Proceeding of the IEEE
International Conference on Robotics and Biomimetics
Dali, China, December 2019

SuperDock: A Deep Learning-Based Automated Floating Trash Monitoring System

Guanchong Niu, Jie Li, Sheng Guo, Man-On Pun†
The Chinese University of Hong Kong, Shenzhen
Guangdong, China

Leo Hou and Lujian Yang
StrawBerry Innovation, Ltd
Guangdong, China

***Abstract*— Floating trash in the river presents a major environmental problem that endangers the lives of river inhabitants. However, conventional trash monitoring relies on labor-intensive manual inspection by dispatching inspectors to the filed, which is highly cost-ineffective. Despite that Unmanned Aerial Vehicles (UAVs) have been widely proposed for many real-time monitoring applications, limited power supply remains the most challenging bottleneck for such UAV-assisted monitoring applications. In this work, we propose an automated river trash monitoring system called SuperDock. SuperDock consists of a remote processing unit, a docking station and a UAV. SuperDock lands the UAV precisely onto the docking station followed by performing automated battery replacement. As a result, the UAV can continue the monitoring task after the battery replacement. Furthermore, SuperDock includes a deep learning-assisted river trash detection module based on YOLOv3 that runs much faster than the conventional Convolution Neural Networks (CNN). In addition, a data set has been generated for training and testing the deep learning network, specifically for the floating trash detection application. Finally, SuperDock enables the UAV to communicate wirelessly with a remote computer in a real-time manner. Experimental and simulation results show that SuperDock is highly effective in monitoring the floating trash in the river.**

***Index Terms*— UAV, SuperDock, Floating Trash Monitoring, YOLOv3, Transfer Learning.**

I. Introduction

Clean and vibrant rivers are indispensable for the environmental sustainability. However, more and more wastes and pollutants have been discarded to rivers. In particular, floating trash in the river has polluted the river and subsequently, presented a major threat to the lives of river inhabitants. It has been reported that at least one billion birds, one million mammals and inestimable numbers of fish are killed by floating trash every year [1]. Thus, it is an urgent task to design a low-cost floating trash monitoring system.

In the literature, both manual inspection and USV (Unmanned Surface Vehicle)-assisted automated inspection have been proposed to monitor the floating trash in the river. Despite that the manual inspection produces more accurate detection, it is both labor-intensive and time-consuming. In contrast, USV-assisted automated inspection detects trash by running vessels on the river in a more cost-effective manner [2]. However, since the visual detection system is mounted on the USV, the coverage of the system is severely limited. As a result, USV-assisted monitoring systems are not suitable for large-area monitoring. In addition, such USV-assisted systems are highly susceptible to weather conditions, e.g. rain or wind. Finally, it is highly desirable for the monitoring system to be able to operate on a daily basis in both urban and rural areas. To address the aforementioned challenges, this work proposes an Unmanned Aerial Vehicles (UAV)-assisted monitoring system by taking advantages of the flexibility, the low cost and large view angle of UAV.

Fig. 1: SuperDock.

Recently, UAVs have been proposed to deliver parcels and perform surveillance. For instance, [3] has first proposed to utilize UAVs to monitor the floating trash in the river. However, the system developed in [3] is handicapped by the UAV flight time. Since UAVs are battery power-limited, the system in [3] can fly for less than 30 minutes, which renders the system unsuitable for practical monitoring tasks. To cope with this drawback, [4] proposes a joint problem of flight mission planning and recharging optimization for UAVs, using the calibrated power consumption model of a UAV,

This work was supported, in part, by the Shenzhen Science and Technology Innovation Committee under Grant No. ZDSYS20170725140921348, the Robotic Discipline Development Fund (2016-1418) from the Shenzhen Government and the National Natural Science Foundation of China under Grant No. 61731018.

† Corresponding author, email: SimonPun@cuhk.edu.cn.

