence stationary objects are white. Obtaining satisfactory results, further training is redundant and unnecessary.

$$MeanVelocity \leftarrow \frac{\sum Velocity}{MotionMask} \quad (1)$$

- Results fusion

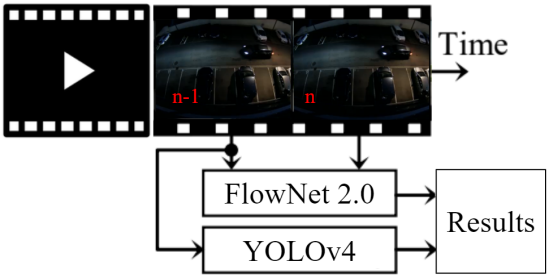

Fig. 3. The video processing procedure.

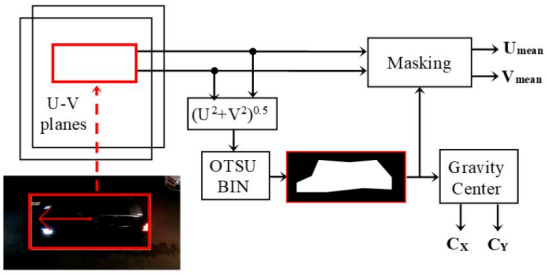

Fig. 4. Block diagram of the fusion.

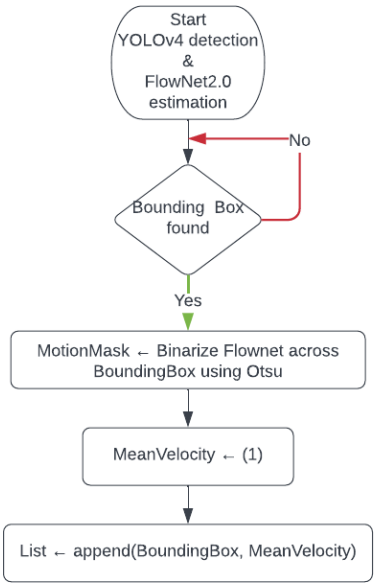

Fig. 5. Flowchart.

FlowNet 2.0 needs two consecutive frames to infer the optical flow. YOLOv4 inference executes on the second-to-last frame. This process is depicted in Fig. 3 and 4. U and V are the components of the displacement vector calculated for each pixel. For each object found by YOLOv4, a "Motion-Mask" is created. If the velocity vector is present in this mask, i.e. the object is moving, the mean velocity is calculated by (1). Then the bounding box and mean velocity information are appended to a list which serve for visualization, the flowchart of this process is given in Fig. 5.

- Actual vehicle speed estimation

In a first approximation for the speed estimation, we used the standard rectilinear projection model (2). "g" is the magnification factor between image pixel coordinates (x,y) and real ground coordinates (X,Y) of a given point "M", demonstrated in Fig. 6. In this geometrical representation, "O" is the optical center of lens; "R" is the distance from point "M" to the projection of "O" on the ground; "r" is the projection of "M" to the center of image; "H" is the height of the lens (camera), and "θ" is the angle of view. We also tested another approximation method to correct the measurement results, using an equidistant projection model (5-6), which is more consistent with fish-eye lenses.

$$g = \frac{x}{X} = \frac{y}{Y} = \frac{f}{H} \approxeq 80\ pixels/m \tag{2}$$

$$R = \sqrt{X^2 + Y^2} = H \times tg(\theta) \tag{3}$$

$$r = \sqrt{x^2 + y^2} \tag{4}$$

The equidistant projection model yields (5) and (6). f' is the equivalent of focal length f, in equidistant model and "w" is the width of the image in pixels.

$$r = f' \times \theta\ with\ f' = \frac{w}{FoV_H} \approx 688\ pixels \tag{5}$$

$$\frac{Y}{X} = \frac{y}{x} \tag{6}$$

By introducing (5) and (6) in (3) we obtain (7).

