Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/32/20230219

YOLO model-based target detection algorithm for UAV images

Anqi Wei

School of Communication & Information Engineering, Shanghai University, Shanghai, 200444, China

waq99@shu.edu.cn

Abstract. The increasing popularity of drones has paved the way for their utilization in various sectors, including civil, commercial, and government agencies. These unmanned aerial vehicles have proven to be invaluable in capturing images and videos from vantage points that were once difficult to access, leading to a wide range of applications. Images captured by drones often have target objects that are small in the frame and a large number of photos or videos captured, so that it is difficult for people to find the target objects in the photos. Nowadays, target detection of images captured by drones through deep learning methods, such as the YOLO algorithm, can greatly help people's work. In this paper, the authors of this paper have investigated for the last three years, for target detection of UAV images, optimization based on the original YOLO algorithm to achieve improved detection results. The research in this paper summarizes the existing research results and is of great significance to the subsequent research and application of UAV image processing.

Keywords: yolo, drone image, UAV image, target detection.

1. Introduction

Commercial small aerial vehicles, also known as drones, have the advantage of being portable and more flexible during flight. Drones on the market today are often equipped with high-definition cameras and have real-time sharing capabilities that allow users to capture images and analyze them at any time. Nowadays, using drones for real-time monitoring is becoming increasingly common, and the technology can be applied to human flow monitoring, road traffic flow assessment, forest fire inspection, and large motor equipment inspection. The technology reduces the need for personnel for such tasks or reduces the pressure on personnel, and personnel are less likely to need to travel to dangerous areas, providing more security for staff. The use of drones for inspection and finding, in addition to the hardware equipment requirements, also needs to be able to picture quickly to complete the target detection and the need for relatively high recognition accuracy.

There are the following problems to be solved in target recognition of images using UAVs. The first one is the dataset. In early image target detection algorithms, such as face recognition, license plate recognition, etc., most images are portraits or the front of the object. Much research is also based on the training of such datasets obtained. However, due to the flight characteristics of UAVs, the images captured are all top-down views, and the algorithms also need to be retrained on the data under this characteristic. Second, the size of the target detection object captured by the UAV, even at high resolution, will be smaller than the target object captured by the flat camera. The algorithm needs to enhance the detection performance based on this property, such as improving the characterization of target features, which can improve the correct rate of target object detection. In addition, images captured by UAVs often have the problem of rotational angle because it is impossible to restrict the UAV's camera during flight [1]. Therefore, the algorithm's detection cannot be limited to horizontal or vertical situations but must have good detection for any rotation angle.

As a popular mainstream target detection algorithm, the YOLO algorithm is fast, small, and guarantees a good detection rate [2]. Many researchers have based on this algorithm framework and further optimized the framework for the characteristics of UAV images. In this paper, the authors investigate the research results in this direction in the last three years, so that the subsequent research can be further deepened.

2. Research results in the past three years

The YOLO algorithm is one of the most popular deep learning-based target detection algorithms and has now been released in YOLOv7. YOLOv4 builds on YOLOv3, a major breakthrough in real-time target detection and improved accuracy. YOLOv5 builds on the success of YOLOv4, further improving accuracy and enhancing the model's generalization capabilities. YOLOv7 is the latest version, optimized for speed and accuracy.

Luís Augusto Silva et al. used YOLOv4, YOLOv5, and YOLOv7 for target detection in UAV images, and achieved 59.9% mAP in YOLOv5, 65.70% mAP in YOLOv5 with Transformer Prediction, and 73.2% mAP in YOLOv7 [3]. For UAV images, Transformer Prediction Head (TPH) is added to YOLOv5, improving the large-scale change of target objects in UAV images. And effective methods for target detection in UAV images were screened, and self-trained classifiers were used to improve the classification accuracy under too vague criteria. The project also improved the classification of the dataset of pavement damage, which made the algorithm more effective in detection through more detailed classification.

Jinsu An et al. proposed a method to improve the YOLOv5 network with CBAM to implement an algorithm for target detection for UAV vision and achieved a mAP of 22.56% [4]. The addition of CBAM to the network of the original YOLOv5 realized the introduction of combinatorial fast structures. Because CBAM combines channel attention and spatial attention, the improved YOLOv5 has a convolutional block attention module. CBAM strengthens the performance of the attention module based on the BAM module while remaining lightweight and generalized, and can be flexibly added within any CNN architecture and train the model end-to-end. During the optimization of the three parts of YOLOv5, CSPDarknet53 was used for the backbone, PANet for the neck, and B×(5+C) for the output layer of the head. With this optimization, the algorithm has a better performance for the extraction of feature information.

