

Enhancing Trust in AI Techniques for Object Detection and Tracking in Drone Video Streams through Temporal Information

Executive Summary

The deployment of Artificial Intelligence (AI) for object detection and tracking in Unmanned Aerial Vehicle (UAV) video streams presents transformative opportunities across diverse sectors, from search and rescue to autonomous navigation. However, the inherent complexities of drone imagery—including variable altitudes, low resolution, occlusions, and rapid motion—pose significant challenges to the reliability and trustworthiness of these AI systems. This report details how the strategic application of temporal information, which involves analyzing the continuity and evolution of data across successive video frames, fundamentally enhances the accuracy, robustness, and ultimately, the trustworthiness of AI models. By leveraging techniques such as Kalman filters, optical flow, and recurrent neural networks, temporal AI addresses critical issues like object re-identification, motion blur compensation, and identity consistency. Furthermore, the integration of Explainable AI (XAI) is crucial for interpreting these complex temporal decisions, thereby fostering human understanding and confidence in autonomous drone operations. This comprehensive analysis underscores that temporal information is not merely an enhancement but a foundational requirement for developing reliable and trusted AI solutions in dynamic drone environments.

1. Introduction: The Landscape of AI in Drone Vision

1.1. Overview of Object Detection and Tracking in UAVs

Object detection is a foundational computer vision task focused on identifying and localizing objects within a single image frame. This typically involves drawing bounding boxes around detected entities and classifying them into predefined categories, answering the fundamental

questions of "what" an object is and "where" it is located.¹ Complementing this, object tracking extends the capabilities beyond single-frame identification by continuously monitoring and following objects across successive frames in a video sequence. This process aims to maintain the object's unique identity and trace its trajectory over time, providing answers to "what," "where," "when," and "how" an object moves.¹

The integration of Artificial Intelligence, particularly deep learning methodologies, has revolutionized these capabilities. AI-driven systems enable efficient, cost-effective, and automated processes that are critical for a wide array of real-world applications.⁴ Unmanned Aerial Vehicles (UAVs), commonly known as drones, are at the forefront of this technological integration. Their inherent advantages, such as flexible flight altitudes and diverse imaging perspectives, make them invaluable tools for applications spanning surveillance, search and rescue, traffic monitoring, and comprehensive environmental analysis.³ Nevertheless, these dynamic operational scenarios introduce a unique set of challenges for AI vision systems, which must be meticulously addressed to ensure reliable and trustworthy performance.

1.2. Inherent Challenges of Drone Video Streams

Drone video streams present a highly complex and dynamic environment that significantly challenges both traditional and advanced AI algorithms for object detection and tracking. These challenges directly impact the reliability and, by extension, the trustworthiness of AI systems in real-world deployments.

One prominent challenge stems from **variable flight altitudes and diverse perspectives**.⁵ UAVs operate across a wide range of altitudes, leading to substantial changes in object scale. An object appearing medium-sized or large at low altitudes might become significantly smaller when captured from higher elevations. This variability necessitates detection algorithms with robust spatial generalization capabilities. Furthermore, different viewing angles, such as top-down versus oblique views, provide distinct visual cues, requiring models to maintain consistent and accurate detection across varying perspectives.⁵

Low resolution and the presence of small objects pose another critical hurdle. In low-resolution settings, small objects like pedestrians and bicycles often appear too blurred to be reliably detected, resulting in degraded performance of existing methods. Enhancing dataset resolution is crucial for providing richer, clearer object features, as small-scale objects inherently possess fewer distinguishable features, making them difficult to differentiate from the background.⁵

The issue of **object category enrichment and domain gap** is also significant. Many existing UAV object detection datasets are predominantly focused on vehicle categories, failing to encompass the diverse range of objects encountered in practical scenarios. Datasets like DOTA, while offering oriented bounding box annotations, are often captured from high-altitude imaging perspectives that differ substantially from typical UAV viewpoints. This creates a significant domain gap, leading to degraded model performance when applied to actual drone imagery.⁵

Background clutter and the presence of similar objects further complicate detection and tracking. Drone videos frequently capture complex environments with numerous objects, making it arduous to distinguish the target from its surroundings. Visual similarities between the target and background elements or other non-target objects can easily distract the tracking algorithm, leading to misidentification or loss of track.⁷

Occlusion and target disappearance represent a major impediment to continuous tracking. Objects can be partially or fully blocked from view, resulting in a loss of critical visual and motion information.⁷ This disruption makes it challenging for the AI to maintain a precise appearance model or accurately predict an object's movement. If a target leaves the field of view or remains fully occluded for an extended period, the system's ability to re-detect and re-establish its identity becomes paramount.⁷

Motion blur and fast motion are common in drone footage due to rapid target movement or swift camera maneuvers. Such conditions can render visual details indistinct, making it difficult to maintain a consistent appearance model for the object. Extreme speeds can also cause objects to move across frames too quickly, leading to a loss of critical visual features like shape and color.⁷

Scale and aspect ratio variation are inherent to dynamic drone operations. An object's size changes dynamically with its distance from the camera, complicating feature extraction. Similarly, an object's aspect ratio can shift due to rotations or perspective changes, requiring the tracker to adjust its bounding box, which is particularly challenging for extreme variations.⁷

Lighting changes further exacerbate these issues, as variations in illumination can drastically alter an object's appearance, affecting its color, texture, or brightness, and potentially obscuring vital motion information.⁷

Computational constraints pose a practical limitation. Deep learning algorithms, while powerful, are computationally intensive, presenting significant challenges for drones with limited onboard processing capabilities.⁹ Offloading computation to cloud or edge resources can alleviate this, but it introduces its own set of drawbacks, including increased latency and substantial energy costs for communication. Real-time performance is critical for autonomous drone applications, as delayed results can render the information useless or even dangerous.¹⁰

Furthermore, **dataset bias and annotation challenges** are frequently encountered in aerial object detection. Datasets specifically tailored for drone imagery require meticulous annotation, and models trained on standard images may not be appropriate for detecting objects in aerial contexts.⁹ Finally, certain drone types, such as flapping-wing drones, can introduce

severe jitter between shots due to their aerodynamic structure, further impacting tracking stability.⁷

The quality of input data has a profound and cascading effect on the reliability of AI systems. When image resolution is low, object categories are limited, or datasets are biased, the resulting model performance is inherently compromised. For instance, blurred small objects may not be reliably detected, and models trained on mismatched data domains will exhibit

degraded performance.⁵ This diminished performance directly undermines the reliability of the AI system. If an AI frequently misses objects, misidentifies them, or loses track, human operators will inevitably lose confidence in its capabilities, especially in high-stakes applications like search and rescue or security. This highlights that fundamental data challenges, if left unaddressed, create a direct bottleneck for operational reliability and trust, irrespective of the sophistication of the algorithms employed.

