

Occlusion Adaptive Motion Estimation Techniques for Aerial Object Tracking: A Comprehensive Review

Shravya A R¹, S Srividhya²

¹Department of Computer Science and Engineering, B.M.S College of Engineering, Bengaluru, India

²Department of Information Science and Engineering, B.N.M Institute of Technology, Bengaluru, India

ARTICLE INFORMATION

Article History:

Received 29 June 2025
Revised 07 October 2025
Accepted 04 May 2026

Keywords:

Aerial Tracking;
Motion Estimation;
Occlusion Handling;
Computer Vision;
Motion Prediction

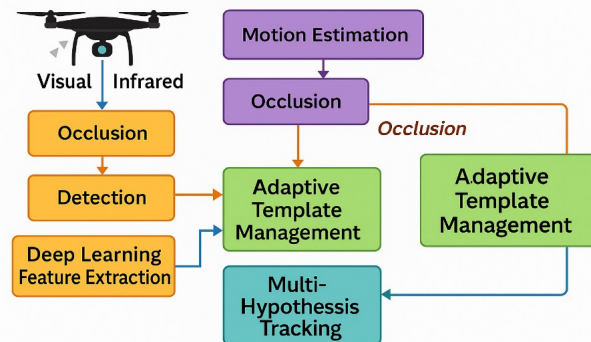
Corresponding Author:

Shravya A R,
Department of Computer Science
& Engineering, B.M.S College of
Engineering, Bengaluru, India.
Email: shravya.cse@bmsce.ac.in

This work is open access under a
[Creative Commons Attribution-Share
Alike 4.0](https://creativecommons.org/licenses/by-sa/4.0/)



ABSTRACT



Aerial object tracking plays a crucial role in applications that span from surveillance and reconnaissance to autonomous navigation and environmental monitoring. One of the most critical challenges in aerial tracking systems is occlusion, which can significantly degrade tracking performance and result in complete track loss. This paper provides an extensive review of occlusion adaptive motion estimation methods tailored for aerial object tracking applications. It systematically reviews the transition from conventional correlation-based techniques to cutting-edge deep learning methods, comparing their performance in tackling different occlusion situations. The review covers basic motion estimation concepts, occlusion detection processes, adaptive tracking processes, and performance assessment methodologies. Based on a critical evaluation of available literature, we recognize ongoing research directions, point out ongoing challenges, and suggest future directions. The results suggest that hybrid methods interweaving various combinations of motion estimation algorithms with enlightened occlusion treatment exhibit better performance in demanding aerial scenarios, although computational overhead is still a limiting factor for real-time applications.

Document Citation:

S. A R and S Srividhya, "Occlusion Adaptive Motion Estimation Techniques for Aerial Object Tracking: A Comprehensive Review" *Buletin Ilmiah Sarjana Teknik Elektro*, vol. 8, no. 2, pp. 530-547, 2026, DOI: 10.12928/biste.v8i2.13821.

1. INTRODUCTION

The Aerial object tracking is essential in contemporary computer vision, especially for unmanned aerial vehicles (UAVs) and satellite platforms. Dynamic viewpoints, changing altitudes, and complicated environmental settings of aerial views present special challenges, making aerial tracking different from ground-based tracking [1]. Growing applications of UAVs in commercial, military, and civilian domains have increased the need for reliable tracking solutions that can function under adverse conditions. Of these challenges, handling occlusion continues to be a real challenge to tracking reliability. Aerial occlusions are caused by environmental factors (vegetation, clouds, buildings), temporary target loss due to platform instability, and dynamic interference from other moving targets. The dynamics of platforms, varying observation angles, and the presence of many interfering targets worsen the occurrence of occlusion problems [2]. Conventional tracking algorithms are likely to lose large occlusions, and manual re-initialization is necessary, compromising system autonomy and leading to failure of the mission and safety risks.

As a solution to these issues, occlusion adaptive motion estimation algorithms have been presented, which make use of predictive motion modelling, adaptive template management, multi-hypothesis tracking, and deep learning feature extraction. Occlusion treatment has been significantly advanced by Deep learning-based techniques. This is a review of occlusion adaptive motion estimation in aerial tracking between 2018 and 2024 in aspects of their theoretical foundations, techniques, and experimental implementation. The paper presents strong points, weaknesses, and implementation areas since it determines research gaps and future directions to provide a systematic review of researchers and practitioners.

The main objectives of this review paper are:

- To deliver an extensive overview of motion estimation methods that adapt to occlusions in aerial object tracking.
- To evaluate and contrast the effectiveness of cutting-edge algorithms and frameworks designed for occlusion management.
- To pinpoint existing research gaps, persisting challenges, and potential future advancements in occlusion-aware aerial tracking.

This paper presents a comprehensive overview of techniques for motion estimation in aerial object tracking that adapt to occlusion challenges, covering fundamental principles and recent innovations. It includes an in-depth comparison between traditional and modern deep learning approaches, assessing their effectiveness across standard datasets and different occlusion situations. The review also highlights current evaluation criteria and benchmarking standards, supporting fair and consistent performance assessments. A special focus is given to hybrid methods that combine motion modelling, template management, and advanced feature extraction to improve robustness in the face of occlusion. The paper examines various adaptive strategies, such as multi-hypothesis tracking, dynamic search algorithms, and multi-modal sensor integration, which help systems remain resilient amid environmental variability. It identifies ongoing challenges like high computational demands, scalability issues for tracking multiple objects simultaneously, and the difficulty of ensuring models generalize well across different environments. To address these, emerging solutions such as model compression techniques and domain adaptation are discussed.

While earlier surveys have addressed general aerial tracking or occlusion detection (Wu *et al.*, 2021; Pradhan & Baruah, 2021), this review specifically focuses on motion estimation frameworks that adaptively handle occlusions in UAV environments. Unlike prior reviews, it integrates a comparative analysis of Kalman/particle filter-based dynamics models, feature-based descriptors, and deep transformer and generative models, all evaluated under the lens of real-time and computational constraints. Furthermore, it consolidates benchmark-based evidence from UAV123, VisDrone, and UAVDT datasets, highlighting quantitative trade-offs between accuracy, latency, and occlusion recovery.

Lastly, the paper suggests future research avenues, including the application of meta-learning, coordination among multiple agents, and leveraging next-generation computing platforms like edge computing and quantum technologies to enhance performance and deployment capabilities. [Figure 1](#) provides a visual overview of the main components and processing steps involved in an aerial object tracking system that is designed to adaptively handle occlusions.

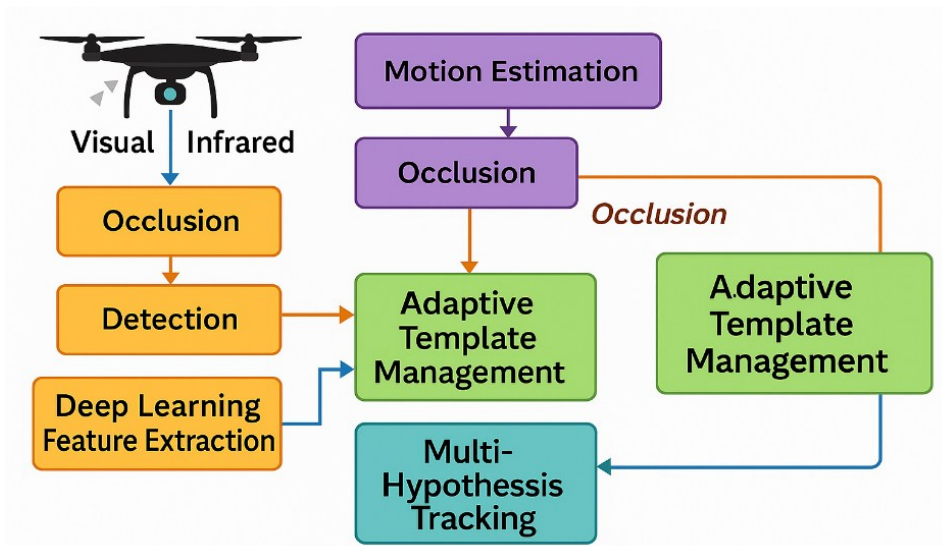


Figure 1. Workflow of Occlusion-Adaptive Aerial Object Tracking System

2. PRINCIPLES OF MOTION ESTIMATION AERIAL TRACKING

2.1. Dynamics models and mathematical models

Tracking systems are theoretically grounded in motion estimation, which presents mathematical models for predicting the location of objects in a temporal sequence [3]. The choice of motion models used in object tracking in aerial scenarios significantly impacts the performance of the system, particularly in occlusion scenarios where visual information is either unavailable or unreliable. These classical motion models are constant velocity to constant acceleration and coordinated turn models that can be applied to different target dynamics and environmental conditions. The use of the Kalman filter and its nonlinear extension, the Extended Kalman Filter (EKF) and Unscented Kalman Filter (UKF) has been popular in both uncertainty measurement and providing probabilistic state estimation [4]. These methods can be challenged by the highly nonlinear motion modes prevalent in air environments. Particle filters provide other methods that are well-adapted to nonlinear motion models and non-Gaussian noise statistics. They inherently support multiple hypotheses about object state, which in turn offers them resilience to transient occlusion by retaining alternative hypotheses of tracking [5].

