Theory and Practice of Science and Technology
2026, VOL. 7, NO. 2, 97-100
DOI: 10.47297/taposatWSP2633-456919.20260702

Survey of High-Speed Vehicle Tracking Technology for Unmanned Aerial Vehicles Based on Neural Networks and YOLO

Kaizhuo Liu

School of Electromechanical Engineering, Guangdong University of Technology, Guangzhou, Guangdong, 511400, China

ABSTRACT
Due to the rapid development of unmanned aerial vehicle (UAV) technology and the increasing maturity of artificial intelligence algorithms, tracking of high-speed moving targets by UAVs has become a research hotspot in computer vision, pattern recognition, and intelligent control. Therefore, this paper provides a well-structured and logical systematic survey of UAV high-speed vehicle tracking technology based on neural networks and YOLO models: first, the background and significance of the discussed problem are presented; then, the latest methods of four core technical modules, namely target detection, state estimation, trajectory planning, and tracking control, are naturally introduced; subsequently, the application principles and implementation methods of key technologies such as YOLO target detection algorithm, adaptive neural network control, disturbance observer, finite-time control, and barrier Lyapunov functions in UAV tracking systems are analyzed in detail. Finally, existing shortcomings are objectively summarized, and future directions are reasonably proposed.

KEYWORDS
Unmanned aerial vehicle (UAV); Target tracking; You Only Look Once; Neural network

1 Introduction

As a new type of aerial mobile platform, UAVs have outstanding advantages such as strong maneuverability, flexible deployment, and low cost; therefore, they have been sufficiently and successfully applied in military reconnaissance, traffic monitoring, disaster rescue, precision agriculture, and other fields. More importantly, in recent years, the hardware cost of UAVs has continuously decreased, and flight control technology has become increasingly mature; thus, UAVs have naturally and smoothly penetrated from military special equipment into various civilian fields. Tracking and monitoring of high-speed moving vehicles by UAVs is one of the key technologies for realizing intelligent traffic management, border security, and target tracking applications. However, it must be acknowledged that UAV tracking of high-speed vehicles is accompanied by several distinct and difficult technical problems: First, when vehicles are traveling at high speeds, their position and attitude change extremely fast, so extremely high real-time performance is required for target detection; second, because of the relative motion between UAVs and vehicles, the tracking system must have extremely strong anti-interference capability; third, the dynamic characteristics of UAVs themselves and external factors such as wind disturbance will introduce uncertainties in tracking control. Fortunately, deep learning technology has achieved breakthrough progress in the computer vision field in recent years, and the YOLO series target detection algorithms have become a very reliable tool for real-time target detection. Complementing this extremely well are advanced control methods such as adaptive neural network control, finite-time control theory, and barrier Lyapunov functions, which have provided novel and powerful ideas for solving uncertainty problems in UAV tracking systems.

2 Deep Learning-Based Target Detection Technology

2.1 Evolution of YOLO Series Algorithms

Target detection is the first and most fundamental link in UAV tracking systems, and YOLO series algorithms are currently one of the most popular and mature real-time target detection algorithms. By adopting a single-stage detection architecture, target detection problems are naturally and reasonably transformed into regression problems, thereby well balancing speed and accuracy. Specifically, YOLOv1 was first proposed by Redmon et al. in 2016, where the input image is divided into grids, and bounding boxes and category probabilities are directly predicted for each grid. YOLOv2 introduces batch normalization, anchor box mechanisms, multi-scale training, and other techniques. YOLOv3 adopts a multi-scale feature map fusion architecture and achieves 33.0% mAP in COCO dataset experimental results listed in Table 1. YOLOv4 uses CSPDarknet53 backbone network, PANet feature pyramid, Mish activation function, and other techniques to make very solid and clever improvements to the algorithm, reaching 43.5% AP in COCO dataset experimental results given in Table 1. YOLOv5 was developed by Ultralytics company, providing various plug-and-play model variants from nano to xl, achieving 66.4% mAP@0.5 in COCO dataset experiments listed in Table 1. YOLOv7 introduces trainable bag-of-freebies techniques. YOLOv8 makes very elegant optimizations to the model structure, adopting Anchor-Free design with more

advanced label assignment strategies; therefore, it achieves 72.0% mAP@0.5 in COCO dataset experiments listed in Table 1. YOLOX is a very typical and excellent attempt in anchor-free design, achieving very good accuracy with decoupled head design and dynamic label assignment strategies. YOLOR proposes a complete multi-task learning framework, using unified representation learning to simultaneously and naturally improve the performance of detection, segmentation, and other tasks.

