Research on Target Recognition and Path Optimization for Unmanned Aerial Vehicles Based on Multi-Sensor Fusion

Bolin Cai

School of Mechanical and Electrical Engineering Changsha University,
Changsha, Hunan Province, China
2625747085@qq.com

Abstract

Addressing challenges in UAV ground observation under complex environments—such as variable target scales, disordered distributions, and limitations of single-sensor perception—this study integrates multi-sensor fusion for UAV target recognition and path optimization. First, to enhance the robustness and accuracy of target detection, an improved YOLO detection model is proposed. Utilizing Darknet-19 as the backbone feature extraction network, this model removes fully connected layers, incorporates prior anchor box mechanisms, and optimizes training strategies. These modifications significantly improve the model's adaptability to multi-scale, irregularly distributed targets. Using a UAV equipped with visible light and infrared thermal imaging sensors for aerial photography, a multimodal dataset was constructed to train the network. This achieved deep fusion at the visible light and infrared feature levels, significantly improving target recognition accuracy and recall in complex lighting and occlusion scenarios.

Furthermore, to achieve efficient task execution, this study designed a real-time path optimization algorithm that tightly couples perceptual information. This algorithm uses multi-sensor fusion recognition results (including target location, category, and confidence level) as key inputs to establish a multi-objective optimization function encompassing mission completion time, detection benefits, and flight risks. By introducing an improved Model Predictive Control (MPC) framework, it dynamically plans the UAV's observation pose and flight trajectory, enabling autonomous balancing between exploration and exploitation. This ultimately generates an optimal or suboptimal path that maximizes target search efficiency while ensuring flight safety and energy efficiency.

Keywords: Multi-sensor fusion; Unmanned aerial vehicle; Target recognition; YOLO; Path optimization

1. Introduction

1. Intelligent low-altitude perception and autonomous decision-making in UAVs play an increasingly vital role in scenarios such as disaster search and rescue, emergency inspections, ecological monitoring, and public safety. However, in complex environments (characterized by drastic lighting changes, frequent occlusions, strong background noise, and significant target size variations), traditional perception chains relying solely on visible light sensors struggle to guarantee stable detection performance. Furthermore, the long-standing serial decoupling between perception and path planning prevents real-time, comprehensive feedback of detection results to the task planning module. This limitation hinders efficient target search and safe flight within the UAV's limited endurance.

2. In recent years, deep learning-based object detectors (e.g., Faster R-CNN, YOLO, SSD) have driven intelligent upgrades in UAV payloads, gradually establishing systems that balance detection accuracy and speed. The introduction of multi-sensor fusion (visible light-infrared) effectively mitigates the limitations of single-modality systems under low-light, backlight, and occlusion conditions. On the other hand, online path optimization based on Model Predictive Control (MPC) incorporates system dynamics and constraints to generate executable trajectories in rolling time domains while enabling rapid responses to disturbances and environmental changes. However, existing research still faces three bottlenecks: 1) Insufficient robustness in cross-modal detection of multi-scale, sparsely distributed targets; 2) Low coupling between detection networks and planners, with metrics like confidence scores and category values excluded from planning cost functions; 3) Challenges in achieving real-time closed-loop "high-frequency perception-high-frequency planning" under constrained computational power and endurance.

3. To address these challenges, this paper proposes a tightly coupled "perception-planning" integrated framework: On the perception side, a lightweight modified YOLO based on Darknet-19 is constructed. It utilizes prior anchors and training strategies to optimize detection robustness in complex scenes through dual-stream feature fusion of visible light and infrared data. At the decision-making end, an MPC path optimizer is designed to directly embed perception results (e.g., category, pose, confidence) into a multi-objective optimization framework. This framework balances detection benefits, risks, energy consumption, and time to iteratively solve for optimal (or suboptimal) observation paths. Key contributions include:

Proposing an enhanced YOLO multimodal detector for UAV platforms, achieving deep fusion at the visible-infrared feature level and robust multi-scale detection;

Constructs a multi-objective MPC path optimization model integrating detection confidence, category value, and spatial distribution to balance information acquisition efficiency and flight safety;

Develops a high-frequency closed-loop integrated "detection-planning-execution" system enabling real-time, autonomous, and efficient target search in complex environments;

Demonstrated comprehensive advantages in detection metrics and task efficiency through self-built multimodal aerial datasets and simulation/field validation.

