

# Scan From the Sky: A Path Planning Method with Perception Optimization for UAVs

Songhe Yuan (✉ [sevenyuan6@gmail.com](mailto:sevenyuan6@gmail.com))

Institute of Advanced Manufacturing Technology, HeFei Institutes of Physical Science, Chinese Academy of Sciences

**Kaoru Ota**

Muroran Institute of Technology

**Jianghai Zhao**

HeFei Institutes of Physical Science, Chinese Academy of Sciences

---

## Research Article

**Keywords:** Unmanned aerial vehicle (UAV), path planning, scene understanding

**Posted Date:** September 1st, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-827549/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Scan From the Sky: A Path Planning Method with Perception Optimization for UAVs

Songhe Yuan<sup>1\*</sup>, Kaoru Ota<sup>2</sup>, Jianghai Zhao<sup>3</sup>

<sup>1</sup>*Institute of Advanced Manufacturing Technology, HeFei Institutes of Physical Science, Chinese Academy of Sciences.* <sup>2</sup>*Muroran Institute of Technology.* <sup>3</sup>*HeFei Institutes of Physical Science, Chinese Academy of Sciences*

<sup>1</sup>Changzhou, Jiangsu Province, China. <sup>2</sup>Muroran City, Hokkaido Japan. <sup>3</sup>Hefei City, Anhui Province, China. <sup>1</sup>email:sevenyuan6@gmail.com

**Abstract**—Unmanned aerial vehicles (UAVs) are frequently adopted in disaster management. The vision they provided is extremely valuable for rescuers. However, they face severe problems in their stability in actual disaster scenarios, as the images captured by the on-board sensors cannot consistently give enough information for deep learning models to make accurate decisions. In many cases, UAVs have to capture multiple images from different views to output final recognition results. In this paper, we desire to formulate the fly path task for UAVs, considering the actual perception needs. A new convolutional neural network (CNN) model is proposed to detect and localize the objects, such as the buildings, as well as an optimization method to find the optimal flying path to accurately recognize as many as possible objects with a minimum time cost. The simulation results demonstrate that the proposed method is effective and efficient, and can well address the actual scene understanding and path planning problems for UAVs in the real world.

**Index Terms**—Unmanned aerial vehicle (UAV), path planning, scene understanding.

## 1 INTRODUCTION

The unmanned aerial vehicle (UAV) is often used in disaster management. For example, after a disaster, people can use the UAV to detect the affected area to find the affected buildings and residents; and in the disaster recovery stage, the UAV can be used to assess the severity of the disaster and the feasibility and cost of the reconstruction, in order to assist people in the disaster reconstruction planning, as shown in Figure 1. However, a serious problem of UAV is that its battery is limited and cannot be used for long-term, large-scale survey tasks. Therefore, how to develop a feasible and effective flight path for it to detect and identify objects efficiently and accurately has become a very important issue.

This task mainly includes two aspects, namely path planning [1] and scene understanding [2]. Although people have done a lot of research work on both sides and got some results, they all have some problems. For example, most of these methods consider this task by two totally-separated problems, i.e., path planning and scene understanding, rather than as a whole. Therefore, they may not be well applied to actual detection tasks in disaster management. In fact, for path planning, the result of scenario understanding is a very important feedback. In one single shot, some of the buildings or objects in the picture are not easily discernible, and it is likely to cause the failure of the scene understanding model. The unclear zone of this area can be used as the key consideration area for the next shooting, and based on that, plan the fly path for the UAVs.

For the flight planning of drones, the previous methods mainly considered coverage and total flight time. And in our approach, we also need to consider the confidence of the recognition. This is to ensure that the deep learning

network [3] can accurately and reliably identify buildings, pedestrians and other objects in the video data collected by the drone. In the optimization process, the traditional method mainly uses a heuristic algorithm based on the traveling salesman method. The heuristic algorithm optimizes the performance of the entire algorithm by minimizing the cost of reaching the current state and the cost of going from the current state to the next state. In our approach, we also chose a similar approach to optimize the path, but since we added the elements of confidence, the overall optimization algorithm is more complex.

In this article, we present a UAV path planning approach to target detection and location tasks using depth models. The main contributions of our work include:

- A drone planning algorithm. In this algorithm, the flight time of the drone, the coverage of the area to be detected and the confidence of object detection can be guaranteed and optimized.
- An object recognition deep learning model. This model can simultaneously detect and locate pre-selected objects, such as buildings, pedestrians, etc., in video frames captured by drones.
- A drone image dataset. The dataset was collected from real drones, and all buildings in each video frame were manually tagged and the exact location was marked. It can be used for performance measurement and comparison of drone detection and positioning methods.

The rest of the paper is organized as follows. Section 2 introduces the existing research in related area. Section 3 presents some notations and the problem definition, and gives an overview of the proposed system. In section 4, we present a scene understanding method and propose the

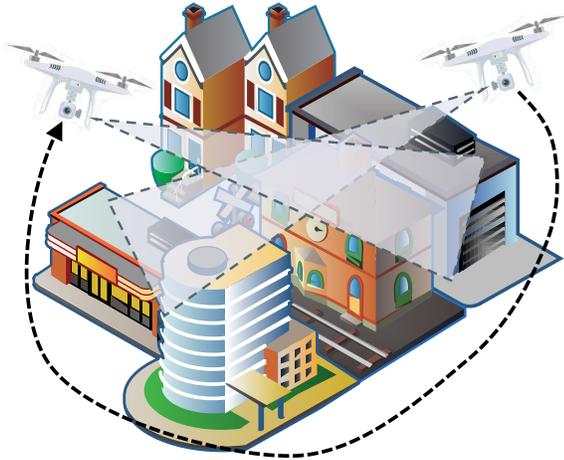


Fig. 1. Path Planning for Scene Understanding with UAVs.

core algorithm of path planning. Experimental evaluation is shown in Section 5. Finally, conclusions are drawn in Section 6.

