Pet cat behavior recognition based on YOLO model

Hsiu-Te Hung
Department of Information Management Chaoyang Univerity of Technology
Taichung, Taiwan
s10814620@cyut.edu.tw

Rung-Ching Chen
Department of Information Management Chaoyang Univerity of Technology
Taichung, Taiwan
crching@cyut.edu.tw

***Abstract*—With the progress of the times and the rapid development of science and technology, machine learning and artificial intelligence are increasingly used in transportation, logistics, and homes. In terms of pets, pet monitoring has also become very popular in recent years. Therefore, this study for the real-time monitoring of home pets, using the raspberry pie as a monitoring system, proposed a raspberry pi-based YOLOv3-Tiny identification system YOLOv3-Tiny method with rapid detection and better boundary frame prediction. One thousand one hundred twenty-eight pictures of cats' movements were collected in the room for marking and training. Finally, according to the input image categories, the results of six cat action categories were output. They were sleeping, eating, sitting down, walking, going to the toilet, and search on a trash can. The average accuracy was 98%. Through image recognition, the images were sent to the user's mobile phone app and computer. When the cat goes to the toilet for too long or flips through the trash can, the system will instantly send a message to the owner's mobile phone to achieve an instant preventive remote pet monitoring system.**

Keywords: Raspberry Pi, Image recognition, YOLOv3-Tiny

I. INTRODUCTION

Nowadays, most people keep pets, and dogs' ratio to cats is the highest among pets in Taiwan. In recent years, more and more people have held pets. According to the Taiwan Agricultural Commission [1], the population under 15 years of age has fallen at a rate of 4% every year. In contrast, the number of cats and dogs increased from 2.789 million in 2011 to 2.51 million in 2017. The growth rate is quite fast. It is worth noting that in 2017, although the number of people who raised dogs was still higher than that of people who raised cats, from the growth data in recent years, cats grew by 27%, while dogs only grew by 2%. The estimation is based on the number of dogs and cats. The average annual rate of increase and decrease in the child population is from 2011 to 2017. TrendSight [2] company predicts that the number of dogs and cats in Taiwan will exceed the number of children under 15 for the first time in the second half of 2020. Based on the above analysis, we can know that more and more people are keeping pets, so the business opportunities of the pet industry will gradually rise.

Most pet owners will keep their pets alone at home for a long time, and pets may eat by mistake or bite wires. These behaviors may cause pets to be injured, so observe animals is also a part that cannot be ignored. There are many types of pet monitors on the market, such as the PAWBO [3] interactive pet camera. The owner can connect to the pet monitor by applying the mobile phone, use the dot game's function to interact with the pet, or use the ringtone to call the pet to eat. Most pet monitors include voice interaction and automatic feeder functions. Although you can see your pet in real-time through the image, you cannot detect its current behavior.

This research collects photos of cats' various behaviors at home and uses Raspberry Pi and YOLO deep learning to analyze the accuracy of the cat's behavior. Through image recognition and deep learning, when the cat search on a trash can or goes to the toilet for too long, the message is transmitted to the owner's mobile phone. Finally, a pet monitoring system for real-time prevention is realized. The rest of this article is organized as follows. Section 2 introduces related work, and Section 3 describes the system architecture and YOLOV3-Tiny algorithm. In Section 4, the relevant experimental settings are pointed out, and the experimental results are discussed. Finally, the summary and future of this article are introduced.

II. RELATED WORK

A. Pet Care

Although more and more families keep pets nowadays, many owners still neglect the health of their pets. According to the Animal Protection Department [4], the most common cause of death in cats is kidney failure. The leading cause of kidney failure in cats is chronic water shortage and urinary tract infection. If a cat has a urinary tract infection, the most apparent symptoms are frequent visits to the toilet and no urine discharge. If there are too many items in the house, it will cause the pet to have pica, which indirectly increases the cat's risk of kidney failure. How to be effective The care of pets is critical.

B. YOLO

YOLO was first mentioned in 2015 by Joseph Redmon in the paper You Only Look Once: Unified, Real-Time Object Detection[5], YOLO is a target detection system based on a single neural network. Unlike other algorithms, YOLO divides the picture into multiple units and predicts each unit's bounding box coordinates and the probability of its category. Afterward, Joseph Redmon et al. improved YOLO and proposed YOLOv2 and YOLOv3. The full text of YOLOv2 is YOLO 9000: Better, Faster, Stronger[6]. In the paper, it is mentioned that the model can detect 9000 kinds of objects. A new feature extractor called Darknet-19 is used to improve the detection speed and accuracy of the model. To achieve better classification results, YOLOv3 has made improvements based on Darknet-19 proposed by YOLOv2 and adopted a more personal convolutional layer neural network Darknet-53, which is mainly composed of 1X1 and 3X3 convolutional layers, a total of 53 layers. As the number of network layers continues to deepen, the ResNet[7] structure is used to solve the gradient problem, significantly reducing the difficulty of training deep networks and improving the accuracy more obviously.