Moreover, the challenges encountered in drone video streams, such as scale variation, viewpoint changes, motion blur, and occlusion, are rarely isolated phenomena; they frequently co-occur. For example, a fast-moving drone might capture small objects that are simultaneously blurred and briefly disappear behind obstacles.⁵ Addressing any one of these challenges in isolation, such as merely improving small object detection, may not yield significant overall performance gains if the other interconnected issues are not simultaneously tackled. This necessitates holistic, multi-faceted solutions capable of handling these compounded challenges, rather than piecemeal improvements. The inherent complexity arising from these intertwined issues naturally makes the process of building trust more arduous, as system failures can originate from a confluence of factors.

A critical consideration is the trade-off between onboard processing capabilities and the need for real-time performance and trustworthiness. Drones operate with limited computational resources.⁹ While offloading intensive computations to cloud or edge servers can enhance processing power, this approach introduces increased latency and significant energy costs for communication.¹⁰ However, real-time performance is paramount in dynamic drone applications, as delayed results can render the information obsolete and potentially hazardous.¹⁰ This creates a fundamental dilemma: highly accurate, complex models might necessitate offloading, thereby increasing latency and dependence on external infrastructure, which can erode trust in autonomous, on-the-fly decision-making. Conversely, lightweight onboard models, while faster, might sacrifice accuracy. The implication is that achieving trust in real-time drone AI requires a delicate balance between computational efficiency, latency, and model complexity, potentially driving the development of optimized lightweight models or intelligent hybrid offloading strategies that dynamically adapt to available resources and network conditions.¹¹

Table 1: Key Challenges in Drone Video Object Detection and Tracking

Challenge Category	Specific Challenge	Description	Impact on AI Performance	Relevant Sources
Visual Characteristics	Low Resolution & Small Objects	Small objects appear blurred, lack features.	Poor detection, low reliability ⁵	⁵
	Scale & Aspect Ratio Variation	Object size/shape changes with distance/orientation.	Complicates feature extraction, inaccurate bounding boxes ⁷	⁷
	Motion Blur & Fast Motion	Rapid movement blurs details, loses	Difficult to maintain	⁷

		visual features.	appearance model, loss of tracking ⁷	
Environmental Factors	Occlusion & Disappearance	Objects are partially/fully blocked or exit view.	Loss of visual/motion data, tracking failure, re-detection needed ⁷	⁷
	Background Clutter & Similar Objects	Target difficult to distinguish from complex surroundings.	Distraction, misidentification ⁷	⁷
	Lighting Changes	Illumination variations alter object appearance.	Affects color/texture, obscures motion ⁷	⁷
Operational Constraints	Variable Altitudes & Viewpoints	Objects appear differently based on drone position.	Demands strong spatial generalization, robust detection across perspectives ⁵	⁵
	Computational Limitations	Deep learning is resource-intensive for onboard processing.	Increased latency, energy costs, limits real-time capability ⁹	⁹
	Dataset Bias & Annotation	Datasets lack diversity or are not UAV-specific.	Degraded model performance due to domain gap, unsuitable for aerial images ⁵	⁵
	Jitter	Unstable camera movement (e.g., flapping-wing drones).	Impacts tracking stability ⁷	⁷

2. Fundamentals of Temporal Information in Computer

Vision

2.1. Defining Temporal Cues and Context in Video Analysis

Temporal information in computer vision refers to the rich data and intricate relationships derived from the sequential progression and evolution of frames over time within a video stream.² Unlike static image analysis, which processes each frame as an isolated entity, temporal analysis inherently considers the continuity, dynamics, and changes that occur between consecutive frames. This approach allows for a more profound understanding of the visual content.

A key application of temporal information is **temporal video segmentation**, a process that divides a video into distinct, meaningful segments based on temporal boundaries or content changes. This involves identifying transitions between different shots, events, or activities, often utilizing algorithms, including machine learning techniques, to detect subtle or abrupt changes in the video's content over time.¹² For instance, in a surveillance feed, this process could differentiate between periods of activity and inactivity, or in a sports broadcast, it could separate gameplay from a halftime show.¹²

Temporal understanding is also crucial for **action detection and localization**. Action detection goes beyond merely classifying actions; it also localizes their temporal segments, providing precise timing and duration information within a video sequence.¹³ Action localization further refines this by identifying and localizing both the temporal ("when") and spatial ("where") extent of specific actions. This dual understanding is vital for comprehending the context of an action and distinguishing it from similar actions that might occur immediately before or after.¹³

In the context of object tracking, temporal understanding is inherent and indispensable. Object tracking fundamentally focuses on monitoring an object's spatial and temporal changes throughout a video sequence, encompassing its presence, position, size, and shape.² It involves continuously tracking an object's trajectory over time, taking into account its motion patterns and appearance evolution across multiple frames.¹

At its core, **Temporal AI** refers to an AI system's advanced capability to consider and reason about time as a fundamental element within its decision-making processes. In contrast to traditional AI models that often treat time as a static input, temporal AI systems comprehend that decisions made at one point in time can profoundly influence future outcomes, and that past events must be meticulously factored into current predictions.¹⁴ This paradigm shift enables temporal AI to leverage time-dependent data, such as timestamps, event sequences, and dynamic temporal patterns, to construct more accurate, adaptive, and dynamic models.¹⁴

2.2. The Distinct Advantage of Temporal Analysis over Static Frame

Processing

The primary distinction between static frame processing and temporal analysis lies in their approach to video data. Static image processing treats each video frame independently, effectively losing the crucial context and continuity provided by the sequence of events.

Conversely, temporal analysis capitalizes on the close inter-frame relationships³, thereby providing continuity and a significantly richer understanding of scene dynamics.²

This contextual awareness facilitates a powerful **predictive capability**. Temporal models are uniquely positioned to learn historical patterns and extrapolate them to predict future states or movements of objects.¹⁴ This predictive power is indispensable for anticipating object trajectories, particularly in dynamic environments like those encountered by drones, where real-time decision-making is critical for safety and operational efficiency.¹⁴

Furthermore, temporal analysis offers superior **robustness to transient issues**. Single-frame methods are highly vulnerable to momentary disruptions such as partial occlusion, motion blur, or sudden lighting changes in individual frames. Temporal analysis, however, can effectively "look past" these transient disturbances by leveraging information from preceding and succeeding frames. This ability allows the system to maintain object identity and tracking continuity even when visual information is temporarily compromised.¹⁵

Finally, temporal analysis is essential for **understanding dynamic behavior**. Static images provide only snapshots and cannot capture evolving changes or dynamic behaviors, such as the subtle vibrations of machinery or the erratic movements of a target.¹⁸ By analyzing motion patterns across video frames, temporal analysis enables the detection of abnormal patterns, subtle changes, or deviations from normal operating states, which is crucial for applications like predictive maintenance.¹⁸

2.3. General Benefits of Incorporating Temporal Information for AI Model Performance

The integration of temporal information into AI models yields substantial benefits, significantly enhancing their performance in dynamic environments.

