ISBN: 978-981-18-5852-9
WCSE 2022 Spring Event: 2022 9th International Conference on Industrial Engineering and Applications
doi: 10.18178/wcse.2022.04.161

Real-time Detection System of Illegal Behaviors in Urban Management Based on Deep Learning

Yanjun Fan [1], Zhuo Cheng [1], Chao Wang [1] and Peng Tian [1]

[1] Research and Development Center
Suzhou Vortex information Technology Co., LTD.
Suzhou, China, 215124

Abstract. The paper presents a real-time detection system of illegal behaviors in urban management based on deep learning. First of all, we collected and annotated a large number of video images, and established object detection dataset and ground image classification dataset, which serve as the data basis for the training of the system's deep learning algorithm. Then, the algorithm framework of the system is proposed, and some new methods are proposed in the key steps of the algorithm, such as upsampling method, filtering rules of object tracking, ground ROI region extraction and illegal behavior determination. Experimental results indicate that the proposed system can meet the requirements of practical applications in real time and performance.

Keywords: Machine learning, Computer vision, Deep learning, Object detection, Object tracking, Image classification, Dataset

1. Introduction

Today, with the rapid process of global urbanization, there are more and more challenges in urban management, such as illegal parking of motor vehicles, illegal parking of non-motor vehicles, unlicensed vendors, littered bags, etc. In the traditional urban management pattern, it takes a lot of people to patrol to detect these illegal behaviors, which is inefficient and costly. In this paper, we proposed and developed an intelligent detection system based on computer vision and deep learning technology, which can effectively detect various illegal behaviors, such as illegal parking, unlicensed vendors and littered garbage bags. The intelligent system can complete the task of urban management patrol in a more efficient and economical way.

There has been considerable amount of research related to special tasks of illegal parking detection [1-5]. Background modeling and motion tracking approaches, such as Gaussian mixture model [1] and Kalman filter [2], have been widely used in illegal parking detection systems with stationary cameras. In recent years, more and more researchers have applied deep learning to the illegal parking detection system with stationary cameras. Alon and Dioses [3] used MobileNet SSD algorithm to detect illegally parked vehicles, realizing an economical outdoor illegal parking detection system. In paper [4], the Faster-RCNN combined with the VGG-16 detection framework was used to detect illegal Parking. In paper [5], illegally parked vehicles were located and classified using an optimized SSD method. Although good performance has been achieved, the above illegal parking detection systems based on deep learning can only be applied to stationary camera monitoring scenarios, not with moving camera. Object detection with mobile cameras has more challenges, such as cluttered environment, various illumination conditions, weather conditions, occlusion, shadows of buildings or trees, etc. The goal of our system is to detect a variety of objects and violations with moving cameras, such as vehicles, bicycles, electric bicycles, people, unlicensed vendors, garbage bags, etc.

In the paper, a real-time illegal behaviors detection system based on deep learning is proposed, which can not only detect motor vehicles illegal parking, but also detect non-motor vehicles illegal parking, unlicensed vendors and littered garbage bags. The other parts of the paper are organized as follows. Section 2 presents the hardware composition of the system, as well as the datasets we collected and annotated to train the deep learning algorithm in the system. Section 3 introduces the algorithm flow of the system and several new methods we put forward in the main steps of the system. Section 4 illustrates the experimental results of

the system on our datasets and the experimental results of real road tests, and conclusions are made in Section 5.

2. Hardware and datasets

2.1. Hardware

The system we proposed is mainly composed of the following hardware: 1) An SUV (Changan Auchan X5) is used as the main infrastructure to improve the mobility of the whole system. 2) The on-board camera, Hikvision DS-2CD2T25FD-I3, installed on the roof of the car, which has a resolution of 1920*1080P and a frame rate of 12 frames per second. 3) The CHCNAV CGI-210, a high-precision GPS device installed inside the car, is used for positioning the car itself and estimating its own motion. 4) In order to meet the requirements of low power consumption in on-board conditions, we chose a Laptop computer as the computing infrastructure of the system. The laptop is powered by an AMD 5900HS 8-core processor, 16GB of 3200MHz ram and an NVIDIA RTX 3060 laptop GPU graphics card. The modified experimental vehicle used in the illegal behaviors detection system is shown in Fig. 1.