978-1-7281-6321-5/19//$31.00 ©2019 IEEE

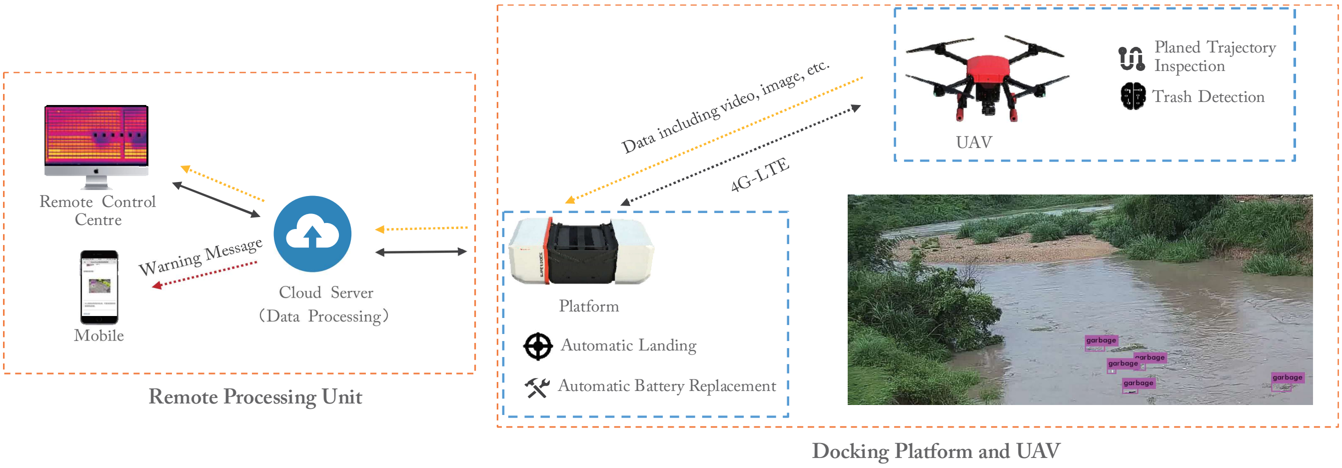

Fig. 2: Block diagram of proposed automated inspection system: SuperDock.

with an objective to complete a tour mission for a set of sites of interest in the shortest time. However, the UAV flight time in [4] remains to be battery power-limited. Furthermore, [5] develops a prototype to automatically replace the battery by using a battery case and a carriage to assist battery swapping with the battery carriage being attached to the bottom of the UAV.

In this work, we propose a UAV-assisted trash monitoring system with novel hardware architecture and detection algorithm design. It has three distinct advantages in hardware architecture as compared to the UAV-assisted systems proposed in the literature. First, a novel docking station is designed and implemented for battery replacement. Second, the proposed system is specifically designed for floating trash monitoring by taking into account many practical design issues such as wireless communications, waterproof and system stability. Finally, empowered by an intelligent chip named Jeston TX2 module (TX2), the system is able to perform object detection in a real-time manner. In addition to the novel hardware architecture, this work also implements a novel deep learning network specifically designed for floating trash detection. More specifically, this work first develops a training and testing data set specifically designed for trash detection purposes. Then, a modified YOLOv3 [6] network is established for trash classification and localization. Compared to conventional Convolutional Neural Networks (CNN) algorithms such as YOLO [7], [8], [9] and Regions with CNN features (R-CNN) [10], [11], [12], YOLOv3 has been shown to outperform Faster R-CNN in sensitivity and processing time in the context of car detection from aerial images [6].

The main contributions of this paper are summarized as follows:

- An automated trash monitoring system named SuperDock is proposed and implemented. The system comprises a remote processing unit, a UAV and a docking station. The low-cost docking station is designed to reliably perform automated battery replacement. In addition, the docking station can protect the UAV from bad weather and perform real-time wireless communications using its built-in modules HUAWEI ME909s-821 and DJI Lightbrige;
- The proposed UAV is empowered by TX2 to achieve fast object detection, remote flight control and various data processing tasks. As a result, the UAV can autonomously monitor the river without daily manual supervision. Furthermore, the UAV flight mission can be scheduled remotely;
- To apply deep learning in trash detection and localization, a data set specifically designed for floating trash detection is created. By performing the YOLOv3-assisted object detection algorithm, the UAV can accurately detect floating trash in a real-time manner.

II. Design of SuperDock

In this section, we introduce the system design. As shown in Fig. 2, the system contains three parts, namely a Remote Processing Unit, a docking station and a UAV. The docking station of size 1.7×1.7 m^2 is equipped with a position measurement system while the UAV bottom size is measured as 0.6×0.6 m^2. The remote processing unit can process the collected data and communicate with users. In the following, the description of the system will be elaborated.

