Deep Learning and Visual SLAM for Autonomous Navigation of UAVs: Status, Challenges, and Future Perspectives

Keyu Long *

Glasgow College Hainan of University of Electronic Science and Technology of China, Chengdu, China

* Corresponding Author Email: 2022360901018@std.uestc.edu.cn

Abstract. This paper discusses the application and integration of deep learning and visual SLAM technology in UAV autonomous navigation. With the wide application of UAVs in the fields of transportation, agriculture, military and environmental monitoring, improving the autonomous navigation capabilities of UAVs has become an important demand. This paper first analyses the advantages and disadvantages of deep learning and visual SLAM. Then, this paper emphasizes the core role of deep learning in feature extraction, target recognition and path planning. The key role of visual SLAM in real-time localization and environment mapping is highlighted as well. This paper discusses the combined application of these two technologies and demonstrates how to enhance the stability and accuracy of visual SLAM systems through deep learning in complex dynamic environments. Although the current technology still faces high computational requirements and real-time challenges, this paper proposes future research direction to promote the continued development of UAV autonomous navigation technology.

Keywords: Deep learning, SLAM, unmanned aerial vehicle, autonomous navigation.

1. Introduction

With the rapid development of modern science and technology, unmanned aerial vehicles (UAV) have been widely used in transportation, agriculture, the military and the environment. However, to achieve autonomous navigation of UAVs, technologies such as deep learning and visual recognition need to be deeply integrated with UAV technology. The advantages of UAV autonomous navigation are real-time, accuracy and efficiency. It allows the UAV to implement real-time and accurate navigation in complex and changing environments. Deep learning is a machine learning method based on artificial neural networks, which is particularly good at processing object recognition and scene understanding in images and videos [1]. In the field of autonomous navigation of UAVs and autonomous vehicles, visual Simultaneous Localization and Mapping (SLAM) technology can use the data collected by visual sensors such as cameras to generate real-time environment maps and determine the location of devices [2, 3]. In UAV autonomous navigation, deep learning is mainly applied to tasks such as target recognition and path planning [1]. For deep reinforcement learning, through Deep Q-Networks (DQN) Proximal Policy Optimisation (PPO) and other models, UAVs can process environmental information more efficiently and achieve more accurate autonomous navigation [4]. Visual SLAM technology allows UAVs to use sensors such as cameras to build maps and determine their current location in unknown environments in real time. This function of visual SLAM is critical for UAV navigation in both dynamic and static environments. While deep learning excels in feature extraction and image processing, its high demand for computing resources is a challenge for small UAVs [1]. In addition, training deep learning models usually takes a lot of time. Although visual SLAM can realize real-time positioning and map construction in a new environment, it is highly dependent on lighting conditions and susceptible to interference from moving objects [3]. Visual SLAM enables the UAV to perceive the environment and position, and deep learning enables the UAV to make decisions. Therefore, these two technologies are the key technologies to realize UAV autonomous navigation. This paper aims to comprehensively analyse the application status, advantages and disadvantages of deep learning and visual SLAM technology in UAV autonomous navigation. In addition, this paper proposes the possibility of combining the two technologies and evaluates the development prospect and direction of UAV autonomous navigation technology in the future. It is expected that the research of this paper can provide some insights for the research direction and development prospect of UAV autonomous navigation technology.

2. Application of deep learning and visual SLAM in UAV autonomous navigation

2.1. Application of deep learning in UAV autonomous navigation

Deep learning technology has important application value in the field of UAV autonomous navigation, especially in the aspects of environment perception, path planning and target detection. Different deep learning algorithms have their strengths and limitations and are suitable for different application scenarios.

You Only Look Once (YOLO) is a popular real-time algorithm in the field of object detection. It is widely known because of its high accuracy and fast inference speed. YOLO can directly predict bounding boxes and class probabilities in the process of processing a complete image. This ability makes it particularly suitable for obstacle detection and target tracking in UAV navigation [5]. YOLO has certain limitations when dealing with complex backgrounds and small targets such as localization errors and insufficient detection accuracy of small targets. However, its advantages in speed and computational efficiency, especially in application scenarios with high real-time requirements, still make it a very practical tool [5].