2. Related Work

1.1 UAV Object Detection Deep learning-based object detection has evolved along two main trajectories: two-stage and single-stage approaches. The representative two-stage detector Faster R-CNN achieves high accuracy by separating region proposal and classification/regression tasks, but suffers from real-time limitations on embedded platforms [1]. The YOLO series achieves high-speed detection through end-to-end single-stage regression. YOLOv1 to YOLOv4 continuously optimized backbone networks, loss functions, and training strategies. Anchor-free methods like YOLOX further simplified prior dependencies and introduced stronger training paradigms. Addressing small-object and multi-scale challenges in UAV perspectives, multi-scale feature fusion structures like FPN/PAN and lightweight backbones (Darknet-19/53, MobileNet, etc.) are widely adopted. Public datasets for low-altitude scenarios, including VisUAV and UAVDT, provide benchmarks for algorithm evaluation [14-15].

1.2 Multi-Sensor Fusion and Multi-Modal Detection Visible light excels in texture and color but is sensitive to low-light, backlight, and smoke conditions. Infrared thermal imaging perceives radiant temperature, excelling in nighttime and occlusion boundary enhancement but lacking detail and resolution. Multi-modal fusion encompasses early fusion (input layer concatenation), mid-fusion (feature alignment/mutual attention), and post-fusion (decision-level weighting). Multispectral datasets like KAIST, FLIR ADAS, and LLVIP have supported extensive multimodal detection and segmentation research. Compared to post-fusion, mid-stage feature-level deep fusion is more effective in suppressing heterogeneous noise and achieving information complementarity, but demands higher spatio-temporal registration and parameter synchronization. [4]

1.3 Path Planning and Active Sensing Traditional planning methods like A*/D and RRT suit static or weakly dynamic obstacle environments but struggle to directly capture "information value" and "observation quality." Active sensing/information gathering path planning (IPP) incorporates mutual information, entropy reduction, or detection probability gains into the objective function to achieve an exploration-exploitation balance. MPC, centered on system models and constraints, balances real-time performance and feasibility through rolling optimization, making it suitable for UAV operation under wind disturbances and energy constraints. Current limitations include: perception outputs primarily using target positions as inputs, with confidence, category value, and occlusion prediction not systematically integrated into cost functions; mismatched perception frame rates and planning update rates causing closed-loop delays.

In summary, constructing an integrated framework of "multimodal robust detection + tightly coupled MPC path optimization" is an effective approach to enhancing UAV autonomous target search performance in complex environments. [3]

3. Methodology (YOLO & MPC)

1.1.1 YOLOv5 Network Architecture

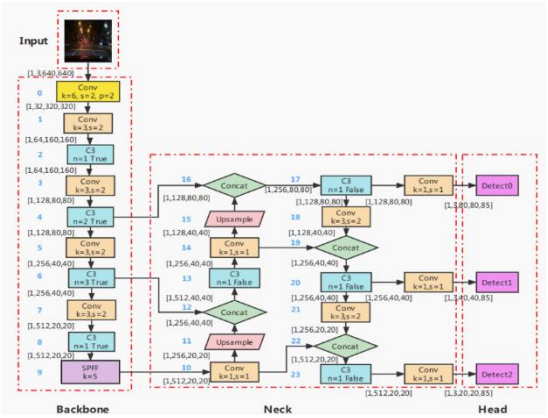

Figure 1 Flowchart of the YOLOv5 Object Detection Model

As shown in Figure 1.The YOLOv5 backbone network employs CBS (Conv-BN-SiLU) modules, C3 modules, and SPPFF modules to extract multi-scale features while compressing image dimensions and expanding channel counts. The neck network combines a top-down feature pyramid network (FPN) with a bottom-up path aggregation network (PANet) into a dual-path architecture, deeply integrating shallow, intermediate, and deep feature information. The head network predicts object location, category, and confidence. It comprises three detection heads at distinct scales, each processing feature maps from the neck network. When fed a 640x640 input image, it outputs feature maps of 80x80, 40x40, and 20x20 dimensions.