Firstly, it leads to **enhanced accuracy and robustness**. By incorporating time-dependent data and temporal logic, AI systems can construct more accurate and dynamic models.¹⁴ This improves overall performance in complex, dynamic environments, making models more resilient against real-world challenges such as occlusions, illumination changes, and scale variations.¹⁵

Secondly, temporal context facilitates **improved object re-identification**. Maintaining the identity of an object across multiple frames is a persistent challenge, especially after temporary disappearances or significant appearance changes. Temporal information provides the necessary context to consistently re-identify objects, which is paramount for multi-object

tracking where assigning unique and persistent IDs is crucial.²

Thirdly, temporal models enable a **better understanding of motion dynamics**. Models utilizing 3D convolutions or recurrent neural networks are particularly adept at capturing the dynamics of an action and how spatial elements change and interact over time.¹³ This leads to a deeper, more nuanced understanding of object movement and behavior.¹⁶

Fourthly, temporal information contributes to **efficient video processing and analysis**.

Temporal segmentation simplifies video editing by breaking content into manageable segments, enhances retrieval and indexing through metadata tagging, and enables automatic content summarization by extracting key scenes.¹² For data analytics, it provides valuable insights into patterns and trends over time, which is beneficial in fields like surveillance and sports analytics.¹²

Finally, temporal information provides crucial **contextual awareness**. It allows AI systems to reason about time, influencing how decisions are made and how data is interpreted within a dynamic environment. This contextual understanding is particularly valuable for applications such as surveillance systems and video summarization, where understanding the temporal flow of events is critical.¹²

Temporal information serves as the fundamental bridge from mere "detection" to comprehensive "understanding." Object detection, in its basic form, provides a static snapshot of "what" is present and "where" it is located.¹ However, temporal information, through its application in tracking and sequence modeling, augments this by adding the critical dimensions of "when" an event occurs, "how" an object moves, and even "why" it exhibits certain behaviors.² This transformative shift from static identification to dynamic understanding is indispensable for higher-level AI applications in drones, including predictive maintenance, where anticipating equipment failures is key¹⁸; anomaly detection, which relies on identifying deviations from normal patterns⁶; and autonomous navigation, where real-time adaptation to changing environments is paramount.¹⁴ The implication is that temporal data elevates AI from a passive observer to an active interpreter and predictor of events, a capability fundamental for building trust in autonomous decision-making systems.

Furthermore, temporal AI offers a significant "memory" advantage, crucial for robustness. Traditional frame-by-frame processing inherently lacks a memory of past events, rendering it highly vulnerable to transient issues such as temporary occlusion or brief motion blur.⁷

Temporal AI, particularly when employing architectures like Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTMs), inherently incorporates a "memory" of past states within its processing.¹⁴ This internal memory enables the system to infer an object's likely state or trajectory even when direct visual information is temporarily compromised, much like human visual persistence allows us to track an object even when it momentarily disappears from view. This capability directly translates to improved robustness and a significant reduction in tracking failures, which in turn builds confidence and trust in the system's continuous and reliable operation. The ability to "remember" past movements and "predict" future positions is a core component of overall system reliability.

3. Advanced Temporal AI Techniques for Robust Object Detection and Tracking

3.1. Leveraging Motion and Trajectory Modeling

The effective application of temporal information in drone video analysis relies heavily on sophisticated techniques for modeling motion and predicting trajectories.

Kalman Filters for State Estimation and Prediction: Kalman filters are fundamental mathematical tools widely utilized for motion prediction and correction in object tracking systems.² These filters operate by first predicting an object's motion based on established velocity and acceleration models, and then iteratively refining these predictions with new observations obtained from each successive frame.¹⁵ They are particularly effective in scenarios where computational resources are limited or where deep learning methods might be excessively demanding.¹⁵ In many modern systems, Kalman filters are integrated into hybrid approaches, such as SORT and Deep SORT algorithms. These hybrid models combine Kalman filters for robust motion estimation with Convolutional Neural Network (CNN)-based feature extraction, thereby enhancing accuracy in real-time video streams. Notably, Kalman filters can play a crucial role in maintaining stable tracking during periods of occlusion by providing reliable trajectory predictions.⁸

Optical Flow for Pixel-Level Motion Analysis: Optical flow is a powerful computer vision technique that quantifies the apparent motion of objects or individual pixels between consecutive video frames.²¹ It is mathematically defined by the optical flow equation, which considers both spatial and temporal gradients of pixel intensity.²¹ This technique is indispensable for understanding scene dynamics and the movement of objects within it.²¹ In drone applications, optical flow is leveraged for critical functions such as flight stabilization, precise obstacle avoidance, and maintaining the drone's position relative to the ground, even in environments where GPS signals are unavailable.²¹ While traditional algorithms like Lucas-Kanade and Horn-Schunck laid the groundwork for optical flow, modern deep learning models such as FlowNet, FlowNet 2.0, LiteFlowNet, and PWC-Net have significantly advanced its accuracy and computational efficiency.²¹ Optical flow can also be effectively employed to differentiate between moving objects like UAVs and birds by analyzing their distinct spatiotemporal features.¹⁶

Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTMs) for Sequence Learning: Recurrent Neural Networks (RNNs) are specifically designed to process sequential data, where the activation functions from previous data points in a sequence influence the current output. A hidden layer within an RNN acts as a form of memory, retaining and passing relevant features to the subsequent sequence elements.¹⁴ Long Short-Term

Memory (LSTMs) networks are a specialized type of RNN developed to overcome the vanishing gradient problem, which can impede standard RNNs from effectively learning long-term dependencies in sequential data.¹⁴ LSTMs employ sophisticated "gates" (input, forget, and output gates) that dynamically regulate the flow of information, allowing them to selectively retain or discard past information as needed.²⁰ In object tracking, LSTMs are widely used to model the complex trajectory of objects over time.¹⁵ They excel at capturing intricate temporal dependencies and are highly effective for processing dynamic patterns of motion.¹⁶ For instance, a Bio-Inspired CNN integrated with Gated Recurrent Units (GRUs), a variant of RNNs, can effectively capture temporal motion dynamics for distinguishing between UAVs and birds, with velocity-conditioned gates adapting to varying motion characteristics.¹⁶

3.2. Addressing Critical Challenges with Temporal Cues

Temporal information provides powerful mechanisms to address many of the inherent challenges in drone video object detection and tracking.

