# Unmanned Aerial Vehicle Target Tracking Systems in Complex Environments Based on Visual Enhancement Technology

Tianlin Guo

School of International Education, Beijing University of Chemical Technology, Beijing BJ 102200, China

2024090082@buct.edu.cn

**Abstract.** Unmanned aerial vehicle target tracking is one of the key areas of study in the field of computer vision and intelligent control, widely used in aerospace, intelligent driving, security surveillance, smart agriculture, and search and rescue, which can also carry out aerial multi-view multi-platform long-distance real-time dynamic object tracking through its own high maneuverability, good adaptability to the environment, etc. Although UAVs (Unmanned Aerial Vehicles, UAVs) have good maneuverability, they also have shortcomings; the complex climate can easily cause target deformation, partial occlusion, similar objects interfering, and even illumination changes will lead to poor performance of visual system, the amount of data of UAV on board is huge, and then the computing power of the onboard UAV is small, difficult to achieve high precision algorithm for real-time performance and high energy consumption. Therefore, this paper designs a UAV target tracking system that integrates multimodal visual enhancement methods to utilize multimodal perception and realize scene-adaptive switching, as well as lightweight feature fusion, thereby reducing computational resources, improving tracking accuracy, and enhancing robustness under complex backgrounds. The system also leverages the latest communication technology.

**Keywords:** UAV, Visual Enhancement, System Algorithms, Target Tracking

## 1 Introduction

The rapid growth in drone technology and its increasing commercialization are driving expansion in applications, from ordinary aerial photography to public services, disaster relief, traffic supervision, and public security. With drones capable of conducting versatile tasks at favorable prices, most of them play an indispensable role in a wide range of complex tasks. However, various environmental factors, such as rough weather conditions, complex terrains, dense traffic flows, and frequent external interference, can cause issues like visual image blur, missing target features, and overlapping target signboards, which can significantly impact target tracking or recognition errors. The drones cannot complete their objectives securely under these


K. Subramanian (ed.), *Proceedings of the International Workshop on Advances in Deep Learning for Image Analysis and Computer Vision (IWADIC 2025)*, Advances in Computer Science Research 128,
https://doi.org/10.2991/978-94-6239-648-7_43

circumstances, rendering them unsafe and unsound work methods. If proper solutions are not found, this kind of issue will definitely become a significant barrier that hinders the use of drones in key business areas, restricting their practical use in real operational environments.

To address this problem, researchers have increasingly started integrating vision enhancement technology into the field of drone target tracking using visual enhancement technology to effectively deal with low light, adverse weather, and complex backgrounds via image preprocessing, feature enhancement, and multi-source information fusion; improving the target recognition accuracy and stability of drones in a complex scene, compared with traditional tracking methods which can only use visual signals as guidance, not only enhancing the adaptability of the drone in multi-interruptive environments but also supplying contactless, real-time, and robust feature signals. To assist drones in performing such long-term and complicated tasks, this paper believes they need to be supported.

Developmentally, drone target tracking technologies have evolved from traditional image-matching-based techniques to integrating CNNs(Cable News Networks,CNN) , attention mechanisms, and lightweight deep learning models. While early algorithms performed well under ideal conditions, they failed to meet their objective when light levels changed or targets were occluded. Furthermore, as development progressed, with deep learning and multi-modal perceptions being applied, lightweight architectures from CNNs, Siamese networks, and Transformers have gradually become commonplace. They are increasingly utilised for highly effective feature extractions as well as the ability to better adapt to the varying environments presented on space-limited ONB platforms. At the same time, the incorporation of IR cameras, LiDAR, and visual information enables the expansion of drone functions for various tasks conducted under challenging conditions, such as poor lighting, adverse weather, and rugged terrain.

Through this, the paper aims to focus on visual-enhanced drone target tracking in complex environments, where it will explain the disturbances and aftermath caused by the above factors on the drone's visual tracking process. Secondly, this paper will explore the principles, adaptability, and key algorithms of visual enhancement techniques, and then combine multimodality fusion with the lightweight model scenario to conduct an integrated assessment. Finally, a summary of existing research shortcomings and a prospecting of future development trends will be provided, aiming to clarify the blueprint of complex environments. This will lay a theoretical basis and supply practical references for optimising and applying drone visual tracking technology in complex environments.