A. *Visual guidance-assisted UAV autonomous landing*

UAV landing is generally based on Real-time kinematic (RTK) centimeter high-precision positioning technology. However, during the landing process, the UAV will

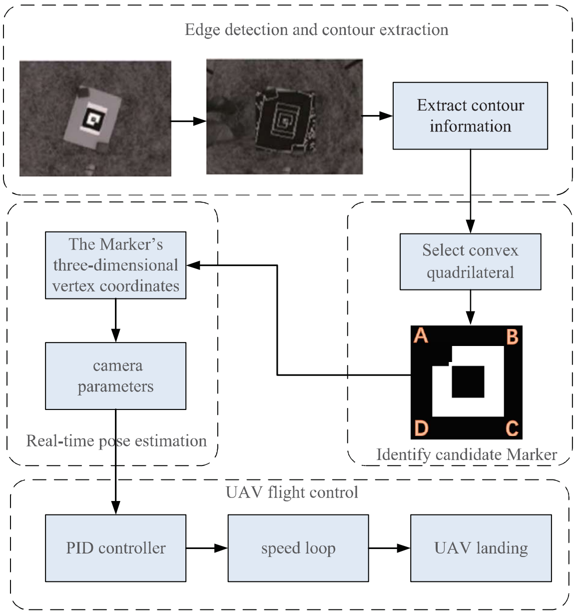

Fig. 3: Flowchart for autonomous landing.

be affected by both ground effect and lateral gust, which may cause poor landing accuracy. Therefore, in order to ensure that the UAV can accurately dock on the charging station, we combine RTK and visual guidance to achieve autonomous landing. Specifically, when the drone receives the flight control landing command, the UAV automatically flies to the designated area (i.e., airspace near the apron) through the RTK real-time feedback of the UAV position and the known apron position. Next, the UAV uses visual guidance to achieve precise landing. As the flowchart shown in Fig. 3, the specific steps are as follows:

1) *Edge detection and contour extraction*: After the UAV reaches the airspace near the apron, it captures the real-time image with the on-board camera at a fixed rate (5 Hz) and transforms it into a grayscale image. The Canny operator or adaptive threshold method is used to obtain the binarized edge. The distribution map is filtered by the corrosion expansion algorithm before the contour information of the image is extracted.
2) *Identify candidate Marker*: Marker is the exact identifier of the docking position of the UAV in the apron. In order to get accurate contour information, we need to remove the contour inside the Marker and the excessive pixels in the contour. Subsequently, the outer contour of the convex quadrilateral is selected as the candidate parking space Marker, and the coordinates of the four vertices in the image are obtained.
3) *Real-time pose estimation*: The spatial position and size of the Marker is the known information, i.e., the Markers three-dimensional space vertex coordinates are known. According to the correspondence between a series of 3D points and 2D points of the image, combined with the camera parameters, the position, and pose of the UAV relative to the Marker coordinate system can be estimated.
4) *UAV flight control*: In order to achieve stable landing, the UAV operates in different control modes according to the landing conditions. After the UAV descends to a certain altitude, the UAV is designed to check the image positioning effect. If the image shows acceptable quality, then the UAV will activate the proportional-integral-differential (PID) controller to control the speed loop and adjust its speed. In contrast, if UAV's altitude is too low, the PID controller is used to control the acceleration loop to ensure that the UAV lands on the apron smoothly and accurately.

B. Battery Replacement Mechanism

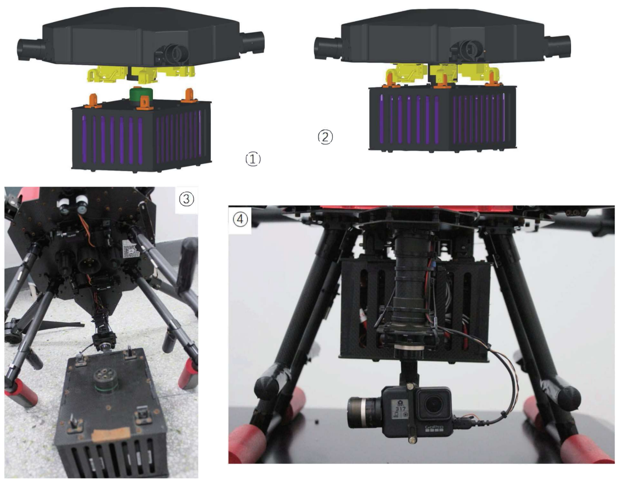

Fig. 4: The mechanism for automated battery replacement: 1) The battery is detached from UAV. 2) The battery is attached to UAV. 3) The battery replacement part. 4) UAV with battery.