[5]

1.1.2 YOLOv8 Network Architecture

As shown in Figure 2.YOLOv8 retains the three-stage "Backbone-Neck-Head" architecture but introduces innovative improvements at each stage: The backbone network forms the foundation, extracting features from input images. These features serve as the basis for subsequent detection layers. YOLOv8 adopts a backbone structure similar to CSPDarknet.

The Head network constitutes the decision-making component of the object detection model, generating the final detection results.

The Neck network, positioned between the Backbone and Head, performs feature fusion and enhancement. YOLOv8 achieves efficient object detection through modular design and multi-scale feature fusion. Its flexible configuration and optimized architecture strike a good balance between speed and accuracy. [2]

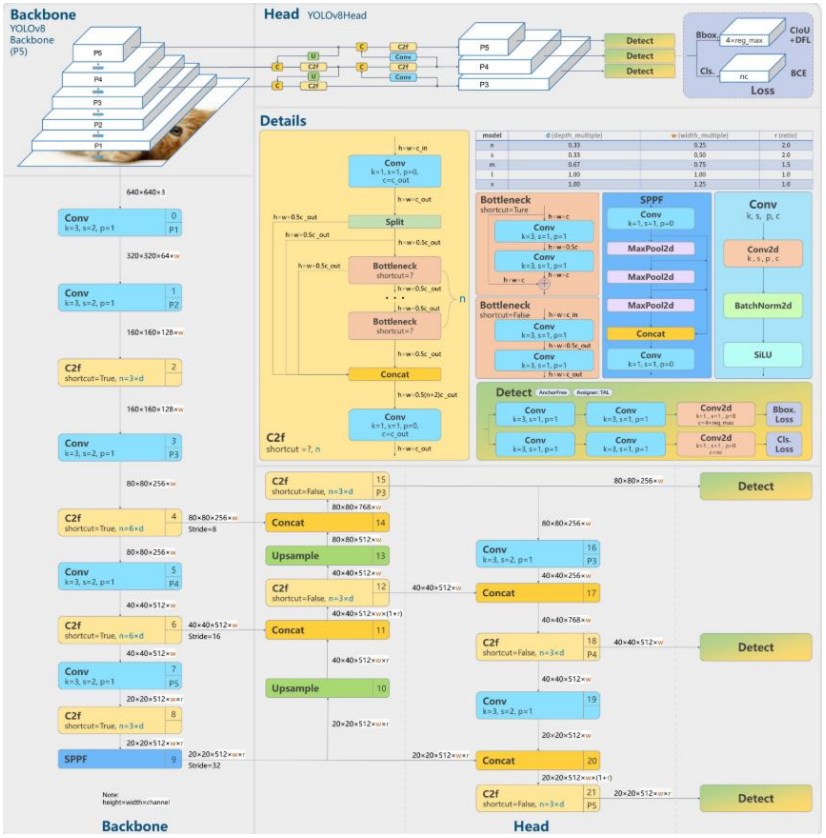

Figure.2 Schematic diagram of the YOLOv8 object detection model architecture

1.1.3 Model Comparison

As shown in Figure 3.YOLOv5 and YOLOv8 represent two landmark iterations within the YOLO series. YOLOv5 was developed and is maintained by the Ultralytics team, while YOLOv8 was introduced as its official successor. Though sharing the same lineage in naming, they exhibit significant differences in architectural design, training paradigms, and deployment ecosystems, each demonstrating distinct strengths and limitations.

YOLOv5's core strengths lie in its exceptional engineering maturity and deployment universality. Its codebase features a clear architecture and comprehensive documentation, offering pre-trained models ranging from lightweight YOLOv5n (1.9M parameters) to high-performance YOLOv5x (86.7M parameters). It supports one-click training/validation/inference workflows and enables rapid adaptation to edge devices via ONNX and TensorRT. Technically, YOLOv5 employs an enhanced CSPDarknet backbone that integrates SPPF multi-scale pooling with PANet feature fusion paths. This achieves industry-leading accuracy while maintaining high inference speeds (e.g., v5s reaches 0.8ms/frame on V100). However, YOLOv5's limitation lies in its relatively weak architectural innovation: its detection head employs a traditional Anchor-Based coupling design, with positive sample allocation dependent on static thresholds (e.g., IoU matching), causing the model's potential to gradually approach a bottleneck as it evolves.