Fig. 1: The modified experimental vehicle used in our illegal behaviors detection system

2.2. Datasets

The quality of datasets has a significant impact on the performance of deep learning models. Due to the lack of public datasets for the application scenarios we proposed, we decided to collect video data and annotate images by ourselves.

1) Video data collection

Using the modified experimental vehicle shown in Fig. 1, we collected a large amount of real street landscape video data from August 2021 to September 2021 in Suzhou, China. After the video data was collected, the video was segmented into image frames, and the images that may contain the target object were selected through rapid manual screening and provided to the next images annotation step.

2) Images annotation

In the stage of images annotation, there are two types of annotation tasks: object-level annotation and image-level annotation. Object-level annotation provides the training dataset for the object detection task, which require marking object categories and a tight box around the object in the image. Image-level annotation is relatively simple, only need to mark a binary label for the presence or absence of an object class in the image. The results of image-level annotation are usually used to train the object classification model.

In order to improve the efficiency of image annotation, we developed a software system that can help annotation workers to complete image-level annotation and objective-level annotation more quickly. After images annotation processing, two kinds of datasets were obtained: object detection dataset and ground image classification dataset.

3) Object detection dataset

The object detection dataset contains 11 object categories that may be associated with illegal behaviors. In the image, each object is marked with its corresponding object category, and its position is marked with a tight bounding box. The object categories and the number of objects of each category in the object detection dataset are shown in Table. 1. Typical images of object categories in the object detection dataset are shown in Fig. 2.

Table 1: The number of objects of each category in the object detection dataset

Object categories	Number
Car	29221
Person	18961
Electric bicycle	12491
Minibus	6352
Garbage bag	5806
Tricycle with signboard (Unlicensed vendor)	3219
Bicycle	3728
Trike	3280
Pedicab	3138
Van	2956
pickup truck	2731
Total	**91883**

