Vision-Based Drone Detection, Tracking, and Payload Identification using RGB and Infrared Imagery: A Review of Research from Premier Conferences

1. Introduction

1.1. The Rising Significance of Drone Perception

Unmanned Aerial Vehicles (UAVs), commonly known as drones, have witnessed an exponential increase in adoption across a multitude of sectors, including industrial inspection, environmental monitoring, precision agriculture, geographical surveying, search and rescue, security, surveillance, and logistics.[1] Their ability to operate remotely or autonomously, capture aerial perspectives cost-effectively, and navigate complex environments makes them invaluable tools.[1] However, this proliferation brings forth significant security and privacy concerns stemming from potential misuse, ranging from unauthorized surveillance and smuggling to espionage and potential terrorist activities.[8] Consequently, the development of robust perception systems capable of reliably detecting, tracking, and characterizing drones is paramount for both leveraging their benefits and mitigating associated risks. Key tasks within this domain include drone detection (determining the presence of a UAV), tracking (estimating its trajectory over time), and payload identification (understanding what the drone is carrying).

1.2. Role of RGB and Infrared (IR) Sensing

Vision-based systems, primarily utilizing RGB (visible spectrum) and Infrared (IR) cameras, are central to drone perception research. RGB cameras provide rich color and texture information, beneficial for detailed object recognition under sufficient lighting conditions.[20] IR cameras, specifically thermal IR, detect heat signatures, enabling operation in low-light or nighttime scenarios and potentially identifying drones based on heat generated by motors or batteries.[9] They can also offer advantages in adverse weather conditions like fog or smoke where visible light is obscured.[22] The complementary nature of these sensors motivates research into multi-modal fusion approaches, aiming to combine their strengths for more robust and versatile perception capabilities across various operating conditions.[1] Both sensor types are increasingly integrated onto UAV platforms themselves or utilized in ground-based counter-UAV systems.[1]

1.3. Focus: State-of-the-Art from Premier Conferences

This report provides a comprehensive literature review focused specifically on

advancements in drone detection, tracking, and payload identification using RGB, IR, and fused data, as presented in **top-tier Computer Vision (CV), Robotics, and Artificial Intelligence (AI) conferences**. The selection of these venues—including flagship conferences like the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), the IEEE/CVF International Conference on Computer Vision (ICCV), the European Conference on Computer Vision (ECCV), the IEEE International Conference on Robotics and Automation (ICRA), the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), the Conference on Neural Information Processing Systems (NeurIPS), the AAAI Conference on Artificial Intelligence (AAAI), the International Conference on Machine Learning (ICML), and the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), along with their associated workshops (CVPRW, ICCVW, WACVW)—is deliberate.[14] Publication in these highly selective, peer-reviewed forums signifies work that is considered novel, rigorously evaluated, and representative of the cutting edge in the field. While remote sensing conferences like the International Geoscience and Remote Sensing Symposium (IGARSS) also contribute [33], the primary focus here remains on the core CV, Robotics, and AI literature where foundational perception algorithms are typically introduced and advanced. This stringent focus ensures the review captures the most impactful and validated methodologies shaping the state-of-the-art.

1.4. Report Structure and Objectives

This review is structured as follows: Section 2 delves into drone detection methods using RGB, IR, and fusion techniques. Section 3 examines drone tracking approaches, including trajectory estimation and motion analysis. Section 4 discusses the challenges and methods related to payload identification. Section 5 addresses real-time processing considerations crucial for practical deployment. Finally, Section 6 provides a discussion synthesizing the findings and outlining future research directions. The core objectives are to identify and analyze state-of-the-art methods presented in the target premier conferences for each task (detection, tracking, payload ID) across different modalities (RGB, IR, Fusion), evaluate trajectory analysis techniques (specifically approaching/receding inference), discuss payload classification approaches, and assess the real-time performance capabilities reported in this body of literature.

2. Drone Detection via RGB, IR, and Fusion

Drone detection forms the initial step in most counter-UAV pipelines, aiming to answer the fundamental question: "Is a drone present in the sensor's field of view?". Vision-based detection, while offering rich information, faces numerous hurdles.

2.1. Challenges in Vision-Based Drone Detection

Research presented across various conferences and surveys consistently highlights significant challenges inherent to detecting drones using visual sensors. Drones often appear as **small objects**, particularly when observed from a distance, possessing few pixels and lacking distinct features.[2] This is exacerbated by **complex or cluttered backgrounds** (e.g., urban environments, foliage) where the drone can easily blend in.[1] Performance is further complicated by **varying illumination and adverse weather conditions** (e.g., low light, glare, rain, fog) which can degrade image quality.[9] Drones can exhibit **fast and erratic motion**, making localization difficult.[9] Moreover, distinguishing drones from other small, flying objects like **birds** poses a significant classification challenge.[10] Finally, the **varying perspectives and altitudes** inherent in UAV operations add further complexity.[2] While other sensor modalities like radar, Radio Frequency (RF) analysis, and acoustic sensors exist, they have their own limitations (e.g., difficulty with small radar cross-sections, reliance on active RF communication, susceptibility to noise, limited range), underscoring the continued importance and research focus on vision-based solutions.[8]

2.2. RGB-Based Detection Approaches

The field has largely transitioned from traditional computer vision techniques (e.g., background subtraction [10], Hough transforms [42]) towards deep learning, primarily Convolutional Neural Networks (CNNs), due to their superior performance in complex recognition tasks. Advances in general object detection models, frequently presented at major CV conferences like CVPR, ICCV, and ECCV, are rapidly adapted and evaluated for the specific task of drone detection, often appearing in subsequent conference workshops or specialized publications.[14] This indicates that drone detection heavily leverages foundational progress in the broader computer vision community.