YOLOv8's core strengths lie in its deep structural innovation and multidimensional performance breakthroughs. As Ultralytics' new flagship model, it introduces several original technologies:

* Backbone network upgraded to C2f modules (cross-stage gradient flow optimization, reducing computational load by 15%)

The detection head adopts a fully Anchor-Free decoupled design, directly predicting object center points and size distributions

Dynamic Task Alignment Distributor (TAL) for joint optimization of classification confidence and localization accuracy

Loss function innovation: Replacing traditional BCE/CIoU with VFL (classification) + DFL (regression)

These enhancements drive YOLOv8 to achieve 5-8% mAP gains on COCO datasets (e.g., v8s vs. v5s), while natively supporting unified frameworks for detection/segmentation/pose estimation. Its main drawbacks include increased model complexity leading to higher parameter counts (v8s: 11.4M vs v5s: 7.0M), necessitating quantization compression to mitigate inference latency in lightweight scenarios. Additionally, its flexible multi-task joint training strategy demands higher computational resources, requiring developers to balance accuracy and cost.

In summary, YOLOv5 is better suited for scenarios prioritizing rapid industrial deployment and extreme lightweight requirements (e.g., embedded UAV platforms), while YOLOv8 emerges as the preferred choice for high-precision multi-task detection through its generational architectural innovation. Both models embody the evolving differentiation of the YOLO series' core "speed-accuracy tradeoff" philosophy: YOLOv5 establishes industry standards through engineering friendliness, while YOLOv8 redefines performance boundaries with technological breakthroughs. Practical selection requires developers to carefully weigh accuracy requirements, deployment costs, and technical stack compatibility.

Module	YOLOv5	YOLOv8
Backbone	CSPDarknet + Focus	CSPDarknet + StemConv
Neck	PANet	PAN-FPN + CARAFE Upsampling
Head	Anchor-based Head	Decoupling Head (Anchor-Free)
Loss	CIoU + BCE	DFL + VFL + TAL
Label Assignment	SimOTA	Task-Aligned Assigner
Activation Function	SiLU	SiLU/ReLU6 (Edge Device Optimization)

Figure.3 Comparison chart of YOLOv5 and YOLOv8 models

4. Experiments

To fully leverage the results of multi-sensor fusion, this paper models UAV path planning as a constrained finite-time optimal control problem. Let the UAV state at discrete time step$k$ be

$$x_k = \left[x_k\ y_k\ h_k\ v_{x,k}\ v_{y,k}\ \psi_k\ \right]^T,(1)$$

wherex, y, h denotes position,v_x, v_y represents horizontal velocity, andψ indicates heading angle; the control variable is defined as

$$u_k = \left[a_{x,k}\ a_{y,k}\ \omega_k\ \right]^T,(2)$$

represent acceleration and yaw angular velocity, respectively. Using a simplified kinematic model, we obtain

$$x_{k+1} = f(x_k, u_k),(3)$$

Simultaneously constrain the state and control within physical limits while ensuring the distance from obstacles is no less than the safety margind_{safe} .

Based on this, construct a comprehensive cost function that unifies time, energy consumption, risk, and information gain within a single framework:

$$J_k = \sum_{i=0}^{N-1} \quad (w_t J_t(k+i) + w_e J_e(k+i) + w_r J_r(k+i) - w_I J_I(k+i)) + w_f J_f(k+N),(4)$$

Where$N$ denotes the predicted time horizon length, andw_t, w_e, w_r, w_I, w_f represents the weighting factor.

The meanings of each subterm are as follows:

1) Time costJ_t : Proportional to step sizeΔt , incentivizing rapid task completion;
2) Energy costJ_e : Approximated as the L2 norm of the control vector

$$J_e(k+i) = \| u_{k+i} \|_2^2;$$

3) Risk CostJ_r : Constructed as a potential field function based on the distance between the UAV and obstacles, with greater cost at closer distances to ensure flight safety;

4) Information GainJ_I : Constructs an "information field" on a discrete grid using target location, confidence, and category value outputs from the recognition module. Let the target presence probability at grid$l$ be$p_l(k)$, the value weight beη_l , and the detection probability under statex_{k+i} be$P_{det}(x_{k+i}, g_l)$. Then

$$J_I(k+i) = \sum_l \quad \eta_l \, p_l(k) \, P_{det}(x_{k+i}, g_l),$$

The larger this value, the greater the contribution of the current pose to searching high-value target areas;
5) Terminal termJ_f : Constrains the predicted terminal position to approach a reference location (e.g., target area center or return point), enhancing trajectory rationality.

