YOLO-Based Unmanned Aerial Vehicle Technology

Subjects: Engineering, Aerospace | Robotics | Computer Science, Interdisciplinary Applications Contributor: Chunling Chen, Ziyue Zheng, Tongyu Xu, Shuang Guo, Shuai Feng, Weixiang Yao, Yubin Lan

YOLO-based unmanned aerial vehicle (UAV) technology is proposed in the context of interdisciplinary or multidisciplinary collaborative research, by the cross-fertilization of You Only Look Once (YOLO) target detection algorithms and UAV technology. In this YOLO-based UAV technology, the UAV can provide more application scenarios for the YOLO algorithm, while the YOLO algorithm can assist the UAV with more novel tasks. The two can complement each other to further facilitate people's daily lives while contributing to the productivity of their respective industries.

YOLO UAV

object detection

interdisciplinary

application

1. Early Development of YOLO-Based Unmanned Aerial Vehicle Technology

At the beginning of the research on YOLO-based unmanned aerial vehicle technology (YBUT), the technology was proposed because of a technological fusion between UAV technology and You Only Look Once (YOLO) algorithms in the context of a trend towards cross-disciplinary development. Among other things, UAV technology research began in the 1920s and has been developed, to date, with successful applications in agriculture, surveillance, monitoring, traffic construction, system transportation, system inspection, etc. The YOLO algorithm was proposed in 2016 ^[1], and after several improvements, it has reached the forefront of the object recognition field in terms of detection speed, detection accuracy, and recognition classification.

The application of YBUT in real production operations started in 2017. In the early stages of the YBUT application, the main working mechanism was image or video data acquisition by UAVs, followed by object detection, identification, and classification by computers running YOLO-based object detection algorithms. To explore methods to detect vehicles from UAV-captured images for application in traffic monitoring and management, and as deep learning algorithms have shown significant advantages in target detection, researchers have tried to apply YOLO-based object detection algorithms to vehicle detection in UAV images. Jiang et al. ^[2] constructed a multisource data acquisition system by integrating a thermal infrared imaging sensor and a visible-light imaging sensor on a UAV, corrected and aligned the images through feature point extraction and single response matrix methods, and then performed image fusion on the multisource data. Finally, they utilized a deep learning YOLO algorithm for data training and vehicle detection (see **Figure 1**). The experimental results found that the inclusion of a thermal infrared image dataset could improve the accuracy of vehicle detection and verified that the YOLO framework is an advanced and effective framework for real-time target detection. The first attempt to combine and

apply the YOLO algorithm with UAV technology by Jiang et al. ^[2] demonstrated the usability of YOLO-based UAV technology. Although the detection performance of the early YOLOv1 algorithm was not very good, the experimental results were relatively satisfactory as the first exploration of the technology and the innovative incorporation of thermal infrared image data.



Figure 1. Flowchart of the proposed method by Jiang et al. ^[2].

Based on this research, Xu et al. ^[3] proposed an improved algorithm for small vehicle detection based on YOLOv2, whose detection structure model is shown in **Figure 2**. Compared with the YOLOv2 model structure, the algorithm adds an additional feature layer that can reach 1/32 of the input image in size, making the algorithm more adept at detecting small targets and having higher localization accuracy than YOLOv2. This research has greatly contributed to researchers' understanding of YBUT and has also inspired researchers to make targeted improvements to the YOLO algorithm structure when carrying out applications in this field. Since then, Ruan et al. ^[4] and Yang et al. ^[5] have further explored the application of YBUT in other fields.



Figure 2. Dense YOLO network model ^[3].

In addition, Ruan et al. ^[4] attempted to use a deep learning and vision-based drogue detection and localization method to address the accurate detection and localization of fog droplets for autonomous aerial refueling of UAVs in complex environments. They used the trained YOLO algorithm for cone trace detection, the least squares ellipse

fit to determine the long semiaxis of the ellipse after determining the fiducial location, and, finally, a monocular vision camera for vertebral drogue localization (see **Figure 3**). The simulation experimental results show that the method can not only correctly identify cones in complex environments but also accurately locate cones in a range of 2.5–45 m, indicating that the YOLO method has good results for target object detection and localization in various complex environments. Yang et al. ^[S] investigated a method to achieve real-time pedestrian detection and tracking on a mobile platform with multiple disturbances; they attempted to use a UAV hovering in the air for data acquisition of special targets while using a ground station deployed with YOLOV2 to accept video streams from the UAV for analysis and detection. The results of outdoor pedestrian detection experiments showed the robustness of the method when the brightness varied and pedestrians continued to interfere, demonstrating that this is a stable method for exceptional pedestrian tracking on UAV platforms. Most of these early studies explored simple applications of the fusion of the two technologies due to the lack of maturity of the technology fusion application, but the information gained from the research is of greater reference value for subsequent research on YBUT. An increasing number of researchers are focusing on and exploring the field of YOLO-based UAVs, continuing to drive progress in the development of the field, and a new generation of YBUT has emerged as the performance of high-performance computer processors increases while the size of the hardware decreases.