Faster R-CNN, as an object detection algorithm including two modules, is known for its excellent detection accuracy and reliable region proposal methods. It is especially suitable for accurate object detection and classification in complex environments. However, because of the high computational complexity of Faster R-CNN, although its Region Proposal Network (RPN) has greatly reduced the computing time, its overall computing process still requires high resource consumption [6]. This computational complexity limits the application of Faster R-CNN in high-speed moving UAVs, and it is difficult to meet the real-time requirements.

U-Net is a widely used image segmentation algorithm that is used for segmentation tasks of medical images. It is known for its high performance on a small number of datasets. In UAV navigation, U-Net is used for scene understanding and path planning to identify feasible paths and obstacles through the segmentation of images. Although U-Net performs well in processing complex image tasks, its high demand for computing resources makes its application on embedded platforms challenging, especially in scenarios with high real-time requirements [7].

Dhinakaran ET. Al. evaluated the performance of Deep Reinforcement Learning (DRL) algorithms including Deep Q-Networks (DQN), Proximal Policy Optimization (PPO) and Trust Region Policy Optimization (TRPO). The results show that the success rate of DQN is 75.2%, PPO is 82.6% and TRPO is 78.1%. The results also show that the navigation results are greatly affected by the optimization parameters [4].

Despite the promising application of deep learning in UAV navigation, there are still some key challenges that need to be addressed in future developments. First, the demand for computing resources continues to increase as well as the complexity of deep learning models. To address this challenge, future research directions may include model compression, quantization techniques, and the application of edge computing to reduce dependence on hardware resources [1]. Secondly, the robustness of deep learning algorithms under different environmental conditions also needs to be improved. Finally, how to further improve the real-time performance of the algorithm while maintaining high precision is an important direction of future research. By optimizing the model structure, reducing redundant calculations, or using dedicated hardware accelerators, more efficient real-time processing power is expected.

2.2. Application of visual SLAM in UAV autonomous navigation

Visual SLAM technology enables UAVs to automatically locate and build maps in unknown environments, so it is widely used in navigation and path planning in dynamic environments. The following are several main visual SLAM algorithms and their application status and advantages and disadvantages analysis in UAVs.

ORB-SLAM is a visual SLAM algorithm based on feature points, which uses ORB (Oriented FAST and Rotated BRIEF) features for positioning and mapping. The algorithm performs well in static environments due to its high accuracy and fast processing capability. However, when there are few feature points in the environment, such as in dimmer non-textured regions or under drastic environmental changes, ORB-SLAM may encounter local localization errors and global mapping interruption. To improve the robustness of the system, researchers have proposed the algorithm improvement of fusing IMU (inertial measurement unit) and visual information, which significantly improves navigation accuracy and stability in dynamic environments by combining multi-sensor information [8].

Large-scale direct Monocular SLAM (LSD-SLAM) is a visual SLAM algorithm that directly uses a monocular camera. Its main technical characteristic is the direct monocular SLAM method, rather than relying on traditional feature point extraction. LSD-SLAM has efficient trajectory tracking, depth map estimation and optimization capabilities, which is especially suitable for 3D mapping. However, LSD-SLAM also has some disadvantages, such as the robustness of the algorithm is weak and error accumulation may occur in certain situations and environments. In addition, since this algorithm relies on a monocular camera, it may not be able to correctly estimate the true scale of the surrounding environment [9].

Murwantara et al. used SLAM and Signal Reference Points to build a UAV navigation system [2]. An experiment was conducted to verify the effectiveness of this system. Experimental results show that the system can make the robot return to the starting position autonomously after executing the task. The system also significantly reduces the complexity of autonomous navigation.

Although the application of visual SLAM in UAV autonomous navigation has great potential, it still faces several challenges, especially how to deal with illumination changes, dynamic objects, and high computational complexity [3]. Future directions may include the combination with sensors, or the development of adaptive algorithms to improve the stability of SLAM systems under different lighting conditions. In addition, the combination of scene understanding, and object detection algorithm can identify and eliminate the interference of dynamic objects, to enhance the robustness of the SLAM system. To implement real-time SLAM on embedded platforms, the algorithm structure must be optimized to reduce unnecessary computational overhead or hardware accelerators must be used to improve the processing speed. With these measures, visual SLAM is expected to play a greater role in future UAV autonomous navigation.