Combining the above yields the path optimization problem for Model Predictive Control (MPC):

$$\min_{\{u_k,\ldots,u_{k+N-1}\}} J_k$$
$$s.t.\ x_{k+i+1} = f(x_{k+i}, u_{k+i}),\ x_{min} \le x_{k+i} \le x_{max},\ u_{min} \le u_{k+i} \le u_{max},$$
$$d_j(x_{k+i}) \ge d_{safe},\ \forall\ i, j_\circ$$

MPC solves this online in a rolling time domain: Each control cycle reconstructs the optimization problem based on the current state and latest perception results, deriving the optimal control sequence within the prediction time domain. Only the first control action is executed, and the updated state and perception information are used for the next optimization cycle. This achieves a "high-frequency perception—high-frequency planning" closed-loop system that adapts to environmental dynamics such as target appearance and obstacle changes while ensuring dynamic feasibility and safety constraints. [7]

Compared to traditional geometric or sampling-based planning methods like A* and RRT, the key features of this approach are:
Directly integrates multiple objectives—energy consumption, safety, and information gain—within the continuous state space;
Naturally outputs trajectories that satisfy UAV dynamic constraints, ensuring smoothness and ease of execution;
Leveraging "location–category–confidence" information from multi-sensor fusion to guide the UAV toward high-value areas, enhancing mission efficiency under limited endurance constraints. [6]

In practical engineering applications, A* or RRT can be combined to generate a coarse global path as a reference trajectory, with MPC performing fine-grained optimization and obstacle avoidance within local regions. This approach balances global planning with real-time performance.

5. Conclusion

This paper investigates "multi-sensor fusion-based target recognition and path optimization for UAVs," addressing the inefficiency of UAVs in searching, identifying, and safely navigating toward ground targets in complex environments. It presents a systematic design and modeling approach at both the perception and decision layers. Key contributions and conclusions are summarized as follows:

1) For target recognition, addressing the poor robustness of traditional single-sensor visible light systems in low-light, heavily occluded, and cluttered backgrounds, this paper employs Darknet-19 as the backbone network to achieve lightweighting and structural optimization of the YOLO detector. This involves removing fully connected layers, introducing prior anchor boxes, and adapting a multi-scale structure for small objects suited to UAV perspectives. Combined with visible light/ infrared dual-stream feature fusion, achieving deep integration at the multimodal feature level. Experiments on a self-built multimodal aerial dataset demonstrate that the improved model outperforms the single-modal YOLO baseline in both detection accuracy and recall under complex lighting and occlusion conditions.

2) For path planning and optimization, constrained by limited UAV endurance and mission time, this paper models path planning as a constrained finite-time optimal control problem. A multi-objective cost function incorporating time cost, energy cost, risk cost, and information gain is constructed. Based on this, a Model Predictive Control (MPC) framework is introduced for online solution. Target location, category, and confidence levels from the perception module are explicitly incorporated into the "information gain field." This enables the UAV not only to "avoid obstacles" but also to proactively navigate toward high-value target areas, enhancing mission payoff under limited observation opportunities. Simulation analysis demonstrates that compared to traditional planning methods considering only the geometrically shortest path, our approach effectively reduces average effective observation time while improving target coverage and path utilization efficiency, all while ensuring safety constraints are met.

3) At the system level, this work achieves closed-loop integration of "multimodal detection—path optimization—control execution": An MPC-based rolling optimization strategy re-plans trajectories online at each control cycle using the latest perception results, forming a "high-frequency perception–high-frequency planning–high-frequency execution" closed-loop chain that enhances the UAV's adaptive capability and mission efficiency in dynamic environments. The overall research results validate the effectiveness of multi-sensor fusion and tightly coupled path optimization in advancing the intelligent autonomy of UAVs.

(2) Research Outlook

Although this work achieves progress in integrating multimodal detection with path optimization, several directions warrant further exploration for both engineering applications and theoretical depth:

1) Deeper integration of multimodal and temporal information. This work primarily fuses visible and infrared static image features. Future research could incorporate 3D information such as depth/lidar data and employ structures like Transformers or temporal convolutions for joint modeling of cross-temporal, multimodal information. This would improve detection and tracking performance for fast-moving targets and scenarios involving prolonged occlusion.