Table 1 Performance Comparison of YOLO Series Algorithms

Algorithm	Backbone Network	mAP@0.5(%)	FPS	Characteristics
YOLOv3	Darknet-53	55.3	45	Multi-scale detection
YOLOv4	CSPDarknet53	62.8	65	Balance of accuracy and speed
YOLOv5	CSPDarknet53	66.4	140	Edge deployment
YOLOv7	CSPDarknet53	69.7	161	Training techniques
YOLOv8	CSPDarknet53	72.0	155	Multiple versions

2.2 Improvements of YOLO in UAV Target Detection

Because UAV aerial images have characteristics such as large target scale variations, complex backgrounds, and frequent occlusions, the academic community has already proposed very well-structured and logical improvement schemes for related problems: In terms of lightweight improvements, YOLO-tiny series directly and naturally achieves smaller model volumes and faster inference speeds by reducing the number of network layers and the number of channels; knowledge distillation, model pruning, and other techniques have also been well applied to YOLO lightweight improvements; in addition, YOLO variants based on lightweight backbone networks such as MobileNet and EfficientNet are being actively explored. In terms of small target detection improvements, feature pyramid enhancement is beneficial for small target detection; attention mechanisms can make networks more naturally focus on regions of interest; multi-scale training objectively enhances adaptability to targets of different scales. In terms of complex environment adaptation, data augmentation methods reasonably simulate various environmental conditions by randomly disturbing image brightness and contrast; domain adaptation methods are used to reduce distribution differences between source and target domains; robust training methods directly strengthen the model's robustness to interference through adversarial examples and noise injection.

3 Target Tracking and State Estimation Methods

3.1 Classical Tracking Algorithms

Kalman filtering is a classical linear state estimation method, where a target motion model is first established, and then optimal estimation of the system state is performed using observed data. Therefore, standard Kalman filtering is naturally divided into two steps: prediction and update, with advantages such as high computational efficiency and rigorous theoretical derivation; however, it must be acknowledged that its processing capability for nonlinear systems is limited. Therefore, extended Kalman filtering was subsequently developed, which performs first-order Taylor expansion of nonlinear functions at the current estimation point to approximate nonlinear systems. Complementing this extremely well is unscented Kalman filtering, which uses deterministic sampling strategies to directly and reasonably approximate nonlinear transformations. Another important branch is particle filtering, which uses Monte Carlo sampling to approximate posterior probability distributions and does not require strict assumptions about the probability distribution of system noise; therefore, it naturally becomes a powerful tool for handling state estimation problems in nonlinear non-Gaussian systems. Mean shift algorithm is a tracking method based on kernel density estimation, which estimates target position by iteratively solving the gradient of probability density function; therefore, it has high computational efficiency, but it inevitably brings the defect of being sensitive to target appearance changes.

3.2 Deep Learning-Based Tracking Methods

Since SiamFC designs tracking methods from the perspective of learning target similarity metrics, it naturally uses two shared-weight convolutional networks to process search images and template images respectively, and then uses cross-correlation operations to directly and efficiently obtain similarity response maps; therefore, SiamFC has extremely fast tracking speed (reaching 86 fps) and does not require online updating. Complementing this extremely well, SiamRPN introduces region proposal networks and adds bounding box regression branches, so it can more accurately and reliably estimate target positions. Detection-based tracking naturally and reasonably combines target detection with data association: first, target detectors are used to detect all targets in each frame, and then data association algorithms are used to match detection results with existing tracks.