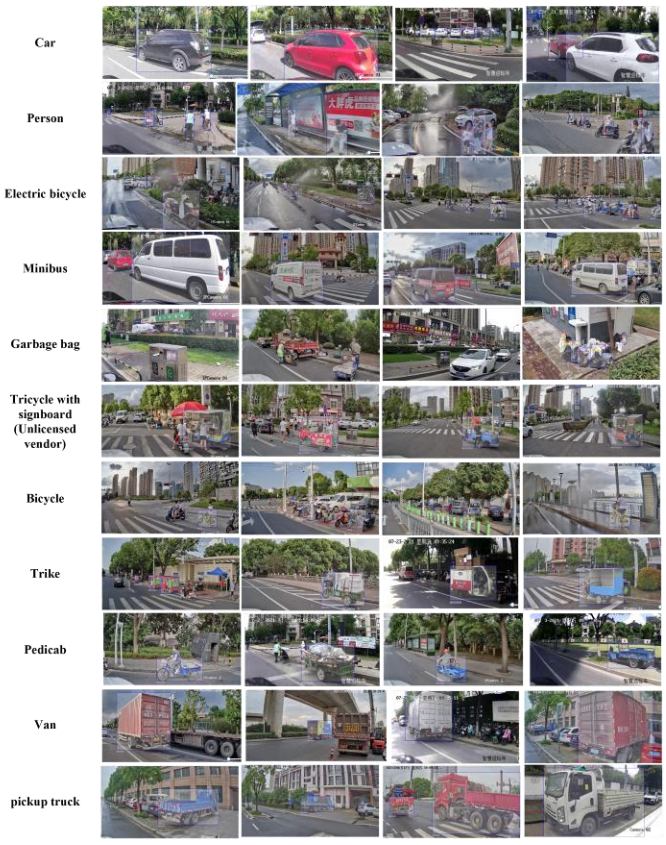

Fig. 2: Typical images of object categories in the object detection dataset

4) Ground image classification dataset

After the object that may be associated with illegal behavior is detected from the image, in order to judge whether there is illegal behavior, we also need to determine whether the object exists in the legitimate area. If the object exists in an illegal area, the system should judge it as an illegal behavior. The features of ground images, such as roads with parking lines, sidewalk with parking lines, and sidewalk with bicycle parking stubs, can be used to determine whether it is a legitimate area. To this end, we set up the ground image classification dataset with 7 categories by manual annotation. A total of 26,487 images were collected and annotated. Table. 2 shows the number of images under each category, and the typical category images are shown in Fig. 3.

Table 2: The number of images of each category in the ground image classification dataset

Image categories	**Number**
Sidewalk	3275
Sidewalk with parking lines	4238
Sidewalk with bicycle parking stubs	3884
Tree lawn	3394
Lawn	1860
Road	6326
Road with parking lines	3510
Total	**26487**

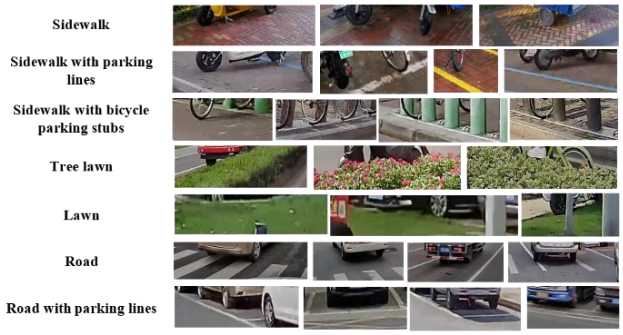

Fig. 3: Typical images of categories in the ground image classification dataset

3. The proposed system

As illustrated in Fig. 4, the algorithm flow of the system we proposed is mainly composed of six steps as follows: video image capture, preprocessing, object detection, object tracking, ground ROI classification and illegal behavior determination.

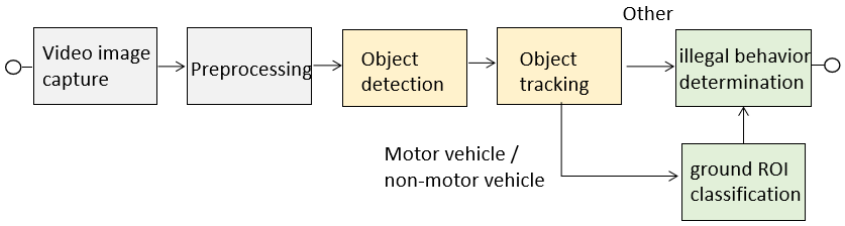

Fig. 4: The algorithm flow chart of our system

3.1. Video image capture and preprocessing

First of all, the real-time street view video images were captured by the on-board camera in our system. Furthermore, a series of preprocessing such as video decoding, frame segmentation and image resizing were carried out to obtain the normalized image for the next step.

3.2. Object detection

In this step, YOLO [6] approach was selected to perform the object detection task. Compared to computer servers, on-board hardware resources are more limited. This requires the deployed deep learning algorithm model to be small in size and fast in execution under the premise of high precision. Because the YOLO approach can integrate gradient changes into feature maps, it can reduce the number of model

parameters and FLOPS. YOLO model not only ensures high performance and precision, but also reduces the size of the model. It is particularly suitable for deep learning scenarios where hardware resources are limited.

In our system, YOLOv5 algorithm was used to detect 11 object categories in the image. If the target object was detected, the category label, bounding box position and the original image were transmitted to the object tracking processing step as input data. In our experiment, it was found that when YOLOv5 model was used to detect large objects in the image, such as motor vehicles or non-motor vehicles, the model had high precision and recall. But for smaller objects, such as garbage bags, the model had a lower recall and a large number of positive cases were predicted to be negative. In order to solve this problem, the upsampling method was used in the 17th layer of YOLOv5 model network, which was commonly known as small object detection layer. The upsampling method can enlarge feature maps and enhance the dimension of data, thus improving the model's ability to detect small objects.