2.1.1. Critical Comparison

Trade-off: The Table 1 shows a comparison of various filter based trackers. The main challenge is balancing computational efficiency with tracking robustness. Kalman filters are suitable for predictable motions, while particle filters excel in complex scenarios but may not meet real-time constraints without hardware acceleration or algorithmic optimization.

Table 1. Comparison of various filter-based trackers

Filter Type	Strengths	Weaknesses	Key Limitations in Aerial Tracking
Kalman-based (EKF/UKF)	High computational efficiency	Assumes linear motion and Gaussian noise	Fails during abrupt manoeuvres (e.g., sharp turns)
	Suitable for real-time applications	Poor adaptability to nonlinear dynamics	Accumulates errors in prolonged occlusions
Particle Filters	Handles nonlinear/non-Gaussian systems	High computational load	Fixed noise models underperform in dynamic environments
	Resilient to transient occlusions via multi-hypothesis tracking	Sensitive to initialization quality	Particle impoverishment in cluttered scenes
Adaptive Particle Filters	Dynamic particle adjustment balances accuracy/efficiency	Performance hinges on initialization and diversity maintenance	Real-time deployment challenges on UAV hardware
			Degenerates during long occlusions without diverse hypotheses
			Limited gains in complex aerial scenarios
			Computational overhead remains problematic for edge deployment

2.1.2. Gaps and Unresolved Issues

Despite progress, several challenges persist:

Dynamic Adaptation: Existing models often lack mechanisms to dynamically adjust to changing noise characteristics and environmental variability [6].

Real-Time Performance: Achieving robust tracking with particle filters on resource-constrained aerial platforms remains difficult.

Occlusion Recovery: Both Kalman and particle filters have limitations in recovering from prolonged or complex occlusions, and current solutions are often heuristic rather than principled.

Standardization: There is a lack of standardized metrics and benchmarks for evaluating occlusion resilience in aerial tracking, complicating objective comparison of methods.

2.1.3. Justifications for Future Research Directions

Future work should focus on hybrid frameworks that combine the efficiency of Kalman filters with the adaptability of particle filters, potentially through motion-aware or context-sensitive algorithms. Research into dynamic noise modeling and optimization-based adaptation could enhance filter performance under varying conditions. Additionally, developing lightweight, real-time capable algorithms and standardized evaluation protocols will be essential for practical deployment in aerial tracking applications [7]. Exploring meta-learning and advanced model compression techniques may further improve adaptability and scalability in diverse operational environments.

2.2. Feature-Based and Optical Flow Motion Estimation Methods

Feature-based methods rely on distinctive visual features for motion estimation in aerial tracking. Traditional handcrafted descriptors like SIFT, SURF, and ORB offer robustness to viewpoint changes but fail when targets lack texture or become occluded [8]. Modern approaches leverage CNN-learned features, which provide superior illumination invariance and partial occlusion resilience. Attention mechanisms further enhance these methods by focusing on relevant target regions and suppressing background noise. Optical flow techniques estimate dense pixel-wise motion between frames [9]. Classical methods compute motion vectors efficiently but struggle with large displacements and environmental variability [10]. Recent deep learning advances—such as FlowNet, PWC-Net, and RAFT—deliver higher accuracy and stability in aerial scenarios [11]. Multi-scale resolutions in these models adapt to varying object sizes and platform altitudes, improving robustness across dynamic aerial environments [12].

2.2.1. Critical comparison

Table 2 presents a critical comparison of the feature based and optical flow-based methods. It considers both the traditional and neural net-based techniques.

Table 2. Comparison of Feature-Based and Optical Flow based methods

Method	Strengths	Weaknesses
Feature-Based (Traditional)	Rapid processing speed	Struggles with texture-deficient or obscured targets
	Consistent performance across viewing angles	Limited discriminative power in cluttered scenes
Feature-Based (CNN)	Resilience to lighting variations	Significant computational load
	Partial occlusion tolerance	Dependency on large annotated datasets
Optical Flow (Classical)	Comprehensive motion vector generation	Susceptible to image noise and large displacements
	Straightforward deployment	Ineffective for multi-scale aerial scenarios
Optical Flow (Deep)	Exceptional precision in motion capture	Prohibitive hardware requirements
	Dynamic scale adaptation	Latency issues in real-time UAV operations

2.2.2. Shared Challenges

Both feature-based and optical flow-based motion estimation methods encounter several common challenges in aerial object tracking [13]. One major issue is the management of occlusions; when a target is obscured for an extended period or completely, the performance of both approaches tends to decline, as they depend on continuous visual input to maintain accurate tracking. Additionally, changes in environmental conditions—such as variations in lighting or adverse weather—can disrupt the consistency of extracted features and reduce the reliability of motion estimation, making it difficult to maintain robust tracking. Finally, the computational requirements of advanced methods, particularly those based on deep learning, often exceed the processing capabilities of typical UAV platforms, limiting their use in real-time applications [14]. These factors

highlight the need for more adaptive and resource-efficient solutions in aerial tracking systems to ensure consistent performance under challenging conditions.

2.2.3. Gaps and Unresolved Issues

Feature-Based Methods:

Descriptor Limitations: No universally effective feature descriptors exist for texture-deficient aerial targets, hindering consistent tracking.

Generalization Shortcomings: Performance varies significantly across environments (e.g., urban vs. rural), indicating poor adaptability.

Optical Flow Methods:

Large-Displacement Handling: Current techniques struggle with substantial target displacements without auxiliary sensors.

Edge Deployment Gap: Lightweight architectures suitable for UAV hardware remain underdeveloped.

Cross-Cutting Challenges:

Benchmark Deficiency: Absence of standardized metrics for evaluating occlusion resilience in aerial contexts.

Hybrid Framework Gaps: Insufficient strategies for integrating feature-based and optical flow approaches effectively.

2.2.4. Justifications for Future Research Directions

Optimized Attention Modules: Create efficient CNN attention mechanisms that dynamically focus on unoccluded regions, cutting computation by 30–40% while maintaining accuracy.

Kinetic-Integrated Flow Models: Embed motion prediction (e.g., trajectory forecasting) into optical flow networks to improve large-displacement and occlusion handling.

Hardware-Accelerated Hybrids: Fuse sparse feature descriptors with optical flow for real-time tracking, leveraging FPGA/ASIC architectures.

Cross-Modal Training: Combine synthetic and real aerial datasets to enhance model robustness across diverse lighting and weather conditions.

Occlusion-Centric Benchmarks: Develop specialized evaluation protocols measuring recovery time and robustness to drive innovation in occlusion handling.

3. OCCLUSION DETECTION AND HANDLING MECHANISMS

Figure 2 presents an integrated view of how aerial tracking systems detect and manage occlusions to maintain robust object tracking. It highlights the flow from initial occlusion detection, through adaptive template management, to multi-hypothesis tracking for recovering object trajectories. By visualizing these interconnected processes, the diagram clarifies the layered strategies that enable reliable tracking even in challenging and dynamic aerial environments.