Mitigating Occlusion and Target Disappearance through Re-identification and Predictive Models: Occlusion, whether partial or full, leads to a significant loss of appearance and motion data, making continuous tracking difficult.⁷ Temporal information is paramount in overcoming this. Algorithms like Deep SORT integrate deep appearance features to distinguish visually similar objects and maintain their tracks even after temporary occlusion or sudden trajectory changes.¹⁵ Similarly, Re-identification (ReID) models leverage deep learning-based appearance features to recognize objects upon their reappearance after occlusion.¹⁵ Advanced approaches include query-guided redetection trackers (QRDTs) that utilize Siamese neural networks to dynamically update target appearance through inter-frame information transfer, enabling identification of targets even when partially occluded.⁸ A cross-fusion layer within such systems can model semantic correlations between search and query features, focusing attention on the occluded target.⁸ Furthermore, Kalman filters are invaluable for reliably predicting trajectories during periods of full occlusion, thereby preventing tracking failure and misidentification due to similar distracting objects.⁸ The inherent temporal memory of RNNs and LSTMs, which maintain a hidden state incorporating information from previous frames, allows them to "remember" an object's trajectory and characteristics during temporary occlusion, greatly aiding re-identification upon reappearance.¹⁶

Compensating for Motion Blur and Handling Fast-Moving Objects: Motion blur, caused by rapid movement, renders visual details indistinct.⁷ Temporal methods are crucial for inferring true motion despite individual blurred frames. Optical flow, by its very nature, is designed to capture motion, even in the presence of blur.¹⁶ Modern deep learning models for optical flow, such as FlowNet, significantly improve accuracy and robustness in challenging scenarios like motion blur.²¹ Some advanced models employ dynamic receptive fields that adjust their focus based on object velocity, which can help in processing blurred images more effectively.¹⁶

Additionally, temporal attention mechanisms within AI models can track object movement over time, leading to improved real-time detection accuracy for small, fast-moving objects.¹⁶

Ensuring Object Identity Consistency and Reducing ID Switches: Identity switching, where a tracker mistakenly assigns a new ID to an existing object or reassigns an old ID to a different object, is a major challenge in Multi-Object Tracking (MOT).¹⁹ Temporal information is key to mitigating this. Temporal association networks, exemplified by improved FairMOT models, utilize temporal embedding structures to integrate historical frame information with the current frame. This integration enhances robustness and significantly reduces the occurrence of ID switches.¹⁷ Separating the detection and re-identification (ReID) heads within these models can also improve overall performance by allowing each component to optimize for its distinct objective.¹⁷ Deep association metrics, which combine motion predictions with deep learning-based appearance descriptors, are employed to distinguish similar objects and further reduce ID switching. The Hungarian algorithm is commonly used for data association, effectively matching object detections across frames based on both their location and appearance.¹⁵ For large-scale surveillance, multi-camera tracking (MTMC) systems leverage temporal and spatial information, along with visual embeddings, to maintain unique IDs for objects tracked across multiple camera views.²³

Improving Generalization Across Varying Scales and Viewpoints: Drone videos inherently involve significant variations in object scale and viewing angle.⁵ Temporal coherence plays a vital role in addressing this. Approaches like Temporal Coherent Object Flow (OFTrack) process multiple consecutive frames as a "section," effectively boosting optical flow to object flow. This methodology incorporates object-aware sampling and scale-aware correlation for precise target discrimination, combined with temporal-spatial attention for robust association in long video sequences.²⁴ Multi-resolution analysis, which involves processing videos at different resolutions, can robustly detect transitions; lower resolutions highlight global changes, while higher resolutions refine segment boundaries.¹² Furthermore, temporal fusion architectures integrate multi-sensor data across different time steps to construct dynamic environmental models. This enhances the continuity of perception and decision-making by using historical information to improve the interpretation of the current frame, helping to mitigate issues like sensor noise and object occlusion while effectively capturing object dynamics through spatiotemporal modeling.²²

3.3. Real-time Processing and Computational Efficiency in Drone Environments

Achieving real-time performance in drone-based object detection and tracking is critical but challenging, primarily due to the high-speed processing requirements, the computational complexity of neural network architectures, inherent frame rate constraints, and the limited hardware capabilities of drones.¹⁰ Delayed results in dynamic drone applications can render the information useless or even detrimental.¹⁰

A fundamental dilemma exists between **onboard processing and offloading**. While offloading computationally intensive tasks to remote servers or edge devices enables more sophisticated functionality, it introduces significant drawbacks such as increased latency, substantial energy costs for communication, and a dependence on centralized infrastructure.¹⁰ Consequently, lightweight onboard methods are actively being developed as alternatives to achieve real-time performance and conserve energy by minimizing reliance on external resources.¹⁰

To enhance computational efficiency, researchers focus on **optimized architectures**. This includes the development of lightweight CNN architectures (e.g., MobileNet, YOLO), efficient region-based detectors (e.g., Faster R-CNN), and optimization techniques such as pruning, quantization, and model distillation.¹⁵ YOLO models, in particular, are recognized for their real-time processing capabilities, which are crucial for time-sensitive UAV applications.²⁵

Leveraging **temporal correlation for efficiency** is another key strategy. Compared to analyzing static images, utilizing temporal correlations inherent in video data can yield additional efficiency gains.¹⁰ This approach allows for more intelligent resource allocation by providing context that reduces redundant computation or enables models to focus attention on relevant areas.⁸ This suggests that temporal information, while adding complexity, can also serve as a mechanism for computational optimization, which is vital for resource-constrained drone environments.

Advanced **computation offloading frameworks** like "DroneCOCONet" are designed to coordinate the intelligent processing of large video datasets. These frameworks employ edge computation offloading to minimize total computation costs and latency while ensuring that application requirements for processing time are met.¹¹ This includes incorporating heuristic decision-making processes and reinforcement learning approaches for dynamic offloading, allowing systems to adapt to changing conditions and resource availability.¹¹

Furthermore, **control networking (CoNet)** plays a vital role in ensuring high bandwidth and low transmission delay. Tools such as QUICer can significantly accelerate video data transmission, directly contributing to higher video processing speed and accuracy by leveraging advanced networking protocols.¹¹

For robust real-time systems, **multi-threading and multi-processing** are essential. Multi-threading can continuously capture video frames, while multi-processing handles intensive video processing tasks concurrently. This prevents degradation of the Graphical User Interface (GUI) and ensures a smooth user experience.⁴ Techniques like frame skipping can be employed to prevent blocking behavior and allow processing to catch up, maintaining system responsiveness in real-time.⁴

The effectiveness of temporal AI, particularly for real-time applications, is profoundly dependent on the underlying hardware capabilities, sensor quality, and network infrastructure. Even the most advanced temporal algorithms will falter if the data acquisition is poor, for example, due to severe jitter from the drone's physical structure⁷, or if data transmission introduces prohibitive latency. This highlights a symbiotic relationship where advancements in drone hardware, imaging sensors (e.g., global shutter sensors to reduce

motion blur ²⁶⁾, and communication protocols (e.g., QUICer for reduced latency ¹¹⁾ are as crucial as algorithmic improvements for realizing the full potential and trustworthiness of temporal AI in drone vision. The overall reliability and trustworthiness of the system are not solely built upon the AI component but on the seamless integration and robust performance of the entire technological stack.