## 2 Analysis of Visual Tracking Issues for Unmanned Aerial Vehicles in Complex Environments

The interference of complex environments on uncrewed aerial vehicle visual tracking presents diverse characteristics, which can be roughly divided into environmental challenges and target detection challenges. From the ecological perspective, extreme

weather conditions are the primary source of interference. Weather conditions such as rain, wind, strong gusts, and sandstorms can cause speckle noise, unstable UAV attitude, and reduced contrast between the target and the background. Dynamic changes in lighting conditions can lead to a decrease in the signal-to-noise ratio of the visual sensor, image overexposure, and loss of edge features. From the target perspective, the dynamic characteristics of the tracking object itself can also affect the tracking performance. Deformation or attitude changes of the target during movement can lead to a decrease in the matching degree of the target's appearance features with the initial template. In contrast, changes in the distance between the UAV and the target can cause significant scaling of the target, posing challenges for the adaptive adjustment of the tracking box. Overall, multiple interference factors in complex environments often coexist and superimpose on each other, not only increasing the difficulty of target feature extraction and retention but also significantly raising the robustness requirements for tracking algorithms. These environmental challenges set higher standards for subsequent system design, requiring a balance to be struck between algorithm stability, real-time response, and environmental adaptability.

# 3 Principles and Adaptability Analysis of Visual Enhancement Technology

Visual enhancement technology provides abundant information for the target tracking of uncrewed aerial vehicles in complex environments. By extracting and processing this information correctly, along with analysing technical principles, adaptability, and learning from the system's optimisation, the target tracking of uncrewed aerial vehicles in complex environments can be achieved.

## 3.1 Visual enhancement technology

To eliminate restrictions when UAVs perform visual tracking, one approach is to enhance visual technologies. For low-light, blurry videos or moving targets partially hidden, techniques such as image preprocessing, feature enhancement, and multisource information fusion can be employed to address visual defects caused by illumination conditions or occlusions.

Specifically, the CNNs play a dominant role in visual enhancement tasks; because multiple layers of convolution and pooling operations can automatically extract information from images with various sizes and levels to retain essential targets and suppress the background noises, therefore visual enhancement can leverage its technique that can be employed for feature extraction in a unified pipeline, such as the one provided by a CNN; in view of this technique, visual enhancement offers a powerful source for supporting UAV to track targets under complicated circumstances.

## 3.2 Adaptability to complex environments

Visual information degradation in complex environments can be varied. To achieve stable target tracking performance while enhancing it, this paper should select, optimise, and combine effective methods tailored to different environments. This not

only has good detection accuracy but also meets the requirements of real-time operation and low resource consumption in actual scenarios.

In low-light conditions or at night, images are prone to insufficient brightness, increased noise, and missing details, which makes it necessary to develop corresponding low-light image enhancement algorithms, noise suppression technology, and the fusion of infrared and visible light images. In conditions where direct light is too strong, due to overexposed and high-contrast light intensities under unbalanced illuminance, target features can become invisible. Under these conditions, the application of high dynamic range imaging, an illumination equalization algorithm, and reflection suppression technology yields good results in addressing this problem. The meteorological conditions have a specific influence on the image, particularly causing image blurring, reducing image contrast, and distorting colors. Rain/fog/snow removal algorithms, as well as robust cross-modal fusion detection frameworks, are available for mitigation. Prior modelling of weather degradations can enhance image clarity through a convolutional network, thereby achieving better results.

By conducting a concentrated analysis of environmental elements and leveraging highly adaptive technologies across multiple fields, this paper aims to maximise UAV target tracking reliability under conditions of weak computing power and power. This approach provides theoretical evidence to inform subsequent system design and optimisation efforts.

# 4 Design of Unmanned Aerial Vehicle Target System Based on Visual Enhancement

In complex environments, visual enhancement technology plays a core role in unmanned aerial vehicle target tracking systems. Current research generally adopts the overall approach of "perception - enhancement - tracking", that is, by using sensors to collect multi-modal data, the quality of the images is optimised in the visual enhancement stage, and the enhanced images are input into the target detection and tracking module to improve the robustness in low-light conditions, harsh weather, and dynamic scenes.

## 4.1 Methods for Dealing with Insufficient Light

In the perception layer, RGB cameras remain the primary source of visual information, which can be complemented by methods such as infrared imaging, depth cameras, or LiDAR sensors to provide additional details in low-light or obstructed line-of-sight conditions (Table 1) [1,2].