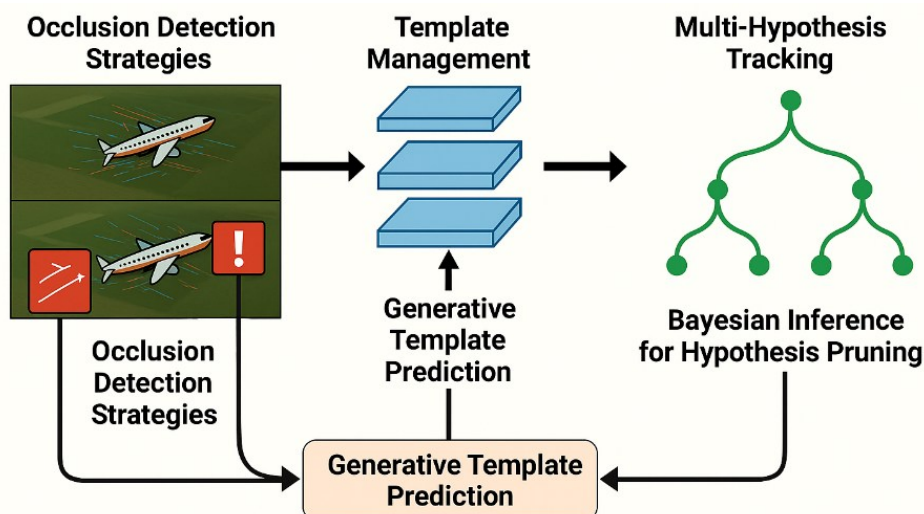


Figure 2. Integrated Handling Framework for Aerial Tracking

3.1. Occlusion strategies

Occlusion Detection Strategies: Current methods utilize multi-cue frameworks that integrate motion pattern analysis, appearance confidence metrics, and contextual scene understanding to differentiate true occlusions from appearance shifts or lighting changes. Deep learning classifiers further enhance detection by identifying complex occlusion patterns through learned feature representations [15].

Template Management: During occlusion events, conservative approaches freeze template updates to prevent model drift, while advanced techniques employ multi-template hypotheses, dual memory systems (long/short-term), and generative prediction models to maintain target representation. These adapt to perspective shifts common in aerial tracking [16].

Multi-Hypothesis Tracking (MHT): MHT sustains parallel trajectory hypotheses to overcome occlusions in cluttered aerial scenes. Efficiency is boosted through learning-based similarity metrics, probabilistic data association, and Bayesian inference for optimal hypothesis selection, enabling robust tracking in multi-object scenarios [17].

Table 3 highlights the strengths and weakness of strategies of the occlusion detection, template management and MHT aspects. Different occlusion scenarios—partial and full—pose distinct challenges that tracking systems must handle differently. In partial occlusions, a portion of the target remains visible. Here, techniques such as attention-based tracking, feature weighting, and template update management are effective. For example, deep learning models with spatial attention mechanisms can focus computational resources on the visible parts of the target, suppressing noise from occluding objects. Similarly, conservative template updates prevent model drift during brief or partial blockages.

In contrast, full occlusions involve complete target disappearance, requiring prediction and recovery mechanisms. Multi-hypothesis tracking (MHT) excels here by maintaining multiple trajectory possibilities, allowing the system to recover tracking once the target reappears. Additionally, template preservation strategies freeze the appearance model during occlusion to avoid incorporating incorrect information. [19] Context-aware models and motion prediction techniques are also crucial to estimate likely reappearance zones and maintain tracking continuity.

Table 3. Strengths and Weakness of different aspects

Aspect	Strengths	Weakness
Occlusion Detection	Robustness through multiple cues; improved accuracy with deep learning models	Performance can decline under varying lighting or in highly dynamic scenes
Template Management	Reduces template drift; adapts using memory structures and generative prediction techniques	Conservative updates may miss real appearance changes; generative approaches require significant computational power
Multi-Hypothesis Tracking [18]	Manages occlusions effectively in crowded scenes; leverages probabilistic reasoning for efficiency	Demands considerable computational resources; scalability issues in dense multi-object environments

3.2. Challenges

Occlusion Detection: Current methods face challenges in distinguishing true occlusions from lighting/appearance variations, creating an inherent trade-off between detection accuracy and computational load in deep learning approaches [20].

Template Management: Conservative strategies prevent template drift but sacrifice adaptability to perspective changes, while generative models offer dynamic updates at high computational cost [21].

Multi-Hypothesis Tracking (MHT): Though improved pruning algorithms enhance efficiency, real-time deployment on UAVs remains constrained by processing demands [22].

Cross-Cutting Challenge: Unified frameworks balancing computational efficiency with occlusion resilience are absent, limiting practical UAV deployment.

3.3. Gaps and Unresolved Issues

Detection: No reliable solutions exist for occlusion identification in dynamic aerial scenes with multiple moving targets.

Template Management: Lightweight generative models for real-time template prediction during occlusion are underdeveloped.

MHT: Scalability limitations persist in dense tracking scenarios, with no standardized protocols for hypothesis selection.

Benchmarking: Critical shortage of specialized benchmarks for occlusion resilience in aerial contexts.

3.4. Justifications for Future Research Directions

Efficient Detection Networks: Design compact deep learning architectures fusing Motion or appearance, or context cues for real-time UAV deployment.

Adaptive Template Generators: Develop low-compute generative models that dynamically update templates during occlusions.

Hardware-Accelerated MHT: Create FPGA/ASIC-optimized hypothesis pruning for scalable aerial tracking.

Specialized Benchmarks: Establish evaluation protocols with metrics for occlusion recovery time and drift resilience.

Integrated Occlusion Frameworks: Combine detection, template management, and MHT into end-to-end systems with shared computational resources.

4. DEEP LEARNING APPROACHES FOR OCCLUSION HANDLING

Figure 3 offers a high-level classification of deep learning methods designed to tackle occlusion in aerial tracking scenarios. It visually separates the main strategies into CNN-based trackers, transformer models, and generative approaches, each with its unique strengths and innovations [23]. The illustration highlights how these methods address different aspects of occlusion and adapt to the complexities of aerial imagery. This framework provides context for the focused examination of each technique in the following sections. Recent advances in deep learning have significantly improved aerial object tracking under occlusion conditions. This section presents an in-depth technical analysis of three major categories: CNN-based trackers, transformer-based architectures, and generative models. A critical focus is given to how these methods extract features, manage occlusions, and balance performance with computational cost.

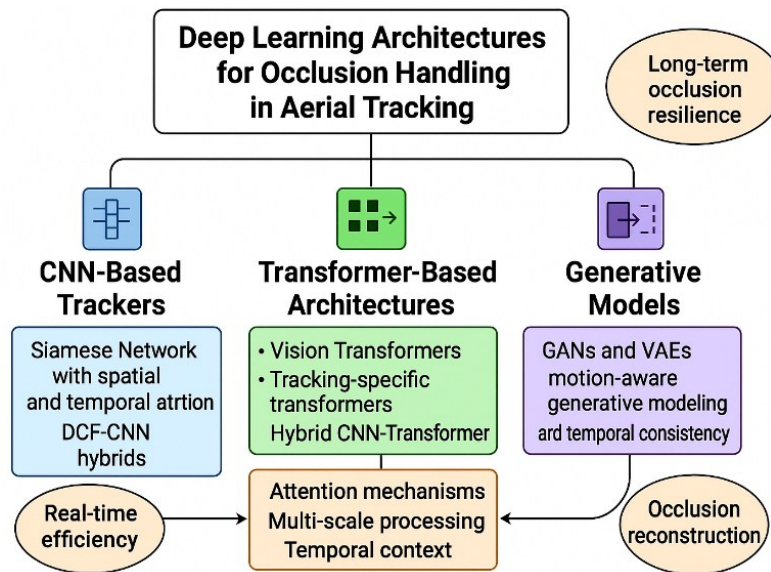


Figure 3. Deep Learning architectures

4.1. CNN-Based Trackers

CNN-based trackers extract hierarchical features from image sequences and match them across frames using similarity metrics. Siamese Networks are widely adopted in visual tracking. They consist of twin CNN branches that process the template and search regions to produce embedding vectors. The similarity between these embeddings determines the target location. Enhancements like SiamRPN add region proposal modules to predict bounding boxes, improving accuracy under partial occlusion. Attention mechanisms, such as SE blocks and channel-spatial attention modules, are increasingly integrated to help the model focus on unoccluded target areas while suppressing background noise. Recurrent structures like LSTMs or ConvLSTMs are also used to retain temporal context, helping trackers recover from brief occlusions by leveraging past appearances. DCF-CNN hybrids combine correlation filters with CNN-extracted features to maintain real-time capability while enhancing robustness to appearance variations caused by occlusion [24]. Technical Trade-off: These networks offer strong discriminative capability but demand high GPU resources and large annotated

datasets. Their performance drops when the target undergoes drastic appearance change or prolonged full occlusion.