Figure 3. Drogue detection method ^[4].

2. YOLO-Based Unmanned Aerial Vehicle Technology Develops by Leaps and Bounds

As YBUT continues to evolve, a new generation has emerged in which the UAV is equipped with a highperformance processor rich in computing resources, within which YOLO-based object detection algorithms are deployed, allowing the processor to detect, to identify, and to classify mission objects in real time as the UAV collects data. Zhang et al. ^[6], to explore the feasibility of a new generation of technology, embedded the YOLOv3 algorithm into the resource-limited NVIDIA Jason TX1 platform environment (see **Figure 4**) and had the UAV carry the embedded platform for real-time pedestrian detection experiments. The experimental results demonstrated the feasibility of implementing YOLO-based real-time target detection on a resource-limited mobile platform and provided a reference for the development of next-generation YBUT. Alam et al. ^[2], to alleviate the computational pressure on the onboard embedded processor of the UAVs and to enhance the practicality of YBUT, proposed a cost-effective aerial surveillance system that reserves the limited Tiny-YOLO computational requirements on the onboard embedded processor Movidius VPU, shifts the large Tiny-YOLO computational tasks to the cloud, and maintains minimal communication between the UAV and the cloud. Experimental results showed that the system is six-times faster in target detection processing at frames per second compared to the speed of other state-of-the-art approaches, while the application of airborne-embedded processor technology reduces end-to-end latency and network resource consumption (see **Figure 5**). Similar research was conducted by Dimithe et al. ^[8]. The new generation of YBUT brings the YOLO algorithm and drone technology closer together. Although the new generation of YBUT does not show higher performance than previous technologies due to the limited computational resources of the embedded processor on board, the advantages of the new generation of YBUT were demonstrated with practical results by Zhang et al. ^[6]. It is sufficient to show that the future development of YBUT will tend towards a high degree of integration of YOLO algorithms with UAV technology.



Figure 4. Four-rotor monitoring UAV ^[6].



Figure 5. Complete system design by Alam et al. ^[7].

Based on this research, Cao et al. ^[9] proposed a target detection and tracking method based on the YOLO algorithm and the PID algorithm by using a new generation of high-performance embedded processor, NVIDIA Jason TX2, in combination with the pixhawk4 flight control processor. The PID algorithm performs UAV flight control, and the YOLO algorithm is used to identify objects, to extract pixel coordinates and then to convert the pixel coordinates to actual coordinates, where pixel coordinates were the coordinates of the target object relative to the camera image. The actual coordinates were the relative coordinates of the target in a spatial coordinate system constructed with the camera lens as the coordinate origin. Experimental results showed that the method can effectively detect flight targets and perform real-time tracking tasks. Doukhi et al. ^[10] used a UAV equipped with an Nvidia Jetson TX2 high-performance embedded processor and a PID controller. Then, they deployed the YOLOv3 algorithm in the embedded processor to intuitively guide the UAV to track the detected target by using the YOLO-based target detection algorithm, while the PID controller was used to control the UAV flight. Experimental results showed that the proposed method successfully achieves a visual SLAM for localization and UAV tracking flight through the fisheye camera only without external positioning sensors or the introduction of GPS signals (see **Figure 6**). Afifi et al. ^[11] proposed a robust framework for multiscene pedestrian detection, which uses YOLOv3

object detection as the backbone detector (see **Figure 7**) and runs on the Nvidia Jetson TX2 embedded processor onboard the UAV. Experimental results from multiple scenarios of outdoor pedestrian detection showed that the proposed detection framework showed better performance in terms of mAP and FPS, as the computational resources of the embedded processor increase compared to the YOLOv3 algorithm. To facilitate the development of a new generation of YBUT, Zhao et al. ^[12] improved YOLOv3-tiny, resulting in an 86.1% decrease in the model size, a 19.2% increase in AP50, and a 2.96-times faster detection speed than YOLOv3. The experimental results demonstrated that the improved algorithm is more suitable for low-end performance embedded processors in UAV target detection applications.