Table 2: Temporal AI Techniques and Their Contributions to Object Tracking Robustness

Temporal AI Technique	Mechanism/How it Works	Challenges Addressed	Contribution to Robustness	Relevant Sources
Kalman Filters	Predicts motion based on models, corrects with new observations.	Trajectory Prediction, Occlusion	Smooth, stable tracking; maintains identity during brief occlusions.	⁸
Optical Flow	Quantifies pixel-level motion between frames.	Motion Blur, Fast Motion, Object Separation	Captures motion despite blur; aids in distinguishing moving objects.	¹⁶
RNNs/LSTMs	Process sequential data, retain "memory" of past states.	Occlusion, ID Switching, Complex Motion	Remembers trajectory during occlusion; models dynamic motion patterns.	¹⁴
Deep SORT	Integrates deep appearance features with motion prediction.	Occlusion, ID Switching, Re-identification	Distinguishes similar objects; tracks through temporary occlusions.	¹⁵
Query-Guided Redetection	Dynamically updates target appearance via inter-frame transfer.	Occlusion, Target Disappearance	Improves re-identification discrimination; focuses attention on occluded target.	⁸
Temporal Association Networks	Integrates historical frame info with current frame; separates detection/ReID heads.	ID Switching, Identity Consistency	Enhances robustness; reduces misassignments of object IDs.	¹⁷
Temporal Coherent Object	Processes multi-frame	Scale Variation, Viewpoint	Precise target discrimination	²⁴

Flow	"sections"; uses object-aware sampling & scale-aware correlation.	Changes, Long-term Association	across varying scales/views; robust association in long videos.	
Temporal Fusion Architectures	Integrates multi-sensor data across time steps to build dynamic models.	Sensor Noise, Occlusion, Motion Blur, Object Dynamics	Enhances continuity of perception; mitigates transient issues.	²²

4. Fostering Trust in AI Systems through Temporal Information and Explainability (XAI)

4.1. The Multifaceted Nature of Trust in Autonomous AI Systems

Trust in Artificial Intelligence systems is a complex construct, fundamentally involving an agreed-upon understanding between human operators and machines regarding the system's expected performance.²⁷ Humans cultivate trust through consistent, reliable, and explainable interactions with the systems they utilize.²⁷

However, several factors present significant challenges to establishing this trust. The "black-box" nature of many traditional AI and Machine Learning models, particularly deep learning architectures, inherently obstructs transparency. This opacity creates a lack of understanding regarding their internal decision-making processes.²⁷ High entropy, or a state of disorder and randomness, within AI systems can further diminish human trust, especially in uncertain and competitive operational environments.²⁷

The importance of trust becomes even more pronounced in high-stakes domains, such as autonomous drone operations for critical applications like search and rescue, security, or disaster response. In these scenarios, trust is paramount, as operators require absolute assurance regarding the reliability and rationale behind UAV operations.²⁸ Research indicates that the timing and severity of AI errors significantly influence user trust, highlighting the need for not just accuracy, but also predictable and understandable error patterns.²⁹

4.2. How Temporal Consistency and Predictive Capability Enhance Reliability and Trust

Temporal information directly contributes to building trust by enhancing the fundamental attributes of consistency and reliability in AI systems. By maintaining object identity across successive frames and facilitating smooth, continuous trajectories, temporal information significantly reduces erratic behavior and identity switches, which are major inhibitors of trust.² A system that consistently tracks an object, even through complex occlusions, is inherently perceived as more reliable by human operators.

The **predictive capability** afforded by temporal AI further augments reliability. The ability of temporal AI to understand and forecast future outcomes based on historical data¹⁴ is a powerful trust-building feature. For instance, in autonomous vehicles, and by direct extension, in drones, temporal AI enables real-time adaptation to dynamic environments, allowing for proactive adjustments in speed, route, and actions based on time-sensitive factors like changing traffic or weather conditions.¹⁴ This proactive, rather than merely reactive, capability instills confidence in the system's foresight and enhances overall safety perception.

Temporal methods also lead to **reduced false positives and false negatives**. By leveraging rich spatiotemporal features, AI models can significantly decrease the occurrence of false-positive detections¹⁶ and minimize tracking failures caused by missed detections (false negatives).¹⁵ This improvement in accuracy directly correlates with increased reliability and, consequently, a higher level of trust from human users.

Finally, the **robustness in adverse conditions** imparted by temporal methods is a critical factor in fostering trust. AI systems that can maintain high performance under challenging circumstances such as motion blur, severe occlusion, or drastic lighting changes¹⁵ are inherently more trustworthy. This resilience demonstrates the system's capacity to operate effectively in unpredictable real-world scenarios.

4.3. The Role of Explainable AI (XAI) in Interpreting Temporal Decisions

Explainable AI (XAI) plays a pivotal role in bridging the gap between complex AI performance and human trust by addressing the "black-box" limitations of traditional machine learning models. XAI aims to provide interpretable and human-understandable explanations for AI decisions.²⁷ In the context of UAVs, XAI can elucidate how sensor data is interpreted to avoid obstacles or prioritize specific routes, thereby enhancing transparency.²⁸

XAI facilitates **human-understandable insights from temporal context**. Techniques such as saliency maps and Gradient-weighted Class Activation Mapping (Grad-CAM) can highlight regions of interest within captured images that significantly influence the UAV's perception and subsequent actions.²⁸ This allows operators to verify that the UAV correctly identifies and prioritizes relevant objects, such as obstacles or points of interest, thereby improving the accuracy and reliability of autonomous operations.²⁸ While current XAI for object detection is still in early stages of development³⁰, methods like D-CLOSE, D-RISE, and G-CAME are being

developed to generate saliency maps and explain localization and classification decisions.³⁰ XAI can also articulate the **rationale for decisions** made based on visual inputs, such as why a UAV altered its flight path to avoid an obstacle.²⁸ This transparency is crucial for debugging AI models, ensuring compliance with safety standards, and fostering trust among users and stakeholders.²⁸ For temporal AI, explanations would ideally extend beyond static justifications to dynamic, time-aware rationales. This would involve showing *why* a particular trajectory was predicted, *how* an object was successfully re-identified after an occlusion (e.g., based on its consistent motion pattern or reappearing features), or *what* specific temporal cues led to a certain behavioral classification. Such dynamic explanations would move beyond static interpretations to provide a more comprehensive and trustworthy understanding of the AI's temporal reasoning.