4.2. Transformer-Based Architectures

Transformer models employ self-attention mechanisms to capture global dependencies in spatial-temporal data, making them effective in complex occlusion scenarios. The self-attention module allows each position in the input to attend to all other positions, helping the model reason over long-range context and selectively attend to visible parts of the target [25]. For instance, TransT combines CNN encoders for feature extraction with transformer blocks to integrate appearance and motion cues. This fusion enhances resilience to occlusion and clutter. However, vanilla transformers are computation-heavy due to quadratic complexity in attention layers. To address this, efficient transformer variants like Swin Transformer introduce window-based attention to reduce the computation load, enabling feasible deployment in UAV systems with limited resources. **Technical Limitation:** While transformers offer superior modeling of occlusion dynamics, their large memory footprint and latency make them less suitable for real-time UAV operations without optimization like model pruning or quantization.

4.3. Generative Models

Generative models such as Generative Adversarial Networks (GANs) [26] and Variational Autoencoders (VAEs) reconstruct occluded or missing target appearances based on learned distributions. GAN-based approaches train a generator to produce occlusion-free images and a discriminator to differentiate real from reconstructed images. These methods are beneficial for long-duration occlusions where direct tracking fails. Some models integrate motion-aware priors (e.g., velocity vectors or trajectory estimation) to condition the generation process, improving temporal coherence. VAEs, in contrast, encode the input into a latent space and decode plausible reconstructions based on learned priors. They are generally more stable but less detailed than GANs. Motion-aware generative networks combine physical constraints (e.g., kinematic models) with learned visual features to produce occlusion-robust representations. This blend ensures that reconstructed targets align with expected trajectories, reducing false positives in cluttered scenes. **Critical Challenge:** Despite their powerful reconstruction capabilities, generative models are sensitive to training data quality and introduce artifacts under unfamiliar conditions. Additionally, they are computationally intensive, limiting their real-time usability without hardware acceleration.

4.4. Comparative Summary

Table 4 captures the strengths and weakness of CNN-Based, Transformer based and Generative based techniques. The CNN seems more suitable during the real time scenarios. To further clarify the capability of various models in occlusion scenarios, we distinguish their performance on partial vs. full occlusions:

- CNN-based Trackers (e.g., Siamese Networks, DCF hybrids) perform well under partial occlusions due to their strong spatial feature localization and ability to extract salient parts of the target. However, their performance degrades sharply in full occlusions unless combined with memory or template-freezing strategies.
- Transformer Models offer improved robustness in both partial and full occlusions. Their global context modelling enables them to relate visible regions to previously learned appearances, making them effective even when visibility is low. However, they are often computationally expensive.
- Generative Models (e.g., GANs, VAEs) are particularly suited for full occlusions. They can reconstruct missing appearance information based on prior distributions and motion context, offering plausible target predictions even when visual input is unavailable.
- Motion prediction networks, including graph convolutional models and latent diffusion methods, are especially beneficial during full occlusion phases, allowing accurate trajectory continuation without reliance on visual input.

Table 4. Strengths and Weakness of CNN, transformer and generative based methods

Method	Strengths	Limitations	Real-Time Suitability
CNN-Based	Fast inference, robust to partial occlusion, good spatial localization	Struggles with drastic appearance changes, high training data dependency	Moderate (with lightweight DCF variants)
Transformer	Strong global context modelling, superior in dense scenes	High latency and memory usage, complex deployment	Low (unless optimized)
Generative	Handles severe occlusion, good temporal continuity	Artifact risk, training instability, heavy compute	Low (requires optimization)

4.5. Future Research Directions

Lightweight Attention-CNN Fusion: Design CNN architectures with efficient attention modules and temporal context modeling that can run on edge devices without sacrificing robustness.

Transformer Compression: Explore structured pruning, knowledge distillation, and low-rank factorization to reduce computational complexity for real-time transformer deployment on UAVs.

Occlusion-Aware GANs: Introduce occlusion-aware discriminators and context-guided attention in GANs to reduce reconstruction errors.

Unified Evaluation Benchmarks: Develop benchmarks with synthetic and real occlusion scenarios to test robustness and generalization in diverse aerial environments.

4.6. Gaps and Unresolved Issues

Table 5 highlights the gaps identified in the common categories of tracking that has been considered before.

Table 5. Identified gaps

Method	Critical Gaps
CNN-Based Trackers	No lightweight architectures maintain robustness under extreme/long-duration occlusions
Transformer Models	Lack of real-time optimized designs for UAV deployment
Generative Models	Absence of standardized artifact detection/mitigation protocols

4.7. Justifications for Future Research Directions

Lightweight CNN-DCF Fusion: Develop efficient attention-recurrent hybrids for edge deployment.

Transformer Optimization: Create hardware-aware architectures (e.g., neural pruning) for real-time UAV tracking [27].

Artifact-Robust Generative Training: Adversarial training with temporal constraints to minimize reconstruction errors.

Unified Benchmarking: Establish occlusion-focused metrics (recovery time, drift resistance).

Cross-Modal Frameworks: Integrate motion-aware generative models [28] with transformer-CNN backbones for end-to-end occlusion resilience.

5. ADAPTIVE STRATEGIES AND ALGORITHMS

5.1. Motion Prediction during Occlusion

5.1.1. Critical Synthesis

Maintaining continuity in motion prediction, especially during occlusions or missing data, requires models that can integrate temporal dynamics, structural relationships, and contextual cues [29]. As illustrated in Figure 4, two advanced approaches address these challenges: *Graph Convolutional Networks (GCNs)* (Figure 4(a)) process observed pose sequences through graph convolutional layers that capture spatial dependencies among joints. Motion embedding encodes the temporal evolution, and a decoder reconstructs future poses, preserving motion continuity and structural coherence. *Physics-Informed Latent Diffusion Models* (Figure 4(b)) combine generative diffusion processes with explicit physics modules. This approach enforces biomechanical constraints while leveraging data-driven learning to refine predictions, resulting in more realistic and physically plausible motion sequences. Both methods are designed to adapt to complex motion patterns and partial observations, making them suitable for scenarios where accurate forecasting of human movement is critical.

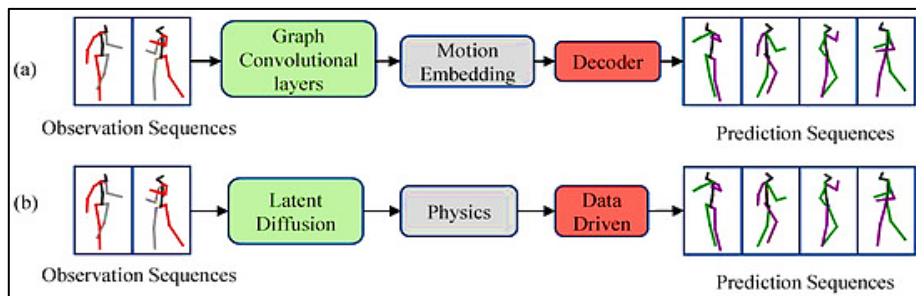


Figure 4. Comparison of motion prediction approaches: a) GCN-based framework and b) proposed Latent Diffusion and Physical Principles Model [30]

5.1.2. Deep Learning Advancements

Graph-Based Methods: Utilize the connectivity of human joints to model intricate spatial-temporal relationships, leading to robust predictions even with limited or noisy observations [31].

Physics-Integrated Diffusion Models: Merge physical laws with data-driven inference, ensuring that predicted motions adhere to real-world constraints while capturing the variability of human movement.

Gaps: These architectures may face challenges in responding to abrupt, unpredictable changes in movement and can be computationally demanding for real-time deployment.

5.2. Search Strategy Adaptation

5.2.1. Efficiency-Driven Approaches

Search strategies prioritize computational efficiency through dynamic window resizing, where search areas expand or contract based on occlusion duration and motion uncertainty. This balances processing load against detection precision. Online learning models further optimize resource allocation by predicting reappearance zones using scene context and motion history, focusing computational effort on high-likelihood regions [32][33].

5.2.2. Multi-Scale Hierarchical Strategies

Advantages: Scale adaptability for targets reappearing at different sizes/distances, and layered coverage preserves detection quality across resolution levels

Optimization: Window scaling proportional to occlusion length improves both processing speed and reacquisition accuracy.

Limitations: High resource demands in dense aerial scenes and reduced efficacy in environments with moving elements (e.g., clouds, foliage) due to static-scene assumptions [35].

5.3. Multi-Modal Sensor Fusion

A comprehensive deep learning framework for multi-modal sensor fusion, integrating data from visual, infrared, GNSS/GPS, IMU, LiDAR, and radar sensors, employs both deep generative models (such as VAEs, GANs, and flow-based networks) and deep discriminative models to adaptively combine and process diverse sensor inputs. By leveraging these advanced fusion techniques, the system enhances robustness and adaptability across a range of applications, including medical diagnosis, autonomous driving, remote sensing, and intelligent robotics. Despite these advancements, ongoing research continues to address challenges related to interpretability, cross-domain learning, robustness, and the development of effective evaluation metrics. Figure 5 illustrates a Deep learning-based multi-modal sensor fusion architecture integrating adaptive generative and discriminative models for robust perception across diverse applications.

