Compact automatic solution to detect and warn abandoned children on school buses

Ngo Thi Thu Huong
Dept. of Electrical and Electronics
Campus in Ho Chi Minh City
University of Transport and Communications
Ho Chi Minh City, Viet Nam
huongntt1_ph@utc.edu.vn

Van Binh Nguyen
School of Electrical Engineering
International University
Ho Chi Minh City 700000, Viet Nam
Vietnam National University
Ho Chi Minh City 700000, Viet Nam
nvbinh@hcmiu.edu.vn

***Abstract*—This paper proposes a compact automatic system, which could be installed easily on vehicles, to detect and send warnings if pupils are accidentally left on school buses. The detection and identification method is mainly based on MobileNets - the new structure of convolutional neural networks. The information is then fed to an embedded processing module and alarms are provided to relevant people via sound, text messages, phone-calls and/or emails in order to avoid unfortunate consequences. The device is tested on an real bus to show the efficiency and safety with various situations.**

Keywords—Human detection, CNN neural network, MobileNets, embedded system

I. Introduction

In recent years, with the increasing school transportation services, related unfortunate incidents appear more often. Most of the people in danger are children who travel by school bus every day. Around the world, school buses for students are designed and equipped with special equipment to ensure student safety during the time on the vehicle.

In personal cars, there are modern technologies to prevent the neglect of young children. These can be mentioned as: Hyundai rear passenger warning system, General Motors rear seat reminder system, ... Depending on the vehicle manufacturer, these systems may be built into the vehicle during production. There are also companies specializing in the production of smart devices that help protect children, such as Sensorsafe safety technology developed by Evenflo, Drivers Little Helper – sensor system to assist the driver. These devices are placed under the seat or attached to the seat belt when the child is in the car. These technologies work well for personal cars, but when it comes to school buses, it is quite expensive and puts additional pressure on the driver in terms of device management.

The school bus system in the world has also applied many methods to help protect children. In addition, there are more separate designs with specific colors, specified in the law and given priority in traffic. It is also equipped with alarm systems to prevent students from being forgotten. Bell-flag-empty vehicle signs are installed at the end of the vehicle, forcing the driver to check the entire vehicle at the end of the journey (Australia; Canada; Korea). In more modern systems, students have to swipe their personal cards when getting on and off the bus (Zonar Z-Pass device, USA). Besides, a smart sensor system alarms when it detects that students and belongings are left behind (developed United Arab Emirates).

In Viet Nam, there is no requirement or any specific regulation for school buses. In reality, each school has its own set of rules. There have also been some studies to help protect children after recent accidents. Two students at Bai Chay High School (Quang Ninh) proposed an automatic warning device with an ultrasonic sensors being used to recognize driver having left his car. After a certain period of time, if the driver has not returned, the system would activate alarms when there is a child in the car. A text message will be sent to driver or via a phone call and the driver as well as parents could talk with the child. This device has been successfully tested on a car, but when considering a school bus (more than 16 seats), the number of sensors will have to be more so the reliability will not be high.

This research carries out a solution to monitor and manage school bus. This paper is divided into three main sections. Section II introduces a solution to detect and warn abandoned children on school buses. Section III presents a children recognition algorithm by using CNN – Mobilenet Network. Section IV discusses results from the child recognition program and from real system installed on a 16-seat car.