Zhengwei Li et al. proposed R-YOLOv5, a lightweight rotating target detection algorithm, which achieves an mAP of more than 80% on different datasets and shows good generalizability [5]. The algorithm is based on YOLOv5 and optimized for the backbone feature extraction network, the neck feature fusion network, and the prediction head. The angle prediction branch is combined in the part of the prediction head, and the Circular Smooth Labeling (CSL) angle classification method is introduced to be able to measure the distance between the angle labels, which makes YOLOv5 able to detect the scenes with unknown rotation angles. The problem of tiny objects in UAV images not being able to retain feature information to higher level feature information after multiple convolution operations is also partially solved by feature fusion embedded in the Swin Transformer block (STrB). In UAV images, there is also the problem that similar objects cannot be detected correctly due to more noise in the feature information. Using the Feature Enhanced Attention Module (FEAM), which incorporates the Multihead Self Attention (MHSA) module, enhances the ability of the network to capture the information and ensures detection accuracy. There is a problem of object scale distortion in the images captured by the UAV. By adding the Adaptive Spatial Feature Fusion Structure (ASFF) to the head of YOLOv5, the algorithm can adapt to objects of different scales and no longer loses object information.

The algorithm reduces the excessive computational complexity during feature fusion in the backbone network, increases the utilization of detailed information, and improves multi-scale feature fusion in the head.

Oyku Sahin et al. In response to the problems of high viewing angles, large target scale variations and unstable image quality of UAV images, and an improved network structure was used [6]. The network structure utilizes a combination of Convolutional Neural Networks (CNNs) and Feature Pyramid Networks (FPNs) to effectively capture target information at different scales in the image, thus improving target detection accuracy. The team also proposed a new loss function for optimizing the training process of the network. This loss function combines the position loss and confidence loss of the target frame, as well as the categorization loss, which integrates multiple aspects of target detection and enables the network to better learn the position and category information of the target. In addition, to cope with the problem of changing target scales in UAV images, a target scaling technique is introduced in the paper for appropriate processing of targets at different scales in the image, improving the robustness and performance of target detection. The experimental results show that YOLODrone can achieve higher target detection accuracy and faster detection speed in UAV images, proving the superiority and practicality of the method.

Sushil Kumar et al. utilized deep learning techniques, based on the YOLO V5 algorithm, to improve the accuracy and efficiency of target detection in Unmanned Aerial Vehicle (UAV) surveillance images [7]. The algorithm adopts a one-stage detection strategy, transforms the target detection problem into a regression problem, and utilizes a feature pyramid network (FPN) for multi-scale feature fusion, which results in an excellent performance in dealing with target scale variations. Optimizing the YOLO V5 algorithm for the characteristics of UAV surveillance images, the researchers introduced a series of improvements. The method can better adapt to the complex scenes of UAV surveillance images by introducing special convolutional layers, attention mechanisms and data enhancement techniques. Meanwhile, applying pre-training and migration learning enables the model to be trained on small sample data with better generalization ability. To verify the method's effectiveness, the researchers constructed a UAV surveillance image dataset containing various types of targets and conducted extensive experiments on the dataset. The experimental results show that the target detection and recognition method based on the YOLO V5 algorithm achieves significant performance improvement in UAV surveillance images. Compared with traditional methods and baseline models, the method has significant advantages in target detection accuracy and efficiency, and is able to identify and localize various types of targets in surveillance scenes more rapidly.

Weibiao Chen et al. proposed the DSM-YOLO v5 algorithm, which aims to improve the accuracy and efficiency of target detection in UAV aerial images [8]. The paper chose the YOLO v5 algorithm as the basic framework, and the algorithm adopts a one-stage (one-stage) detection strategy, which performs well in dealing with target scale variations by transforming the target detection problem into a regression problem, as well as utilizing a feature pyramid network (FPN) to achieve multi-scale feature fusion. Optimized for the characteristics of UAV aerial images, the paper proposes the DSM (Digital Surface Model) mechanism. The DSM technique further improves the accuracy and robustness of target detection by acquiring the surface elevation information and fusing it into the YOLO v5 algorithm. The introduction of the DSM helps to better localize and recognize the target. The experimental results show that the DSM-YOLO v5 algorithm is able to achieve significant performance improvement in UAV aerial images. Compared with the traditional methods and benchmark models, the algorithm has obvious advantages in target detection accuracy and detection speed.

Songyun Zhang proposed a fast target detection method for UAV imagery based on MobileNet-YOLO V4 model [9]. The authors used MobileNet as the basic network structure, which is a lightweight convolutional neural network with fewer parameters and computational complexity, suitable for target detection on devices with limited resources. Combining MobileNet with YOLO V4, which is an advanced model in the field of target detection, a series of technological improvements, such as the CIOU loss function, the SAM module, and PANet, are used to improve the accuracy and robustness of target detection. The authors used a series of optimisation measures to further improve the speed of target detection. For example, the number of parameters of the model is reduced and the network structure is optimized by network pruning and quantization techniques, which enables the model to perform target detection in UAV imagery quickly and efficiently. In addition, this paper also carries out data enhancement and preprocessing for the characteristics of UAV images, which increases the diversity of samples and improves the generalization ability of the model. Experiments are conducted on UAV image datasets containing various types of targets and different complex scenes, and the results show that the method achieves significant performance improvement in the UAV image target detection task. Compared with traditional methods and other target detection models, the method based on the MobileNet-YOLO V4 model has obvious advantages in terms of speed and accuracy.