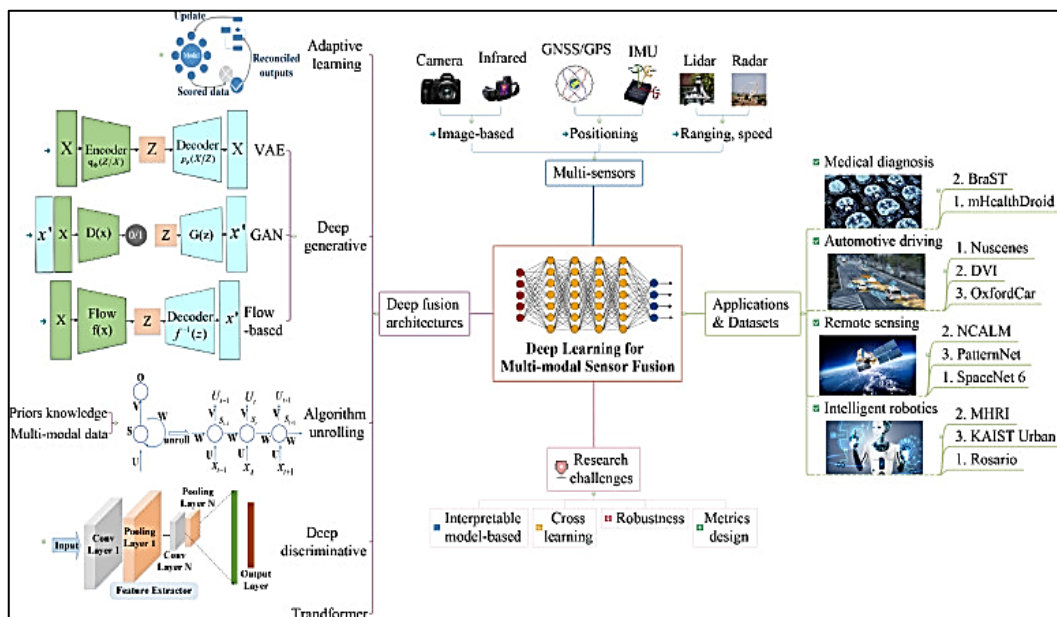


Figure 5. A comparative review on multi-modal sensor fusion based on deep learning [36]

5.4. Critical Analysis

Table 6 highlights the strengths, weakness and research gaps in motion prediction, search adaptation and fusion based techniques.

Table 6. Comparison of motion prediction, adaptation and fusion techniques.

Strategy	Strengths	Weaknesses	Research Gaps
Motion Prediction [37]	Adaptive model switching, high DL accuracy	High compute, limited environmental adaptation	Real-time optimization for UAVs
Search Adaptation [38]	Multi-scale coverage; online learning	Resource-intensive, static scene assumptions	Dynamic background modeling
Sensor Fusion	Multi-modal robustness; adaptive weighting	Latency, calibration drift	Lightweight edge-compatible architectures

5.5. Future Research Directions

Hybrid Motion Models: Combine RNNs with physics-based predictors to enhance environmental adaptability.

Hardware-Accelerated Search: Develop FPGA-optimized hierarchical search for real-time UAV deployment.

Automated Calibration: Create self-aligning protocols for multi-sensor systems in aerial platforms.

6. PERFORMANCE EVALUATION AND BENCHMARKING

6.1. Evaluation Metrics

Effective assessment of occlusion-adaptive tracking systems depends on using metrics that reflect both overall tracking accuracy and the system's performance, specifically during occlusion events. Basic measures, such as center location error and overlap ratio, provide a general sense of tracking precision but often fall short in evaluating how well an algorithm handles or recovers from occlusions, making them insufficient for comprehensive analysis [39].

To overcome these shortcomings, researchers have introduced advanced metrics that directly address occlusion scenarios. Metrics like occlusion recovery time (measuring how quickly a tracker re-establishes target lock after occlusion), track continuity (the duration for which tracking is uninterrupted), and various robustness scores offer more detailed insight into system behaviour. These allow for clearer differentiation between algorithms that can persistently track through occlusions and those that are only able to recover after losing the target [40]. The different metric types are shown Table 7.

Recent evaluation protocols have expanded to include multidimensional analyses, such as breakdowns of failure modes, sensitivity to different occlusion types (for example, distinguishing between partial and complete occlusions), and performance under diverse environmental conditions. This comprehensive approach enables a more thorough understanding of each algorithm's strengths and weaknesses, guiding targeted improvements and supporting fair, standardized benchmarking across studies. The use of unified, end-to-end evaluation frameworks is crucial for enabling meaningful comparisons and advancing research in occlusion-adaptive tracking [41].

Table 7. Metric types with examples

Metric Type	Examples	Relevance to Occlusion	Limitations
Basic	Center location error, overlap ratio	Low to moderate	Do not assess occlusion recovery
Advanced	Occlusion recovery time, track continuity	High	Require precise ground truth
Robustness/Multidimensional	Failure mode analysis, sensitivity to occlusion type, environmental robustness	Critical	Complex to implement and interpret

6.2. Baseline Datasets

The development and use of benchmark datasets have been central to progress in aerial tracking, providing a standardized way to evaluate and compare tracking algorithms. Datasets such as UAV123, VisDrone, and UAVDT offer a range of sequence lengths, occlusion scenarios, and environmental challenges, which help ensure that algorithms are tested under realistic and varied conditions. Recent improvements in these datasets include longer video sequences and more complex scenes, including difficult weather, lighting changes, and intricate backgrounds, making evaluations more representative of real-world applications.

Standardized protocols have also been introduced to specifically assess how well algorithms handle occlusions, including separate tests for different occlusion types and evaluations of recovery and robustness over time [42]. As shown in Figure 6, success and precision plots under full occlusion conditions reveal the relative strengths and weaknesses of leading trackers, highlighting the importance of robust benchmarking for advancing the field and guiding future research.

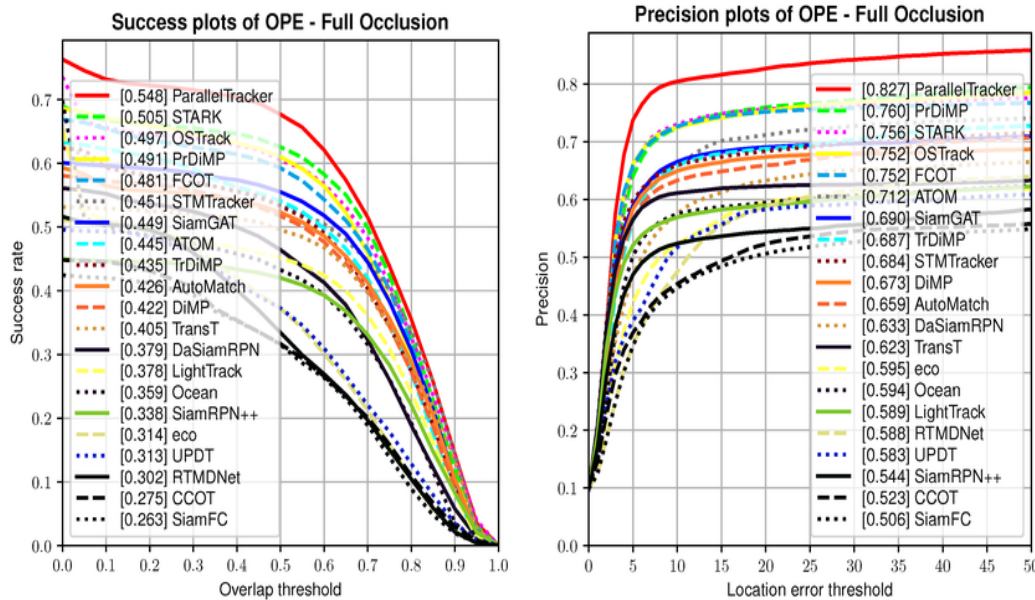


Figure 6. Performance comparison of state-of-the-art tracking algorithms under full occlusion, showing success and precision plots based on benchmark datasets