## 2 RELATED WORKS

### 2.1 Scene Understanding for UAVs

Liu et al. propose a framework [4] combining trajectory detection based on support vector machine and tracking and tracking tracker to realize tracking direction estimation and tracking with low computational cost and real-time. In addition, in their system framework, a simple linear iterative clustering super-pixel segmentation algorithm is adopted to ensure the accuracy of scene segmentation. The visual detection of important objects or people is realized by a single-shot multi-box detector algorithm. Minaeian et al. propose a novel target detection and location scheme, which is based on vision [5] and adopts each UAV as a different function of the cooperation team. In this paper, authors build a team of UAV and multiple unmanned ground vehicles (UGVs) that track and control the population in the border field. A algorithm, which can detect the custom motion, is used in the crowd detection by a mobile camera installed on a drone. Since the drone the lower analysis ability but the wider range for detection, UGV has better analysis and higher accuracy, so a separate body detector is used and the landmark is moved to locate an unknown independent movement at each point in time. The detected population of patterns. The UAV positioning algorithm proposed in this paper uses perspective transformation to translate the crowd position in a image into a position in the real world. Moreover, a thumb positioning method determined by UGV is introduced, which makes the prediction for the geographical location in the detection. Combined with the low-quality attitude-heading reference system (AHRS) on UAVs, Zhang et al. [6] propose a new vision-based positioning method to determine the three-dimensional position of the target. The usual positioning method must rely on many requirements (ie geo-referenced terrain database, precise attitude sensor). Therefore, if the drone system does not meet these requirements, the geographic location of the target cannot

be achieved. On the contrary, the geolocation method proposed in this paper only uses computer vision technology to accurately estimate the target height and yaw angle measurement deviation. The purpose of this method is to eliminate these requirements in current systems while maintaining high target positioning accuracy. Zhu et al. present a superior estimation method for urban traffic density [7] which can efficiently deal with ultra-high resolution video shots obtained from UAVs. They first shot the traffic video in ultra-high resolution for nearly one hour of ultra-high resolution traffic video at five crowded areas in big cities, driving drones during peak hours. They then randomly sample pixel patches, and annotate vehicles to form the data set for the research in the paper, which can also be used to other researches. In the innovative method of urban traffic estimation, it uses deep neural network to detect vehicle and obtain the information, such like location, identification of the vehicle. In addition, they claim that there is other information included in a ultra-high resolution video, making vehicle detection and recognition more accurate than low-resolution content. Fan et al. [8] propose a plant detection method using UAVs. This method consists of three phases. In the first stage, some candidate tobacco plant areas are extracted from the drone image by morphological operation and watershed segmentation. There are tobacco plants or non-tobacco plants in each candidate area. In the second phase, a deep convolutional neural network is established and trained to classify candidate areas as tobacco growing areas or non-tobacco plant areas. In the third stage, post-treatment is carried out to further remove non-tobacco plant areas. There are also many other existing approaches in the area of scene understanding [9], [10], [11], [12], [13], [14], [15], [16], [17]. However, most of them are not designed for the UAVs nor optimized for detection and localization tasks during flying on the air.

### 2.2 Path Planning of UAVs

Yu et al. [18] introduce a solution named as collaborative path planning, which utilizes UAV and UGV to track moving targets in urban environments. The most significant advantage of this algorithm is that it considers the visual occlusion caused by obstacles in the environment. The algorithm models the target state using a dynamically occupied grid, which is refreshed according to the data obtained from the Bayesian filter. Hence, the current behavior and its prediction can be analyzed, based on which, a single vehicle path planning method is presented. The method can maximize the sum of detection probabilities, and has been applied to various scenarios due to its portability. In this scenario, the auction-based decentralized programming algorithm is designed to plan a limited forward-looking path, maximizing the combined probability of detection and the vehicle. Kumar et al. [19] make statistical comparisons between existing UAV path planning methods to determine the best benchmark function. Using the approximate optimization technique determined in the first step, namely the multi-verse optimizer (mvo), they formulate the path planning problem of determining the minimum deviation trajectory of the minimum collision of the drone from the mathematical point of view. They also compare with other

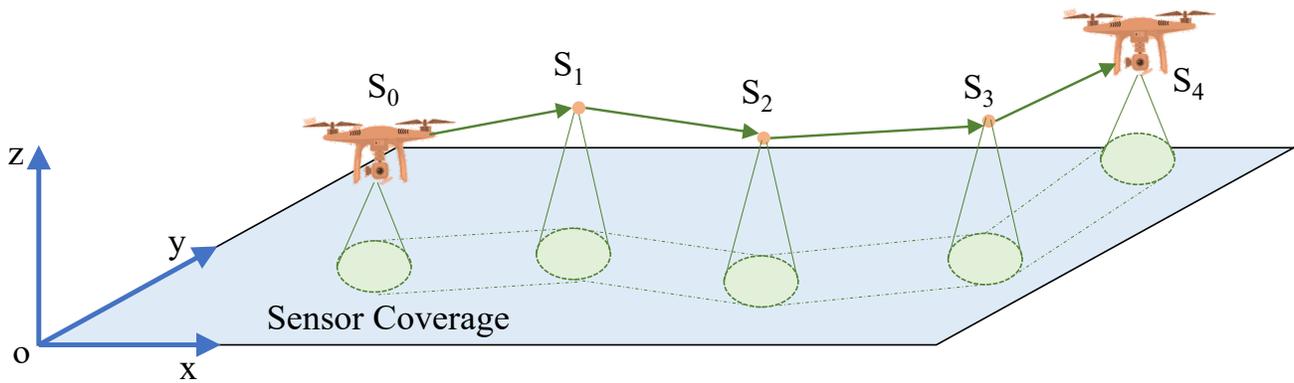


Fig. 2. Path Planning for UAVs.