Key deep learning architectures featuring prominently in drone detection benchmarks within the target conference literature include:

- **Two-Stage Detectors:** Architectures like Faster R-CNN and its derivatives are often employed or used as baselines in studies presented or cited in conference proceedings.[14] They typically achieve high accuracy by first proposing regions of interest and then classifying them, but often at the cost of lower processing speed, making them less suitable for real-time applications.[44]
- **One-Stage Detectors:** Models like You Only Look Once (YOLO) in its various iterations (e.g., YOLOv3, YOLOv5, YOLOv8) and Single Shot MultiBox Detector (SSD) are frequently cited for drone detection, particularly when real-time

performance is critical.[6] These models perform localization and classification in a single pass, offering significant speed advantages.[43] Conference papers often discuss enhancements to these architectures specifically for improving the detection of small drone targets, such as incorporating multi-scale feature fusion or attention mechanisms.[15] For instance, GL-YOMO enhances YOLO with multi-scale fusion and attention for small UAVs.[15]

- **Transformer-Based Detectors:** The Detection Transformer (DETR) and its variants represent a newer class of detectors utilizing attention mechanisms.[47] DETR has been included in drone detection benchmarks reported in conference workshops (e.g., ICCVW [14]) and related studies [15], reflecting the broader trend of adopting transformer architectures in computer vision.

Specific examples from conference workshops include benchmarking these detectors on datasets like Anti-UAV [14] or investigating performance under challenging conditions.[16]

2.3. IR-Based Detection Approaches

Infrared cameras offer distinct advantages for drone detection, particularly **effectiveness in low-light, nighttime, or adverse weather conditions** where RGB sensors falter.[9] They detect thermal signatures, which can originate from drone components like motors or batteries, providing a detection modality independent of visible illumination.[9]

However, IR sensing also presents challenges. Thermal imagery typically has **lower resolution and lacks the rich texture and color information** found in RGB images.[21] Detecting the often **subtle thermal signatures of small drones at long ranges** can be difficult [9], and **thermal clutter** from the background can interfere with detection.

Methods presented in or relevant to conference literature often involve **adapting standard object detectors** (like YOLO or Faster R-CNN) for use with IR data.[11] Research presented at workshops like ICCVW has benchmarked performance on IR-specific datasets such as Anti-UAV IR.[14] Some work focuses specifically on detecting small infrared targets, which is highly relevant to drone detection.[11] While specific IR-only drone detection papers from the absolute top-tier conferences might be less common than RGB or fusion approaches, the problem is actively researched, particularly within specialized workshops and related journals.[11]

2.4. RGB-IR Fusion Strategies for Detection

Fusing RGB and IR data aims to leverage their complementary strengths, creating

detection systems that are more robust across varying conditions (day/night, clear/adverse weather).[21] Research presented at premier conferences explores various fusion strategies:

- **Fusion Levels:**
  - *Pixel-Level Fusion:* Early approaches might involve combining images at the pixel level, for instance, by creating a four-channel input (RGB + IR) for a detector.[49] However, this strategy is highly sensitive to pixel-perfect alignment, which is difficult to achieve in practice with UAV-mounted sensors.
  - *Feature-Level Fusion:* A more common and generally robust approach involves extracting features independently from RGB and IR streams using parallel backbones (e.g., CNNs) and then fusing these features at one or multiple stages in the network.[11] MCFNet, for example, employs a dual-stream architecture with additive fusion at multiple scales.[28] MFFN also proposes multimodal feature fusion.[11]
  - *Decision-Level Fusion:* This strategy involves running independent detectors on each modality and then fusing their outputs (bounding boxes, class scores).[49] The Category Probability Representation Output Strategy (CPROS), detailed in remote sensing literature [49], provides a sophisticated example. It fuses category probability distributions from each detector, explicitly handling confidence levels and conflicts between modalities to arrive at a more reliable final decision.
- Addressing Misalignment: A critical bottleneck for effective fusion, especially feature-level, is the spatial misalignment between RGB and IR images caused by parallax, different sensor resolutions, and imperfect calibration.26 Research at top venues like CVPR directly tackles this. The Offset-guided Deformable Alignment and Fusion (ODAF) module, presented at CVPR 2024 26, offers a state-of-the-art solution. Instead of requiring strict pixel alignment, ODAF uses deformable convolutions guided by initially estimated spatial offsets (from a Cross-modality Spatial Offset Modeling module) to adaptively sample features from the RGB stream at locations corresponding to the IR features. The alignment is learned implicitly to optimize the detection task. The core alignment mechanism uses the equation:
  $y(p)=k=1\Sigma K\omega k\cdot x(p+\phi c+pk+\Delta pk)\cdot\Delta mk$
  where $y(p)$ is the aligned feature, x is the original feature, ϕc is the guiding offset, pk are kernel offsets, Δpk are learned implicit offsets, and Δmk are learned modulation scalars. Bilinear interpolation handles fractional sampling locations. ODAF then employs a decoupled fusion strategy, processing common and specific features separately before concatenation to reduce redundancy.26 The

prominence of such sophisticated alignment techniques in top-tier conference papers underscores that robust fusion necessitates explicit mechanisms to handle the geometric and temporal discrepancies inherent in multi-sensor UAV platforms. Simple concatenation or averaging is often insufficient for SOTA performance.