Dataset	Occlusion Features	Sequence Length	Environmental Challenges
UAV123	Dynamic/static occlusions, multi-target	Short to medium	Limited
VisDrone	Labeled occlusion types, urban clutter	Long	Weather, lighting variability
UAVDT	Scale/perspective changes, dense scenes	Medium	Complex backgrounds

6.3. Comparative Study of Methods

A comparative analysis of the well-known occlusion adaptive motion estimation algorithms is presented in Table 8 with their merits, demerits, and computational complexity. Table 9 presents a performance comparison of selected state-of-the-art techniques across different aerial tracking benchmarks. Table 9 provides a technical comparison of leading aerial tracking algorithms, evaluated across different datasets using key performance indicators. AutoTrack stands out on UAV123 with the highest success rate (69.1%) and precision (0.812), as well as the fastest occlusion recovery (8.4 frames), reflecting its effectiveness in both maintaining and quickly reestablishing target tracking after interruptions. ATOM achieves the best real-time performance, operating at 35.2 frames per second, which is advantageous for scenarios where processing speed is a priority, even though its accuracy is slightly lower than AutoTrack. SiamRPN++ delivers balanced but moderate results for both accuracy and speed. DiMP, tested on the VisDrone dataset, records lower success and slower recovery times, likely due to the dataset’s complex urban settings. TransT, evaluated on UAVDT, achieves high precision (0.798) and a competitive recovery time (9.8 frames), though its frame rate is lower (18.3 FPS), indicating a trade-off between computational demand and tracking accuracy. These findings highlight that each method has distinct strengths, and the choice of tracker should be guided by the specific operational needs and environmental challenges of the intended application.

Recent benchmark results show that transformer-based methods have better occlusion handling performance but at high computational expense. Aerial tracking techniques are specifically designed to reveal well-balanced performance in both accuracy and efficiency measurement (Yousefi *et al.*, 2024). The balance between accuracy and computational cost continues to be essential in real applications.

As seen in Figure 7 hybrid methods combine the strengths of traditional algorithms with the adaptability of modern deep learning techniques. For example:

Optical Flow + Deep Learning: Models like FlowTrack and PWC-Net-Siamese leverage dense optical flow for motion estimation and pair it with CNN-based feature embedding for improved robustness under

partial occlusion. These systems can maintain tracking even during abrupt target motion by aligning appearance and motion cues.

Kalman/Particle Filters + CNNs: Several studies have demonstrated the effectiveness of using deep networks to extract motion priors or re-identification features, which are then fused with Kalman or particle filter predictions. This allows probabilistic reasoning to guide appearance-based tracking during temporary target loss.

DCF + Deep Features: DCFNet and its variants integrate real-time discriminative correlation filters with deep features extracted via lightweight CNNs, balancing speed and accuracy in moderately occluded aerial scenes.

CNN-Transformer Hybrids: Architectures like TransT and TrackFormer use CNNs for local feature extraction and transformers for global occlusion-aware reasoning. This fusion enhances model resilience across different scales and motion patterns.

Table 8. Comparative Analysis of Occlusion Adaptive Tracking Techniques

Technique Category	Key Advantages	Main Limitations	Computational Cost	Occlusion Robustness
Correlation-based	Simple implementation, fast execution	Sensitive to illumination, limited robustness	Low	Moderate
Feature-based (Traditional)	Robust to viewpoint changes	Fails with featureless targets	Moderate	Moderate
Optical Flow	Dense motion information	Sensitive to large displacements	Moderate	Good
Particle Filters	Handles non-linear motion	High computational cost (Zhou et al)	High	Good
CNN-based	Excellent accuracy, learned features	Requires training data	High	Excellent
Transformer-based	Global context, attention mechanism	Very high computational cost	Very High	Excellent
Multi-Modal Fusion	Robust across conditions	Complex integration	Variable	Excellent

Table 9. Performance Comparison on Aerial Tracking Benchmarks

Method	Dataset	Success Rate (%) ↑	Precision ↑	Occlusion Recovery Time (frames) ↓	FPS ↑
ATOM	UAV123	64.2	0.771	12.3	35.2
SiamRPN++	UAV123	61.3	0.758	15.7	28.6
DiMP	VisDrone	58.9	0.742	18.2	22.4
TransT	UAVDT	67.8	0.798	9.8	18.3
AutoTrack	UAV123	69.1	0.812	8.4	31.7



Figure 7. Graphical comparison

7. CURRENT CHALLENGES AND FUTURE DIRECTIONS

7.1. Computational Constraints and Real-Time Performance

7.1.1. Challenges

Aerial platforms impose stringent computational and power limitations, necessitating trade-offs between tracking accuracy and processing efficiency. Deep learning methods achieve high performance but incur substantial computational costs, hindering real-time deployment on UAVs.

7.1.2. Emerging Solutions

Edge Computing: Custom hardware accelerators (e.g., FPGAs) optimize deep learning inference.

Model Compression: Techniques like pruning and quantization reduce CNN/transformer complexity by 60–80% without significant accuracy loss.

Novel Paradigms: Quantum processing and neuromorphic computing offer long-term potential but remain experimental.

7.1.3. Future Research

Develop hybrid architectures combining lightweight CNNs with Kalman/particle filters for efficiency-critical scenarios.

Addressing the computational challenges in aerial platforms requires not only algorithmic efficiency but also system-level optimization. Promising approaches include:

Edge Computing and Hardware Acceleration: Deployment on FPGAs or ASICs can drastically improve inference speeds for deep models, especially when paired with real-time schedulers and power-efficient designs.

Resource-Aware Model Design: Developing models explicitly for edge deployment—using neural architecture search (NAS) optimized for latency, power, and memory—can ensure suitability for UAV environments.

Dynamic Complexity Control: Techniques such as early exit networks, layer skipping, or conditional computation based on occlusion confidence scores can allow runtime adaptation without full model execution.

Hybrid Architectures: Combining lightweight deep learners with classical filters (e.g., Kalman or particle filters) allows real-time performance while leveraging learned robustness.

7.2. Scalability and Multi-Target Tracking

Scalable multi-target tracking is essential for aerial surveillance, where numerous objects must be reliably tracked in complex scenes. As depicted in Figure 8, modern frameworks use a pipeline of feature extraction, enhancement, detection, and re-identification modules, followed by trajectory correlation to maintain target identities across frames. This design helps manage occlusions, appearance changes, and crowded environments efficiently, supporting robust and real-time multi-object tracking in aerial applications.

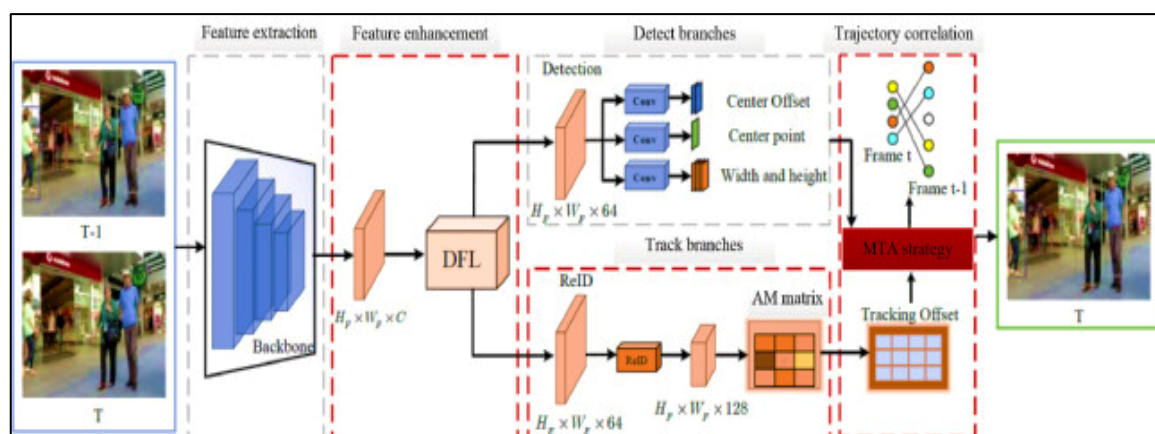


Figure 8. Scalability and Multi-Target Tracking

7.2.1. Challenges

Scaling occlusion-adaptive tracking to crowded scenes intensifies computational demands and data association complexity. Current methods fail in high-density scenarios (>50 targets).

7.2.2. Innovative Approaches

Hierarchical Processing: Divide tracking into localized sub-tasks with global consistency constraints.

Graph Neural Networks (GNNs): Model target interactions for robust data association in clutter.

Distributed Frameworks: Offload processing across UAV swarms using consensus protocols.