existing approaches to verify the proposed path planning method. Yang et al. [20] propose a new approach to individual evaluation and evolution methods. By using this new idea, people can take advantage of high quality waypoints. In the evaluation phase, a new set of evaluation functions is derived from existing targets and constraint functions to evaluate each waypoint. Basically, the derivation can only be made if the original function is separable on the waypoint. To further improve the performance of the proposed planner, the waypoints are encoded in a rotating coordinate system with external constraints. In order to test the ability of new planners in planning barrier-free paths, five scenarios with increasing barriers were constructed. Planners using 3 existing planners and 4 alternatives for comparison, which can prove that it can be executed efficiently and effectively. Wen et al. [21] propose a new way to obtain a feasible and safe path. First, based on the intuitionistic fuzzy set (IFS), static threats (STs) is modeled to represent the uncertainty in STs. Introduced the evaluation and synthesis method of st. Based on the fast detection of the random tree (RRT), the dt reachability set (RS) is supposed to predict the value of threat. Secondly, the main purpose that putting a sub-target selector into the planning scheme is to reduce planning costs and improve the efficiency of searching. Furthermore, a back horizon (RH) is introduced, aiming to dealing with online path planning in more complex environments. Therefore, local planners are designed based on the dynamic domain fast detection random tree (DDRRT). And RRT is adopted to achieve the optimization of the path in the planning program. Yin et al. [22] introduce a scheme for the multi-objective path planning (mopp), which searches the appropriate path in a complex urban scenarios, taking into account the level of safety within the proposed scheme to ensure the security of drones. In particular, the security index maps (SIM) is first utilized to detect various obstacles in geographic maps. Thus, the offline search and online search method based on static sim card are proposed. Offline search is supposed to detect the static obstacles, thereby reducing the driving time; however, online search is supposed to deal with other dynamic obstacles. There are also many other existing approaches in the area of path planning for UAVs [23], [24], [25], [26], [27], [28], [29], [30]. However, most of them do not consider the prediction confidence when choose the future path. Therefore, their prediction

performance are not so satisfactory during actual tasks.

### 3 PATH PLANNING

The path planning problem of UAVs in 3D environment is a hotspot in the development of robotics. Particularly, path planning is one of the basic links in autonomous navigation of drones. It refers to the environment in which obstacles are used according to certain evaluation criteria, and the optimal path should be established. Generally, path planning can be summarized as follows: modeling the environment, executing the path search, and building the optimal path. Modeling the environment means abstracting the actual spatial environment information mathematical model so that it can be understood by computer algorithms. Therefore, environmental modeling is path planning, as shown in Figure 2. The basis for rational environmental modeling has an important impact on the eventual results. The modeling methods for UAVs in 3D scenarios mainly include geometric modeling method and unit decomposition modeling, which considers the flight area of the drone as various units. Actually, the traditional grid method plays the most basic role among the various modeling methods. When modeling the environment, the scenarios is transformed into the traditional grids. The information is beneficial to be preserved in a computer. In addition, the adjacency relationship among the grids is pretty intuitive, so that it is easy to write a program implementation in practical applications, and can obtain better planning results when combined with the planning algorithm. In the field of robot path planning, traditional grid method has been widely concerned by researchers as a typical environment modeling method. With the continuous complexity of environmental information and the continuous improvement of robot autonomy, the traditional grid method is used. When there are three-dimensional environment modeling problems, there are still some problems. Compared with the two-dimensional environment, the traditional grid method is used to directly model the environment in the three-dimensional environment. In order to improve the accuracy of modeling, the grid has to be designed into small size, which leads to a large increase in the number of grids, resulting in a planning algorithm. Processing a large amount of raster data, planning efficiency is significantly reduced.

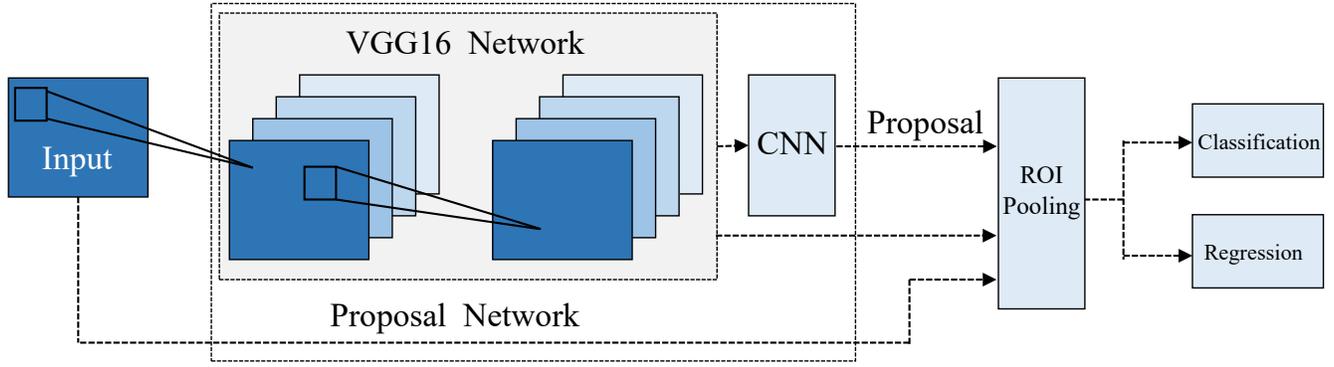


Fig. 3. Structure of the adopted deep model.

This paper mainly considers three kinds of planning indicators, namely the total information density of sub-areas, the coverage time of sub-areas, and the time of transition. Suppose a drone or unmanned boat  $i$  is assigned  $K_i$  sub-areas, and the total information density of the  $k$ -th sub-area is:

$$R_k = \sum_{o \in Objects} 1 - Confidence(o) \quad (1)$$

The sub-area coverage time represents the ideal time required for the drone or unmanned boat to completely cover the sub-area:

$$T_{C,k} = \frac{S}{C} \quad (2)$$

Where  $C$  represents the sensor coverage of the drone, and  $S$  represents the subarea area.