Conference papers presenting or evaluating fusion strategies can be found across CVPR, ICCV, ECCV, WACV, IROS, ICRA, and potentially related venues like IGARSS.[26]

2.5. SOTA Performance & Benchmarks

Evaluating drone detection algorithms relies on standardized datasets and metrics. Datasets frequently cited in conference papers and related benchmarks include VisDrone [2], UAVDT [2], the Anti-UAV challenge datasets (often featuring RGB and IR modalities) [14], DroneVehicle [3], Real World [23], and MIDGARD.[23] Performance is typically measured using metrics like mean Average Precision (mAP), F1-score, precision, and recall.

State-of-the-art results reported in conference workshops or papers utilizing methods presented at premier venues show high performance. For example, benchmarks at ICCVW'21 using detectors like Faster R-CNN and YOLOv3 achieved mAP scores up to 98.6% on subsets of the Anti-UAV dataset.[14] Fusion methods like ODAF demonstrate significant performance gains over single-modality baselines on challenging datasets.[26] MFFN reported an average recognition rate (Pavg) of 92.64% and mAP of 92.01% [11], while MCFNet achieved 67.92% mAP on DroneVehicle and 96.4% on a custom dataset.[28]

Table 2.1: Comparative Analysis of Representative Drone Detection Methods (from or relevant to Top Conferences)

Method Name	Modality	Key Innovation / Paradigm	Target Conference/Venue (Example)	Reported Performance (Metric)	Key Dataset Used	Real-Time Capability (Reported FPS)	Snippets
Faster R-CNN (Bench	RGB, IR	Two-Stage Detectio	ICCVW'21	98.6% (mAP)	Anti-UAV RGB	18.0 FPS	[14]

marked)		n					
YOLOv3 (Benchmarked)	RGB	One-Stage Detection	ICCVW'21	98.6% (mAP)	Anti-UAV RGB	36.0 FPS	[14]
SSD512 (Benchmarked)	RGB	One-Stage Detection	ICCVW'21	98.2% (mAP)	Anti-UAV Full	32.4 FPS	[14]
DETR (Benchmarked)	RGB, IR	Transformer-Based Detection	ICCVW'21	98.0% (mAP)	Anti-UAV IR	21.4 FPS	[14]
YOLOv8 (Applied/Enhanced)	RGB, IR	One-Stage Detection (SOTA Base)	WACVW'24, Remote Sens. J.	High mAP (baseline)	Custom, VEDAI	Fast (varies with version)	[16]
GL-YOMO	RGB + Motion	YOLO + Multi-Frame Motion, Multi-Scale Fusion, Attention	ArXiv (Cites Conf. Work)	High Accuracy (claimed)	Custom ARD-MAV	23.6 FPS	[15]
MFFN	RGB-IR Fusion	Multimodal Feature Fusion Network	Sensors J. (Cites Conf. Work)	92.01% (mAP), 90.52% (F1)	Custom Target Seq.	-	[11]
ODAF-Net	RGB-IR Fusion	Offset-guided Deformable	CVPR'24	SOTA on VEDAI, LLVIP	VEDAI, LLVIP	-	[26]

		Alignment & Fusion		(mAP)			
MCFNet	RGB-IR Fusion	Dual-Stream Late Fusion, Lightweight Transformer	Expert Sys. App. J.	67.92% (mAP), 96.4% (mAP)	DroneVehicle, Custom	-	[28]
CPROS (with YOLOv8)	RGB-IR Fusion	Category Probability Decision Fusion	Remote Sens. J.	+8.6%/+16.4% mAP vs single	VEDAI	Depends on base detector	[49]
F-UAV-D	Event Camera	DVS-based Detection	ArXiv (Cites WACV/ICCV)	High Accuracy (claimed)	Custom	< 50ms (< 15W)	[9]

Note: Performance metrics depend heavily on dataset splits, evaluation protocols, and specific model variants. This table provides representative examples based on the provided snippets.

3. Drone Tracking via RGB, IR, and Fusion

Once a drone is detected, tracking its movement over time is crucial for situation assessment, prediction, and potential interception. Research presented at premier conferences explores various paradigms and techniques for robust drone tracking.

3.1. Tracking Paradigms in Conference Literature

Several approaches to visual object tracking have been presented and evaluated in the context of drone tracking at top conferences and workshops:

- **Tracking-by-Detection (TBD):** This remains a common and practical approach. It involves running an object detector (like YOLO, Faster R-CNN, SSD) on each frame and then associating these detections across frames using algorithms like

Simple Online and Realtime Tracking (SORT), DeepSORT (which adds appearance features), or Tracktor.[14] Benchmarking studies, often presented in conference workshops (e.g., ICCVW Anti-UAV challenge [14]), evaluate the performance of different detector-tracker combinations. While effective, the performance heavily relies on the detector's quality, and association can be challenging with occlusions or similar objects.

