The Principle of UAV Navigation Based on Computer Vision

Junlu Zhang

[1]Beijing 21st Century School, Beijing, 100000, China
zjunlu15@gmail.com

Abstract. Nowadays, unmanned aerial vehicles (UAVs), commonly referred to as drones, have attracted a lot of attention from varies fields. In the development of recent years, one of the most noticeable developments is the application of computer vision in drones, which provides key technologies for visual navigation, positioning and obstacle avoidance, so that drones can truly realize unmanned control. However, because of the limited self-navigation capabilities, UAVs may not be available in all environments. In recent years, visual positioning and mapping, obstacle avoidance and path planning have become the basic components of visual navigation. According to the comparison and classification between 2 major existing machine vision models, Single Shot MultiBox Detector (SSD) and You Only Look (YOLO) Once model, this article will provide a literature review of UAV navigation technologies based on computer vision. This article will also discuss issues and research challenges in the design and implementation of UAV visual navigation technology base on the previous papers. Finally, according to the induction and summary of existing models, give out the suitable application scenarios of different models.

Keywords: Autonomous Unmanned Aerial Vehicles (UAV), Computer Vision, You Only Look Once (YOLO) Framework, Single Shot Multibox Detector (SSD) Model

1 Introduction

With the development of modern technology, drones, because of their characteristics such as lightness, speed and simplicity of manufacture compared to other vehicles, have become the integral part to modern society and the civilian application. It has a profound impact on many fields. In agriculture, they were revolutionary. Drones with high-resolution imaging and sensors enable farmers to precisely monitor the situation of crops, soil conditions and irrigation needs. For rescuing after disaster, drones are invaluable. After a disaster, thanks to the high mobility of drones, they can reach and scan affected areas rapidly, find survivors in dangerous places, with the speed far greater than traditional methods [1]. and even provide emergency supplies, such as food, medicine and water. Commercially, they have transformed aerial photography, mapping, and have shown great potential for logistics delivery in cities. This is largely attributed to advantages that UAVs have while traditional logistics methods do not

J. Iqbal (ed.), *Proceedings of the 2025 2nd International Conference on Mechanics, Electronics Engineering and Automation (ICMEEA 2025)*, Advances in Engineering Research 272,
https://doi.org/10.2991/978-94-6463-821-9_64

have, such as high mobility, easy deployment, and enough flexibility. Despite of that, drones have many drawbacks, such as not being completely out of human control. Therefore, in order to overcome this problem, scientists have developed a method of using artificial intelligence to ensure that drones can operate independently without the full intervene of humans.

In most commercial drones, the UAV usually works according to the following process, as depicted in Figure 1, begins with mission planning and preparation. This stage involves defining the flight path, setting the objectives, and ensuring the drone's systems are in optimal condition. Once ready, the drone proceeds to flight operation. During this phase, it captures images and collects data while executing the planned route.

The captured images and data are then transmitted back to a ground - based station for processing and analysis. This step helps in extracting valuable information related to the mission, such as terrain details or object detection. After analysis, an effect evaluation is carried out, comparing the results with the initial mission goals. Feedback is generated based on this evaluation to make any necessary adjustments for future missions. Finally, the task is summarized, and all relevant data is recorded for reference and further study. This comprehensive workflow ensures the efficient and effective operation of drones in various applications.

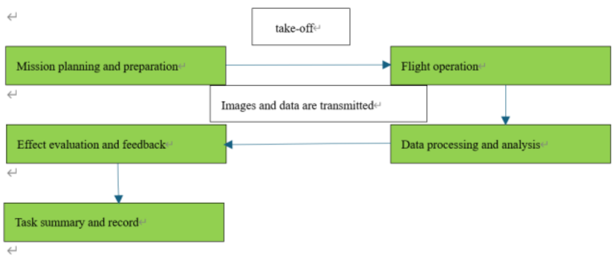

Fig. 1. Workflow of UAV in application scenarios.

In the whole process, most of the process automation is relatively easy to achieve, but in the task execution, the autonomous navigation of the UAV and the adaptive ability to the environment will be greatly challenged, so enhancing the performance of the UAV in the above aspects becomes crucial.

According to the research, over the past few years, several drone navigation technologies have been proposed, including GPS, reference frames, and models. Although these methods vary, they all have a common feature is that they are all based on the detection and analysis of the local terrain to formulate an optimal route.