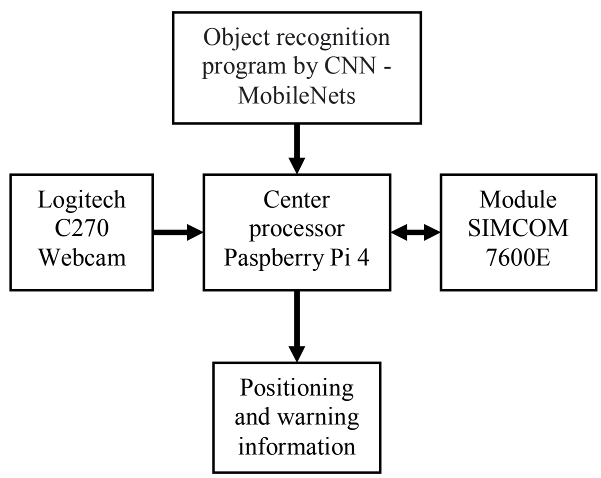

Fig. 1. Structure of warning and detection system

II. Detecting and Warning Abandoned Children

From the mentioned reasons, the authors propose a warning and detection system for students being left on vehicles with the structure diagram in Fig. 1. Battery power supplies all devices, including Raspberry Pi 4 central processing unit, Logitech C270 webcam camera, SIMCOM7600E module and other necessary actuators. The central processing unit is Raspberry Pi 4 [1, 2] which takes image from the camera and follows a special program to process it. The object recognition algorithm using the network architecture of CNN – MobileNets compares the image information with the built-in training results. If there are still children on the school bus, the system will send action commands to the warning message block.

The child body recognition program uses a deep learning model - convolutional neural network [3-8] with MobileNet structure. MobileNets [9, 10] are networks that can be easily used with low hardware configurations such as embedded computers. The training phase is programmed in Python on the OpenCV and Tensorflow library platforms. The environment for data training is used on Ubuntu 18.04 LTS operating system. This is an open sources operating system so it is flexible and strongly supported by the community. However, this environment requires more experience and is more difficult to use than the windows operating system.

The camera which collects images in cars is placed at an appropriate position with good viewing angle. In fact, for the system to work well, we should use a specialized camera with advanced features such as an infrared camera. SIMCOM7600E module is used to support calling, texting, and providing GPS information about the last pick-up and drop-off point of the journey. Besides, the module also is integrated Wi-Fi and Bluetooth function so that the system can be expanded to develop other features. Positioning and warning information block includes a power control module. This module helps providing audible and telephone alarms.

A. Center processor

Raspberry Pi 4 has a compact design, as the size of an ATM card (Fig. 2). There is a CPU, GPU, RAM, micro-SD card slot, Wi-Fi, Bluetooth on board. Installing the operating system can be done easily. Pi 4 has suitable technical features to solve the problems of the proposed system.

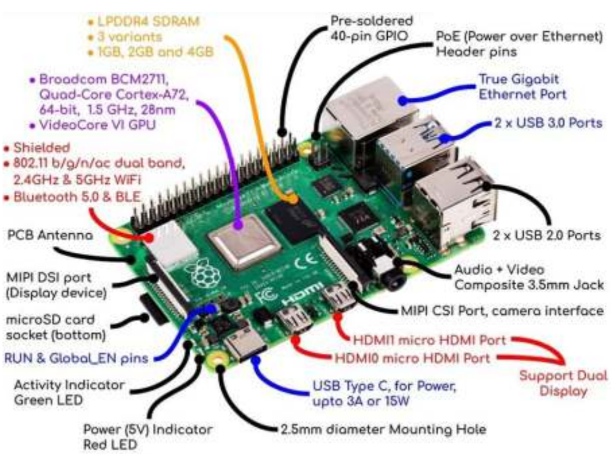

Fig. 2. Raspberry Pi 4 embedded computer

Raspberry Pi 4 uses the open-source operating system Raspberry Pi OS 64bit which is developed from the Linux kernel. The operating system is stable and should be consistent with the design goals.

B. Power supply

One very important block is the power supply. To ensure fire safety, the system operates on a separate power source from that of the car. The battery used is a 3.7V 30Ah Lithium type. Because the Raspberry Pi 4 uses a 5V input voltage, a circuit is provided to increase the voltage from the battery's 3.7V to 5V. The advantage of Lithium batteries is that they have a high discharge current and can be reused many times. However, the battery life depends on the charge-discharge cycle, so it is necessary to follow the correct technical process.

C. Camera

Logitech C270 camera could be a suitable choice for this proposed system because of its reasonable price, good quality. The resulting images are of HD quality (1280 x 720p resolution) with automatic light adjustment, auto focus features. This camera can be easily connected to Raspberry Pi 4 via USB 2.0 port.