4 UAV Tracking Control Algorithms

4.1 Traditional Control Methods

PID control is the most commonly used and mature control method in the industry, where the control input is determined based on the proportional, integral, and derivative three error terms between the target value and actual value; therefore, it has a simple structure and is easy to implement. However, it must be acknowledged that its robustness to system uncertainties and external disturbances is relatively poor. Complementing this extremely well is backstepping, which is a nonlinear control method based on Lyapunov stability theory; virtual control quantities are designed recursively, and system nonlinearities are naturally and reasonably handled.

4.2 Adaptive Neural Network Control

Because adaptive neural network control can learn unknown system dynamics online, it has extremely strong self-learning and self-adaptive capabilities. Moreover, neural networks can approximate any nonlinear function with arbitrary precision; therefore, it naturally becomes an ideal tool for solving uncertainty problems in UAV systems. More strictly, the universal approximation theorem guarantees that three-layer feedforward neural networks can approximate any continuous function with arbitrary precision. A typical adaptive neural network control structure can be naturally and reasonably divided into two parts: neural network controller and adaptive law; the latter is generally designed based on Lyapunov stability theory, which directly guarantees that all signals in the closed-loop system are bounded and tracking errors converge.

4.3 Advanced Robust Control Techniques

Disturbance observer technology has a very clear and natural physical meaning: that is, estimating parameter perturbations and external disturbances in the system, and then compensating the estimated values to the control input; therefore, it can effectively improve the system's anti-interference capability. Specifically, the basic idea of disturbance observer is to treat parameter uncertainties and external disturbances in the system as equivalent disturbances, and then design observers to estimate these disturbances and perform feedforward compensation. Finite-time control can guarantee that the system strictly converges to the desired state within finite time; therefore, compared with traditional asymptotic convergence control, it has faster response speed and better transient performance. Complementing this extremely well is barrier Lyapunov function, which is an extremely effective tool for handling system state constraints; that is, special Lyapunov functions are constructed to ensure stable operation of the system within the constraint range, so it is naturally suitable for the minimum safe distance problem between UAVs and targets.

Table 2 Comparison of UAV Tracking Control Methods

Control Method	Advantages	Disadvantages	Application Scenarios
PID Control	Simple, easy to implement	Poor robustness	Simple environments
Backstepping	High precision, well-established theory	Complex computation	Nonlinear systems
Adaptive Neural Network	Strong adaptability	Difficult convergence analysis	Uncertain systems
Finite-time Control	Fast convergence, good transient performance	Difficult parameter tuning	High-performance requirements
BLF	Constraint guarantee	Complex design	Safety-critical systems

5 Datasets and Performance Evaluation

VisDrone is currently one of the largest UAV aerial datasets, released by the Machine Learning and Data Mining Laboratory of Tianjin University, containing more than 100,000 images and more than 2.4 million annotated boxes. The selected scenes include urban roads, suburban roads, parks, seaside areas, etc., and also cover various weather conditions such as sunny, cloudy, and rainy days. Complementing this extremely well is UAVDT, which is a dataset specifically designed for UAV vehicle detection and tracking, released by Shanghai Jiao Tong University, containing 80 video sequences, explicitly including different lighting conditions such as daytime, nighttime, and foggy days. CARPK is another distinctive UAV dataset, shot from UAV perspectives in large parking lots, with finely annotated approximately 4,000 vehicles. In addition, there are classic UAV tracking datasets such as UAV123, DTB70, and MEAVS. Commonly used evaluation metrics include mAP (mean Average Precision), success rate, precision, center location error, etc.

6 Challenges and Future Trends

Because UAV tracking systems themselves have very clear and interrelated technical difficulties: First is real-time requirements, that is, the computing power of embedded computing units used by UAV platforms is limited; therefore, high-frame-rate and high-accuracy tracking must be achieved under limited computing power. Second is complex environment adaptability; illumination changes, weather factors, and target occlusion will directly interfere with tracking performance. Third is computational resource limitations; deep learning models generally require large amounts of parameter storage and large amounts of computational operations, which are extremely burdensome for embedded devices.