$$R = X\sqrt{1 + \frac{Y^2}{X^2}} = X\sqrt{1 + \frac{y^2}{x^2}} = H \times tg(\frac{r}{f'}) \tag{7}$$

This finally, results in (8) and (9) for image-to-scene transformation.

$$X = \frac{x}{r} \times H \times tg(\frac{r}{f'}) \tag{8}$$

$$Y = \frac{y}{r} \times H \times tg(\frac{r}{f'}) \tag{9}$$

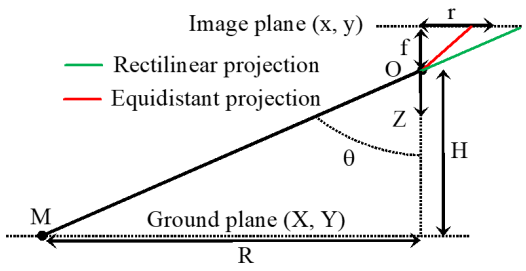

Fig. 6. The geometry of the camera and the ground (seen in half)..

IV. Results

The visualization of the final results is presented in Fig. 7 with some indications for the real distance measurements in the image. The real speed estimation is presented in Fig. 8.

At the beginning of the video sequence (frames 8 to 25), the vehicle moves between the two parking lines (230 cm) in 12 $\pm$ 1 frames. This corresponds to an average speed of 5.75 $\pm$ 0.5 m/s, which is consistent with the estimate provided by our detector. From frame 26 to the end of the video the vehicle reduces its speed before reaching the gate.

The characteristics of our model are compared to other approaches in Table IV.

Fig. 7. A car moving slowly towards the exit gate of a parking.

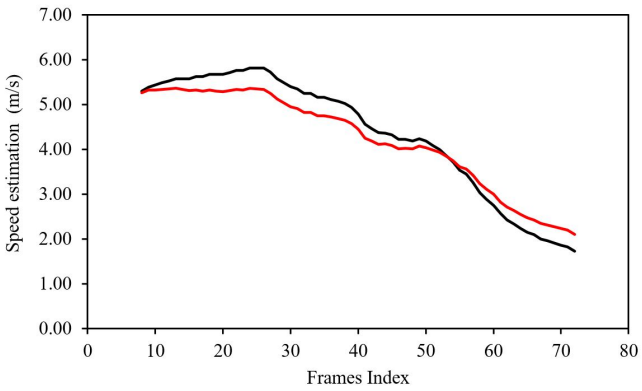

Fig. 8. Real speed estimation. Black: the linear projection model, Red: the equidistant projection model.

V. Discussion

Although one can estimate the speed of a moving object by tracking the object, the height and width of a moving object bounding box are variable during the movement. This can introduce considerable noise in the speed estimation. Also, displacement estimation based on tracking the center of each bounding box requires regular object detection without gaps.

In our approach, the estimation of the velocity does not target a particular point. The velocity calculation is independent of the inevitable false negative object detection.

TABLE IV
COMPARISON

Capacity	[17]	[18]	This paper
Object detection	YOLOv2	Faster RCNN	YOLOv4
Optical flow estimation	FlowNet2.0	Landmark scanlines	FlowNet2.0
Dataset	Kinetics	Nvidia AI City Challenge	Custom dataset
Application	Human action detection	Vehicles on highways	Road user detection and speed estimation
Video shooting mode	Horizontally fixed camera	Horizontally fixed camera	Vertical camera

On the other hand, headlight reflections and sliding shadows on the bodies of parked vehicles result in estimating a non-zero speed for the latter. Fig. 9 provides an example. This is the main drawback of our method. Several solutions to solve this problem are currently under study, such as automatic inhibition of velocity calculation if the area of the moving zone does not exceed a certain percentage of the bounding box area. The center of gravity of the moving area generated by a projected moving shadow is not correlated with the center of the bounding box. Our tests show that generally this center of gravity is found close to the bounding box edges. A criterion based on this deviation can also be exploited to reduce the risk of estimating the velocity of a stationary object. Furthermore, it is possible to restrain the calculation of the speed of a bounding box if the center of the box is quasi-stationary. This solution requires the unambiguous identification of each bounding box by tracking. Since it is very unlikely that the displacement of a vehicle over a few successive images will exceed the width of a bounding box, it is possible to avoid the tracking operation by recording the successive positions of the center of the bounding boxes in a buffer matrix of the same size as the image. A criterion based on the immobility of the center will make it possible to decide whether the object is moving.