3.3. Object tracking

After objects were detected in the previous step, Deep SORT [7] was applied to realize multi-object tracking, and the system records all image frames of the object from appearance to disappearance. By tracking the movement of objects and analyzing the image features around them, it is possible to filter out objects that are clearly unlikely to be doing anything illegal. For example, a car driving normally on the left side of our experimental vehicle, or a car temporarily parked on a street with the driver still on it, should not be judged as illegal parking.

The filtering out objects that are unlikely to be illegal helps our systems improve performance and precision. To this end, we propose a series of rules to filter out objects that are unlikely to be illegal. Assume that the image frame of the object appears is F_s, and the image frame of the object disappears is F_e, then the images between them form a frame sequence $\{F_s, \ldots\ldots, F_e\}$. The filtering rules we proposed are as follows:

- If the tracked object appears to the left of the center line in the start frame F_s, it should be filtered out.
- If the tracked object disappears to the left of the center line in the last frame F_e, it should be filtered out.
- If tracing an object from appearance to disappearance in sequence $\{F_s, \ldots\ldots, F_e\}$ does not conform to the rule from left to right, it should be filtered out.
- If the area of the bounding box of the tracked object does not conform to the increment rule in the sequence $\{F_s, \ldots\ldots, F_e\}$, it should be filtered out.
- With the help of on-board high-precision GPS equipment, the system calculated the moving distance of the experimental vehicle during $\{F_s, \ldots\ldots, F_e\}$. If the distance is greater than the threshold value Δ (indicating that the tracked object appears in the scene for too long), the tracked object should be filtered out. The value of Δ is closely related to the driving speed of the experimental vehicle. In our experiment, we took the threshold value Δ=60m.
- Similar to the previous rule, if the experimental vehicle is stationary during $\{F_s, \ldots\ldots, F_e\}$, all tracked objects should be filtered out;
- Rules for determining when someone is near a vehicle. Set a global counter S = 0. When both people and cars appeared in a frame F_t, the IoU (Intersection over Union) between the two objects was calculated. If the IoU is greater than 10%, it indicates that the person is sitting on the vehicle, then S = S + 1. Repeat the above decision rules and counts for all image frames in sequence $\{F_s, \ldots\ldots, F_e\}$. Let Count be the total number of image frames in $\{F_s, \ldots\ldots, F_e\}$ and calculate the value of $S/Count$. If the value is greater than the threshold θ, it indicates that there are people on the tracked object. So, the tracked object should be filtered out. The choice of threshold θ should be based on experience. In our system, the value was θ=0.2.

3.4. Ground ROI classification

After the processing of the previous step, if the input object obtained is a motor vehicle or a non-motor vehicle, the ground ROI image near the location of the vehicle should also be extracted. In this step, the system classified the ground ROI image to judge whether the vehicle was illegally parked.

EfficientNet [8] was selected to classify ground ROI images in our system. With EfficientNet, our system was able to determine whether the vehicle was parked in a legal area (i.e., white or yellow parking lines on the ground near the wheels) or in an illegal area (i.e., sidewalk, lawn, etc.).

1) Ground ROI region extraction

After object detection and object tracking, a tight bounding box was marked around the detected object. Assume that the width and height of the bounding box are W and H, respectively, and the midpoint coordinates of the lower edge are (x,y). The extracted ground ROI region is rectangular ABCD (as shown in Fig. 5). The coordinates of point A in the upper left corner of rectangle ABCD are $\left(x-\frac{3W}{4}, y-\frac{H}{4} \right)$, and the coordinates of point C in the lower right corner are $\left(x+\frac{3W}{4}, y+\frac{H}{4} \right)$.