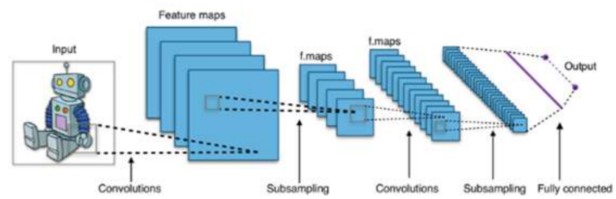

Fig. 3. Typical CNN architecture

III. CNN-MOBILENETS BASED CHILDREN RECOGNITION

Convolutional Neural Network (CNN) is one of the advanced deep learning models that helps building intelligent systems with high precision. CNN establishes the connections between the network's layers based on convolutional calculus, thereby highlighting features to identify objects [1, 3].

The basic structure of CNN consists of convolutional layers, nonlinear layers, and pooling layer (Fig. 3). Convolutional layers combine with nonlinear classes that use nonlinear functions such as ReLU to generate higher-level information for the next layer.

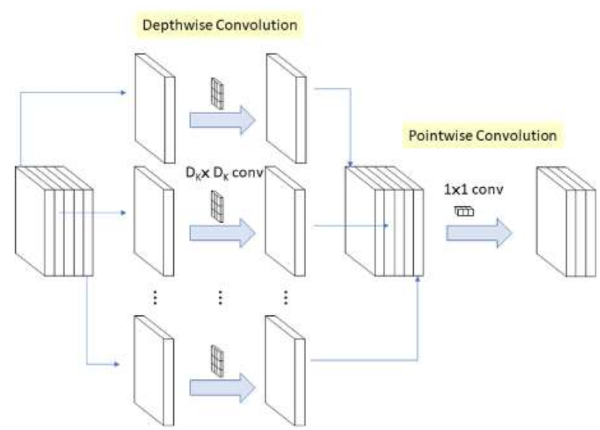

Fig. 4. Depth-wise Separable Convolution

The layers of CNN are divided into two main parts, including image feature filter layer (Conv, Relu, and Pool) and classification layer (FC and softmax). Depending on the number and order of layers in the CNN network, there could be different network architectures such as LeNet, AlexNet, VGGNet, GoogleNet, ResNet, MobileNet,... Although having high precision, they are not suitable for mobile applications or embedded systems with low computing capacity.

MobileNet is a new architecture in the CNN network researched and developed by Google since 2017. MobileNet has small in size with fast processing speed, relatively high precision, easy to adjust to fit with user data. The first layer is the standard convolutional layer (used only once). After the first convolutional layer is the depth-wise separable convolution layers. All layers are followed by trigger functions - ReLU.

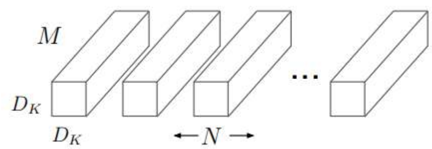

Fig. 5. Standard Convolution Filters

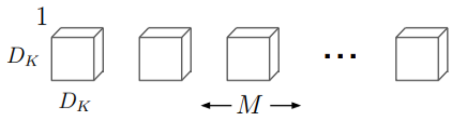

Fig. 6. Depth-wise Convolution Filters

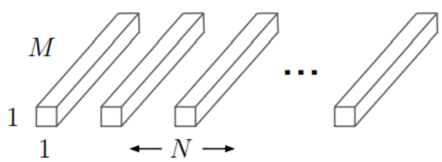

Fig. 7. Point-wise Convolution

The layers using Depth-wise separable convolution include two parts: 3x3 depth-wise convolution and 1x1 point-wise convolution. Depth-wise convolutions and point-wise convolutions are arranged alternately (Fig. 4). Depth convolutions perform convolution on each channel separately to filter the input channels to extract the characteristics in each channel. Point-wise convolutions combine the output channels of the depth convolution to create a new feature (Fig. 5-7), where *M* being the number of input channels, *N* being the number of output channels, *DF* being the spatial width and height of a square input feature map and *DK* being the spatial dimension of the kernel.

Depth-wise separable convolutions cost is calculated by:

$$DK.DK.M.DF.DF + M.N.DF.DF \qquad (1)$$

Standard convolutions have the computational cost of:

$$DK.DK.M.N.DF.DF \qquad (2)$$

Depth-wise convolution has the computational cost of:

$$DK.DK.M.DF.DF \qquad (3)$$

Using depth-separated convolution filters out the data and generates new features with a less amount of computation when compared to standard convolution.