Because YOLO has good detection speed and detection accuracy in object detection, many scholars use the YOLO model for image recognition research. Cuixiao Liang[8] et al. used the YOLOv3 model to detect litchi fruits in the natural environment. The study used cross-validated the brightness level and distance range of night searchlights to determine the best combination of brightness and lighting for litchi detection at night. The final detection results also have good accuracy.

2020 International Symposium on Computer, Consumer and Control (IS3C) | 978-1-7281-9362-5/20/$31.00 ©2020 IEEE | DOI: 10.1109/IS3C50286.2020.00107

978-1-7281-9362-5/20/$31.00 ©2020 IEEE
DOI 10.1109/IS3C50286.2020.00107

Fan Wu [9] et al. improved YOLOv3, using DenseNet[10] is advantages in model parameters to replace the backbone network of YOLOv3 for feature extraction, alleviating the problem of inaccurate detection and overlapping bounding boxes in the original network, and forming YOLO-Densebackbone convolutional neural network. The traditional YOLOv3 reference, the improved algorithm detection accuracy, increased by 2.44%.

III. METHODS

A. System structure

In this paper, considering that training the YOLOv3-Tiny model on the Raspberry Pi will cause a burden and take too long training time, we first train the YOLOv3-Tiny model on the computer and then put the model into the Raspberry Pi, test the collected cat pictures. The system architecture is shown in Figure1. Use the model built on the Raspberry Pi for image recognition, use the IFTTT (If this, then that[11]) network service platform to send the detected behavior images to the mobile phone. This research also uses VNC (Virtual Network Computing[12]) Remotely monitor images, set the Raspberry Pi as the server, and the computer and mobile phone as the client. When a behavior is detected, it will automatically send a message to the mobile phone immediately. You can also view the Raspberry Pi image through the computer and mobile phone.

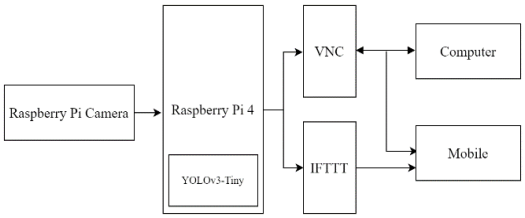


Fig. 1. System architecture

B. YOLOv3-Tiny

In this research, considering the Raspberry Pi hardware configuration, YOLOv3-Tiny is used as the object detection model. YOLOv3-Tiny is simply a simplified version of YOLOv3. Although it is not as good as YOLOv3 in detection speed and accuracy, the results of the research and experiment in this article show good performance. Compared to YOLOv3, the Tiny version compresses the network a lot, omits some feature layers, and does not use the residual layer, and only uses two different scale output layers of 13*13 and 26*26. Because the backbone network of YOLOv3-Tiny is relatively shallow, higher-level semantic features cannot be extracted. However, the advantage of YOLOv3-Tiny is that the system is simple, and the computational burden is small, which is convenient for use on mobile devices such as Raspberry Pi. Run-on the YOLOv3-Tiny network, as shown in Figure2.

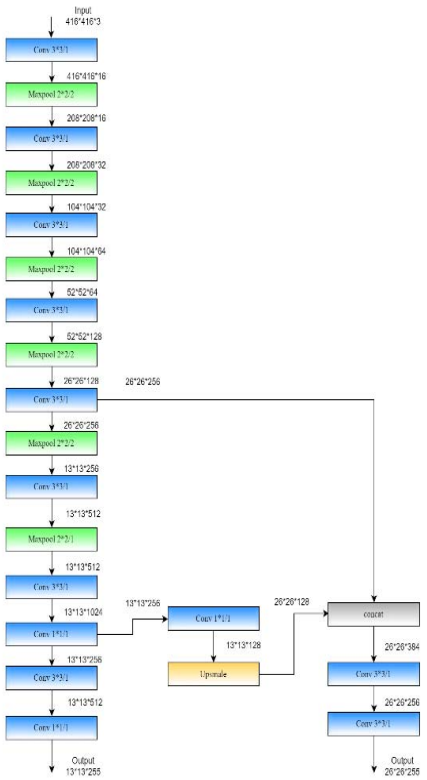


Fig. 2. YOLOv3-Tiny model

In the frame prediction model, YOLOv3 continues the practice of YOLOv2. t_x, t_y, t_w, and t_h are the predicted output of the model, t_x and t_y are the predicted coordinate offset values; t_w and t_h are scale scaling; b_x, b_y, b_w, and b_h are the center coordinates of the final predicted bounding box, width, and height.