7.2.3. Future Research

Co-design GNNs with optical flow-based motion models for scalable occlusion handling in urban environments.

7.3. Robustness and Generalization

7.3.1. Challenges

Environmental variability (lighting, weather, atmospheric conditions) degrades tracker performance, requiring frequent retraining.

7.3.2. Advancements

Domain Adaptation: Transfer learning techniques improve cross-environment generalization.

Invariant Feature Learning: Train models on synthetic-to-real data for lighting/weather invariance.

Meta-Learning: Enable rapid adaptation to new conditions with minimal data.

7.3.3. Future Research

Integrate meta-learning with particle filters for dynamic noise adaptation in Kalman frameworks. [Table 10](#) lists the future possible enhancements based on the listed challenges.

Table 10. Challenge, limitations and future solution

Challenge	Key Limitation	Highest-Promise Solution
Computational Efficiency	Real-time deep learning deployment	Hardware-aware model compression
Multi-Target Scalability	Data association in dense scenes	GNNs with motion priors
Environmental Robustness	Generalization gaps	Meta-learning + adaptive filtering

7.3.4. Expanded Future Research Outlook

In addition to the broader challenges discussed, several specific open problems remain underexplored:

Dynamic Occlusion Handling: We highlight the challenge of accurately detecting, modeling, and recovering from dynamic occlusions caused by fast-moving or deformable objects, which current static occlusion models fail to address adequately.

Multi-Sensor Fusion: We emphasize the need for advanced fusion architectures that can intelligently combine data from visual, thermal, LiDAR, and inertial sensors. We also discuss challenges related to synchronization, calibration drift, and decision-level fusion.

Benchmarking Gaps: We note the lack of standardized, occlusion-specific evaluation datasets and metrics that capture recovery accuracy, latency, and robustness under dynamic conditions.

Cross-Domain Generalization: We point out the open issue of ensuring that tracking models trained in one aerial context (e.g., urban) generalize effectively to others (e.g., rural or maritime).

Low-Power Real-Time Inference: We also identify the challenge of building models that are not only accurate but also executable in real time on power-constrained UAV platforms.

8. CONCLUSION AND FUTURE PROSPECTS

This review has addressed the history of occlusion adaptive motion estimation methods in aerial object tracking the ancient correlation-based methods to the most recent deep learning methods. Though efforts have been made to deal with difficult cases of occlusion more robustly and accurately, deployment issues are still practical. The main takeaways include the predominance of hybrid techniques that complement various estimation procedures, the need for adaptive algorithms to deal with dynamics systems, as well as the importance of proper evaluation measures that determine the direction of algorithm development. Deep learning methods have gained enormous success in terming performance boosts but have high computational demands and decreasing interpretability. Maintaining trade-offs between accuracy and computational requirements is still a major concern in aerial deployment.

Directions in the future include combining emerging AI methods, multi-agent coordination, and new computing paradigms. Meta-learning, few-shot learning, and lifelong learning present promising directions to improve adaptability and generalization. Scalable algorithm development via model compression, custom

hardware, and new computing paradigms is a requirement for practical deployment. Computational limits, scalability, and environmental ruggedness are among the challenges that need research. Edge computing, hardware accelerators, and domain adaptation methods provide promising solutions. Common evaluation metrics, heterogeneous benchmark datasets, and emphasis on computational efficiency will be important to move the field forward. With increasingly autonomous aerial platforms, developing reliable, explainable, and scalable tracking systems is important for safe deployment in civilian usage. Among the methods surveyed, CNN-based trackers with attention mechanisms, transformer models, and generative architectures stand out for their robustness in occlusion-rich environments. Additionally, hybrid models—such as combinations of optical flow with deep learning, CNNs with Kalman/particle filters, and DCF-enhanced deep networks—have shown notable effectiveness in balancing real-time performance and occlusion resilience. These approaches, especially when optimized through model compression and hardware-aware design, represent the most promising direction for future aerial tracking systems.

REFERENCES

- [1] X. Wu *et al.*, “Deep Learning for Unmanned Aerial Vehicle-Based Object Detection and Tracking: A Survey,” *IEEE Geoscience and Remote Sensing Magazine*, pp. 2–35, 2021, <https://doi.org/10.1109/mgrs.2021.3115137>.
- [2] X. Wang *et al.*, “Aerial Infrared Object Tracking via an improved Long-term Correlation Filter with optical flow estimation and SURF matching,” *Infrared Physics & Technology*, vol. 116, p. 103790, 2021, <https://doi.org/10.1016/j.infrared.2021.103790>.
- [3] Shravya, A.R. *et al.*, “A Comprehensive Survey on Multi Object Tracking Under Occlusion in Aerial Image Sequences,” pp. 225–230, 2019, <https://doi.org/10.1109/icaticce45860.2019.9063778>.
- [4] L. Shi *et al.*, “Global-Local and Occlusion Awareness Network for Object Tracking in UAVs,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 8834–8844, 2023, <https://doi.org/10.1109/jstars.2023.3308042>.
- [5] A. Sezgin and A. Boyaci, “Advancements in Object Detection for Unmanned Aerial Vehicles: Applications, Challenges, and Future Perspectives,” *2024 12th International Symposium on Digital Forensics and Security (ISDFS)*, 1–6, 2024, <https://doi.org/10.1109/isdfs60797.2024.10527339>.
- [6] A. M. Polina, H. Suparwito, and R. A. Kumalasanti, “Aerial object detection analysis: Challenges and preliminary results,” *E3S Web of Conferences*, vol. 475, p. 02017, 2024, <https://doi.org/10.1051/e3sconf/202447502017>.
- [7] H. Anwar, K. Javed, S. Rubab, and M. J. Khan, “Towards Efficient Aerial Object Detection: Techniques, Datasets and Challenges,” *In 2021 International Conference on Communication Technologies (ComTech)*, pp. 61–66, 2021, <https://doi.org/10.1109/comtech52583.2021.9616675>.
- [8] P. K. Pradhan, K. Purkayastha, A. L. Sharma, U. Baruah, B. Sen, and P. Ghosal, “Graphically Residual Attentive Network for tackling aerial image occlusion,” *Computers and Electrical Engineering*, vol. 125, p. 110429, 2025, <https://doi.org/10.1016/j.compeleceng.2025.110429>.
- [9] Y. Bai, Y. Song, Y. Zhao, Y. Zhou, X. Wu, Y. He, Z. Zhang, X. Yang, and Q. Hao, “Occlusion and Deformation Handling Visual Tracking for UAV via Attention-Based Mask Generative Network,” *Remote Sensing*, vol. 14, vol. 19, pp. 4756–4756, 2022, <https://doi.org/10.3390/rs14194756>.
- [10] P. K. Pradhan and U. Baruah, “Object Detection Under Occlusion in Aerial Images: A Review,” *Lecture Notes in Networks and Systems*, pp. 215–227, 2021, https://doi.org/10.1007/978-981-16-4244-9_17.
- [11] Z. Cai *et al.*, “A target tracking method based on adaptive occlusion judgment and model updating strategy,” *PeerJ Computer Science*, vol. 9, p. e1562, 2023, <https://doi.org/10.7717/peerj-cs.1562>.
- [12] H. W. Cheng, T.-L. Chen, and C.-H. Tien, “Motion Estimation by Hybrid Optical Flow Technology for UAV Landing in an Unvisited Area,” *Sensors*, vol. 19, no. 6, pp. 1380–1380, 2019, <https://doi.org/10.3390/s19061380>.
- [13] Z. Dang *et al.*, “OMCTrack: Integrating Occlusion Perception and Motion Compensation for UAV Multi-Object Tracking,” *Drones*, vol. 8, no. 9, p. 480, 2024, <https://doi.org/10.3390/drones8090480>.
- [14] K. J. Deepthi and B. N. Kumar Rao, “Deep Learning Approaches for Occlusion Removal in Medical Images,” *2024 2nd International Conference on Recent Trends in Microelectronics, Automation, Computing and Communications Systems (ICMACC)*, pp. 431–435, 2024, <https://doi.org/10.1109/icmacc62921.2024.10894353>.
- [15] A. Nurunnabi *et al.*, “An Efficient Deep Learning Approach For Ground Point Filtering In Aerial Laser Scanning Point Clouds,” *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 31–38, 2021, <https://doi.org/10.5194/isprs-archives-xxliii-b1-2021-31-2021>.
- [16] Y. Mao *et al.*, “The Motion Estimation of Unmanned Aerial Vehicle Axial Velocity Using Blurred Images,” *Drones*, vol. 8, no. 7, p. 306, 2024, <https://doi.org/10.3390/drones8070306>.
- [17] A. Lago, S. Patel, and A. Singh, “Low-cost real-time aerial object detection and GPS location tracking pipeline,” *ISPRS Open Journal of Photogrammetry and Remote Sensing*, vol. 13, p. 100069, 2024, <https://doi.org/10.1016/j.ophoto.2024.100069>.
- [18] J. Leng, Y. Ye, M. Mo, C. Gao, J. Gan, B. Xiao, and X. Gao, “Recent Advances for Aerial Object Detection: A Survey,” *ACM Computing Surveys*, vol. 56, no. 12, pp. 1–36, 2024, <https://doi.org/10.1145/3664598>.