Fig. 5: Extraction method of ground ROI region

2) Enhancement and normalization of ground ROI images

After the ground ROI image was obtained, the symmetric expansion method was used to normalize the image and expand the image size to 380*380, as shown in Fig. 6. The symmetric expansion method has two advantages: the size of the image was expanded to the normalized size required by EfficientNet, and the detailed features of the image were enhanced.

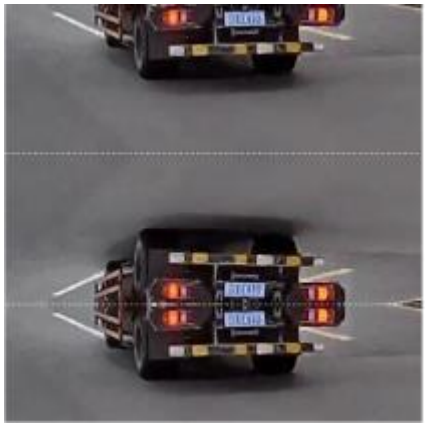

Fig. 6: Enhancement and normalization of ground ROI images

3.5. Illegal behavior determination

In this step, according to the results of object detection, object tracking and ground ROI image classification, we needed to judge whether there were illegal parking, unlicensed vendors, littered garbage bags and other illegal behaviors in the video images. The decision rules of these illegal behaviors include the decision rules of illegal parking, unlicensed vendors and littered garbage bags. The most complex decision rule is the illegal parking decision rule, which is described below.

- If the output of the object tracking stage is motor vehicles (such as cars, minibuses or vans), and the ground ROI image classification result is the road with parking lines, it should be judged as legal parking of motor vehicles.

- If the output of the object tracking stage is motor vehicles (such as cars, minibuses or vans), and the ground ROI image classification result is sidewalk, tree lawn, lawn or road, it should be judged as illegal parking of motor vehicles.
- If the output of the object tracking stage is non-motor vehicles (such as bicycle, electric bicycle or trike), and the ground ROI image classification result is sidewalk with parking lines or sidewalk with bicycle parking stubs, it should be judged as legal parking of non-motor vehicles.
- If the output of the object tracking stage is non-motor vehicles (such as bicycle, electric bicycle or trike), and the ground ROI image classification result is sidewalk, tree lawn, lawn or road, it should be judged as illegal parking of non-motor vehicles.

4. Experimental Results

4.1. Training and test results on datasets

1) Experimental results of object detection

The YOLOv5 object detection model was trained and tested on our own object detection dataset, which contains 91883 labels covering 11 object categories. The iterative training and testing strategies were used to verify the performance of object detection model in our experiment. The mAP curve of iterative training and testing process is shown in Fig. 7. It can be seen that with the increase of iteration steps, the mAP value of the model increases gradually and finally approaches the value of 0.945, indicating that the model has good performance.

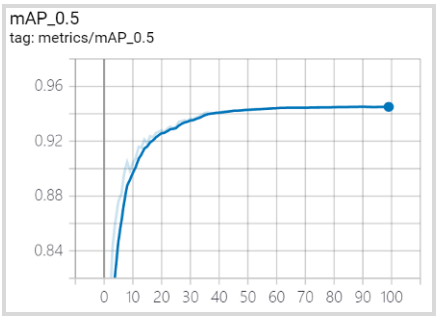

Fig. 7: The mAP curve of iterative training and testing process

The PR curve of the model with a confidence of 0.5 is shown in Fig. 8. It can be concluded that when the confidence is 0.5, the integral area value under the PR curve of all detected object categories is 0.945, at which point the model has high precision and recall.