4.4. Ethical Considerations and the Importance of Human-in-the-Loop Systems

The integration of temporal AI into autonomous drone systems, while offering immense benefits, also necessitates careful consideration of ethical implications. Like all AI, temporal AI carries the potential for **bias and fairness** issues if trained on biased or incomplete datasets, which could perpetuate existing inequalities.⁶ Establishing ethical guidelines is essential to ensure that models are trained on representative and unbiased data.⁶ Continuous monitoring of AI system performance and regular updates with new data are crucial to maintain fairness and accuracy over time.⁶

Privacy concerns are also significant. Temporal data, especially in surveillance or healthcare applications, often contains personal and sensitive information.¹⁴ The use of AI to analyze and predict future events based on this data raises critical questions about data security, informed consent, and the potential for misuse.¹⁴

The question of **accountability** becomes pressing when AI-driven decisions lead to negative outcomes, particularly in high-stakes sectors like autonomous transportation (drones).¹⁴

Developing robust legal frameworks that clearly define liability in the event of an AI failure will be crucial for addressing this concern.¹⁴

To navigate these ethical complexities and enhance overall trust, **Human-in-the-Loop (HITL) systems** are indispensable. XAI actively supports HITL decision-making, where human operators collaborate with UAVs. In scenarios demanding real-time intervention, such as search and rescue missions, XAI-powered UAVs can explain their navigational choices, enabling human operators to assess the underlying reasoning and intervene if necessary.²⁸ This approach ensures that autonomy is balanced with human oversight, thereby reinforcing trust and reliability.⁶ Continuous monitoring by human operators and proactive addressing of emerging biases are also vital components of this collaborative framework.⁶

The relationship between performance, explainability, trust, and adoption forms a critical feedback loop. Enhanced AI performance, particularly through the integration of temporal

information, is a prerequisite for initial reliability. However, for deeper human trust and widespread adoption, especially in sensitive drone applications, explainability becomes crucial.²⁸ When users understand *why* the AI made a particular decision, even if that decision is based on complex temporal patterns, their trust in the system increases.²⁸ This heightened trust, in turn, facilitates greater adoption and reliance on autonomous drone systems. This creates a positive feedback loop: superior temporal performance enables more effective XAI, which cultivates trust, leading to broader deployment and further refinement of the AI systems.

However, a significant challenge remains in addressing the "black box" nature of temporal AI for comprehensive XAI. While XAI aims to open the "black box" of AI models ²⁸, temporal AI models, especially those employing recurrent neural networks (RNNs), Long Short-Term Memory (LSTMs), or complex spatiotemporal convolutions, are inherently more complex due to their sequential and contextual processing capabilities.¹⁶ Explaining a static decision is already challenging; explaining a decision that evolves over time, influenced by a "memory" of past events and intricate temporal patterns, is significantly more difficult. This implies that current XAI methods, which are still in their nascent stages for object detection ³⁰, will require substantial advancements to adequately interpret the nuanced temporal reasoning of drone AI. This area represents a critical frontier for future research to truly foster deep and enduring trust in these sophisticated autonomous systems.

Table 3: Impact of Temporal Information on Trust and Explainability Metrics

Temporal AI Characteristic	Impact on AI Performance	Contribution to Trust	Contribution to Explainability (XAI)	Relevant Sources
Trajectory Consistency	Smooth, continuous object paths; reduced erratic behavior.	Increased Reliability, Operator Confidence	Provides context for decisions; enables post-hoc analysis of trajectories.	²
Predictive Capability	Accurate future state prediction; proactive adaptation.	Enhanced Safety Perception, Greater Predictability	Allows verification of predicted outcomes; clarifies proactive decision rationale.	⁶
Occlusion Handling	Maintained tracking through obstacles; seamless re-identification.	Improved Operational Continuity, Reduced Frustration	Shows how identity is maintained through occlusions (e.g., motion patterns).	⁷
ID Continuity	Minimized ID switches; persistent object identity.	Higher Data Integrity, More Reliable Analytics	Allows verification of object identity over time; clarifies tracking logic.	¹⁵

Real-time Responsiveness	Low latency; timely decision-making.	Enhanced Operational Efficiency, Immediate Feedback	Demonstrates system's ability to react promptly to dynamic changes.	⁴
---------------------------------	--------------------------------------	---	---	--------------

5. Real-World Applications and Case Studies

The application of temporal information in AI for object detection and tracking in drone video streams has significant implications across various real-world domains, enhancing operational capabilities and fostering greater trust.

5.1. Search and Rescue Operations

In **Search and Rescue (SAR) operations**, AI, Machine Learning, and deep learning offer immense potential for enhancing victim identification and location tracking.⁴ Drones equipped with thermal imaging capabilities can detect heat signatures in low-light or nighttime conditions, significantly aiding in the precise location of individuals or animals who may be otherwise invisible.²⁵ The ability to continuously track objects in real-time is vital for SAR missions, where every second counts.³¹ Temporal object tracking systems designed for drone-based SAR must maintain the continuity of an object's identity and location over extended periods.⁴

Computational efficiency is paramount for the timeliness required in SAR. Real-time processing, often targeting rates of 30 frames per second, is a critical performance metric for these systems. This is achieved through the use of optimized detectors like YOLO and the implementation of multi-threading and multi-processing techniques, which allow for concurrent execution of tasks and prevent bottlenecks.⁴ An implicit case study, such as the "Real-Time Object Detection System for Drone-Based Search and Rescue" project, highlights how the integration of multi-threading and multi-processing ensures real-time performance and responsiveness, which is crucial for dynamic SAR environments where quick and accurate responses are essential for saving lives.⁴

5.2. Surveillance, Security, and Border Monitoring

For **surveillance, security, and border monitoring**, AI-powered tracking systems enable advanced persistent surveillance, maintaining effective detection and identification with a reduced burden on human operators.⁶ These systems can monitor vast expanses of terrain,

identify potential threats across land, sea, or air, and provide real-time situational awareness to operators.⁶

The integration of **predictive analytics** is transformative in these domains. AI can sift through extensive historical and current data to predict enemy movements and strategies before they materialize, allowing operators to anticipate future threats and plan countermeasures proactively.⁶ Advanced video tracking systems are capable of following targets over long durations, even in highly cluttered environments, which is crucial for tracking adversarial movements, vehicles, or other drones.⁶ Multi-Object Tracking (MOT) is essential for analyzing various objects simultaneously, such as in traffic monitoring.¹⁵ Furthermore, Multi-Camera Tracking (MTMC) workflows track and associate objects across multiple cameras, maintaining unique IDs based on visual embeddings and spatial-temporal information.²³ This capability is foundational for advanced security applications, including autonomous stores and enhanced self-checkout systems.²³ AI systems can also analyze video feeds to detect unusual activities or movements, enabling

anomaly detection.⁶ Optical flow, for instance, aids in crowd monitoring by analyzing the flow of people to detect patterns or anomalies, flagging unusual movements for security personnel.²¹ An implicit case study, such as the C4 software suite for real-time drone surveillance and management, demonstrates high operational efficiency for integrated control of UAV threats, achieving over 96% accuracy in detection, tracking, and classification within no-fly zones.¹⁶