In addition, MobileNet has two additional parameters, including width factor and resolution factor. The width factor allows to thin the network layer, meanwhile the resolution factor helps to change the input size of the image and reduce the internal representation of each layer (Table 1). Such division of convolution greatly reduces the computational volume and the number of parameters of the network. With this change, MobileNet can work well even on low configuration hardware such as mobile phones or embedded computers.

TABLE I. MOBILENET BODY ARCHITECTURE

Type / Stride	Filter Shape	Input Size
Conv / s2	$3 \times 3 \times 3 \times 32$	$224 \times 224 \times 3$
Conv dw / s1	$3 \times 3 \times 32$ dw	$112 \times 112 \times 32$
Conv / s1	$1 \times 1 \times 32 \times 64$	$112 \times 112 \times 32$
Conv dw / s2	$3 \times 3 \times 64$ dw	$112 \times 112 \times 64$
Conv / s1	$1 \times 1 \times 64 \times 128$	$56 \times 56 \times 64$
Conv dw / s1	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 128$	$56 \times 56 \times 128$
Conv dw / s2	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 256$	$28 \times 28 \times 128$
Conv dw / s1	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 256$	$28 \times 28 \times 256$
Conv dw / s2	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 512$	$14 \times 14 \times 256$
$5\times$ Conv dw / s1	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
$5\times$ Conv / s1	$1 \times 1 \times 512 \times 512$	$14 \times 14 \times 512$
Conv dw / s2	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
Conv / s1	$1 \times 1 \times 512 \times 1024$	$7 \times 7 \times 512$
Conv dw / s2	$3 \times 3 \times 1024$ dw	$7 \times 7 \times 1024$
Conv / s1	$1 \times 1 \times 1024 \times 1024$	$7 \times 7 \times 1024$
Avg Pool / s1	Pool 7×7	$7 \times 7 \times 1024$
FC / s1	1024×1000	$1 \times 1 \times 1024$
Softmax / s1	Classifier	$1 \times 1 \times 1000$

The object detection and recognition program is built on the data set that is collected from the real context by the authors. The training process is programmed in Python language, based on OpenCV and Tensorflow support libraries (Fig. 8).

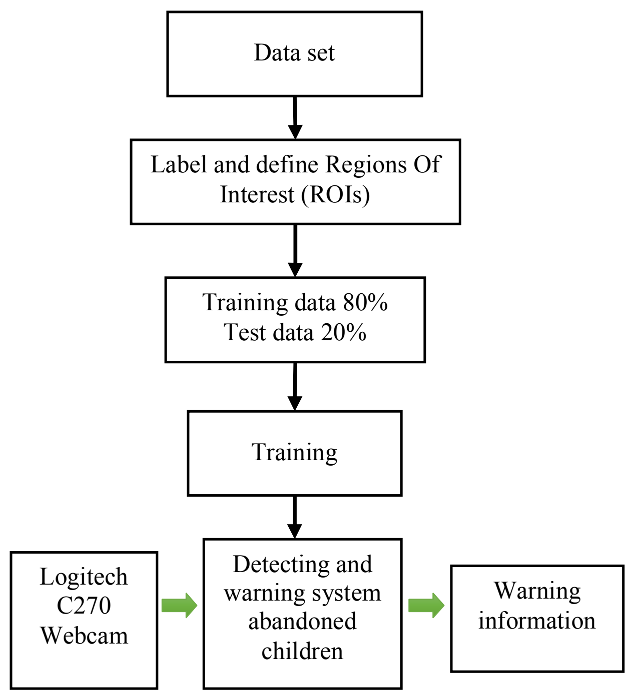

Fig. 8. Identification and detection algorithm

Data sets were collected with different spatial and lighting conditions. The camera used is a normal phone, so the data image quality is average. The dataset consists of 200 images with the resolution of 1365 x 2048p. The LabelImg tool is used to label objects with the identifier name "tre_em" (Fig. 9-10).