- **Discriminative Correlation Filters (DCF):** DCFs were historically significant for real-time tracking.[51] They learn a filter to correlate with the target in subsequent frames. While highly efficient, their reliance on hand-crafted or simpler features often limits their robustness compared to deep learning methods, especially in challenging UAV scenarios.[51] However, they might still appear in conference papers as baselines or integrated with other techniques (e.g., combined with Kalman Filters [52]).
- **Siamese Networks:** This paradigm has gained significant popularity, with numerous methods presented or discussed in conference papers.[18] Siamese trackers learn a similarity metric between an initial target template and candidate regions in subsequent frames. Prominent examples relevant to drone tracking presented at venues like CVPR, WACV, ICRA, and IROS include SiamRPN [53], SiamGAT [55], SiamSTM [56], SiamHSFT [54], and particularly SiamTPN [53], which was presented at WACV'22 and specifically designed for efficient UAV tracking, even claiming real-time CPU performance. SiamSA, presented at IROS'22, focuses on UAV approaching scenarios.[55]
- **Transformers for Tracking:** This represents the most recent and arguably dominant trend in high-performance visual tracking, heavily featured in papers at CVPR, ICCV, ECCV, NeurIPS, and AAAI.[18] Transformers, with their self-attention mechanisms, excel at modeling both spatial and temporal dependencies, capturing long-range relationships between the target template and the search region. This shift towards end-to-end Transformer architectures marks a significant departure from earlier paradigms.

The rapid adoption and SOTA performance of Transformer-based trackers in premier vision and AI conferences strongly suggest this is the current frontier of research. These models often achieve superior accuracy by effectively learning target-specific features and their evolution over time within a unified architecture.[18]

3.2. RGB-Based Tracking Methods (Focus on Transformers)

Given the trend identified in premier conference literature, Transformer-based trackers are central to the current state-of-the-art for RGB visual tracking. Key architectures presented or heavily cited include:

- **Foundation Models:** TransT (CVPR'21 [54]) and STARK (ICCV'21 [55]) were influential early works demonstrating the potential of Transformers for tracking, often using separate encoder-decoder structures or focusing on spatio-temporal modeling.
- **Unified/One-Stream Architectures:** A significant advancement involves using a single Transformer backbone to jointly perform feature extraction and relation modeling between the template and search region. This simplifies the architecture and often improves performance. Notable examples presented at CVPR, ECCV, and NeurIPS include MixFormer/MixFormerV2 (CVPR'22, NeurIPS'23 [18]), OSTrack (ECCV'22 [18]), and SimTrack (ECCV'22 [18]). MixFormerV2, in particular, introduced a fully transformer framework without convolutional heads, further streamlining the architecture and achieving high efficiency.[60]
- **UAV-Specific/Real-Time Focused Transformers:** Recognizing the specific constraints of UAV platforms (limited computation, need for real-time speed), several Transformer-based trackers targeting aerial tracking have been presented, often at robotics conferences like ICRA and IROS or application-focused venues like WACV. Examples include:
  - SiamTPN (WACV'22 [53]): Explicitly designed for CPU-based real-time UAV tracking using a lightweight pyramid Transformer.
  - HiFT (ICCV'21 [57]): Focuses on hierarchical features for aerial tracking.
  - Trackers from ICRA/IROS: ClimRT (ICRA'23 [55]), SGDViT (ICRA'23 [55]), HighlightNet (IROS'22 [55]), LPAT (IROS'22 [55]), SiamSA (IROS'22 [55]), TDA-Track (IROS'24 [61]), Progressive Representation Learning (IROS'24 [61]), DaDiff (IROS'24 [61]).
  - SGLATrack [51]: Specifically tailored for UAV tracking efficiency via layer adaptation.
- **Other Notable Transformers:** A rich ecosystem of Transformer trackers exists, including ToMP (CVPR'22 [57]), GTR (CVPR'22 [57]), UTT (CVPR'22, CVPRW'23 [18]), GRM (CVPR'23 [55]), ROMTrack (ICCV'23 [55]), SwinTrack (NeurIPS'22 [55]), and CTTrack (AAAI'23 [55]).

These Transformer-based methods benefit from powerful pre-trained models, effective modeling of temporal context, and increasingly efficient end-to-end architectures.[55]

3.3. IR-Based Tracking Methods

The tracking paradigms developed for RGB data, particularly Siamese networks and Transformers, can be adapted for IR tracking. The challenges mirror those in IR detection: lower resolution, lack of texture, and potential for low contrast. Conference literature specifically addressing IR drone tracking often emerges from challenges or

workshops focused on anti-UAV applications using thermal infrared (TIR) data. For example, the UTTracker, presented at the CVPR'23 Anti-UAV workshop, utilizes a Transformer baseline (OSTrack) and incorporates additional modules for global detection and background correction specifically for robust TIR UAV tracking.[18] Fusion transformers for RGBT (RGB-Thermal) tracking are also an active area.[59]