Xianghong Cheng et al. improved the YOLO V5 algorithm to detect small targets efficiently and accurately in UAV aerial images [10]. The research team adopted the YOLO V5 algorithm as the basic framework. Aiming at the low-resolution characteristics of small targets, the team introduced a higher-level feature pyramid network to enhance the representation of small targets. To suppress the background interference, this paper adds an attention mechanism, which enables the algorithm to focus more on the important features of small targets, thus improving the accuracy of detection. To verify the effectiveness of the improved YOLO V5 algorithm, the researchers constructed a UAV aerial image dataset containing many small target samples and conducted a series of experiments. The experimental results show that the improved YOLO V5 algorithm significantly improves UAV aerial images' performance. Compared with the traditional method and the benchmark model, the algorithm has obvious advantages in small target detection accuracy and detection speed.

3. Conclusion

In recent years, with the rapid development of UAV technology, UAV image target recognition plays an increasingly important role in military, civil and industrial fields. Among them, the target recognition technology based on YOLO (You Only Look Once) algorithm has attracted much attention. In this paper, we conduct in-depth research on the improvement of the YOLO algorithm for UAV image target recognition in the past three years, in order to understand the performance of the algorithm in solving the problems of rotational angle, small target pixels, etc., and to explore the more detailed enhancement achieved by different versions of YOLO on these improvements.

Our research identified eight important studies that have done a great deal of exploratory work on the particular challenges of UAV imagery. First, researchers have proposed a series of solutions to the problem of target rotation, which is prevalent in UAV imagery. Some of these methods are based on YOLO and introduce a rotation invariance module, which enables the algorithm to better handle targets with inconsistent rotation angles. These improvements effectively improve the accuracy of target recognition and enhance the application of UAVs in dynamic environments.

Second, another group of researchers proposed a series of innovative solutions for the problem of too small target pixels in UAV images. These methods mainly focus on the feature extraction part of the YOLO algorithm, which effectively enhances the perception of small targets by introducing the attention mechanism and image pyramid structure. The results show that these improved algorithms have outstanding performance in recognizing small targets, which greatly improves the detection rate of UAVs on small targets and provides strong support for dealing with complex and changing practical application scenarios.

It is worth noting that although all of these studies improved on the YOLO algorithm, they did not retain the advantages of its inherent light weight and fast detection. This advantage is especially important for UAV image target recognition in today's demands for efficient computation. These researches improve the performance and meet the demand for real-time and practicality in practical applications, enabling UAV technology to work even better in target searching, monitoring and tracking.

In addition, some of the research projects have constructed their own datasets to better validate the performance of the algorithms. By using targeted datasets, these studies can more fully demonstrate the superiority of their improved algorithms and optimize them for specific scenarios. This trend in dataset construction has provided UAV image target recognition research with more reliable evaluation criteria, allowing algorithms to be trained to produce better results and be better adapted to specific mission requirements.

In summary, research on improving the YOLO algorithm for UAV image target recognition has made great progress in the past three years. By improving the algorithm for problems such as rotation angle, too small target pixels, and constructing a customized dataset while retaining the algorithm's advantages, researchers have made positive contributions to the development of UAV technology. However, it is also important to realize that the challenges faced by target recognition in the real world are complex and diverse, and continuous efforts are still needed to further improve the robustness and accuracy of the algorithms in the future to promote the application of UAV technology in a wider range of fields.

References

[1] Li Z, Liu X, Zhao Y, Liu B, Huang Z and Hong R 2021 *Journal of Visual Communication and Image Representation* **77** 103058
[2] Jiang P, Ergu D, Liu F, Cai Y and Ma B 2022 *Procedia Computer Science* **199** 1066–73
[3] Silva L A, Leithardt V R Q, Batista V F L, Villarrubia González G and De Paz Santana J F 2023 *IEEE Access* **11** 62918–31
[4] An J, Putro M D, Priadana A and Jo K-H 2023 *2023 IEEE International Conference on Industrial Technology (ICIT) 2023 IEEE International Conference on Industrial Technology (ICIT) (Orlando, FL, USA: IEEE)* pp 1–6
[5] Li Z, Pang C, Dong C and Zeng X 2023 *IEEE Access* **11** 61546–59
[6] Sahin O and Ozer S 2021 *2021 44th International Conference on Telecommunications and Signal Processing (TSP) 2021 44th International Conference on Telecommunications and Signal Processing (TSP) (Brno, Czech Republic: IEEE)* pp 361–5
[7] Kumar S and Kumar C 2023 *2023 International Conference for Advancement in Technology (ICONAT) 2023 International Conference for Advancement in Technology (ICONAT) (Goa, India: IEEE)* pp 1–5
[8] Chen W, Jia X, Zhu Zh et al. *Computer Engineering and Applications* **1-11**[2023-07-27].http://kns.cnki.net/kcms/detail/11.2127.TP.20230705.2129.004.html
[9] Zhang Song Yun. *Jiangxi Science* 2023 **41(02)** 339-342+355.DOI:10.13990/j.issn1001-3679.2023.02.020.
[10] Cheng X, Cao Y, Hu Y et al. *Flight Control and Detection* 2023 **6(01)** 80-85.