The transition time represents the ideal flight required from the initial point  $P^0$  to the center of the  $K$ th sub-region ( $k = 1$ ), or from the center of the  $k-1$  sub-region to the center of the  $k$ -th sub-region ( $k \geq 2$ ):

$$T_{F,k} = \begin{cases} \frac{\|P^0 - u_k\|}{V}, & K = 1 \\ \frac{\|u_{k-1} - u_k\|}{V}, & K \geq 2 \end{cases} \quad (3)$$

Among them,  $V$  represents the speed of movement of the drone.

Therefore, the expected observed benefit of a drone  $i$  for a sub-area  $k$  is:

$$EP_k^i = \frac{R_k}{T_{C,k} + T_{F,k}} \quad (4)$$

According to the above formulas, the  $K_i$  sub-regions are iteratively sorted to determine the optimal observation order. It is generally expected that the larger the observed benefit, the higher the information density of the region or the less the ideal observation time, the higher the priority.

After determining the optimal observation order for each sub-area, the expected observed benefits of the drone or unmanned boat  $i$  are available:

$$EA_i = \sum_{k'=1}^{K_i} EP_{k'}^i \quad (5)$$

where  $k'$  represents the sub-area number after sorting. In addition, define the ideal observation time for each drone:

$$T_i = \sum_{k'=1}^{K_i} (T_{C,k'} + T_{F,k'}) \quad (6)$$

The total allocation indicator for the drone is:

$$EA = \lambda_1 \sum_{i=1}^N EA_i + \lambda_2 \sum_{i,j=1}^N |T_i - T_j| \quad (7)$$

Among them, the first part represents the total expected observation gain, and the second part is used to balance the task execution cost among the various observation forces such as drones.  $\lambda_1$  and  $\lambda_2$  are weight coefficients. The larger the  $EA$ , the higher the regional observation efficiency, so the optimal allocation objective is as follows:

$$\max(EA) \quad (8)$$

After the sub-areas are allocated, each drone or unmanned boat adopts a parallel receding horizon control (RHC) route planning method aiming at maximizing the observation gain, so that the planned route meets the task time constraint. Assuming that the center position of the sub-area to which the drone is assigned is  $\{S'_1, S'_{K_i}\}$ , the route planning procedure based on the parallel RHC is as follows.

First, each route segment is initialized, including the shortest route segment  $\phi_0$  (time  $t_0$ ) from the starting point  $P^0$  to the center of the sub-region  $S'_1$ , and the coverage observation route  $\phi_j = \{S'_j\}$  of each sub-region ( $j = 1, \dots, K_i$ ) (time  $t_j = 1$ ), the shortest transition route  $\phi_{jj+1}$  ( $j < K_i$ ) (time  $t_{jj+1}$ ) in each sub-area.

Then, if the sum of the time of each of the above-mentioned route segments is less than the task time  $T$ , then a certain route segment needs to be selected and a new waypoint is added. The specific implementation strategy is as follows: adopt the RHC method to pre-plan a new waypoint for each route segment  $\phi_j$ , and select from it. The destination point  $P_j$  of the largest single observation gain is added to the internal route segment  $\phi_j$  of the corresponding sub-area ( $j$ ), and the time  $t_j = t_j + 1$  is updated; if  $j < K_i$ , the slave is updated the point  $P_j$  is the shortest transition

route  $\phi_{j \rightarrow j+1}$  of the center of the sub-area  $S'_{j+1}$ , and the flight time. Repeat the above steps to gradually expand each covered route segment until the sum of the time periods is equal to the mission time  $T$ .

Finally, each route segment is sequentially connected, that is,  $\{\phi_0, \phi_1, \phi_{1 \rightarrow 2}, \dots, \phi_{K_i-1 \rightarrow K_i}, \phi_{K_i}\}$ , as a planning route.

## 4 PERCEPTION METHOD

### 4.1 Deep Learning Model

After obtaining frames of captured video, the system can work on the scene understanding, as shown in Figure 3. Compared to other perception methods, which manually design the feature map for the deep model, we directly map the input frames to object categories and localization, which has the potential to achieve higher network precision.

A multi-layer detection network is designed to find different object characteristics. The input is the captured images. And the main objective is to detect specific objects and localization. Our model contains 7 hidden layers and one output layer. There are five common convolutional layers and two fully-connected layers in the proposed model. We demonstrate that this structure has a good performance on obtaining the relationship between video frames and object information. The distinctiveness is that we design this model as a two-branch structure. There are two different sub models adopted to respectively calculate the category and localization information. Since the features extracted in the lower part of the deep model do not have too much difference, the first 5 convolutional layers are shared between two sub-networks for training efficiency. The remaining layers in each sub-network are used to calculate the higher-level features for the model, because is usually task-related and do not show similarity among different artificial intelligence tasks.

In order to learn two objects at the same time, a multi-task loss is adopted as the cost function of our deep model.

$$L = \lambda_1 L_{\text{category}} + \lambda_2 L_{\text{localization}} + \lambda_3 L_{\text{dec}} \quad (9)$$

where  $L_{\text{category}}$ ,  $L_{\text{localization}}$  and  $L_{\text{dec}}$  are cost functions for two different objectives, and  $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$  can decide their importance for the whole cost value.

$L_{\text{category}}$  represents the loss value to judge the selected data samples as correct category categories, i.e., whether or not there is the specific kind of category in the video frames.  $L_{\text{category}}$  is indeed a softmax cost function.