3.4. RGB-IR Fusion Strategies for Tracking

While fusion for detection is well-explored in the conference literature reviewed, dedicated RGB-IR fusion *tracking* algorithms appear less frequently explicitly detailed in the provided snippets. However, the principles remain similar: combining complementary information for robustness. Potential strategies, extrapolating from detection fusion and general tracking principles, could include:

- **Feature-Level Fusion:** Integrating features from parallel RGB and IR backbones within a Siamese or Transformer tracker architecture. RGBT tracking approaches like the Progressive Fusion Transformer mentioned in [59] likely fall into this category.
- **Decision-Level Fusion:** Combining the outputs (e.g., predicted bounding boxes, confidence scores) of separate RGB and IR trackers.
- **Modality Switching/Gating:** Using one modality (e.g., IR) primarily when the other (RGB) fails (e.g., at night or in fog), or dynamically weighting contributions based on conditions.
- **Multi-Sensor Input:** Systems like the one described in [27] use multiple sensors including thermal IR and fuse data, potentially at the track level, although the specific conference context is missing.

Further investigation into specific RGBT tracking papers presented at CVPR, ICCV, ECCV, ICRA, or IROS would be needed to detail SOTA fusion tracking methods.

3.5. Trajectory Estimation & Motion Analysis

Raw outputs from visual trackers can be noisy or suffer from temporary failures. For applications requiring smooth trajectory estimates, prediction, or analysis of motion dynamics (like determining if a drone is approaching or receding), filtering techniques are commonly integrated, particularly in research presented at robotics-focused conferences like ICRA and IROS, or applied vision workshops like ICCVW.

- **Filtering Techniques:**
    - *Kalman Filters (KF):* Widely used due to their efficiency and optimality for linear systems with Gaussian noise. Standard KF [45], Extended Kalman Filter (EKF) for systems with non-linearities [37], and Unscented Kalman Filter (UKF)

which often handles non-linearities better than EKF [65] are employed. They fuse measurements (tracker output) with a motion model to produce smoothed state estimates (position, velocity), predict future states (useful for handling temporary tracker failures or occlusion), and reduce noise.[45] Variations like iterative multi-model KFs address unknown or changing dynamics [63], and adaptive KFs adjust parameters online.[52] Conference papers like [37] (ICCVW'19) explicitly use EKF for online localization and trajectory smoothing for UAV tracking. Others combine KFs with DCF trackers [52] or Siamese trackers like SiamFC.[45]

- *Particle Filters (PF):* Suitable for non-linear, non-Gaussian problems, representing the state distribution with a set of weighted particles (samples).[46] They can handle complex motion models and multi-modal distributions but are generally more computationally expensive than KFs.[62] Techniques like using evolution strategies (crossover, mutation) can mitigate issues like sample impoverishment.[66] Model-based tracking systems presented in robotics conference contexts often utilize PFs or hybrid approaches like the Unscented Particle Filter (UPF).[65]

The consistent appearance of these filters in applied tracking papers suggests that for robust, smooth trajectory estimation suitable for control or higher-level analysis, visual tracker outputs are often refined using established state estimation techniques.

- **Motion Modeling:** Filters rely on underlying motion models. Simple models like constant velocity are often assumed [67], but more complex models can be incorporated. Additionally, motion cues derived directly from images, such as motion difference maps [40] or optical flow [15], can potentially serve as inputs or complementary information to the tracking and filtering process.
- **Approaching/Receding Analysis:** Direct classification of "approaching" or "receding" behavior is not a standard output of most trackers discussed. However, this information can be readily **inferred** from the state vector estimated by a Kalman or Particle filter. These filters typically estimate both position (p=[X,Y,Z]) and velocity (v=[vx,vy,vz]).[63] By calculating the relative position vector from the observer to the drone and projecting the estimated velocity vector onto this line-of-sight, one can determine the radial velocity. A negative radial velocity indicates the drone is approaching, while a positive radial velocity indicates it is receding. The rate of change of the estimated distance (magnitude of the position vector) also provides this information. Some work, like SiamSA presented at IROS'22 [55], explicitly targets UAV *approaching* scenarios, suggesting motion dynamics are considered in the tracker design itself.

3.6. SOTA Performance & Benchmarks

Evaluating tracking performance relies on datasets specifically designed for UAV scenarios, often featuring challenges like small object size, fast motion, and camera movement. Key benchmarks frequently used in conference papers include UAV123 [3], LaSOT [45], UAVDT [2], VisDrone [2], the Anti-UAV challenge datasets [14], TNL2k [60], UAV20L [54], DTB70 [54], and UAVTrack112.[68] Common evaluation metrics include Area Under Curve (AUC) of the success plot, precision plots, overall success rate (SR), and for multi-object tracking, metrics like Multi-Object Tracking Accuracy (MOTA).[14]

Transformer-based trackers consistently report state-of-the-art results on these benchmarks in recent conference publications. For example, SiamTPN (WACV'22) reported an AUC score of 58.1% on LaSOT.[53] MixFormerV2-B (NeurIPS'23) achieved an AUC of 70.6% on LaSOT and 56.7% on TNL2k.[60] TBD methods benchmarked at ICCVW'21 showed MOTA scores up to 98.7% (using Tracktor with DETR on Anti-UAV IR).[14] SiamFC combined with a Kalman filter reported accuracy/success rates of 66.0%/47.4% on UAV123.[45]