5.3. Autonomous Navigation and Smart City Management

In **autonomous navigation and smart city management**, AI-driven navigation, adaptive flight control, and real-time decision-making enable UAVs to support critical applications such as disaster relief and climate monitoring.²⁸ AI-driven systems process large datasets in real-time, allowing UAVs to adapt to dynamic environments and perform complex tasks like obstacle avoidance and route optimization.²⁸ Monocular vision, leveraging AI and Machine Learning, enables UAVs to estimate depth, detect obstacles, and construct environmental maps in real-time, facilitating precise navigation and responsive control.²⁸

For urban planning and traffic management, object tracking algorithms can analyze hours of traffic footage to derive insights into traffic flow, congestion patterns, and accident hotspots. This transforms raw video data into actionable insights that can inform policy decisions and urban infrastructure improvements.²

Predictive maintenance is another key application where computer vision techniques, including action recognition and leveraging temporal information, are employed. Analyzing motion patterns across video frames can detect deviations from normal operating states, signaling underlying mechanical issues before they lead to critical failures.¹⁸ Integrating sensor data with vision to create multi-modal datasets further enhances predictive models.¹⁸ An implicit case study, such as Tesla's Full Self-Driving (FSD) system, incorporates spatial RNN

modules and temporal/spatial feature queues to enhance robustness to dynamic occlusions and retain information in static scenes, which is crucial for complex autonomous driving scenarios and directly applicable to drone navigation.²² This exemplifies how temporal fusion architectures are applied in sophisticated autonomous systems.

The scalability of trust in real-world deployments presents a significant challenge. While individual case studies demonstrate promising results, such as the C4 suite achieving over 96% accuracy¹⁶ or improved MOTA/MOTP for FairMOT variants¹⁷, scaling these systems to real-world, diverse, and highly dynamic environments, like vast smart cities or extensive border regions, introduces new complexities.²⁸ The ability of AI to generalize effectively across diverse environments is a recognized limitation.³¹ Trust built in controlled experimental settings may not readily translate to the unpredictable chaos of real-world operations. This implies that future research and development must extend beyond mere algorithmic improvements to focus on robust deployment strategies, continuous learning mechanisms, and adaptive models that can maintain consistent performance, and by extension, trust, across a wide spectrum of operational conditions.

Temporal AI acts as a crucial enabler for shifting from reactive to proactive operations. Many critical applications, ranging from Search and Rescue⁴ to security⁶ and autonomous navigation²⁸, benefit immensely from AI's ability to predict and anticipate future events.

Traditional object detection is largely reactive, identifying objects

after they appear. In contrast, temporal AI, with its inherent predictive capabilities¹⁴, allows for proactive decision-making and the implementation of preemptive countermeasures.⁶ This fundamental shift from reactive to proactive operations, made possible by deep temporal understanding, fundamentally transforms the nature of human interaction with and reliance on drone systems. It fosters a significantly higher level of trust by minimizing surprises, maximizing preparedness, and enabling more strategic and effective interventions.

6. Conclusion and Future Directions

6.1. Synthesis of Key Findings

The comprehensive analysis presented in this report underscores that temporal information is not merely an auxiliary feature but a fundamental requirement for developing robust and trustworthy AI systems in drone object detection and tracking. The unique and formidable challenges inherent in drone video streams—including severe occlusions, motion blur, significant scale variations, and computational constraints—cannot be effectively overcome by static, frame-by-frame processing alone. Temporal cues provide the essential continuity, predictive power, and identity consistency necessary for reliable tracking in such dynamic environments.

Advanced temporal AI techniques, such as Kalman filters for state estimation, optical flow for motion analysis, and recurrent neural networks (RNNs) and Long Short-Term Memory (LSTMs) for sequence learning, are pivotal in mitigating these challenges. These methods enable systems to maintain object identity through occlusions, compensate for motion blur, reduce identity switches, and generalize across varying scales and viewpoints. Furthermore, the strategic management of real-time processing and computational efficiency, through optimized architectures and intelligent offloading strategies, is crucial for the practical deployment of these temporal AI models on resource-constrained drones.

Crucially, the enhanced performance and reliability delivered by temporal information form the bedrock upon which human trust is built. A system that consistently and accurately tracks objects, even under adverse conditions, naturally inspires greater confidence. This performance-driven trust is further amplified by the integration of Explainable AI (XAI). XAI plays a vital role in demystifying the "black-box" nature of complex AI models, making their temporal decisions interpretable and understandable to human operators. This transparency is essential for debugging, regulatory compliance, and fostering broader acceptance of autonomous drone capabilities in high-stakes applications.

6.2. Recommendations for Advancing Trustworthy Temporal AI in Drone Applications

To further advance the development and deployment of trustworthy temporal AI in drone applications, the following recommendations are proposed:

- **Develop More Comprehensive and Diverse Datasets:** Future research should prioritize the creation of UAV-specific datasets that are high-resolution, encompass a rich diversity of object categories, and are collected across a wide range of altitudes and viewing angles. This will significantly reduce existing domain gaps and improve the generalization capabilities of AI models.⁵
- **Advance Hybrid Onboard-Offboard Processing Architectures:** Continued optimization of intelligent computation offloading and control networking solutions is essential. The focus should be on achieving an optimal balance between real-time performance, energy efficiency, and computational load, thereby ensuring reliable operation in resource-constrained drone environments.¹⁰
- **Enhance Temporal XAI for Dynamic Explanations:** A critical area for development is XAI techniques specifically tailored to interpret the temporal reasoning of AI models. This involves moving beyond static saliency maps to provide human-understandable, dynamic insights into *why* trajectories are predicted, *how* identities are maintained through occlusions, and *what* temporal patterns drive specific decisions.²⁸
- **Strengthen Long-Term Temporal Modeling:** Research should explore more effective methods for capturing and leveraging long-range temporal dependencies. This is particularly important for long-span tracking tasks and for predicting complex, non-linear object behaviors over extended periods, which current models often struggle

with.¹⁷

- **Integrate Multi-Modal Sensor Fusion with Temporal AI:** Further investigation into combining visual temporal information with data from other sensors (e.g., LiDAR, thermal imaging, Inertial Measurement Units (IMU), audio) is recommended. This fusion can create more robust and context-aware predictive models, especially for challenging conditions like low visibility or complex backgrounds.¹²
- **Prioritize Ethical AI Development and Human-in-the-Loop Systems:** It is imperative to continue developing and implementing robust ethical guidelines, stringent data protection measures, and clear accountability frameworks. Emphasis should be placed on human-in-the-loop systems where AI provides interpretable insights to assist, rather than fully replace, human decision-making, particularly in high-stakes drone applications.⁶
- **Focus on End-to-End Joint Optimization:** Continued exploration and refinement of end-to-end architectures that integrate detection, tracking, and prediction into a unified temporal framework are crucial. This approach aims to minimize information latency and error accumulation across different modules, leading to more seamless and efficient performance.²²