There are  $n$  input images  $\mathcal{X} = \{x_1, x_2, \dots, x_n\}$  in deep model, and some different labels for them  $x_i$ , which are called the ground-truth  $\mathcal{Y} = \{y_{x_1}, y_{x_2}, \dots, y_{x_n}\}$  and localizations  $\mathcal{R} = \{r_{x_1}, r_{x_2}, \dots, r_{x_n}\}$ . The labels can be valued from  $y \in \{y_1^*, y_2^*, \dots, y_m^*\}$ .  $r$  is a real value.

$$L_{\text{category}} = -\frac{1}{n} \left[ \sum_{i=1}^n \sum_{j=1}^m \mathbf{1}\{y_{x_i} = y_j^*\} \log \frac{e^{h_{\text{category}}^{(j)}(x_i)}}{\sum_{\phi=1}^m e^{h_{\text{category}}^{(\phi)}(x_i)}} \right] \quad (10)$$

where  $h_{\text{category}}^{(j)}(x_i)$  belongs to  $[0, 1]$ , representing the confidence that one specific area is a true object, e.g., building, car, human, etc.  $n$  represents the number of input samples,  $i$  is the sample id of one input sample,  $m$  represents the total

category number,  $j$  is the specific category id,  $h_{\text{category}}(x_i)$  is the Softmax result, and  $h_{\text{localization}}(x_i)$  is the Smooth  $L_1$  result.

$L_{\text{localization}}$  is utilized to find the object localization of one object. According to the smooth  $L_1$  loss in of Fast R-CNN,  $L_{\text{localization}}$  can be written as

$$L_{\text{localization}} = \frac{1}{n} \sum_{i=1}^n \text{smooth}_{L_1}(h_{\text{localization}}^{(r)}(x_i) - r_{x_i}) \quad (11)$$

### 4.2 UAV Image Dataset

We created a new drone image dataset ourselves to train our deep neural network. The training set for neural networks comes from the video that is actually captured on the drone. We extract keyframes for it, and then perform image calibration and area marking on each image.

The marking tool we chose is LabelMe. LabelMe is an annotation markup software made by Python and drawn with GUI by Qt. It can greatly improve the efficiency of image marking.

We have marked a total of 1000 drone images, with dozens of buildings to be marked on each image. The entire marking process took three months/person and marked more than 30,000 different types of buildings with different appearances.

The marked image is shown in Figure 4. Each building has clear boundaries and tag values. In the experimental part, we used this database for training and testing, and some of the results are shown in the next section.

## 5 PERFORMANCE EVALUATION

The training process is performed with some state-of-the-art deep learning frameworks and a modernized NVIDIA GTX 1080 GPU. The learning procedure takes about 2 hours due to the super-large data. And we present some instances for the results of confidence value (shown in Fig.5) and visualization (shown in Fig.6).

As mentioned above, the object is divided into different grids and detected respectively. We analyze the data obtained from the images, and make the prediction that if it is a building. In Fig.5, we show 9 results for the grid of ground truth objects and their predictions. Some examples have 25 grids while some examples have 9 grids, 8 grids or other number. The number of grids is depend on the the number of buildings in the backgrounds. The more buildings for ground truth, the more grids should be generated. Generally, the grid colored with blue represents that there is a building in the real world, and the other grid means there is not a building in the real world. As for the value of each grid, it is the confidence value for the prediction. In our method, we estimate that there is a building if the confidence value exceeds 0.6, and possibility that there is a building is higher when the value of confidence is higher. For example, in Fig.5 (a), the confidence value of the first grid in line 1 is 0, which is lower than 0.6, therefore, we make the prediction that there is no building in this grid; the confidence value of the second grid in line 1 is 10.854, which is larger than 0.6, then we make the prediction that there is a



Fig. 4. Labeled UAV dataset.

building in this grid, which is proved to be true according to the ground truth results. In particular, high confidence value in the grid, which includes a building in the real world, means the high performance of the method.

Fig.6 illustrates the visualization outcomes of the building predictions. The background is the ground truth for the real scenario in each instance, and the identified buildings are labeled with different colors. We show the visualization outcomes for 9 scenarios, including the commercial circle, the play ground, the road area. The buildings are dense in some scenarios, and others are not, which is related to the number of grids generated in the method. Among the 9 visualization outcomes, it can be seen that the size and shape of the building can be drawn with the colored

areas. In general, it has recognized all the buildings and has high performance, however, there exist some points to be improved. For example, the buildings are recognized about 10 in Fig.6 (a) and labeled with blue, purple, green and so on. But, some corners and boundaries are not completed in the visualization outcome when we check the results in detail. Moreover, the method results in a error that recognized the wrong object. In Fig.6 (g), for example, it can be seen that the play ground has been colored with purple, but actually, the play ground is not a building. And in Fig.6 (i), the ring building, which is colored with blue, is not recognized accurately enough, because the center area in the ring building is not a part of the building. Hence, the center area should not be colored with blue. In a word,