2.3. Fusion application of deep learning and visual SLAM

In recent years, the combination of deep learning and visual SLAM has shown great application potential in many fields. This combination provides new solutions to improve the robustness and accuracy of visual SLAM systems, especially in dynamic and complex environments [10,11]. By using neural networks to identify and filter dynamic objects, visual SLAM systems can work more stably, thereby enhancing the autonomous navigation ability of mobile robots such as UAVs [11]. Through an end-to-end deep learning model, high-quality features can be extracted directly from raw images, thereby improving the overall accuracy and environmental adaptability of SLAM systems [10, 12].

In practical applications, the fusion of deep learning and visual SLAM has achieved remarkable results. For example, the Deep SLAM system successfully achieves efficient camera pose and depth map estimation by combining deep learning with geometric constraints through unsupervised learning methods. The system constructs a complete SLAM framework through Mapping-Net, Tracking-Net and Loop-Net to maintain high accuracy and robustness in complex environments [13].

Similarly, the USP-SLAM system combines the Superpoint network and the improved U-Net network to complete the feature point extraction and filtering of dynamic objects, thus improving the robustness and accuracy of the SLAM system in dynamic scenes. For dynamic environments, USP-SLAM improves the absolute trajectory error by 96.12% compared with the traditional SLAM method ORB-SLAM2 [14]. These cases show that deep learning has significant advantages in improving visual SLAM performance, especially when dealing with complex and dynamic environments.

The fusion of deep learning and visual SLAM is not only reflected in the improvement of the algorithm level but also can further enhance the system performance by combining with other technologies and hardware. For example, combining deep learning with an Inertial Measurement Unit (IMU) can further improve the stability and positioning accuracy of visual SLAM systems by using the acceleration and angular velocity information provided by the IMU [10]. The method of multi-sensor fusion can make up for the deficiency of a single sensor and maintain the high robustness of the system when the visual information is insufficient or the environment changes dramatically. In addition, the computational efficiency of the fusion system of deep learning and visual SLAM can be greatly improved by using advanced hardware accelerators such as GPU and TPU to meet real-time requirements. This is particularly important for application scenarios that require fast response, such as drones [11].

Although the fusion application of deep learning and visual SLAM has made significant progress, there are still many challenges in practical applications. Firstly, these systems have high requirements on the environment, especially in low-light, textural or highly dynamic environments, and traditional visual SLAM algorithms may not work stably [15]. Secondly, the high computational complexity of deep learning models leads to a large demand for hardware resources in these systems, and it is difficult to achieve real-time operation on resource-constrained devices [10].

Future research will further explore the fusion of deep learning and visual SLAM, focusing on the lightweight design of the system and multi-modal sensor fusion. For example, more lightweight deep learning models are developed, and the network structure is optimized to reduce the computational complexity and energy consumption [11]. Further development may also involve combining visual SLAM with other sensor data such as IMU and lidar, significantly improving the robustness and adaptability of the system. Future research may also include the development of adaptive algorithms to dynamically adjust the parameters of the SLAM system in response to environmental changes, resulting in a more stable and efficient navigation capability.

Overall, the fusion of deep learning and visual SLAM provides strong support for the development of autonomous navigation of UAVs and other mobile robotics technologies. Although current research has demonstrated the potential of these techniques in several domains, there are still challenges in terms of environmental adaptability and computational resource requirements. Future research should focus on the optimization and lightweight design of the system, as well as the further development of multi-sensor fusion technology to overcome the existing limitations and achieve a wider range of practical applications.

3. Conclusion

The integration of deep learning and visual SLAM technology has shown great application potential in the field of UAV autonomous navigation. By deeply discussing the current situation and challenges of these two technologies in UAV application, this paper illustrates their key role in improving the autonomous navigation capability of UAVs. The significant advantages of deep learning in feature extraction, target recognition and path planning make it an important tool for UAVs to achieve accurate navigation in complex environments. However, the high demand for computational resources in deep learning and the time required for model training are challenges to be solved. Similarly, the importance of visual SLAM technology in real-time localization and map construction cannot be ignored, but its sensitivity to lighting conditions, dynamic objects, and computational complexity still limits its wide application. Systems that combine deep learning and visual SLAM also bring a significant increase in computational complexity, especially on resource-constrained embedded platforms. Moreover, high computing requirements limit the energy efficiency of the system. Future research should focus on optimizing deep learning models, improving the adaptability of SLAM algorithms, and achieving multi-sensor fusion to overcome the shortcomings of current technologies. Through model compression, hardware acceleration and collaboration between edge computing and cloud computing, the real-time performance and computing efficiency of the system are expected to be improved, while reducing the dependence on hardware resources. Additionally, optimizing energy efficiency is significant for the long endurance and practical applications of UAVs. With the continuous progress of deep learning and visual SLAM technology, UAV autonomous navigation technology will usher in new development, laying a solid foundation for the future of intelligence and automation.