Future development directions include: starting with lightweight models, first developing lighter deep learning models to facilitate real-time inference on edge devices; then doing multi-modal fusion, naturally and reasonably combining information from various sensors such as visible light, infrared, and radar; then extending to multi-target tracking, solving the problem of multi-UAV collaborative tracking of multiple targets; finally conducting real machine deployment, truly and reliably transferring the obtained achievements to real UAV platforms for verification. Therefore, high-speed vehicle tracking technology by UAVs has excellent application prospects in intelligent traffic management, border security, disaster rescue, and other fields.

References

[1] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 779-788.

[2] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 7263-7271.

[3] Redmon J, Farhadi A. YOLOv3: An incremental improvement[EB/OL]. arXiv:1804.02767, 2018-04-08. https://arxiv.org/abs/1804.02767.

[4] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal speed and accuracy of object detection[EB/OL]. arXiv:2004.10934, 2020-04-23. https://arxiv.org/abs/2004.10934.

[5] Jocher G, et al. ultralytics/yolov5 [CP/OL]. GitHub, 2020 [2024-04-01]. https://github.com/ultralytics/yolov5.

[6] Wang C Y, et al. Scaled-YOLOv4[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 13024-13033.

[7] Ultralytics. YOLOv8 [CP/OL]. 2023 [2024-04-01]. https://github.com/ultralytics/ultralytics.

[8] Wang C Y, Yeh I H, Liao H Y M. YOLOR[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 11530-11539.

[9] Ge S S, Lee T H, Zhu G. Adaptive neural network control of a class of MIMO nonlinear systems[J]. International Journal of Adaptive Control and Signal Processing, 1998, 12(8): 667-692.

[10] Polycarpou M M. Stable adaptive neural control Scheme for nonlinear systems[J]. IEEE Transactions on Automatic Control, 1996, 41(3): 447-451.

[11] Bhat S P, Bernstein D S. Continuous finite-time stabilization[J]. IEEE Transactions on Automatic Control, 1998, 43(5): 678-682.

[12] Bhat S P, Bernstein D S. Finite-time stability of continuous autonomous systems[J]. SIAM Journal on Control and Optimization, 2000, 38(3): 751-766.

[13] Tee K P, Ge S S, Tay E H. Barrier Lyapunov functions for the control of output-constrained nonlinear systems[J]. Automatica, 2009, 45(4): 918-927.

[14] Bertinetto L, et al. Fully-convolutional siamese networks for object tracking[C]//ECCV. 2016: 850-865.

[15] Li B, et al. High performance visual tracking with siamese region proposal network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 8971-8980.

[16] Zhu P, et al. Detection and tracking meet drones under challenges[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(1): 1-18.

[17] Kalman R E. A new approach to linear filtering and prediction problems[J]. Journal of Basic Engineering, 1960, 82(1): 35-45.

[18] Julier S J, Uhlmann J K. Unscented filtering and nonlinear estimation[J]. Proceedings of the IEEE, 2004, 92(3): 401-422.

[19] Arulampalam M S, et al. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking[J]. IEEE Transactions on Signal Processing, 2002, 50(2): 174-188.

[20] Comaniciu D, Ramesh V, Meer P. Kernel-based object tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(5): 564-577.

[21] Krstic M, Kanellakopoulos I, Kokotovic P V. Nonlinear and adaptive control design[M]. Hoboken: John Wiley & Sons, 1995.

[22] Slotine J J E, Li W. Applied nonlinear control[M]. Englewood Cliffs: Prentice Hall, 1991.

[23] Khalil H K. Nonlinear systems[M]. 3rd ed. Upper Saddle River: Prentice Hall, 2002.

[24] Chen W H, et al. A nonlinear disturbance observer for robotic manipulators[J]. IEEE Transactions on Industrial Electronics, 2000, 47(4): 932-938.