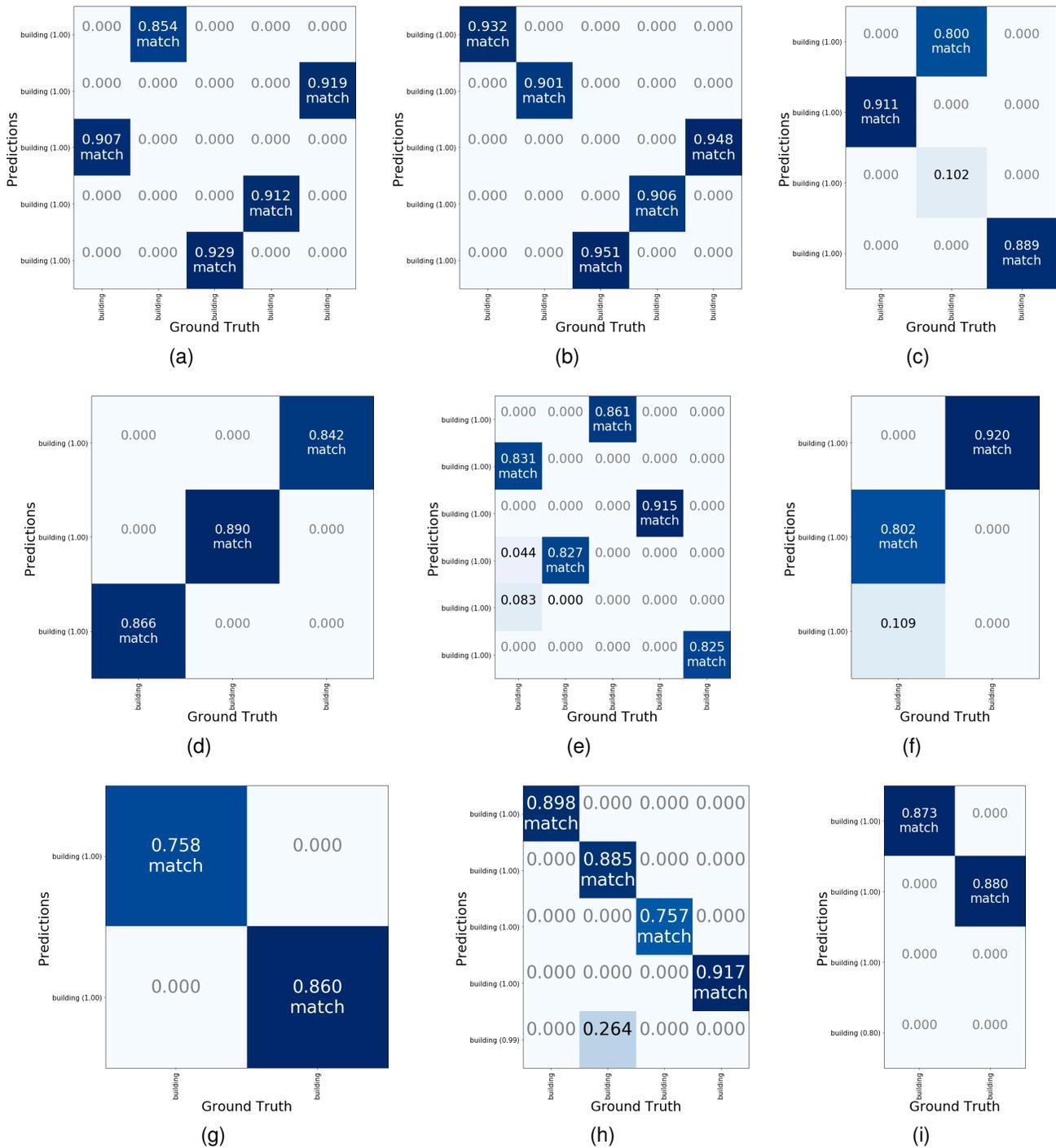


Fig. 5. Grid of ground truth objects and their predictions.

our method can recognize the buildings in various scenarios with high accuracy, but still, we should further improve the accuracy of the recognition and try to avoid the error, in order to apply it in the real applications in a better way.

According to the results of comparison between ground truth and predictions in grids as well as the visualization for building predictions, our method has achieved high performance, such as accuracy, efficiency and reliability.

## 6 CONCLUSION

In this paper, we formulate the fly path task for UAVs, considering the actual perception needs. A new convolutional neural network (CNN) model is proposed to detect and localize the objects, such as the buildings, as well as an optimization method to find the optimal flying path to accurately recognize as many as possible objects with a minimum time cost. The simulation results demonstrate that the proposed method is effective and efficient, and can well address the actual scene understanding and path planning