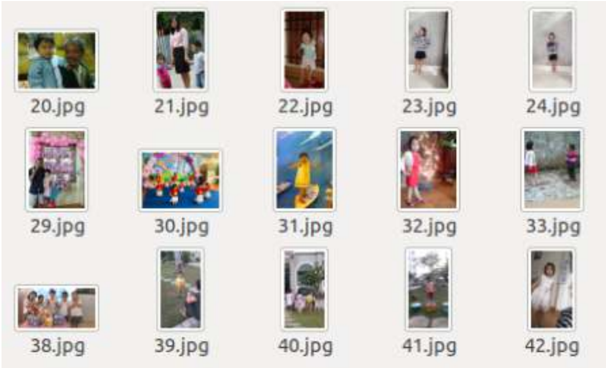

Fig. 9. Data set for labeling

Fig. 10. Labeling process

IV. RESULTS

A. Results from the child recognition program

CNN application utilizes ssdlite_mobilenet model to build the recognition program. The results obtained under the actual conditions are as shown in Fig. 11-13.

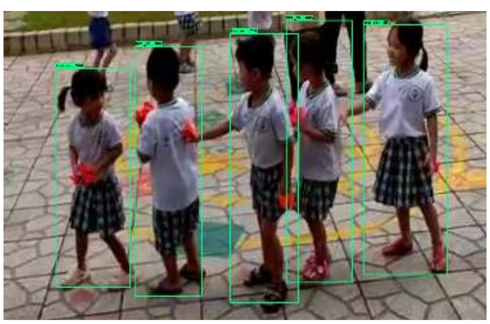

Fig. 11. Results of recognition after training

The device recognizes children in different real-life conditions such as outdoor space, in narrow rooms, and on cars with the accuracy being up to 90.6%. The total number of tested objects is 500.

The detected frame rate on desktop CPU Ryzen 5 1600x Ram 32GB, SSD120, Ubuntu OS is real time and on Rasberry Pi 4 2GB is 2.5 FPS with 720p HD resolution. However, the hardware configuration for network training requires a higher level. To achieve more efficiency and optimize training time, we should run the program on the GPU that is provided with optimal capabilities from hardware to software, mostly taking advantage of the CUDA core to process complex calculations and algorithms.

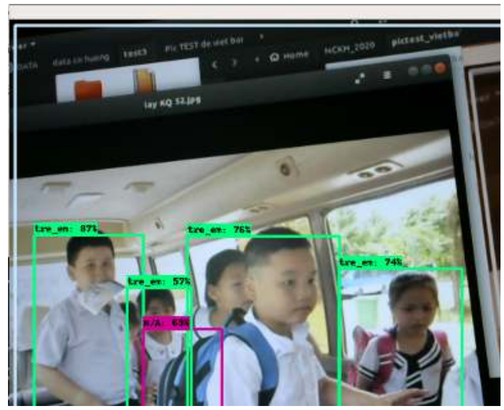

Fig. 12. Recognition results on 16-seat car

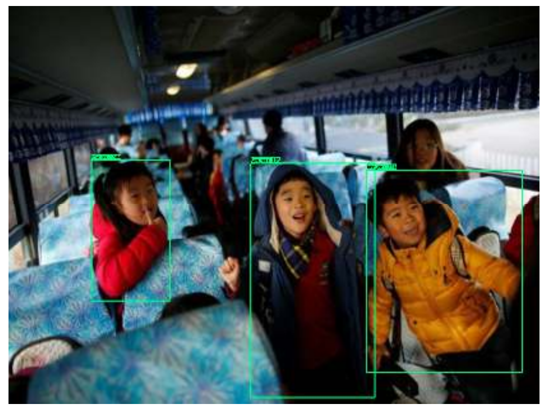

Fig. 13. Recognition results on 45-seat car

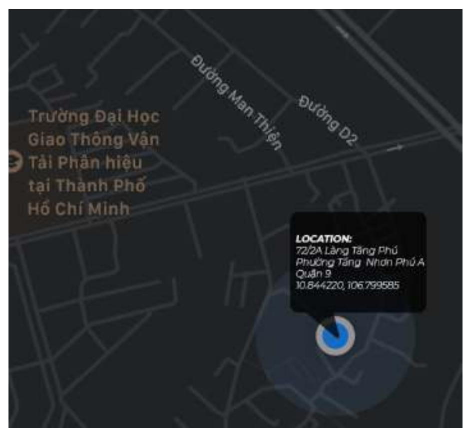

Fig. 14. Result of locating the vehicle's position

With the current accuracy of recognition, it is possible to list several options that can improve the accuracy of child identification using proposed neural network technology.