Table 3.1: Comparative Analysis of Representative Drone Tracking Methods (from or relevant to Top Conferences)

Method Name	Tracking Paradigm	Modality	Key Innovation / Feature	Target Conference/Venue (Example)	Reported Performance (Metric, Dataset)	Real-Time Capability (Reported FPS)	Snippets
TransT	Transformer	RGB	Attention-based feature fusion	CVPR'21	High Accuracy (Generic Benchmarks)	-	[54]
STARK	Transformer	RGB	Spatio-Temporal Modeling, Controll	ICCV'21	SOTA (at time) on LaSOT, TrackingNet,	~40 FPS (GPU)	[55]

			er Head		GOT-10k		
MixFormerV2-B	Transformer (Unified)	RGB	Fully Transformer, Prediction Tokens, Distillation	NeurIPS'23	70.6% (AUC, LaSOT), 56.7% (AUC, TNL2k)	165 FPS (GPU)	[60]
MixFormerV2-S	Transformer (Unified)	RGB	Efficient version of MixFormerV2	NeurIPS'23	Competitive (e.g., >FEAR-L on LaSOT)	Real-time (CPU)	[60]
OSTrack	Transformer (Unified)	RGB	One-stream, ViT-based, Candidate Elimination	ECCV'22	SOTA on LaSOT, TrackingNet, GOT-10k	~110 FPS (GPU, OSTrack-256)	[18]
SiamTPN	Siamese + Transformer	RGB	Lightweight Pyramid Transformer, Cross Attention	WACV'22	58.1% (AUC, LaSOT)	>30 FPS (CPU), 45 FPS (CPU+OpenVINO)	[53]
SGLATrack	Transformer	RGB	Similarity-Guided Layer Adaptation for Efficiency	ArXiv (Targets UAV)	Improved Accuracy-Speed Trade-off	Real-time (Claimed for UAV)	[51]

UTT (OSTrack baseline)	Transformer + Modules	TIR	Unified: Local Tracking, Global Detect, BG Correction, SOD	CVPRW'23 (Anti-UAV)	2nd Place in 3rd Anti-UAV Challenge	-	[18]
SiamFC + KF	Siamese + Filter	RGB	SiamFC + Status Detect + Kalman Filter Relocation	J. Syst. Eng. Electron.	66.0%/47.4% (Acc/SR, UAV123), 72.0%/58.6% (Custom)	55-62 FPS	[45]
Model-Based + PF/UKF/UBiF	Model-Based + Filter	RGB	3D Model Projection, Particle Filter, UKF/UBiF	OCEANS'15/'18 (Robotics Conf. related)	Low error for landing	-	[65]
TBD (DETR + Tracktor)	TBD	IR	Detection + Association	ICCVW'21	98.7% (MOTA, Anti-UAV IR)	Depends on Detector (~21 FPS)	[14]

Note: Performance metrics and FPS are highly dependent on the specific model version, dataset, hardware, and evaluation protocol. This table provides representative examples.

4. Payload Identification and Classification

Identifying what a drone is carrying (its payload) is a critical capability, particularly for security applications (detecting hazardous materials or unauthorized surveillance equipment) and logistics (verifying cargo). Research in this area often leverages general object recognition techniques developed within the broader computer vision

community and presented at premier conferences.

4.1. Leveraging Object Recognition Techniques

Payload identification can often be framed as an object detection or recognition task applied to the specific context of objects attached to or carried by a drone. Therefore, advancements in general object detection models presented at CVPR, ICCV, ECCV, etc., are highly relevant.[20] Standard detectors like YOLO [5], and potentially others like Faster R-CNN, SSD, or DETR, can be trained or fine-tuned to recognize specific types of payloads visible in RGB or IR imagery captured from ground-based cameras or other surveillance platforms. Papers discussing object detection from drones in surveillance or delivery contexts provide insights into applicable methods, even if not explicitly labeled "payload identification".[5] The relative scarcity of papers explicitly titled "drone payload identification" in the premier conference literature, compared to drone detection or tracking, suggests that this task is often addressed by applying state-of-the-art general object detectors rather than developing entirely bespoke algorithms. The primary challenge lies in adapting these general methods to the specific constraints of viewing payloads on drones.

4.2. Specific Methods and Challenges

Applying object detection to payloads presents unique challenges:

- **Appearance Variability:** Payloads can vary significantly in shape, size, and material.
- **Occlusion:** The drone's body or propellers might partially obscure the payload.
- **Small Relative Size:** The payload might occupy only a small portion of the overall drone image or the entire scene.
- **Viewpoint:** The payload might only be visible from specific angles.
- **Fine-Grained Classification:** Distinguishing between similar-looking payloads (e.g., different types of sensors or containers) may be necessary.