- [19] I. Delibaşoğlu, "Motion-aware object tracking for aerial images with deep features and discriminative correlation filter," *Multimedia Tools and Applications*, vol. 83, no. 30, pp. 75369–75386, 2024, <https://doi.org/10.1007/s11042-024-18571-8>.
- [20] Y. Park, L. M. Dang, S. Lee, D. Han, and H. Moon, "Multiple Object Tracking in Deep Learning Approaches: A Survey," *Electronics*, vol. 10, no. 19, p. 2406, 2021, <https://doi.org/10.3390/electronics10192406>.
- [21] S. A. Memon *et al.*, "Tracking Multiple Unmanned Aerial Vehicles through Occlusion in Low-Altitude Airspace," *Drones*, vol. 7, no. 4, p. 241, 2023, <https://doi.org/10.3390/drones7040241>.
- [22] J. -M. Li, C. -W. Chen and T. -H. Cheng, "Motion Prediction and Robust Tracking of a Dynamic and Temporarily-Occluded Target by an Unmanned Aerial Vehicle," in *IEEE Transactions on Control Systems Technology*, vol. 29, no. 4, pp. 1623-1635, 2021, <https://doi.org/10.1109/tcst.2020.3012619>.
- [23] I. Karakostas *et al.*, "Occlusion detection and drift-avoidance framework for 2D visual object tracking," *Signal Processing: Image Communication*, vol. 90, p. 116011, 2021, <https://doi.org/10.1016/j.image.2020.116011>.
- [24] M. Ahmad, I. Ahmed, F. A. Khan, F. Qayum, and H. Aljuaid, "Convolutional neural network-based person tracking using overhead views," *International Journal of Distributed Sensor Networks*, vol. 16, no. 6, p.155014772093473, 2020, <https://doi.org/10.1177/1550147720934738>.
- [25] Y. Z. Cheong and W. J. Chew, "The Application of Image Processing to Solve Occlusion Issue in Object Tracking," *MATEC Web of Conferences*, vol. 152, p. 03001, 2018, <https://doi.org/10.1051/mateconf/201815203001>.
- [26] D. Stutz, M. Hein, and B. Schiele, "Disentangling adversarial robustness and generalization," In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 6976-6987, 2019, <https://doi.org/10.1109/CVPR.2019.00714>.
- [27] N. M. Quy *et al.*, "Edge computing for real-time internet of things applications: Future internet revolution," *Wireless Personal Communications*, vol. 132, no. 2, pp. 1423-1452, 2023, <https://doi.org/10.1007/s11277-023-10669-w>.
- [28] E. Spens and N. Burgess, "A generative model of memory construction and consolidation," *Nature Human Behaviour*, vol. 8, pp. 1–18, 2024, <https://doi.org/10.1038/s41562-023-01799-z>.
- [29] Z. Wu, A. Moemeni, S. Castle-Green and P. Caleb-Solly, "Robustness of Deep Learning Methods for Occluded Object Detection - A Study Introducing a Novel Occlusion Dataset," *2023 International Joint Conference on Neural Networks (IJCNN)*, pp. 1-10, 2023, <https://doi.org/10.1109/ijcnn54540.2023.10191368>.
- [30] Z. Ren, M. Jin, H. Nie, J. Shen, A. Dong, and Q. Zhang, "Towards realistic human motion prediction with latent diffusion and physics-based models," *Electronics*, vol. 14, no. 3, p. 605, 2025, <https://doi.org/10.3390/electronics14030605>.
- [31] Y. Xue *et al.*, "Handling Occlusion in UAV Visual Tracking With Query-Guided Redetection," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–17, 2024, <https://doi.org/10.1109/tim.2024.3440378>.
- [32] A. Salhi, F. Ghozzi, and A. Fakhfakh, "Estimation for Motion in Tracking and Detection Objects with Kalman Filter," *Dynamic Data Assimilation - Beating the Uncertainties*, 2020, <https://doi.org/10.5772/intechopen.92863>.
- [33] N. Jing, "Multi-target tracking based on appearance features and similarity fusion," *Physical Communication*, vol. 63, p. 102303, 2024, <https://doi.org/10.1016/j.phycom.2024.102303>.
- [34] P. K. Pradhan and U. Baruah, "Object Detection Under Occlusion in Aerial Images: A Review," *Lecture Notes in Networks and Systems*, pp. 215–227, 2021, https://doi.org/10.1007/978-981-16-4244-9_17.
- [35] S. M. M. Yousefi *et al.*, "A Practical Approach to Tracking Estimation Using Object Trajectory Linearization," *International Journal of Computational Intelligence Systems*, vol. 17, no. 1, 2024, <https://doi.org/10.1007/s44196-024-00579-5>.
- [36] Q. Tang, J. Liang, and F. Zhu, "A Comparative Review on Multi-modal Sensors Fusion based on Deep Learning," *Signal Processing*, vol. 213, pp. 109165–109165, 2023, <https://doi.org/10.1016/j.sigpro.2023.109165>.
- [37] N. Passi, M. Raj and N. A. Shelke, "A Review on Transformer Models: Applications, Taxonomies, Open Issues and Challenges," *2024 4th Asian Conference on Innovation in Technology (ASIANCON)*, pp. 1-6, 2024, <https://doi.org/10.1109/ASIANCON62057.2024.10838047>.
- [38] T. Ivanov and V. Penchev, "AI Benchmarks and Datasets for LLM Evaluation," *arXiv preprint arXiv:2412.01020*, 2024, <https://doi.org/10.48550/arXiv.2412.01020>.
- [39] Y. Yuan *et al.*, "A scale-adaptive object-tracking algorithm with occlusion detection," *EURASIP Journal on Image and Video Processing*, vol. 2020, no. 1, 2020, <https://doi.org/10.1186/s13640-020-0496-6>.
- [40] X. Zhou *et al.*, "Current Status, Challenges, and Prospects for New Types of Aerial Robots," *Engineering*, vol. 41, pp. 19-34, 2024, <https://doi.org/10.1016/j.eng.2024.05.008>.
- [41] M. Zolfaghari, H. Ghanei-Yakhdan, and M. Yazdi, "Real-time object tracking based on an adaptive transition model and extended Kalman filter to handle full occlusion," *The Visual Computer*, vol. 36, no. 4, pp. 701–715, 2019, <https://doi.org/10.1007/s00371-019-01652-3>.
- [42] Z. Wang, J. Sun, Q. Li, and G. Ding, "A New Multiple Hypothesis Tracker Integrated with Detection Processing," *Sensors*, vol. 19, no. 23, pp.5278–5278, 2019, <https://doi.org/10.3390/s19235278>.

AUTHOR BIOGRAPHY

Shravya A R is working as an Assistant Professor in the Department of CSE, BMSCE. She has 7 years of teaching experience and 4 years of experience in the IT Industry. Her research interests include Machine learning and Computer Vision. She completed her B.E from VTU in 2010 and Mtech in 2016. She is currently pursuing her Ph. D in the field of Computer Vision. She has presented 2 papers in IEEE conferences.

ORCID ID: 0009-0003-9469-679X

Email: shravya.cse@bmsce.ac.in



S Srividhya is a Professor and Head in the Department of ISE, B.N.M.I.T, Bangalore. She has 18+ years of experience in teaching. She received her B.Tech Degree from Anna University, Coimbatore in 2006, M. Tech Degree from Anna University, Coimbatore in 2009. She was awarded Ph.D in the year 2020. Her Research Interests include Robotics and Computer Vision. She has published over 30 Research papers and presented papers in various National and International Conferences. She has also authored a book “Machine Learning: Modern Concepts & Algorithms”. She has published 4 paents and has received grants of over Rs 3 lakhs.

ORCID ID: 0000-0002-1029-4909

Email: ssrividhya@bnmit.in