This is particularly applicable in low-light, nighttime, and complex terrain scenarios, ensuring continuous visibility of targets in the absence of light and providing sound and stable video data synchronisation and depth information, minimising the impact of occlusions. Regarding the accurate estimation of target distance and position, the research by scholars Hyunjin Choi and Youdan Kim on uncrewed aerial vehicles (UAVs) equipped with monocular vision sensors indicates

that target tracking is quite challenging. They researched the sensor's inability to measure the distance between aerial targets and the UAV. Based on specific image processing techniques, they proposed a measurement model for the visual sensor. They designed a nonlinear adaptive observer to estimate states, parameters, and the target position [3]. This effectively enhances the perception capability and improves the accuracy of target recognition.

**Table 1.** Comparison of methods for dealing with insufficient light

| Core methods | Applicable environment | Advantages | Limitations |
|---|---|---|---|
| RGB camera + infrared camera multimodal fusion [1] | Low light, nighttime | Maintain target visibility in low-light conditions and achieve multi-modal complementarity. | High cost and complex |
| RGB + LiDAR fusion [2] | obstruction, undulating terrain | Increase the penetration power and minimise the effect of blocking objects. The weight and power consumption of Lidar are both significant. | Lidar is relatively heavy and has a high power consumption. |

### 4.2 Methods for dealing with rainy and foggy weather

Most studies on enhancement layers are related to adaptive image processing methods under specific environmental conditions (Table 2). The relevant methods encompass low-light enhancement, defogging, de-raining, and HDR correction techniques.

In terms of low-light enhancement, standard methods include brightness restoration and noise suppression, both of which are based on convolutional neural networks. In the field of low-light visual enhancement research, scholars such as Zheng and Shen conducted a systematic review of the low-light image and video enhancement domain, pointing out that the main challenges currently faced include overexposure and underexposure issues under mixed lighting conditions, as well as the lack of high-quality low-light video datasets suitable for training and testing [4]. This study introduced the SICE dataset and its two enhanced variants, SICE_Grad and SICE_Mix. It proposed the large-scale, high-resolution low-light video dataset,

American Night Sky, for evaluating algorithm performance under various lighting and scene changes [4].

**Table 2.** Comparison of Handling Methods for Rain and Foggy Weather

| Core methods | Applicable environment | Advantages | Limitations |
|---|---|---|---|
| Low-light enhancement network based on Convolutional Neural Networks[4] | Nighttime, low illumination | Effective restoration of brightness and details while suppressing noise. | Information may still be lost even in extremely low-light conditions. |
| Unsupervised Multi-branch De-fogging Network (UME-Net)[5] | Foggy and hazy weather | Does not rely on paired data and retains high-frequency details. | Inadequate adaptability to heavy fog and rapidly changing scenes. |
| HDR correction combined with multi-exposure fusion [6] | High-intensity light, high-contrast scenes | Reduce overexposure and underexposure to preserve details. | Multi-frame compositing is susceptible to motion blur. |

For dehazing and deraining, multi-scale feature fusion or unsupervised generative networks are often adopted. Sun, H. et al.'s research addressed the problems of input data distribution confusion and feature loss in unsupervised image dehazing methods based on CycleGAN. It proposed an unsupervised dehazing network (UME-Net) that integrates a multi-branch high-frequency enhancement strategy [5-7]. This method introduces a shared encoding module and a multi-branch decoding module in a multi-branch network to maintain the consistency of generated features. It combines a high-frequency information compensation strategy to enhance the texture and structural details of the image. Moreover, by inverting the atmospheric scattering model, two critical parameters for dehazing photos — namely, the atmospheric light and scene transmission rate — can be obtained. By estimating the scene transmission rate of each pixel, a relatively straightforward dehazing image can be achieved without any training process [5].

The HDR correction technique can suppress overexposure and underexposure in dynamic scenes, thereby improving image detail fidelity. Xiao Jun et al. considered the problem of degrading the HDR image quality in dynamic scenes. They proposed a deep, progressive feature aggregation network to enhance imaging quality in situations involving drastic lighting changes and object movement. It simultaneously optimises the bright and dark details, which exhibit high similarity characteristics and

match and aggregate multiple scale characteristics through sampling. As shown in Figure 1, these characteristics are based on the multi-scale properties derived from the discrete wavelet transform (DWT), which facilitates super-resolution reconstruction. Experimental results demonstrate that the HDR images generated by this approach exhibit superior detail fidelity, texture recovery, and visual quality compared to existing methods.[6].