Figure 6. Software architecture for deep-learning-based motion control ^[10]. The red circles in the diagram represent the input RGB images in the YOLOv3 algorithm, the orange circles represent the calculation process of the YOLOv3 algorithm, and the blue circles represent the target and bounding box data detected by the YOLOv3 algorithm.



Figure 7. Workflow of the pedestrian detection framework [11].

Although the new generation of YBUTs can simplify the operational steps in applications, can improve operational integrity, and can increase the adaptability of technology applications, it is still difficult to obtain satisfactory performance for complex applications based on the limited computing power of existing embedded processors. At the same time, high-performance embedded processors have been slow in development and may not be able to match the computing performance of high-performance computer processors for some time. Therefore, most researchers prefer applications that take the form of a UAV collecting image or video data and a high-performance computer deploying YOLO for object detection ^[13]. Regardless of which of the two approaches researchers take, each study and application drives YBUT research forwards so that YBUT continues to be understood, accepted, and used by researchers in other fields.

References

- Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 27–30 June 2016; pp. 779–788.
- Jiang, S.; Luo, B.; Liu, J.; Zhang, Y.; Zhang, L. UAV-based vehicle detection by multi-source images. In Proceedings of the 2nd CCF Chinese Conference on Computer Vision (CCCV), China Comp Federat, Tianjin, China, 12–14 October 2017.
- Xu, Z.; Shi, H.; Li, N.; Xiang, C.; Zhou, H. Vehicle detection under UAV based on optimal dense YOLO method. In Proceedings of the 5th International Conference on Systems and Informatics (ICSAI), Nanjing, China, 10–12 November 2018; pp. 407–411.
- 4. Ruan, W.; Wang, H.; Kou, Z.; Su, Z.; IEEE. Drogue detection and location for UAV autonomous aerial refueling based on deep learning and vision. In Proceedings of the IEEE CSAA Guidance, Navigation and Control Conference (CGNCC), Xiamen, China, 10–12 August 2018.
- Yang, Z.; Huang, Z.; Yang, Y.; Yang, F.; Yin, Z. Accurate specified-pedestrian tracking from unmanned aerial vehicles. In Proceedings of the 18th IEEE International Conference on Communication Technology (IEEE ICCT), Chongqing, China, 8–11 October 2018; pp. 1256–1260.
- Zhang, D.; Shao, Y.; Mei, Y.; Chu, H.; Zhang, X.; Zhan, H.; Rao, Y. Using YOLO-based pedestrian detection for monitoring UAV. In Proceedings of the 10th International Conference on Graphics and Image Processing (ICGIP), Chengdu, China, 12–14 December 2018.
- 7. Alam, M.S.; Natesha, B.V.; Ashwin, T.S.; Guddeti, R.M.R. UAV based cost-effective real-time abnormal event detection using edge computing. Multimed. Tools Appl. 2019, 78, 35119–35134.
- Dimithe, C.O.B.; Reid, C.; Samata, B. Offboard machine learning through edge computing for robotic applications. In Proceedings of the IEEE SoutheastCon Conference, St Petersburg, FL, USA, 19–22 April 2018.

- Cao, M.; Chen, W.; Li, Y. Research on detection and tracking technology of quad-rotor aircraft based on open source flight control. In Proceedings of the 39th Chinese Control Conference (CCC), Shenyang, China, 27–29 July 2020; pp. 6509–6514.
- 10. Doukhi, O.; Hossain, S.; Lee, D.-J. Real-time deep learning for moving target detection and tracking using unmanned aerial vehicle. J. Inst. Control. Robot. Syst. 2020, 26, 295–301.
- Afifi, M.; Ali, Y.; Amer, K.; Shaker, M.; Elhelw, M. Robust real-time pedestrian detection on embedded devices. In Proceedings of the 13th International Conference on Machine Vision, Rome, Italy, 2–6 November 2020.
- 12. Zhao, L.; Zhang, Q.; Peng, B.; Liu, Y. Faster object detector for drone-captured images. J. Electron. Imaging 2022, 31, 043033.
- Zheng, A.; Fu, Y.; Dong, M.; Du, X.; Chen, Y.; Huang, J. interface identification of automatic verification system based on deep learning. In Proceedings of the 2021 IEEE 16th Conference on Industrial Electronics and Applications (ICIEA), Chengdu, China, 1–4 August 2021; pp. 46–49.

Retrieved from https://encyclopedia.pub/entry/history/show/96731