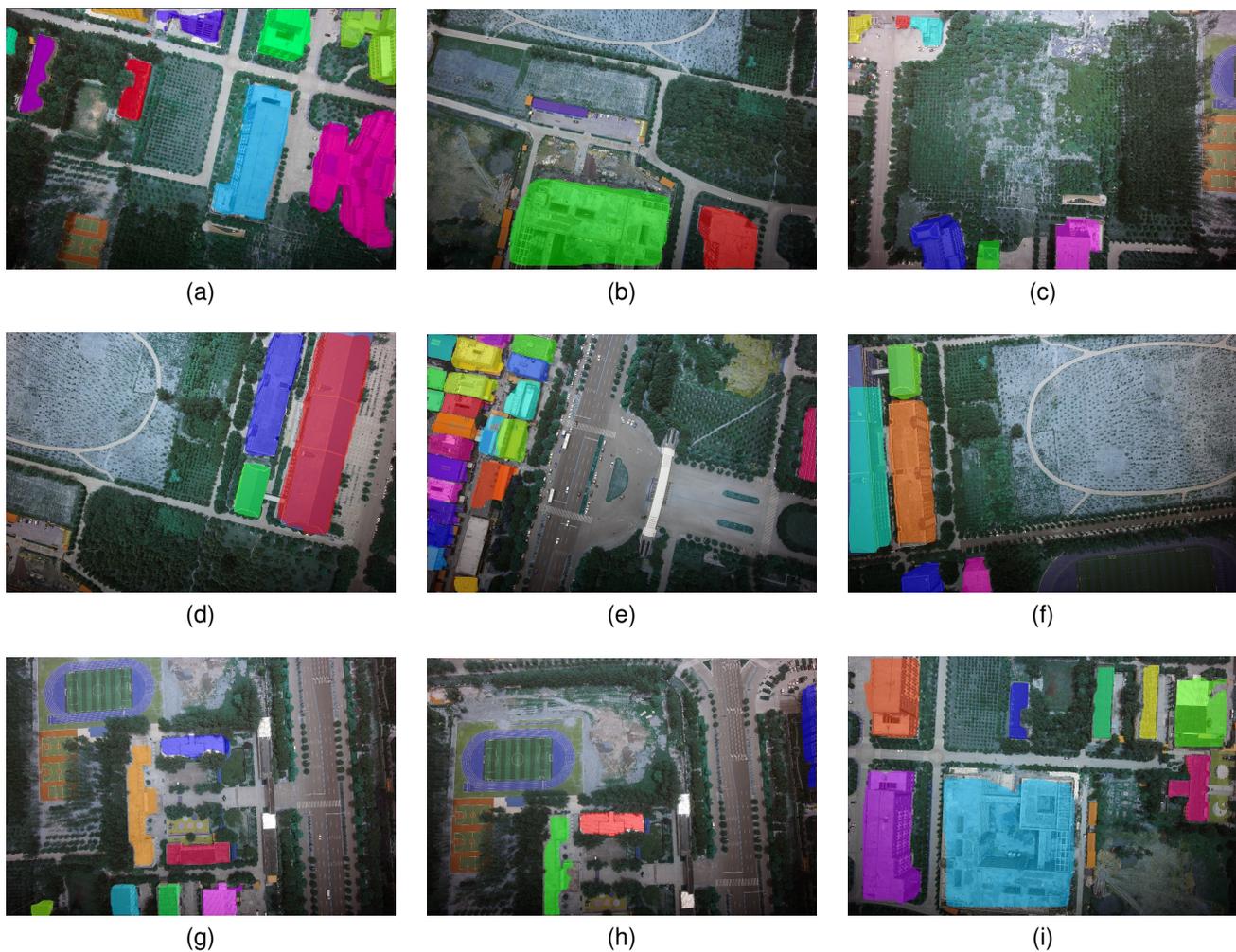


Fig. 6. Visualization of the building predictions.

problems for UAVs in the real world.

## ACKNOWLEDGMENT

This work is partially supported by Jiangsu Provincial Key Research and Development Program, BE2017007-1.

## REFERENCES

- [1] M. Dong, K. Ota, M. Lin, Z. Tang, S. Du, and H. Zhu, "Uav-assisted data gathering in wireless sensor networks," *The Journal of Supercomputing*, vol. 70, no. 3, pp. 1142–1155, Dec 2014. [Online]. Available: <https://doi.org/10.1007/s11227-014-1161-6>
- [2] L. Li, K. Ota, M. Dong, and W. Borjigin, "Eyes in the dark: Distributed scene understanding for disaster management," *IEEE Transactions on Parallel and Distributed Systems*, vol. 28, no. 12, pp. 3458–3471, Dec 2017.
- [3] H. Li, K. Ota, and M. Dong, "Learning iot in edge: Deep learning for the internet of things with edge computing," *IEEE Network*, vol. 32, no. 1, pp. 96–101, Jan 2018.
- [4] Y. Liu, Q. Wang, Y. Zhuang, and H. Hu, "A novel trail detection and scene understanding framework for a quadrotor uav with monocular vision," *IEEE Sensors Journal*, vol. 17, no. 20, pp. 6778–6787, Oct 2017.
- [5] S. Minaeian, J. Liu, and Y. Son, "Vision-based target detection and localization via a team of cooperative uav and uavs," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 46, no. 7, pp. 1005–1016, July 2016.
- [6] L. Zhang, F. Deng, J. Chen, Y. Bi, S. K. Phang, X. Chen, and B. M. Chen, "Vision-based target three-dimensional geolocation using unmanned aerial vehicles," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 10, pp. 8052–8061, Oct 2018.
- [7] J. Zhu, K. Sun, S. Jia, Q. Li, X. Hou, W. Lin, B. Liu, and G. Qiu, "Urban traffic density estimation based on ultrahigh-resolution uav video and deep neural network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 12, pp. 4968–4981, Dec 2018.
- [8] Z. Fan, J. Lu, M. Gong, H. Xie, and E. D. Goodman, "Automatic tobacco plant detection in uav images via deep neural networks," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 3, pp. 876–887, March 2018.
- [9] R. Mottaghi, S. Fidler, A. Yuille, R. Urtasun, and D. Parikh, "Human-machine crfs for identifying bottlenecks in scene understanding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 74–87, Jan 2016.
- [10] A. Geiger, M. Lauer, C. Wojek, C. Stiller, and R. Urtasun, "3d traffic scene understanding from movable platforms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 5, pp. 1012–1025, May 2014.
- [11] C. H. Lim, A. Risnumawan, and C. S. Chan, "A scene image is nonmutually exclusive? a fuzzy qualitative scene understanding," *IEEE Transactions on Fuzzy Systems*, vol. 22, no. 6, pp. 1541–1556, Dec 2014.
- [12] L. Yu, S. Yeung, and D. Terzopoulos, "The clutterpalette: An interactive tool for detailing indoor scenes," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 2, pp. 1138–1148, Feb 2016.
- [13] Y. Gu, Q. Wang, and S. Kamijo, "Intelligent driving data recorder

in smartphone using deep neural network-based speedometer and scene understanding," *IEEE Sensors Journal*, vol. 19, no. 1, pp. 287–296, Jan 2019.