Recently, computer vision algorithms have been successfully tested and improved in reality. A range of functions are realized through the application of this technology, such as attitude estimation, aerial obstacle avoidance, and navigation [2]. As the application of computer vision technology in drones, variety of problems comes up. According to V. Reddy, the challenges for drones include data availability, as datasets vary and self-collected datasets are common. AI approaches need improvement, such as using more advanced models. Computational power and energy consumption are also issues, along with hardware limitations in processing speed and accuracy [3].

According to the essay, although there are many different kinds of model development, the requirement for more sophisticated model is one of the problems at this time [4]. At the same time, some articles also put forward several problems in computer vision navigation, such as illumination variation, occlusion, similar texture areas, etc [5]. However, the drawback of different models is different from that of model to model; the advantages of different types of model are different from the disadvantages of model, which in the paper is about YOLO model and SSD model, that is the popular of commercialized UAV, the analysis of the model of some types of different models based on the research of different model in different articles is carried out by the review of research on different models.

2 Review on latest relevant research

The report "Survey on Computer Vision for UAVs: Current Developments and trends “by Kanellakis and Nikolakopoulos, "Unmanned Aerial Vehicles (UAV)have seen considerable advances in machine vision applications in the past few years; and has been described in the following paragraphs: "Survey on Computer Vision for UAVs: Current Developments and Trends.

One of the key developments is in the area of navigation and control is Vision-based position-attitude control methods. It is developed for enabling UAVs to better navigate in various environments. For example, some systems combine stereo vision and inertial navigation systems, as mention in the article referenced by [6], a real-time system for UAV take-off, hover and landing based on two concentric ground markers is proposed. UAV uses this combination to enhance its navigation accuracy. This allows UAVs to fly more autonomously and precisely, reducing reliance on traditional navigation methods such as GPS, which may be unreliable in certain conditions.

Beside of that, Simultaneous Localization and Mapping (SLAM)techniques have also seen progress. Algorithms like those based on monocular visual – inertial. As described in multiple studies cited in, the algorithm is characterized by the correlation of the detected features with the UAV state, and the detected point database is updated by the EKF in each cycle [6]. Before the update, the image features are matched, hence enable UAVs to build maps of their surroundings while determining their own position [7]. This development is crucial for tasks such as autonomous exploration and inspection in unknown environments. For instance, a quadrotor UAV can use these techniques for autonomous flight and live dense 3D mapping. In addition, obstacle detection and avoidance are another important aspect. Vision-based systems use techniques such as stereo vision, optical flow, and deep learning-based object detection algorithms to identify obstacles. UAVs can then plan their paths to avoid these obstacles [6]. Meanwhile, some approaches use logic controllers for obstacle avoidance, providing a more solution in complex environments.

Target tracking is also one of the big use for application as well. The use of visual information is used by UAVs to track aerial and/or ground targets, and different algorithms are used, including those based on colour profile and feature tracking and model-based tracking. For the case of urban environment, UAV is able to track people or moving vehicles, which is helpful for surveillance and security application [8].

To summarize, the most recent developments in UAV machine vision applications are wide and wide, which include navigation, mapping, obstacle avoidance, and target tracking, all of which are always changing to meet the increasing demands for UAV autonomy and functionality.

3 Comparative analysis of existing navigation models

3.1 Brief Introduction to SSD Model

The SSD (Single Shot Multi-Box Detector) is an advanced object detection model designed to perform object detection in a single pass through the network. It capitalizes on a set of feature maps extracted from different layers of a base convolutional network. These feature maps are crucial as they capture objects of varying sizes and scales. SSD employs prior boxes with different scales and aspect ratios, which are placed on the different feature maps. Each prior box is associated with a convolutional predictor that estimates the class scores and bounding box adjustments [9].

The working principle of SSD model, usually according to this process. First, SSD uses a pre - trained convolutional neural network (CNN) as its backbone, like VGG16. The backbone network extracts hierarchical features from the input image. Then, at multiple different scales of the feature maps produced by the network, SSD adds additional convolutional layers. These layers are designed to predict a set of default bounding boxes, also known as anchor boxes, at each location in the feature maps. For each anchor box, SSD predicts the offset to the actual object's bounding box and the probability of the object belonging to different classes. Finally, after prediction, a non - maximum suppression (NMS) algorithm is applied. NMS filters out overlapping and low - confidence bounding boxes, retaining only the most accurate and confident predictions, which are then output as the detected objects.