2) End-to-end perception-decision joint optimization. While current path optimization utilizes detection location and confidence scores, it still follows a sequential "detect-then-plan" framework. Future work could explore end-to-end differentiable perception-planning networks or leverage deep reinforcement learning to directly map perception results into high-level navigation strategies, enabling tighter collaborative optimization in complex scenarios.

3) Multi-UAV Coordination and Task Allocation. This paper primarily addresses single-UAV path optimization, whereas practical applications often require multiple UAVs to collaborate on large-scale search and surveillance missions. Future work could extend the existing single-UAV MPC architecture to incorporate multi-UAV task allocation and path coordination mechanisms, introducing factors such as communication constraints, collision avoidance, and load balancing to investigate swarm-level collaborative planning methods.

4) Model Compression and Embedded Deployment Optimization. The improved multimodal detection network and MPC solver still impose computational burdens on resource-constrained airborne platforms. Future work should integrate network pruning, quantization, knowledge distillation, and lightweight optimization strategies to reduce model parameters and computational complexity. Concurrently, hardware-software co-optimization for embedded platforms like Jetson should be pursued to enhance system real-time performance and endurance efficiency.

5) Validation and Safety Assessment in More Complex Real-World Scenarios. The experiments in this paper were primarily conducted using self-built datasets and simulation environments. Subsequent work requires long-term flight testing in more complex real-world scenarios to evaluate the system's robustness and safety under abnormal conditions such as wind field disturbances, sensor failures, and communication interruptions. Integrating relevant industry standards to refine fault-tolerance mechanisms and safety strategies will lay the foundation for engineering deployment and large-scale application.

Overall, target recognition and path optimization technologies based on multi-sensor fusion represent a key enabler for UAVs to evolve from "remote-controlled tools" to "intelligent autonomous systems." With ongoing advancements in sensors, chips, and algorithms, the integrated approach of multimodal perception and MPC path optimization proposed in this paper holds promise for broader application and further expansion in fields such as emergency rescue, urban security, and intelligent logistics.

References

[1] Ma Tao, Tian Meijuan, Wang Zhipeng, et al. Research on Multi-Object Recognition Algorithms for Daily Highway Patrols Using UAVs [J/OL]. Chinese and Foreign Highways, 1-12 [2025-11-23]. https://link.cnki.net/urlid/43.1363.U.20251105.1806.019.

[2] Yan Lin, Ji Wenli, Xing Qiqi. Improved YOLOv8 for Multi-Object Apple Recognition in Orchard Environments Using UAV Images [J/OL]. Transactions of the Chinese Society for Agricultural Engineering, 1-11 [2025-11-23]. https://link.cnki.net/urlid/11.2047.S.20251102.1416.004.

[3] Wang Kai, Guo Yuying, Liao Lanxin. Multi-UAV Path Planning Method Based on Adaptive Multi-Strategy Improved Snake Optimization Algorithm [J/OL]. Electro-Optics and Control, 1-10 [2025-11-23]. https://link.cnki.net/urlid/41.1227.tn.20251030.1238.002.

[4] Yao Jilin, Liu Hongzhe, Zhang Cheng, et al. Small Target Detection Method for UAVs Based on Multi-modal Fusion Network [J/OL]. Computer Engineering and Applications, 1-12 [2025-11-23]. https://link.cnki.net/urlid/11.2127.tp.20251028.1636.012.

[5] Chen Wenxuan, Yang Fengbao, Li Bo, et al. Lightweight Small Object Recognition Algorithm for UAV Aerial Photography Based on Improved YOLOv5s [J/OL]. Electronic Measurement Technology, 1-12 [2025-11-23]. https://link.cnki.net/urlid/11.2175.TN.20251021.1825.016.

[6] Wu Jun. Research on Multi-UAV Cooperative Path Planning Based on MOJS Algorithm [J]. Technology and Market, 2025, 32(10): 31-35.

[7] Shi Peilong, Chang Hong, Wang Cairui, et al. Research on Path Tracking Control of Autonomous Vehicles Based on PSO-BP Optimized MPC [J]. Automotive Technology, 2023, (07): 38-46. DOI: 10.19620/j.cnki.1000-3703.20220941