As shown in Fig. 4, the battery can be attached to the UAV through the four screw holders on the top. It is worth noting that the directions of screw holders are designed to be different to enhance stability and prevent trivial installation mistakes. On the bottom of the UAV, the battery replacement mechanism comprises four servomotors to release or grasp the battery. The connector is in the center of battery with soft rubber which can protect the connector from a sharp collision.

While RTK can help facilitate the automated landing, the battery replacement mechanism requires accurate UAV positing. Thus, the UAV position has to be corrected after landing. Inspired by the mechanism in [5], we propose to use the servomotor arms to adjust the UAV position. With the adjustment, the UAV is placed at the desired position for battery replacement. During the replacement, the old battery is firstly released and recharged on the station before a fully charged battery is pushed to the bottom of UAV and grasped by the servomotors as shown in Fig. 5. To

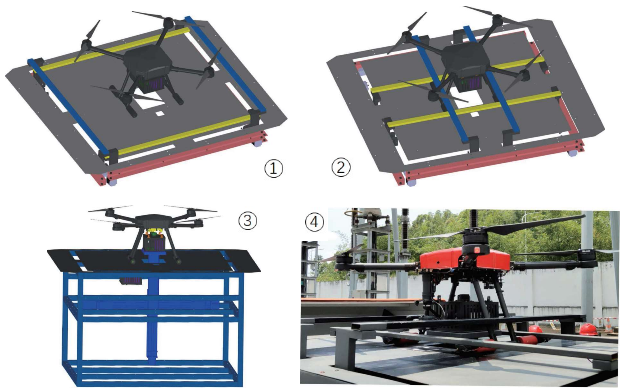

Fig. 5: The procedure for automated landing battery replacement: 1) The UAV lands on the station. 2) The position is adjusted by the servomotor arms. 3) The battery is pushed to the bottom of UAV. 4) The realization for automated battery replacement.

achieve continuous flight missions, multiple batteries should be stored and charged in the docking station. In our system, the charging time is about 120 minutes for one battery while the flight time is about 30 minutes for each fully charged battery. This amounts to at least four batteries required to maintain the UAV on a continuous flight mission. Using less than four batteries is also possible if non-stop flight missions are not required.

C. Weather Condition Monitoring

To enable flight missions in all weather conditions, a waterproof design has been implemented in SuperDock to protect the station and UAV from raining and insolation. The cover on the docking station can be closed while the UAV is out on the flight mission. Before the UAV takes off, the weather conditions are detected by the sensors on the docking station as shown in Fig. 6. Alternatively, the weather forecast can be also obtained from the cloud server while the system decides whether the weather is suitable for taking off or not. We have formulated an empirical formula for automated decision:

$$T = (1 - P - (W + 1)/18) \times B, \qquad (1)$$

where P and W are respectively the probability of precipitation and wind scale ranging from 0 to 17. Furthermore, B is a binary parameter: B is one if the UAV battery is sufficient for take-off; Otherwise, B is zero. Finally, T is the indicator to decide whether the weather conditions are safe for the UAV to take off. The threshold of T depends on the practical environment.

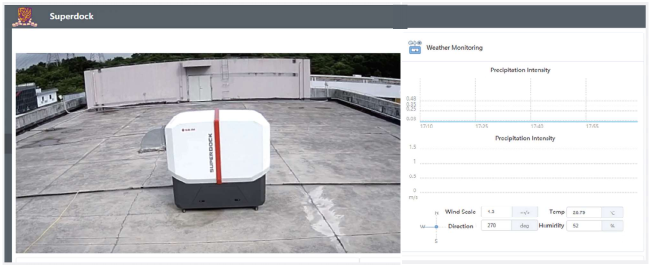

Fig. 6: The monitoring for weather condition.

D. Communications System

The station can connect to a cloud server via a wired or WiFi network. More specifically, the UAV is equipped with a HUAWEI ME909s-821 and a DJI Lightbridge. ME909s-821 is an advanced GSM module for 4G-LTE wireless communications while DJI Lightbridge is a long-range video downlink connection capable of transmitting video. The devices are utilized in different cases as follows:

1) *4G-LTE communications*: In general, the UAV communicates with the docking station via ME909s-821 if the 4G network is available. This feature is particularly useful for countries that are well covered by the 4G network. For instance, the coverage of the 4G network in China is about 95%. Thanks to the line of sight (LoS) environment of UAV communications, the achievable communication distance can be much larger than the conventional ground users. By taking advantages of these hardware and wireless channels, the UAV can communicate with the docking station in long distance with latency about 200 ms.
2) *Lightbridge communications*: For some rural regions where the 4G network is not available, the UAV has to transmit data via Lightbridge. Lightbridge consists of an Air System and a Ground System. It integrates the remote controller module into Ground System that comes with a number of aircraft and gimbal controls as well as some customizable buttons. The transmission distance of Lightbridge is up to 5 km using the unlicensed 2.4G frequency band.