SSDs are very effective because it has one of them as the main features of SSD. It is very fast and is also good at the application in which the process of realisation of the object, however, it is very less of a very fast in the realisation of real-time processing. Also a major addition of this is its capability of dealing with the different sizes of the objects by means of multi-scale detection. This does not mean it does not have any disadvantage. The accuracy in the detection of smaller objects can be very less than some of the existing models and if this increases to huge model size, the detection speed can be quite slow.

3.2 Brief Introduction to YOLO Model

The most well-known object detection model is YOLO, which was created to detect real-time data. It makes use of a single neural network design, which in one forward pass directly forecasts the object classes and bounding box coordinates by itself. Every grid in the input image is in charge of object prediction inside of that area by partitioning it into a grid [10].

As shown in the Figure 2, the YOLO model operates in an integrated process. Initially, image pre - processing resizes the input image to a fixed dimension, say 416×416 pixels, facilitating subsequent operations. Then, its backbone network, like Darknet with convolutional, pooling, and sometimes residual layers, extracts multi - scale feature maps rich in semantic and positional details. These feature maps are partitioned into grids, with each grid tasked to predict objects whose centres lie within it. For every grid, the model estimates multiple bounding boxes along with their confidence scores and class probabilities. Finally, in post - processing, the Non - Maximum Suppression algorithm filters out low - confidence and overlapping bounding boxes, yielding the final detection results that precisely identify the objects' categories, positions, and confidence levels in the image.

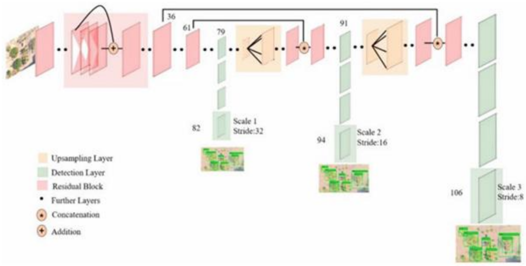

Fig. 2. Architecture of YOLOv3 [11].

YOLO is mainly based on its speed. Because it is so quickly that it works very well and is therefore very suited to uses for instantaneous results such as in the form of a real-time video surveillance system as well because of the very basic model structure which is very easy to do so in the model, so that it is easier to process the information of the whole image, and so it can perform context aware detection. Although, it is too certain of this, and it can be difficult to detect objects of small size, which cannot deal with the object of extreme ratio of objects, because some of these methods.

3.3 Comparison and Application Scenarios

When comparing SSD and YOLO, their different characteristics lead to different application scenarios. SSD’s better performance in handling a wide range of object sizes makes it a good choice for applications where the objects in the scene have significant size variations and a relatively high level of accuracy is required, but the speed requirement is not as strict as real-time video processing. For example, in some industrial inspection tasks where the objects are of different sizes and the detection results do not need to be instantaneous.

On another side, due of the high speed of YOLO, the application's choice of which is the one that wants to be able to perform real-time responses such as live video streaming surveillance, or autonomous driving system where the real time of the object's detection is very important for the safety of the person's object in such situations, hence the capability of detecting with very high accuracy is required; the ability to detect very large number of possible obstacle or the target in such situation than that which is higher than accuracy of detection of the small or odd shaped object.

3.4 Suggestions on Selection of model under different scenarios

When selecting a proper vision model for the UAV, the specialization of the particular application requirement is a must, which needs a lot of attention in the process of selecting the correct model of the UAV.

In industrial inspection work, the SSD model is a better option, such as identifying products or parts of different sizes on the production line, and in such cases the objects usually have large size differences, the detection results do not require instantaneous feedback. Because the SSD model can detect different-sized objects with high accuracy, it can be used to detect multiple-sized objects efficiently. It is a good fit because it has the advantage of being very efficient in dealing with different-sized products, although it might not be very accurate in detecting small objects than some other models. For example, in a factory's part quality inspection process, a UAV with the SSD model can track a certain path to study the appearance and dimensions of parts in a given way. Because of its very high efficiency, large-area inspections can be completed very quickly and the parts sizes can be handled in an efficient way, which satisfies the inspection requirements in industrial production.