VI. CONCLUSION

The method presented in this paper is a combined algorithm that detects the presence of road users and estimates their speed, relying only on video sequences filmed by a fish-eye camera in a vertical position and under low light conditions. Quantitative analyses suggest that the proposed method is capable of detecting movement and estimating speed effectively.

Bounding boxes restrict the type of objects targeted and limit the areas in which apparent motion is estimated. This is an advantage that makes it possible to exclude certain undesirable movements and light variations such as tree leaves, urban displays, and the camera's vibrations. The wide-angle lens is likened to an equidistant projection model that is used to correct velocity measurements. The first tests carried out concern night-time video sequences representing vehicles traveling at low speed at urban parking. The results show speed estimations that are consistent with the ground truth. In the future, we would study the substitution of FlowNet2.0 by the lighter PWC-Net algorithm to allow faster implementation of this method at a video rate higher than 1 fps.

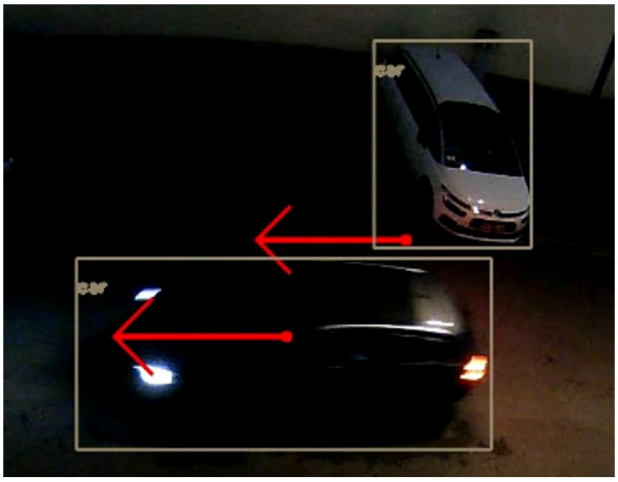

Fig. 9. Incorrect estimation of the speed of a stationary vehicle due to unwanted light reflections.

REFERENCES

[1] T. Yigitcanlar, L. Butler, E. Windle, K. C. Desouza, R. Mehmood, and J. M. Corchado, "Can Building 'Artificially Intelligent Cities' Safeguard Humanity from Natural Disasters, Pandemics, and Other Catastrophes? An Urban Scholar's Perspective," Sensors, vol. 20, no. 10, Art. no. 10, Jan. 2020, doi: 10.3390/s20102988.

[2] Z. Halim, R. Kalsoom, S. Bashir, and G. Abbas, "Artificial intelligence techniques for driving safety and vehicle crash prediction," Artif. Intell. Rev., vol. 46, no. 3, pp. 351–387, Oct. 2016, doi: 10.1007/s10462-016-9467-9.

[3] A. Hbaieb, J. Rezgui, and L. Chaari, "Pedestrian Detection for Autonomous Driving within Cooperative Communication System," in 2019 IEEE Wireless Communications and Networking Conference (WCNC), Apr. 2019, pp. 1–6. doi: 10.1109/WCNC.2019.8886037.

[4] L. Liu et al., "Deep Learning for Generic Object Detection: A Survey," Int. J. Comput. Vis., vol. 128, no. 2, pp. 261–318, Feb. 2020, doi: 10.1007/s11263-019-01247-4.

[5] C.-Y. Wang, A. Bochkovskiy, en H.-Y. M. Liao, "Scaled-yolov4: Scaling cross stage partial network", in Proceedings of the IEEE/cvf conference on computer vision and pattern recognition, 2021, bll 13029–13038.