$$b_x = \sigma(t_x) + C_x \tag{2}$$

$$b_y = \sigma(t_y) + C_y \tag{3}$$

$$b_w = p_w e^{t_w} \tag{4}$$

$$b_h = p_h e^{t_h} \tag{5}$$

Where σ is the sigmoid function, c_x and c_y are the coordinate offsets of the cells, p_w and p_h are the preset anchor box's side lengths.

IV. TRAINING AND RESULTS

A. Data set and Experimental environment

The data set was taken with mobile phones in the living room and room, and the collected pictures were uniformly cropped to 416*416 pixels. The picture's angle is randomly rotated within plus or minus 20 degrees to make the data set more accurate, and the shots are taken in different time periods to make the data set more representative in training. Finally, a total of 1128 photos were collected. During the training process, we used the bounding box labeling tool (BBox Label Tool[13]) to mark the detection target's coordinate position manually. The tool will return the four points of the coordinate and the category label. The data set consists of 6 categories, walking, eating, sleeping, sitting, going to the toilet, and

search on a trash can. The training image and test image of each category are 7:3.

The hardware equipment of this research uses Raspberry Pi 4 Model B V1.2, the memory is 2G LPDDR4-2400 SDRAM, the storage space is microSD 64G, and the camera uses Raspberry Pi Camera Module V2.1, and its specification is 8 million pixels. Raspberry Pi and camera are shown in Figure 3.

Fig. 3. Raspberry Pi 4 and Raspberry Pi Camera

B. *Experimental Results*

After training the model, we tested 432 images in a test set of 6 categories. The loss value is part of the loss function. The smaller the amount is, the better. A total of 12,000 training steps were used to analyze the training process better. In the first 1200 training steps, the loss function decreased rapidly. After 6000 training steps, the loss function's value gradually stabilized, and finally less than 0.03, the mAP is as high as 98.1%, as shown in Figure 4. The test results are listed in Table II. It can be seen from the above tests that the accuracy of the dataset detection is as high as 97% or more; the sleep category and go to the toilet is as high as 99%. The final test result is excellent. The test result is shown in Figure 5. However, in the test, it was found that the undetected image was not detected because the chair blocked the cat's body, and subsequent experiments can continue to strengthen the training picture, as shown in Figure 6. When going to the toilet for more than 30 seconds, and the cat search on a trash can, a message will be sent to the owner's mobile phone, as shown in Figure 6. The owner can remotely view the cat's current behavior through a mobile phone or computer, as shown in Figure 7.

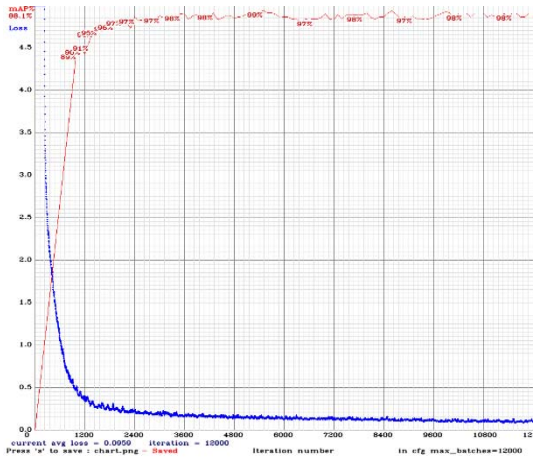
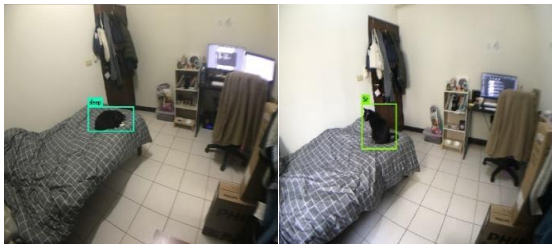