Specific methods mentioned in related contexts include marker detection for drone delivery applications, such as the Vertical Grid Screening (VGS) method using YOLOv5 to identify markers on balconies.[6] While perhaps not from a top-tier CV conference, this illustrates the application of object detection for identifying delivery targets, analogous to identifying a payload's destination or type. Another example involves using PIR sensors within a delivery compartment to detect the presence of an object (payload).[5]

4.3. Role of RGB, IR, and Fusion

Both RGB and IR sensing can contribute to payload identification:

- **RGB:** Provides crucial information about the payload's **shape, color, texture, and any visible markings or labels**.[20] This is essential for recognizing most common objects based on their visual appearance.
- **IR:** Offers the unique capability to detect **thermal signatures**. This could be valuable for identifying payloads that generate heat (e.g., active electronic equipment, chemical reactions) or have distinct thermal properties compared to the drone or background.[1] For security applications, IR might help detect potentially hazardous materials or devices with characteristic thermal profiles. This unique sensing modality makes IR and fusion research particularly relevant for identifying payloads that are ambiguous or invisible in RGB alone.
- **Fusion:** Combining RGB and IR data could enhance robustness. For instance, RGB might identify the shape of a metallic container, while IR could indicate if it contains something significantly warmer or colder than ambient temperature. Fusion strategies discussed for detection (e.g., feature-level fusion like MCFNet [28], adaptive alignment like ODAF [26], or decision-level fusion like CPROS [49]) could potentially be applied or adapted for payload recognition, improving performance across different lighting and environmental conditions.

5. Real-Time Implementation Aspects

For drone detection and tracking systems to be practically useful, especially in dynamic scenarios or for counter-UAV applications, real-time processing is often a mandatory requirement. This typically means achieving inference speeds faster than the incoming frame rate (e.g., < 30-50 ms per frame).[13] Research presented at premier conferences increasingly addresses this need through efficient architectures, model optimization, and consideration of hardware platforms.

5.1. Efficient Architectures in Conference Papers

Achieving real-time performance necessitates the use of computationally efficient model architectures. Conference papers often feature or benchmark lightweight designs:

- **Efficient CNNs:** YOLO variants are consistently highlighted for their speed-accuracy balance.[15] Other lightweight backbones like MobileNet [15] or ShuffleNetV2 (used in the SiamTPN tracker [53]) are employed to reduce computational load. Techniques like depthwise separable convolutions (used in MobileNet) [15] or module designs like the Ghost module (integrated into YOLO [15]) further improve efficiency.
- **Efficient Transformers:** While Transformers can be computationally intensive,

recent research presented at top conferences focuses on developing efficient variants for real-time tracking. Examples include MixFormerV2-S, designed for CPU real-time speed [60], SGLATrack, which uses layer adaptation for efficiency in UAV tracking [51], and SiamTPN, leveraging a lightweight Transformer within a pyramid network for CPU performance.[53]

The frequent presentation of multiple model variants (e.g., small, medium, large versions of YOLO [44] or MixFormerV2 [60]) and explicit benchmarking of both accuracy and speed (FPS) in conference papers [14] demonstrates that the accuracy-speed trade-off is a central consideration in state-of-the-art research. Model design is often driven by the need to balance high performance with the computational constraints of target deployment platforms.

5.2. Model Optimization Techniques

Beyond architectural choices, techniques to optimize trained models for faster inference are crucial and discussed in relevant conference literature:

- **Knowledge Distillation:** Transferring knowledge from a large, accurate "teacher" model to a smaller, faster "student" model. This is explored in contexts like remote sensing [71] and explicitly used in developing efficient trackers like MixFormerV2 (dense-to-sparse and deep-to-shallow distillation [60]) and SGLATrack.[51] TATrack also uses distillation for efficient UAV tracking.[51]
- **Hardware-Aware Optimization:** Designing or optimizing models with the target hardware capabilities in mind.[71] This might involve choosing specific operations that run efficiently on a given processor (CPU, GPU, specialized accelerator).
- **Pruning and Quantization:** While less explicitly detailed in the provided snippets for drone perception, these are standard techniques for reducing model size and computational cost by removing redundant parameters (pruning) or using lower-precision numerical representations (quantization). They are often implied components of achieving efficiency for embedded deployment.

5.3. Hardware Considerations and Benchmarks

The choice of hardware platform significantly impacts real-time feasibility. Research often considers deployment on:

- **GPUs:** Standard platform for deep learning, offering high parallelism. Benchmarks often specify the GPU model (e.g., K40 [44], Quadro M1000M [44]).
- **Embedded GPUs:** Platforms like NVIDIA Jetson (e.g., Jetson AGX Xavier [56]) are common targets for onboard processing on UAVs or edge devices.[24]
- **CPUs:** For scenarios with very limited resources or power, CPU performance is

critical. Trackers like SiamTPN [53] and MixFormerV2-S [60] explicitly target real-time CPU execution.

- **Neuromorphic Hardware / Event Cameras:** Dynamic Vision Sensors (DVS) offer an alternative sensing modality with inherent advantages for low latency (< 10µs response time) and low power consumption (< 10mW sensor power).[13] Research presented at conferences like WACV and ICCV explores using event cameras coupled with Spiking Neural Networks (SNNs) or other efficient processing methods for tasks like fast-moving drone detection (e.g., F-UAV-D system achieving < 50ms inference at < 15W total system power [13]) or tracking.[9] The growing presence of event camera research in premier vision conferences suggests an increasing interest in leveraging their unique properties to overcome limitations of traditional cameras in latency-critical or power-constrained applications like drone perception.