Works cited

1. Introduction to Object Tracking - HackMD, accessed on June 13, 2025, https://hackmd.io/@IKuOpplzSUWLhLim2Z7ZJw/ryTpNXeGn?utm_source=preview-mode&utm_medium=rec
2. Object Tracking in Computer Vision: An In-Depth Exploration and Practical Guide - BasicAI, accessed on June 13, 2025, <https://www.basic.ai/blog-post/object-tracking-in-computer-vision>
3. A Survey on Object Detection and Tracking Methods - SciSpace, accessed on June 13, 2025, <https://scispace.com/pdf/a-survey-on-object-detection-and-tracking-methods-5ekjxnsisa.pdf>
4. Real-Time Object Detection System for Drone ... - IADT OnShow, accessed on June 13, 2025, <https://onshow.iadt.ie/sites/default/files/2025-04/Major%20Project%20Report.pdf>
5. More Clear, More Flexible, More Precise: A Comprehensive Oriented Object Detection benchmark for UAV - arXiv, accessed on June 13, 2025, <https://arxiv.org/html/2504.20032v1>
6. A Step Change in Capability: AI in Security Tracking | Unmanned Systems Technology, accessed on June 13, 2025, <https://www.unmannedsystemstechnology.com/feature/a-step-change-in-capability-ai-in-security-tracking/>
7. Tutorial in 17th ACCV 2024.12.09 - Xin Zhao, accessed on June 13, 2025, <https://www.xinzhaoui.com/PDFs/ACCV2024-slides.pdf>
8. (PDF) Handling Occlusion in UAV Visual Tracking With Query ..., accessed on June 13, 2025,

https://www.researchgate.net/publication/383010370_Handling_Occlusion_in_UAV_Visual_Tracking_with_Query-Guided_Redetection

9. Real-time object detection and video monitoring in Drone System - IRJET, accessed on June 13, 2025, <https://www.irjet.net/archives/V10/i8/IRJET-V10I873.pdf>
10. Real-Time Navigation for Autonomous Aerial Vehicles Using Video - arXiv, accessed on June 13, 2025, <https://arxiv.org/html/2504.01996v1>
11. DroneCOCO: Learning-based Edge Computation Offloading and ..., accessed on June 13, 2025, <https://par.nsf.gov/servlets/purl/10319097>
12. Temporal Video Segmentation | Cloudinary, accessed on June 13, 2025, <https://cloudinary.com/glossary/temporal-video-segmentation>
13. Action Localization: Everything You Need To Know - viso.ai, accessed on June 13, 2025, <https://viso.ai/deep-learning/action-localization/>
14. Temporal AI: Understanding and Modeling Time in Dynamic Decision-Making Systems, accessed on June 13, 2025, https://www.researchgate.net/publication/392145757_Temporal_AI_Understanding_and_Modeling_Time_in_Dynamic_Decision-Making_Systems
15. Deep Learning Object Tracking: Algorithms, Challenges, and ..., accessed on June 13, 2025, <https://flypix.ai/blog/deep-learning-object-tracking/>
16. Bio-inspired motion detection models for improved UAV and bird ..., accessed on June 13, 2025, <https://pmc.ncbi.nlm.nih.gov/articles/PMC12049524/>
17. Multiple Object Tracking of Drone Videos by a Temporal-Association ..., accessed on June 13, 2025, <https://www.mdpi.com/2072-4292/14/16/3862>
18. AI for Predictive Maintenance Using Computer Vision - Voxel51, accessed on June 13, 2025, <https://voxel51.com/learn/ai-for-predictive-maintenance-using-computer-vision>
19. Recent Advances in Embedding Methods for Multi-Object Tracking: A Survey - arXiv, accessed on June 13, 2025, <https://arxiv.org/html/2205.10766v2>
20. Multiple Object Tracking in Deep Learning Approaches: A Survey, accessed on June 13, 2025, <https://www.mdpi.com/2079-9292/10/19/2406>
21. Optical Flow - Everything You Need to Know - viso.ai, accessed on June 13, 2025, <https://viso.ai/deep-learning/optical-flow/>
22. A Survey of Deep Learning-Driven 3D Object Detection: Sensor ..., accessed on June 13, 2025, <https://www.mdpi.com/1424-8220/25/12/3668>
23. The Multi-Camera Tracking AI Workflow - NVIDIA, accessed on June 13, 2025, <https://www.nvidia.com/en-us/ai-data-science/ai-workflows/multi-camera-tracking/>
24. Temporal Coherent Object Flow for Multi-Object Tracking ..., accessed on June 13, 2025, <https://ojs.aaai.org/index.php/AAAI/article/view/32749>
25. Deep Learning Model for Unmanned Aerial Vehicle-based Object ..., accessed on June 13, 2025, <https://www.iicta.org/journals/ria/paper/10.18280/ria.370608>
26. Vision Insights: Finding the Right Vision System for Every Drone - FRAMOS, accessed on June 13, 2025, <https://framos.com/articles/vision-insights-finding-the-right-vision-system-for-every-drone/>

27. Artificial Intelligence (AI) Trust Framework and Maturity Model: Applying an Entropy Lens to Improve Security, Privacy, and Ethical AI - MDPI, accessed on June 13, 2025, <https://www.mdpi.com/1099-4300/25/10/1429>
28. Explainable AI and monocular vision for enhanced UAV ... - Frontiers, accessed on June 13, 2025, <https://www.frontiersin.org/journals/sustainable-cities/articles/10.3389/frsc.2025.1561404/pdf>
29. How Misclassification Severity and Timing Influence User Trust in AI Image Classification - Kapadia Privacy Lab, accessed on June 13, 2025, https://privacy.luddy.indiana.edu/publications/FAccT_2025_preprint.pdf
30. ODExAI: A Comprehensive Object Detection Explainable AI Evaluation - arXiv, accessed on June 13, 2025, <https://arxiv.org/html/2504.19249v1>
31. Real-Time Object Detection and Tracking for Drone Using ... - ijarccce, accessed on June 13, 2025, <https://ijarcce.com/wp-content/uploads/2024/05/IJARCCE.2024.134186.pdf>
32. Top Video Object Detection Algorithm in Computer Vision | Encord, accessed on June 13, 2025, <https://encord.com/blog/video-object-tracking-algorithms/>
33. Learning Temporal Cues by Predicting Objects Move for Multi-camera 3D Object Detection, accessed on June 13, 2025, <https://chatpaper.com/es/paper/12227>