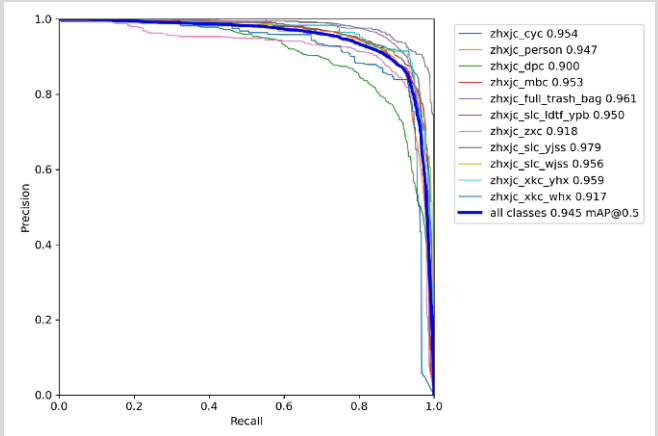

Fig. 8: The PR curve of the model with a confidence of 0.5

2) Experimental results of ground image classification

The EfficientNet object classification model was trained and tested on our own ground image classification dataset, which contains 26,487 ground images. The images in the dataset were divided into training dataset and test dataset in the ratio of 6:4. The relation curve between classification precision and iterative training times of ground image classification model is shown in Fig. 9. As can be seen from the

figure, after more than 300 rounds of training, the Top1 precision of ground image classification finally approached 90.46%

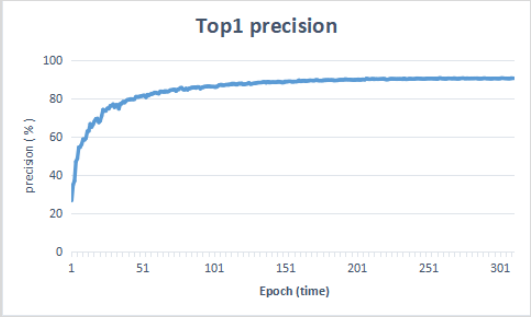

Fig. 9: The relation curve between classification precision and iterative training times of ground image classification model

4.2. Experimental results under real road conditions

To test the overall performance of the system in real street view, we conducted several experiments under real road conditions in Suzhou, China, in November 2021. The experimental results are shown in Table. 3.

In the first experiment on November 2, 2021, there were two defects as follows: 1) Recognition precision and recall of illegal behaviors of unlicensed vendors were lower, only 82.35% and 77.78%, respectively. 2) As there were few littered garbage bags in the real road scene of Suzhou, there were only a few samples of littered garbage bags in our dataset. Therefore, in the first actual road test, the experimental results of the precision and recall were very poor, both of which were zero. In order to solve the above two problems, we have done the following work: 1) Improved the criteria for judging illegal behaviors of unlicensed vendors. For example, a stationary tricycle with a signboard and someone beside it can be considered to be an unlicensed vendor. In addition, we re-annotated the images in the dataset according to the new criteria. 2) In order to solve the problem of less data samples of littered garbage bags, artificial simulation method and image enhancement technology were used to increase the number of data samples. In the actual road test, artificial simulation method was also adopted to increase some littered garbage bags scenarios.

On November 9, 2021, we conducted the second real road test, and the experimental results showed that the recognition precision and recall of unlicensed vendors were greatly improved, reaching 95.45% and 91.30% respectively. However, the recall of littered garbage bags recognition was only 85.71%, which could not meet the requirements of practical application. After analyzing the images in the experiment, it was found that in the actual street scene environment, the size of garbage bags in the image was much smaller than that of motor vehicles or non-motor vehicles, so there were few image features of garbage bags in the whole image. Therefore, we optimized the network structure of YOLOv5s to improve the detection ability of the model for small size objects.