E. UAV Controller

TX2 is the fastest, most-efficient embedded AI computing device available on the market. Its low power consumption (7.5 watt) is particularly suitable for UAV computing. It is built around an NVIDIA Pascal^{TM}-family GPU and loaded with 8 GB of memory and 59.7 GB/s of memory bandwidth.

By sending commands to Pixhawk, we can intelligently control the UAV flight. The images can also be processed by TX2. To simplify the testing, we use Microsoft AirSim (Aerial Informatics and Robotics Simulation) to perform experiments as shown in Fig. 7.

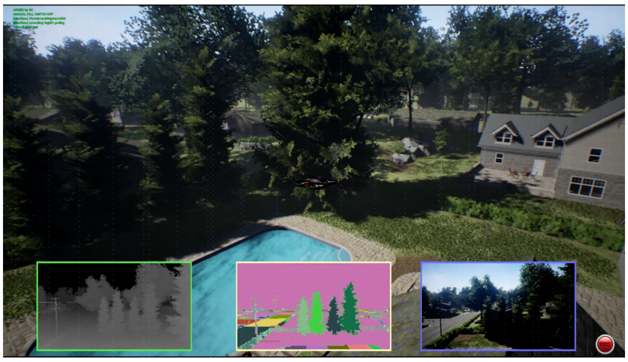

Fig. 7: AirSim simulation.

III. YOLOv3-Assisted Real-time Trash Detection

In this section, we introduce the trash detection algorithm based on YOLOv3.

A. Synergy of Deep Learning and Transfer Learning

Transfer learning enables a system to recognize and apply knowledge and skills learned in previous domains or tasks to new domains and/or tasks. To apply transfer learning to a specific problem, we have to first address the following three questions: (1) What to transfer; (2) How to transfer and (3) When to transfer.

In this paper, we use deep learning to pre-train the network before performing transfer learning to fine-tune the network. More specifically, we first initialize the network with the pre-training model from ImageNet before fine-tuning the network with our original floating trash data set via transfer learning. This combination is shown to generate good results even with a smaller fine-tuning data set.

B. Description of the Data set

To perform the experimental results, we use a training data set containing 80 images and 235 instances of labeled trash. The test data set contains 20 images and 77 instances. The data is collected from the open data station of Shenzhen, China [13]. One of the challenges for creating this data set is the aquatic plant in the river. We have carefully distinguished the floating trash from the aquatic plant in generating the data set.

Due to the limited size of our floating trash data set, it is highly likely that the network may be trained to over-fit such a small data set. To cope with this problem, the following image processing transformation techniques are proposed to generate more data samples from the small data set.

Fig. 8: Data set labelling.

- Changing the image scale, including zooming and adding black borders to the image edges;
- Changing the image contrast;
- Performing image rotation, left-right flip and mirror transformation.

Using the transformation techniques above, the size of the data set can be increased by a factor of more than five. As a result, The enlarged data set can effectively prevent the over-fitting problem caused by limited data set. Fig. 8 shows the images generated in our floating trash data set.

C. The Improved YOLOv3

YOLOv3 is an improved object detection algorithm developed based on YOLOv1 [7] and YOLOv2 [8]. In contrast to its previous generations, YOLOv3 uses a logistic classifier to obtain the likelihood of the detected object, instead of the conventional softmax function [9]. Compared to the Faster R-CNN network, YOLOv3 can process a typical image with improved accuracy using half of the processing time.

The loss function of YOLOv3 mainly includes the loss of x, y, the loss of w, h, the loss of confidence and the loss of classification [9] where x and y are the center coordinates of the detected object while w and h are the weight and height of the bounding box, respectively. In this paper, we propose a modified loss function for the length and width of the rectangle by normalizing the original loss function. The modified width and height loss function is given as follows:

$$\text{Loss}_{wh} = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{obj} [(\frac{w_i - \hat{w}_i}{w_i})^2 + (\frac{h_i - \hat{h}_i}{h_i})^2]. \tag{2}$$

The modified loss function shows good scale consistency in the directions of length and width with better function expression properties and lower computational complexity.