- [14] S. Di, H. Zhang, C. Li, X. Mei, D. Prokhorov, and H. Ling, "Cross-domain traffic scene understanding: A dense correspondence-based transfer learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 3, pp. 745–757, March 2018.
- [15] C. L. Zitnick, R. Vedantam, and D. Parikh, "Adopting abstract images for semantic scene understanding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 4, pp. 627–638, April 2016.
- [16] J. Wald, K. Tateno, J. Sturm, N. Navab, and F. Tombari, "Real-time fully incremental scene understanding on mobile platforms," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3402–3409, Oct 2018.
- [17] Q. Zou, L. Ni, T. Zhang, and Q. Wang, "Deep learning based feature selection for remote sensing scene classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 11, pp. 2321–2325, Nov 2015.
- [18] H. Yu, K. Meier, M. Argyle, and R. W. Beard, "Cooperative path planning for target tracking in urban environments using unmanned air and ground vehicles," *IEEE/ASME Transactions on Mechatronics*, vol. 20, no. 2, pp. 541–552, April 2015.
- [19] P. Kumar, S. Garg, A. Singh, S. Batra, N. Kumar, and I. You, "Mvo-based 2-d path planning scheme for providing quality of service in uav environment," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1698–1707, June 2018.
- [20] P. Yang, K. Tang, J. A. Lozano, and X. Cao, "Path planning for single unmanned aerial vehicle by separately evolving waypoints," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1130–1146, Oct 2015.
- [21] N. Wen, L. Zhao, X. Su, and P. Ma, "Uav online path planning algorithm in a low altitude dangerous environment," *IEEE/CAA Journal of Automatica Sinica*, vol. 2, no. 2, pp. 173–185, April 2015.
- [22] C. Yin, Z. Xiao, X. Cao, X. Xi, P. Yang, and D. Wu, "Offline and online search: Uav multiobjective path planning under dynamic urban environment," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 546–558, April 2018.
- [23] Z. He and L. Zhao, "The comparison of four uav path planning algorithms based on geometry search algorithm," in *2017 9th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, vol. 2, Aug 2017, pp. 33–36.
- [24] T. Morita, K. Oyama, T. Mikoshi, and T. Nishizono, "Decision making support of uav path planning for efficient sensing in radiation dose mapping," in *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, vol. 01, July 2018, pp. 333–338.
- [25] J. Chen, F. Ye, and T. Jiang, "Path planning under obstacle-avoidance constraints based on ant colony optimization algorithm," in *2017 IEEE 17th International Conference on Communication Technology (ICCT)*, Oct 2017, pp. 1434–1438.
- [26] X. Yang, W. Zhou, and Y. Zhang, "On collaborative path planning for multiple uavs based on pythagorean hodograph curve," in *2016 IEEE Chinese Guidance, Navigation and Control Conference (CGNCC)*, Aug 2016, pp. 971–975.
- [27] S. K. al?k, "Uav path planning with multiagent ant colony system approach," in *2016 24th Signal Processing and Communication Application Conference (SIU)*, May 2016, pp. 1409–1412.
- [28] C. Wang, P. Liu, T. Zhang, and J. Sun, "The adaptive vortex search algorithm of optimal path planning for forest fire rescue uav," in *2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, Oct 2018, pp. 400–403.
- [29] and and and, "Path planning for uavs based on improved artificial potential field method through changing the repulsive potential function," in *2016 IEEE Chinese Guidance, Navigation and Control Conference (CGNCC)*, Aug 2016, pp. 2011–2015.
- [30] Z. Zhang, J. Wang, J. Li, and X. Wang, "Uav path planning based on receding horizon control with adaptive strategy," in *2017 29th Chinese Control And Decision Conference (CCDC)*, May 2017, pp. 843–847.



Muroran, Hokkaido, Japan, under the guidance of Prof. Kaoru Ota. Her main fields of research interest include machine learning, UAV image, three-dimensional reconstruction, big data, and data analysis.



**Songhe Yuan** received the B.S. degree in Surveying and Mapping engineering from Shenyang Urban Construction University in 2015. She is currently working toward the M.S. degree in School of Surveying and Mapping, Henan Polytechnic University, under the guidance of Prof. Hailiang Jin. From April 2018 to March 2019, she is a Visiting Student in the Emerging Networks and Systems Lab and Wireless Networks Lab, the Department of Information and Electronic Engineering, Muroran Institute of Technology,

Muroran, Hokkaido, Japan, under the guidance of Prof. Kaoru Ota. Her main fields of research interest include machine learning, UAV image, three-dimensional reconstruction, big data, and data analysis.

**Kaoru Ota** was born in Aizu-Wakamatsu, Japan. She received M.S. degree in Computer Science from Oklahoma State University, USA in 2008, B.S. and Ph.D. degrees in Computer Science and Engineering from The University of Aizu, Japan in 2006, 2012, respectively. She is currently an Assistant Professor with Department of Information and Electronic Engineering, Muroran Institute of Technology, Japan. From March 2010 to March 2011, she was a visiting scholar at University of Waterloo, Canada. Also she was

a Japan Society of the Promotion of Science (JSPS) research fellow with Kato-Nishiyama Lab at Graduate School of Information Sciences at Tohoku University, Japan from April 2012 to April 2013. Her research interests include Wireless Networks, Cloud Computing, and Cyber-physical Systems. Dr. Ota has received best paper awards from ICA3PP 2014, GPC 2015, IEEE DASC 2015, and IEEE VTC 2016-Fall. She is an editor of *IEEE Communications Letters*, *Peer-to-Peer Networking and Applications* (Springer), *Ad Hoc & Sensor Wireless Networks*, *International Journal of Embedded Systems (Inderscience)* and *Smart Technologies for Emergency Response & Disaster Management (IGI Global)*, as well as a guest editor of *ACM Transactions on Multimedia Computing, Communications and Applications* (leading), *IEEE Communications Magazine*, etc. Also she was a guest editor of *IEEE Wireless Communications* (2015), *IEICE Transactions on Information and Systems* (2014), and *Ad Hoc & Sensor Wireless Networks* (Old City Publishing) (2014). She was a research scientist with A3 Foresight Program (2011-2016) funded by Japan Society for the Promotion of Sciences (JSPS), NSFC of China, and NRF of Korea.



**Jianghai Zhao** Dr.ZHAO Jianghai, Professor, vice director of Research Center of Advanced Manufacturing Technology of Institution of Intelligent Machine, Hefei institution of Physical Science Chinese Science Academy of Science. Dr.Zhao is engaged in the research on the design theory and control method of bionic robots. Dr. Zhao undertook some projects supported by the National Science and Technology Support Program, Advanced Research Project of General Armaments Department, Science and Technology Support Program of Jiangsu Province, and the National Natural Science Foundation of China. Dr.Zhao achieve lots of important achievements in scientific field. He had over 40 papers published in scientific journals and has gained the authorization of 16 patents. He awarded 4 prizes, including Third Prize of Anhui Province Science and Technology Innovation Award, Second Prize of Outstanding Contribution Award of Science Island, and outstanding case of Jiangsu institution of industry and technology.