Table 3: Experimental results under real road conditions

Experimental date	Categories of illegal behaviors	TP	FP	TP+FP	TP+FN	Precision	Recall
The first experiment （2021.11.2）	Illegal parking of motor vehicle	100	4	104	105	96.15%	95.24%
	Illegal parking of non-motor vehicle	75	5	80	80	93.75%	93.75%
	unlicensed vendors	14	3	17	18	82.35%	77.78%
	littered garbage bags	0	1	1	1	0.00%	0.00%
The second experiment （2021.11.9）	Illegal parking of motor vehicle	178	4	182	182	97.80%	97.80%
	Illegal parking of non-motor vehicle	169	2	171	171	98.83%	98.83%
	unlicensed vendors	21	1	22	23	95.45%	91.30%
	littered garbage bags	12	1	13	14	92.31%	85.71%
The third experiment （2021.11.16)	Illegal parking of motor vehicle	132	0	132	132	100.00%	100.00%
	Illegal parking of non-motor vehicle	110	2	112	113	98.21%	97.35%

	unlicensed vendors	23	1	24	24	95.83%	95.83%
	littered garbage bags	21	0	21	22	100.00%	95.45%
The fourth experiment （2021.11.19)	Illegal parking of motor vehicle	151	0	151	151	100.00%	100.00%
	Illegal parking of non-motor vehicle	166	1	167	167	99.40%	99.40%
	unlicensed vendors	32	0	32	32	100.00%	100.00%
	littered garbage bags	48	0	48	49	100.00%	97.96%

By analyzing the results of the third and fourth experiments, it can be found that the recognition and detection results of these four types of illegal behaviors are very good, and the precision and recall are above 95%, which can meet the requirements of practical application. In the actual road test environment, the detection speed of the system can reach 12~16 frames/SEC, which can meet the real-time requirements of practical application.

5. Conclusions

In this paper, we proposed a real-time detection system for illegal behaviors in urban management based on deep learning. The proposed system consists of six steps, among which the main algorithm steps are object detection, object tracking, ground ROI classification and illegal behavior determination. In these steps of the system, we put forward some new methods, such as filtering rules of object tracking, ground ROI region extraction and illegal behavior determination, etc. The experimental results show that the proposed system can meet the requirements of practical applications in real time and performance. In addition, we also built an object detection dataset containing 91883 labels covering 11 object categories and a ground image classification dataset containing 26,487 ground images, which is another contribution to this study. Our possible future research directions include: 1) Research methods to improve system precision and recall. 2) Study the technology to improve the real-time performance of the system; 3) Expand the detection scope of the system for illegal behaviors, such as detection of missing manhole covers, unapproved billboards, clutter in front of stores , etc.

6. References

[1] J. T. Lee, M. S. Ryoo, M. Riley, and J. K. Aggarwal. Real-Time Illegal Parking Detection in Outdoor Environments Using 1-D Transformation. *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, pp. 1014-1024, 2009.

[2] R. Akhawaji, M. Sedky, and A. Soliman. Illegal Parking Detection Using Gaussian Mixture Model and Kalman Filter. *2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA)*, 2017, pp. 840-847.

[3] A. Alon and J. Dioses Jr. A machine vision detection of unauthorized on-street roadside parking in restricted zone: an experimental simulated barangay-environment. *International Journal of Emerging Trends in Engineering Research*, vol. 8, pp. 1056-1061, 2020.

[4] J. Xue-Hong, F. Hui-Li, and L. Shi-Yue. The Researches Detection Method of Illegal Parking Based on Convolutional Neural Network. *Proc. of the International Academic Conference on Frontiers in Social Sciences and Management Innovation (IAFSM 2019)*, 2020, pp. 7-11.

[5] X. Xie, C. Wang, S. Chen, G. Shi, and Z. Zhao. Real-Time Illegal Parking Detection System Based on Deep Learning. *Proc. of the 2017 International Conference on Deep Learning Technologies*, Chengdu, China, 2017.

[6] J. a. D. Redmon, Santosh and Girshick, Ross and Farhadi, Ali. You Only Look Once: Unified, Real-Time Object Detection. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[7] N. Wojke, A. Bewley, and D. Paulus. Simple Online and Realtime Tracking with a Deep Association Metric. *IEEE*, pp. 3645-3649, 2017.

[8] M. Tan and Q. Le. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *Proc. of the 36th International Conference on Machine Learning, Proc. of Machine Learning Research*, 2019.