References

[1] H. Samma and S. El-Ferik, et al. Autonomous UAV Visual Navigation Using an Improved Deep Reinforcement Learning. In IEEE Access, 2024, vol. 12, pp. 79967-79977.

[2] I. M. Murwantara, B. Hardjono, H. Tjahyadi and A. S. Putra, et al. A Consolidation of SLAM and Signal Reference Point for Autonomous Robot Navigation. 2020 IEEE 17th International Conference on Smart Communities: Improving Quality of Life Using ICT, IoT and AI (HONET), Charlotte, NC, USA, 2020, pp. 108-112.

[3] W. M. Wong, C. Lim, C. -D. Lee, L. Wang, S. -C. Chen and P. -K. Tsung, et al. KRF-SLAM: A Robust AI Slam Based On Keypoint Resampling And Fusion. 2020 IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, 2020, pp. 296-299.

[4] M. Dhinakaran, R. T. Rajasekaran, V. Balaji, V. Aarthi and S. Ambika, et al. Advanced Deep Reinforcement Learning Strategies for Enhanced Autonomous Vehicle Navigation Systems. 2024 2nd International Conference on Computer, Communication and Control (IC4), Indore, India, 2024, pp. 1-4.

[5] A. Nazir and M. A. Wani, et al. You Only Look Once - Object Detection Models: A Review. 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2023, pp. 1088-1095.

[6] S. Ren, K. He, R. Girshick and J. Sun, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 1 June 2017.

[7] D. Li and X. Chen, et al. Lane Line Segmentation Based on Comparative Analysis of Three Convolutional Neural Networks SegNet, U-Net and Attention-U-Net. 2024 IEEE 2nd International Conference on Control, Electronics and Computer Technology (ICCECT), Jilin, China, 2024, pp. 1002-1009.

[8] C. L. Xu et al., Et al. Research on ORB-SLAM Autonomous Navigation Algorithm. 2019 IEEE 9th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER), Suzhou, China, 2019, pp. 1182-1186.

[9] S. Anandharaman, M. Sudhakaran and R. Seyezhai, et al. A low-cost visual navigation and mapping system for Unmanned Aerial Vehicle using LSD-SLAM algorithm. 2016 Online International Conference on Green Engineering and Technologies (IC-GET), Coimbatore, India, 2016, pp. 1-6.

[10] H. Pu, J. Luo, G. Wang, T. Huang, H. Liu and J. Luo, et al. Visual SLAM Integration with Semantic Segmentation and Deep Learning: A Review. In IEEE Sensors Journal, vol. 23, no. 19, pp. 22119-22138, 1 Oct.1, 2023.

[11] Z. Li, W. GAO, H. Chen and S. Zhang, et al. USP-SLAM: Deep Learning Based Visual SLAM with Robust Feature Extraction under Dynamic Environments. 2023 IEEE International Conference on Robotics and Biomimetics (ROBIO), Koh Samui, Thailand, 2023, pp. 1-6.

[12] S. Mokssit, D. B. Licea, B. Guermah and M. Ghogho, et al. Deep Learning Techniques for Visual SLAM: A Survey. In IEEE Access, 2023, vol. 11, pp. 20026-20050.

[13] R. Li, S. Wang and D. Gu, et al. DeepSLAM: A Robust Monocular SLAM System With Unsupervised Deep Learning. in IEEE Transactions on Industrial Electronics, vol. 68, no. 4, pp. 3577-3587, April 2021.

[14] Z. Li, W. GAO, H. Chen and S. Zhang, et al. USP-SLAM: Deep Learning Based Visual SLAM with Robust Feature Extraction under Dynamic Environments. 2023 IEEE International Conference on Robotics and Biomimetics (ROBIO), Koh Samui, Thailand, 2023, pp. 1-6.

[15] B. GAO, H. Lang and J. Ren, et al. Stereo Visual SLAM for Autonomous Vehicles: A Review. 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Toronto, ON, Canada, 2020, pp. 1316-1322.