D. Performance Evaluation

Fig. 9: Trash detection by YOLOv3.

The training is accomplished on a local personal computer. The batch size is set to 64 and the input image is subdivided to 16×16 grids. The training momentum used in the stochastic gradient descent is set to 0.9 while the anchors that overlap with the ground-truth object by less than a threshold value (0.7) are ignored. Finally, we have trained for $100,000$ steps with a learning rate of 0.0001 in this experiment. The configurations of the local computer used in this research are as follows:

- CPU: Intel Core i7-8700K with six cores
- Graphic card: Nvidia GTX 1060, 6G GDDR5
- RAM: 16GB RAM
- Operating System: 64-bit Ubuntu 16.04

As shown in Fig. 9, the floating trash has been successfully detected and labeled. We have also compared the algorithm with Faster R-CNN. The detailed testing results are shown in Table I.

TABLE I: Performance Evaluation.

Measure	Faster R-CNN	YOLOv3	Improved YOLOv3
Accuracy	76.4%	82.6%	81.2%
Average Processing time	1.392s	0.049s	0.038s

From Table I, it shows that the proposed improved YOLOv3 has good performance on detecting the floating trash. The detection speed is also much faster than Faster R-CNN. Inspection of the mis-detection samples has revealed that some aquatic plants were mis-recognized as floating trash. Furthermore, the wave and shadow in the river also negatively impacted on the detection accuracy.

IV. Conclusion

In this paper, a UAV-assisted automated monitoring system named SuperDock has been designed and implemented to detect floating trash in the river. The intelligent system comprises three parts, namely a remote processing unit, a docking station and a UAV. In particular, the docking station can perform the automated battery replacement for UAV, which enables the UAV to perform continuous flight missions. In addition, the UAV is empowered by faster processing and communication modules. As a result, the UAV can achieve fast object detection and communicate with the docking station or the remote processing unit even on its flight mission. Finally, the UAV is endowed with a modified YOLOv3 network that has been trained with deep learning and transfer learning. Experimental and simulation results have confirmed the effectiveness of SuperDock in monitoring the floating trash in the river. A demo video showing the prototype can be found online in [14].

References

[1] J. Fu, B. Mai, G. Sheng, G. Zhang, X. Wang, P. Peng, X. Xiao, R. Ran, F. Cheng, X. Peng, Z. Wang, and U. W. Tang, "Persistent organic pollutants in environment of the pearl river delta, China: an overview," *Chemosphere*, vol. 52, no. 9, pp. 1411 – 1422, 2003, environmental and Public Health Management. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0045653503004776

[2] C. Su, W. Dongxing, L. Tiansong, R. Weichong, and Z. Yachao, "An autonomous ship for cleaning the garbage floating on a lake," in *2009 Second International Conference on Intelligent Computation Technology and Automation*, vol. 3. IEEE, 2009, pp. 471–474.

[3] M.-H. Chen, C.-F. Chen, and W.-S. Wu, "A river monitoring system using UAV and GIS," in *22nd Asian Conference on Remote Sensing, Singapore*, 2001.

[4] C.-M. Tseng, C.-K. Chau, K. Elbassioni, and M. Khonji, "Autonomous recharging and flight mission planning for battery-operated autonomous drones," *arXiv preprint arXiv:1703.10049*, 2017.

[5] K. Fujii, K. Higuchi, and J. Rekimoto, "Endless flyer: A continuous flying drone with automatic battery replacement," in *2013 IEEE 10th International Conference on Ubiquitous Intelligence and Computing and 2013 IEEE 10th International Conference on Autonomic and Trusted Computing*, Dec 2013, pp. 216–223.

[6] B. Benjdira, T. Khursheed, A. Koubaa, A. Ammar, and K. Ouni, "Car detection using unmanned aerial vehicles: Comparison between Faster R-CNN and YOLOv3," in *2019 1st International Conference on Unmanned Vehicle Systems-Oman (UVS)*. IEEE, 2019, pp. 1–6.

[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[8] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7263–7271.

[9] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," *arXiv*, 2018.

[10] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[11] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[13] "https://opendata.sz.gov.cn/," Open Data Platform, Shenzhen, China.

[14] "https://youtu.be/4UCBMOr26OQ," The Chinese University of Hong Kong, Shenzhen and StrawBerry Innovation, Ltd, China.