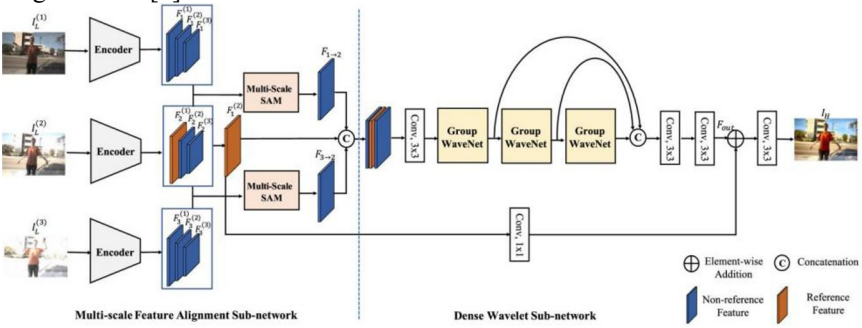


**Fig. 1.** Data trend of the experimental results [6].

In recent years, multimodal fusion methods have emerged in the enhancement layer, significantly improving detection and tracking accuracy in complex scenarios through techniques such as pixel-level registration, feature-level weighted fusion, and decision-level results integration.

### 4.3 System Design for Improving Tracking Accuracy

In the tracking layer, lightweight feature extraction and cross-scale feature fusion structures are becoming a common approach nowadays.

**Table 3.** Comparison of processing methods for improving tracking accuracy

| Core methods | Applicable environment | Advantages | Limitations |
|---|---|---|---|
| Lightweight Siamese Tracking Network[8] | Universal scenarios | High-speed reasoning, suitable for embedded platforms. | Limited adaptability to changes in appearance. |
| Lightweight Transformer Fusion Network[9] | Multiple objectives, dynamic background | The fusion of features across different scales yields excellent results and is highly robust. | The relative computational overhead remains relatively high. |

As shown in Table 3, Siamese series networks and lightweight transformer architectures minimize computational cost without sacrificing accuracy. Additionally, online appearance updating, occlusion handling, and trajectory prediction methods eliminate targets from shadow occlusion, change in target appearance, and loss, respectively. In comparison to single-modality trackers, multiregistration input has greater stability on target maintenance and reidentification. Yet, the balance between real-time performance and computation is an open issue for future investigation.

(UAV)While existing vision-augmented UAV target-tracking systems can perform very well in these challenging situations in terms of target recognition and acquisition, they still face several issues, including difficulties in achieving effective multimodal information fusion, low-power implementation, long-term system stability, and limited onboard computing capacity.

### 4.4 The solution to the problem of limited on-board computing resources

The design of uncrewed aerial vehicles (UAVs) typically requires a balance between miniaturization and lightweight design, which imposes strict limitations on the on-board computing unit in terms of computing power, storage, and power consumption, thereby adversely affecting the performance of the visual tracking system. First, the bottleneck in real-time processing is a prominent issue. Visual enhancement techniques often require substantial support from floating-point operations. To address the problem of high-precision target tracking under limited computing power, Li, XT et al. made refined improvements to the core components, including the feature extractor, target query, and decoder, in the meta-architecture. As shown in Figure 2, this approach can, to a certain extent, balance the model's accuracy and computational efficiency. It has been verified that high tracking performance can still be maintained through structural optimization under limited computing power [10].

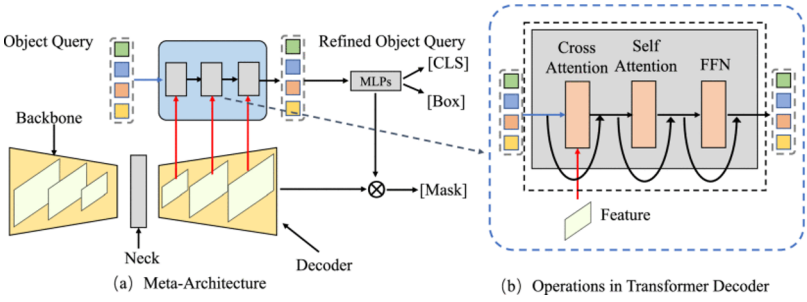


**Fig. 2.** Data trend of the experimental results [10].

Additionally, the limitation of algorithm complexity is a crucial factor that affects the visual tracking performance of UAVs. Due to the limited onboard computing power, the system typically selects lightweight algorithms to reduce computational burden and power consumption, but this comes at the expense of tracking accuracy and robustness. Qian Kun et al. introduced a style recalibration module into the

original SiamRPN network to enhance its ability to perceive image style. They adopted transfer learning to extract more distinguishable features from the reference labelled target images. They thus demonstrated improved performance with higher robustness and stability in the continuous tracking of infrared dim and small targets [11].