- *About the collection of image data for training:* Using specialized cameras to improve the quality of collected photos and videos. If the subjects are children in a car, the sample images must contain clear facial feature. This is more important because schools in some countries often use regular passenger cars (with high backrest of the seats) and the camera can only view on the child's face.
- *Increase the number of sample sets* to at least 1000 images to increase the accuracy of the identification process.

- *Use a powerful computer* to increase processing speed during the training period.

B. Results from real system installations on a 16-seat car

The dimensions of the device kit are 10cm long x 5.5cm wide x 6cm high. The weight including battery is 0.5 kg. This compact size is suitable for mounting on cars.

When the device is active, it can inform the user about the current location of the vehicle (Fig. 14). This study also used the information about the location at the end of the journey to initiate the program to warn the neglected child.

When the device detects that there is still a child in the car, the device will send text messages, make phone calls, and send emails to pre-set phone numbers or mailing addresses (Fig. 15-16).

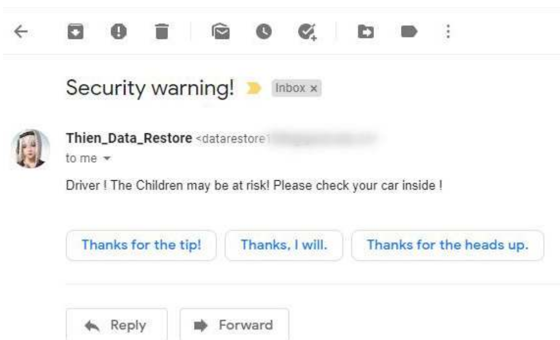

Fig. 15. Warning by email

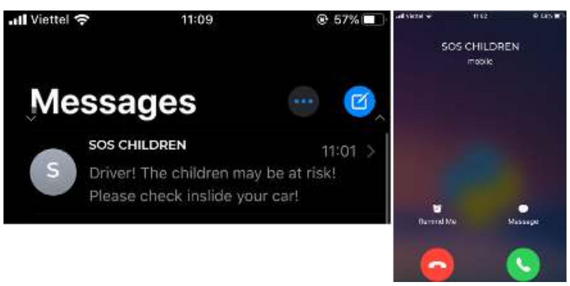

Fig. 16. Warning by message and phone call

V. CONCLUSIONS

This research developed and achieved successfully child identification program. It not only could be used to detect children on the bus but also can detect children - people in other different places. Besides, the identification program is setup into an embedded computer which conducts automatic supporting tasks on a 16-seat car by giving alerts via text message - phone call - email when children are abandoned on school buses.

To ensure the safety of children, authorities need to develop standards of school buses suitable for small stature. Besides, it is not recommended to use old cars or homemade cars for the school bus service. Modern methods and technologies should be encouraged to apply and standardize to maximize effectiveness of management and monitoring procedures.

REFERENCES

[1] J. Marot and S. Bourennane, "Raspberry Pi for image processing education," 25th European Signal Processing Conference (EUSIPCO), IEEE Publishing, 2017.

[2] Agrawal and S. Singhal, "Smart drip irrigation system using raspberry Pi and arduino." Proceedings of International Conference on Computing, Communication & Automation (ICCCA), 2015.

[3] W. Rawat and Z. Wang, "Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review", Neural Computation, June 2017.

[4] V. Plemakova, "Vehicle Detection Based on Convolutional Neural Networks," University of Tartu, Master's Thesis, 2018.

[5] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Advances in neural information processing systems, January 2012.

[6] C. M. Bautista, C. A. Dy, M. I. Mañalac, R. A. Orbe, and M. Cordel, "Convolutional neural network for vehicle detection in low resolution traffic videos," IEEE Region 10 Symposium (TENSYMP), pp. 277–281,2016.

[7] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, December 2015.

[8] Z. Q. Zhao, P. Zheng, S.T. Xu, and X. Wu, "Object detection with deep learning: A review, arXiv:1807.05511, 2018.

[9] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv:1704.04861, 17 Apr 2017.

[10] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks." arXiv:1506.01497, 2015.