[6] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "You Only Learn One Representation: Unified Network for Multiple Tasks," ArXiv210504206 Cs, May 2021, Accessed: Mar. 18, 2022. [Online]. Available: http://arxiv.org/abs/2105.04206

[7] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO Series in 2021," ArXiv210708430 Cs, Aug. 2021, Accessed: Mar. 18, 2022. [Online]. Available: http://arxiv.org/abs/2107.08430

[8] Y. Xiao, A. Jiang, J. Ye, and M.-W. Wang, "Making of Night Vision: Object Detection Under Low-Illumination," IEEE Access, vol. 8, pp. 123075–123086, 2020, doi: 10.1109/ACCESS.2020.3007610.

[9] F. Morgan, P. Hurney, M. Glavin, E. Jones, and P. Waldron, "Review of pedestrian detection techniques in automotive far-infrared video," IET Intell. Transp. Syst., vol. 9, Apr. 2015, doi: 10.1049/iet-its.2014.0236.

[10] A. González et al., "Pedestrian Detection at Day/Night Time with Visible and FIR Cameras: A Comparison," Sensors, vol. 16, no. 6, Art. no. 6, Jun. 2016, doi: 10.3390/s16060820.

[11] D. Fernández-Llorca, A. Hernandez Martinez, and I. garcia daza, "Vision-based vehicle speed estimation: A survey," IET Intell. Transp. Syst., vol. 15, May 2021, doi: 10.1049/itr2.12079.

[12] Z. Teed en J. Deng, “Raft: Recurrent all-pairs field transforms for optical flow”, in European conference on computer vision, 2020, bll 402–419.
[13] D. Sun, X. Yang, M.-Y. Liu, and J. Kautz, “Models Matter, So Does Training: An Empirical Study of CNNs for Optical Flow Estimation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 6, pp. 1408–1423, Jun. 2020, doi: 10.1109/TPAMI.2019.2894353.
[14] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, en T. Brox, “Flownet 2.0: Evolution of optical flow estimation with deep networks”, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, bll 2462–2470.
[15] K. Liu, Y. Ye, X. Li, and Y. Li, “A Real-Time Method to Estimate Speed of Object Based on Object Detection and Optical Flow Calculation,” J. Phys. Conf. Ser., vol. 1004, p. 012003, Apr. 2018, doi: 10.1088/1742-6596/1004/1/012003.
[16] S. Zhang, T. Wang, C. Wang, Y. Wang, G. Shan, and H. Snoussi, “Video Object Detection Base on RGB and Optical Flow Analysis,” in 2019 2nd China Symposium on Cognitive Computing and Hybrid Intelligence (CCHI), Sep. 2019, pp. 280–284. doi: 10.1109/CCHI.2019.8901921.
[17] A. El-Nouby and G. W. Taylor, “Real-Time End-to-End Action Detection with Two-Stream Networks,” ArXiv180208362 Cs, Feb. 2018, Accessed: Mar. 14, 2022. [Online]. Available: http://arxiv.org/abs/1802.08362
[18] M.-T. Tran et al., “Traffic Flow Analysis with Multiple Adaptive Vehicle Detectors and Velocity Estimation with Landmark-Based Scanlines,” in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Jun. 2018, pp. 100–1007. doi: 10.1109/CVPRW.2018.00021.
[19] S. Zhang, T. Wang, C. Wang, Y. Wang, G. Shan, and H. Snoussi, “Video Object Detection Base on RGB and Optical Flow Analysis,” in 2019 2nd China Symposium on Cognitive Computing and Hybrid Intelligence (CCHI), Sep. 2019, pp. 280–284. doi: 10.1109/CCHI.2019.8901921.
[20] K. Liu, Y. Ye, X. Li, and Y. Li, “A Real-Time Method to Estimate Speed of Object Based on Object Detection and Optical Flow Calculation,” J. Phys. Conf. Ser., vol. 1004, p. 012003, Apr. 2018, doi: 10.1088/1742-6596/1004/1/012003.