This paper demonstrates that both model structure optimisation and feature enhancement approaches can enhance the robustness of small target tracking when operating with limited computing power.

### 4.5 Program Design for Unmanned Aerial Vehicle Network Security

Traditional network systems may not be efficient in handling the dynamic demands of drones. As drones are usually deployed in harsh environments and terrains, providing a strong and secure network is of great importance. Based on these facts, Mehta, P. et al. verify the validity of access objects through a perfect identity authentication mechanism to prevent malicious users from invading the drone network. Identity authentication, as the core link to ensure the security of UAV networks, can effectively resist forged identities, session hijacking, and other malicious attacks. In existing technical solutions, two-factor or multi-factor authentication has gradually become mainstream. As shown in Figure 3, by combining access control policies, preventing Wi-Fi attacks and hijacking, and resisting forgery attacks, the security protection capabilities of drone networks in complex environments have been significantly enhanced [12].

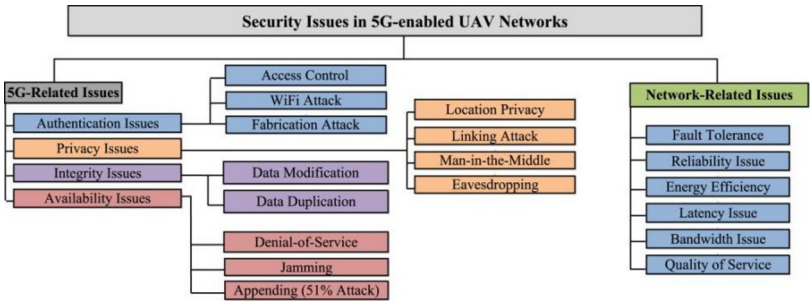


**Fig. 3.** Data trend of the experimental results [12]

## 5 Case Analysis

In the research of uncrewed aerial vehicle (UAV) target tracking, the DJI Matrice 4 series has demonstrated outstanding performance in various scenarios due to its multi-source perception, multi-modal fusion technology, and scalable computing power. In a complex intersection scene in a city with 500 motor vehicles and 200 pedestrians, with a sampling frequency of 30 frames per second and continuous observation for 2 hours, the built-in AI model of the Matrice 4 series achieved a vehicle detection accuracy of 92.3% and a pedestrian detection accuracy of 90.7%,

with only three tracking interruptions and a cumulative duration of less than 10 seconds. In contrast, the DJI Phantom 4 series achieved vehicle and pedestrian detection accuracies of 85.1% and 82.6% respectively, in the same scenario, with 12 tracking interruptions. Additionally, the open AI computing power interface of the Matrice 4 series supports the deployment of third-party models. After a specific team deployed their self-developed multi-target association algorithm, the target tracking delay in traffic congestion areas was reduced from 200 milliseconds to 80 milliseconds. In a mountainous night scene with an illumination intensity of less than 0.1 lux, facing three local occlusions, the average occlusion recovery time of this model, utilizing infrared and visible light fusion technology, was 1.2 seconds, with a tracking stability of 98.5%. The trajectory prediction model was trained using 1000 sets of mountainous night movement trajectory data, achieving a prediction accuracy of 95.3%. This is significantly superior to the average occlusion recovery time of 4.7 seconds and the tracking stability of 76.2% achieved by traditional single visual systems. In a coastal environment with moderate wind and rain, characterized by a wind speed of 15m/s and a rainfall rate of 20mm/h, the LiDAR-visual fusion solution of the Matrice 4 series achieved a target retention rate of 94.2% and a frame rate stability of 96.8%. After 50 simulated wind and rain environment durability tests, the equipment failure rate was less than 2%, significantly outperforming the DJI Phantom 4 series (target retention rate 78.3%, frame rate stability 65.5%) and early traditional visual tracking platforms (target retention rate 62.1%, frame rate stability 48.7%). In summary, advanced UAV platforms represented by the Matrice 4 series, combined with multi-modal perception, visual enhancement, and scalable computing power, have improved core indicators by 10% to 30% in the three scenarios compared to the comparison objects, providing a quantitative reference for equipment selection and algorithm optimisation in the field of UAV target tracking [13].