In contrast, for applications with extremely high real-time demands such as real-time video stream monitoring or autonomous driving systems, the YOLO model is more suitable. In urban traffic monitoring, UAVs need to rapidly identify vehicles, pedestrians, and other obstacles. The YOLO model's outstanding speed ensures the timely detection of potential hazards. Its ability to quickly respond is crucial for maintaining safety. In autonomous driving scenarios, UAVs must swiftly assess the surrounding environment and make decisions promptly. The high-speed processing of the YOLO model enables UAVs to quickly recognize targets in a rapidly changing environment and adjust their flight paths in a timely manner to avoid collisions. Despite its limitations in detecting small or oddly shaped objects, its speed advantage makes it the top choice when real-time response is of utmost importance.

In conclusion, the performance of the UAV in the detection and visual navigation of the SSD can be greatly improved by taking into account the main requirements of various application circumstances, as well as by selecting the SSD and YOLO models in the appropriate order; thus, the UAVs in the visual navigation and detection can be improved much, which will allow to perform more effectively in different fields.

4 Conclusion

The majority of research on UAV navigation uses computer vision to cover this topic is done at the same time. First, this article reviews current state of affairs in the field of

UAV machine vision, including target tracking, obstacle detection and avoidance, and vision-based position-attitude control. The capabilities of UAVs in different fields have improved because to these developments.

In this article, two popular object detection models are compared which is SSD and YOLO model. SSD can detect objects relatively quickly with multi-scale detection ability but has lower accuracy for small objects and may slow down with a large model size. YOLO is extremely fast, has a simple structure and can perform context-aware detection, yet this model has difficulties in detecting small and oddly shaped objects. Their different characteristics lead to different application scenarios. SSD is suitable for applications with size-varied objects and moderate speed requirements, while YOLO is preferred for real-time response applications.

Finally, there are several limitations of present study such as data availability, model improvement, and hardware limitations. Future study may focus on developing more advanced models, optimizing computing efficiency, improving data collecting and preprocessing methods, and combining many sensors and algorithms to achieve more robust and accurate UAV navigation systems.

References

1. Chen, X., Wang, X., Yang, J., et al.: 'Computer vision-based UAV applications: a review', Journal of Intelligent & Robotic Systems, 2020, 99, (1–4), pp. 327–346.
2. Arafat, Y., Alam, M., Moh, S.: 'Vision-based navigation techniques for unmanned aerial vehicles: review and challenges', Drones, 2023, 7, (2), p. 89.
3. Reddy, V., Madan, H., Khatri, N., et al.: 'Autonomous UAV navigation using deep learning-based computer vision frameworks: a systematic literature review' , Array, 2024, p. 100361.
4. Cai, J., Zhang, Y., Wang, Y.: 'A survey of unmanned aerial vehicle navigation: techniques, challenges, and trends', Sensors, 2021, 21, (13), p. 4401.
5. Zhao, H., Sun, Z., and Wang, Y.: 'Challenges and solutions for UAV visual navigation in complex environments', Aerospace, 2022, 9, (7), p. 370.
6. Macario, A., Michel, M., Moline, Y., et al.: 'A comprehensive survey of visual SLAM algorithms', Robotics, 2022, 11, (1), p. 24.
7. Zhao, X., Huang, X., Cheng, J.H., et al.: 'A vision-based end-to-end reinforcement learning framework for drone target tracking', Drones, 2024, 8, (11), p. 628.
8. Kanellakis, C., Nikolakopoulos, G.: 'Survey on computer vision for UAVs: current developments and trends'. Journal of Intelligent & Robotic Systems, 2017, 87, (1), pp. 141–168.
9. Liu, W., Anguelov, D., Erhan, D., et al.: 'SSD: single shot multibox detector', European Conference on Computer Vision. Springer, Cham, 2016, pp. 21–37.
10. Redmon, J., Farhadi, A.: 'YOLO9000: better, faster, stronger', Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017, pp. 6517–6525.
11. Bhavyasree, B., Yashwanthbharadwaj, V., and Neelima, N.: 'An inter-comparative survey on state-of-the-art detectors—R-CNN, YOLO, and SSD', Smart Innovation, Systems and Technologies, 2021, 213, pp. 475–483.