Fig. 4. Average loss and mAP

TABLE I. DETECTION ACCURACY

Category	Training Data	Testing Data	Accuracy
Moving	172	74	98.78%
Eating	126	54	98.33%
Sleeping	162	70	99.57%
Sitting	140	70	97.81%
Search on a trash can	182	78	96.51%
go to the toilet	196	86	98.24%

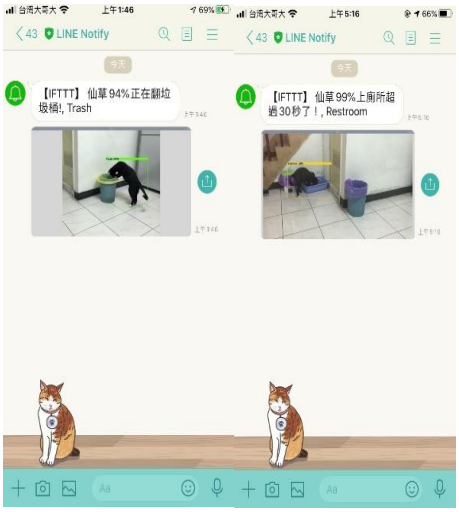
(a) (b)



(c) (d)

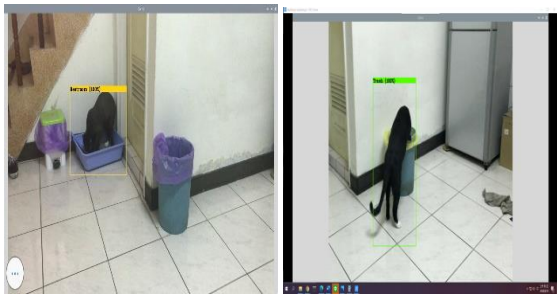
(e) (f)

Fig. 5. Detection results: (a) moving, (b) eating, (c) sleeping, (d) sitting, (e) search on a trash can, (f) go to the toilet.



(a) (b)

Fig. 6. Detection behavior is displayed on the mobile: (a) cat searching on trash, (b) cat go to the toilet for more than 30 seconds.



(a) (b)

Fig. 7. View on device: (a)view on mobile, (b)view on a computer

V. Conclusions and Future

This research proposes a pet behavior detection system based on Raspberry Pi. When pets go to the toilet for too long or search on a trash can, the system will immediately send a message to the mobile phone. Although Yolov3-Tiny is not very good in detection speed, it is still quite a high accuracy. The test results show that the overall accuracy of the six categories is as high as 98%. In the future, I hope to continue to add categories, such as scratching, licking hair, running, etc., to make the entire system more complete.

Acknowledgment

This paper is supported by the Ministry of Science and Technology, Taiwan. The Nos are MOST-107-2221-E-324 -018 -MY2 and MOST-109-2622-E-324 -004, Taiwan. This research is also partially sponsored by Chaoyang University of Technology (CYUT) and Higher Education Sprout Project, Ministry of Education (MOE), Taiwan, under the project name: "The R&D and the cultivation of talent for health-enhancement products."

References

[1] "Taiwan Agricultural Commission," https://www.coa.gov.tw, accessed: 2019/12/15.

[2] "TrendSight," http://www.trendsightinc.com, accessed: 2019/12/16.

[3] "PAWBO," https://lokaloka.tw/pawbo20170405, accessed: 2019/12/16.

[4] "Taipei City Animal Protection Office," https://www.tcapo.gov.taipei/, accessed:2020/ 7/29.

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," arXiv preprint arXiv:1506.02640, 2015.

[6] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger," arXiv preprint arXiv:1612.08242, 2016.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," arXiv preprint arXiv:1512.03385, 2015.

[8] Cuixiao Liang, Juntao Xiong, Zhenhui Zheng, Zhuo Zhong, Zhonghang Li, Shumian Chen, and Zhengang Yang, "A visual detection method for nighttime litchi fruits and fruiting stems," Computer and Electronics in Agriculture, vol. 167 February 2020.

[9] Fan Wu, Guoqing Jin, Mingyu Gao, Zhiwei He, and Yuxiang Yang, "Helmet Detection Based On Improved YOLO V3 Deep Model," in 2019 IEEE 16th International Conference on Networking, Sensing and Control, 2019.

[10] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," arXiv preprint arXiv:1311.2524, 2013.

[11] "IFTTT," http://www.ifttt.com/, accessed: 2020/7/5.

[12] "Virtual Network Computing," http://www.realvnc.com/, accessed: 2020/7/5.

[13] "Bbox label tool," https://github.com/puzzledqs/BBox-Label-Tool, accessed: 2019/12/23.