## 6 Conclusion

The paper examines the UAV target tracking system in complex environments. The paper systematically reviews the core problems, technical routes, and practical applications of the relevant work on UAV target tracking systems. It is indicated by the study that UAV has gained wide use application in fields such as public service and emergency rescue, however, there are still two core obstacles in its visual tracking system: firstly, multi-dimensional interference of complex environment factors such as terrible weather, sudden change of lighting, target distortion and dynamic clutter all reduce the quality and details of vision information and lead to feature loss; secondly, limited onboard computing resources bring significant restrictions on computing capacity, storage space, power supply, these hinder the algorithms of high precision visual enhancement to meet the needs of real-time performance and stability.

This work aims to discuss the application value of visual enhancement technology from the aspect of the following issues. The integration of CNN with a lightweight architecture, multi-modal perception fusion, and scene-adaptive processing enables the system to achieve robust performance in low-light, occlusion, and poor weather

scenarios. The application of the multi-source perception algorithm and anti-interference algorithm improves tracking accuracy and stability, as demonstrated in a real-world case study using the field test of the Matrice 400 series drone. At the same time, the technical solution has passed the verification on a certain level of practicality.

Unfortunately, current research still has certain deficiencies: firstly, the degree of efficient fusion and spatiotemporal calibration of multimodal information is low; the real-time processing ability is not up to par under low-power consumption conditions; and the tracking stability cannot guarantee performance under long-term, complex situations. In view of this situation, future studies should consider the following three main aspects: First, the development of scene-adaptive algorithms that can achieve dynamic disturbance factor sensing and intelligent adjustment of enhancement methods; secondly, to overcome the dilemma of striving for both low computational load and high-precision network structures and solve the accuracy-efficiency tradeoff; thirdly, to formulate a standardized evaluation index and data sharing mechanism for different types of UAV visual tracking systems and benchmark them using standard frameworks to improve technology based on standardized comparative analyses and promote the widely used UAV visual tracking technology.

## References

1. Yue, G., Li, Z., Tao, Y., Jin, T.: Low-illumination traffic object detection using the saliency region of infrared image masking on infrared-visible fusion image. Journal of Electronic Imaging **31**(3), 033029 (2022)
2. Wang, X., Fu, C., He, J., Wang, S., Wang, J.: StrongFusionMOT: A multi-object tracking method based on LiDAR-camera fusion. IEEE Sensors Journal **23**(11), 11241–11252 (2023)
3. Xu, S.W., Shui, P.L., Yan, X.Y., Cao, Y.H.: Combined adaptive normalised matched filter detection of moving target in sea clutter. Circuits, Systems, and Signal Processing **36**(6), 2360–2383 (2017)
4. Zheng, S., Ma, Y., Pan, J., Lu, C., Gupta, G.: Low-light image and video enhancement: A comprehensive survey and beyond. arXiv preprint arXiv:2212.10712 (2024)
5. Liao, M., Lu, Y., Li, X., Di, S., Liang, W., Chang, V.: An unsupervised image dehazing method using patch-line and fuzzy clustering-line priors. IEEE Transactions on Fuzzy Systems **32**(6), 3381–3395 (2024)
6. Xiao, J., Ye, Q., Liu, T., Zhang, C., Lam, K.: Deep progressive feature aggregation network for multi-frame high dynamic range imaging. Neurocomputing 594, 127804 (2024)
7. Sun, H., Luo, Z., Ren, D., Du, B., Chang, L., Wan, J.: Unsupervised multi-branch network with high-frequency enhancement for image dehazing. Pattern Recognition 156, 110763 (2024)
8. Ding, Y., Miao, K.: Lightweight Siamese network with global correlation for single-object tracking. Sensors 24, 8171 (2024)
9. Zheng, W., Lu, S., Yang, Y., Yin, Z., Yin, L.: Lightweight transformer image feature extraction network. PeerJ Computer Science 10, e1755 (2024)
10. Li, X., et al.: Transformer-based visual segmentation: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence **46**(12), 10138–10163 (2024)

11. Qian, K., Zhang, S., Ma, H., Sun, W.: SiamIST: Infrared small target tracking based on an improved SiamRPN. Infrared Physics & Technology 134, 104920 (2023)
12. Mehta, P., Gupta, R., Tanwar, S.: Blockchain envisioned UAV networks: Challenges, solutions, and comparisons. Computer Communications 151, 518–538 (2020)
13. DJI: DJI Matrice 4 Series User Manual. DJI, Shenzhen (2023)