Conference papers often provide concrete Frames Per Second (FPS) or latency benchmarks:

- **Detection:** Faster RCNN (~18 FPS [14]), YOLOv3 (~36 FPS [14]), SSD512 (~32 FPS [14]), DETR (~21 FPS [14]), YOLOv4+attn (~15 FPS [15]), MobileNet-based YOLO (~82 FPS [15]), YOLO+motion (~24 FPS [15]). F-UAV-D (Event-based) < 50ms latency.[13]
- **Tracking:** YOLOv5-based TBD (~5 FPS for Faster RCNN backbone, faster for YOLO [44]), SiamTPN (>30 FPS CPU, 45 FPS CPU+OpenVINO [53]), MixFormerV2-B (~165 FPS GPU [60]), MixFormerV2-S (Real-time CPU [60]), SiamSTM (>35 FPS Jetson AGX Xavier [56]), SiamFC+KF (55-62 FPS [45]), KF tracker (43 FPS [52]).

Table 5.1: Real-Time Performance Benchmarks of Representative Models (from or relevant to Top Conferences)

Model Name	Modality	Task	Target Conference/Venue (Example)	Reported Speed (FPS / Latency)	Hardware Platform (Example)	Accuracy Metric (Example)	Snippets
Faster R-CNN	RGB	Detection	ICCVW'21	~18 FPS	GPU (unspecified)	mAP	[14]

YOLOv3	RGB	Detection	ICCVW'21	~36 FPS	GPU (unspecified)	mAP	[14]
SSD512	RGB	Detection	ICCVW'21	~32 FPS	GPU (unspecified)	mAP	[14]
DETR	RGB/IR	Detection	ICCVW'21	~21 FPS	GPU (unspecified)	mAP	[14]
MobileNet-YOLO	RGB	Detection	ArXiv (Cites Conf. Work)	~82 FPS	GPU (unspecified)	mAP	[15]
F-UAV-D	Event Camera	Detection	ArXiv (Cites WACV/ICCV)	< 50 ms	Embedded (< 15W)	Accuracy	[13]
MixFormerV2-B	RGB	Tracking	NeurIPS'23	~165 FPS	GPU	AUC	[60]
MixFormerV2-S	RGB	Tracking	NeurIPS'23	Real-time	CPU	AUC	[60]
SiamTPN	RGB	Tracking	WACV'22	>30 FPS / 45 FPS (Optimized)	CPU	AUC	[53]
SiamSTM	RGB	Tracking	Remote Sens. J.	>35 FPS	NVIDIA Jetson AGX Xavier	Precision/Success	[56]
SiamFC + KF	RGB	Tracking	J. Syst. Eng. Electron.	55-62 FPS	GPU (unspecified)	Accuracy/Success	[45]

TBD (FasterRCNN+KCF)	RGB	Detection	Sensors J. (Cites Conf. Work)	19 FPS / 33 FPS	2GB / 4GB GPU-RAM	-	[41]

Note: FPS values are highly dependent on hardware, model size, input resolution, and implementation details. This table provides examples reported in the literature.

6. Discussion and Future Research Directions

6.1. Synthesis of State-of-the-Art Findings

This review, focused on premier CV, Robotics, and AI conference literature, reveals several key trends in vision-based drone perception using RGB and IR sensors. Deep learning, particularly CNNs and increasingly Transformers, dominates both detection and tracking tasks, largely supplanting traditional methods in high-impact research. For detection, while efficient one-stage detectors like YOLO are popular for real-time needs, Transformer-based detectors like DETR are also being benchmarked. In tracking, a significant shift towards end-to-end Transformer architectures (e.g., MixFormer, OSTrack, STARK, TransT) is evident, offering powerful spatio-temporal modeling capabilities and achieving state-of-the-art results on challenging UAV benchmarks. Siamese networks remain relevant, especially lightweight versions tailored for UAVs (e.g., SiamTPN).

RGB-IR fusion is recognized as crucial for robust, all-condition performance. Research presented at top venues like CVPR highlights that naive fusion is insufficient; sophisticated techniques addressing sensor misalignment, such as adaptive feature alignment (e.g., ODAF [26]), are critical for unlocking the potential of multi-modal data. For trajectory estimation and robust tracking in practical applications (often presented at ICRA/IROS), classical filtering techniques like Kalman Filters (KF, EKF, UKF) and Particle Filters remain standard tools integrated with visual trackers to smooth estimates, handle noise, and predict motion.[37] Payload identification appears to be primarily addressed by applying general object detection techniques to the drone context, rather than through highly specialized algorithms. Finally, the accuracy-speed trade-off is a persistent theme, driving research into efficient architectures, model optimization (like distillation), and exploration of alternative sensors like event cameras for low-latency performance.

6.2. Persistent Challenges Identified in Literature

Despite significant progress documented in premier conference papers, several

challenges remain focal points of ongoing research:

- **Small, Fast, and Distant Targets:** Reliably detecting and tracking drones that occupy few pixels, move rapidly or erratically, and are observed from long distances remains a fundamental difficulty.[2]
- **Environmental Robustness:** Achieving consistent performance amidst complex backgrounds, dynamic clutter, occlusions, and varying illumination or adverse weather conditions is still challenging.[16]
- **Fusion Reliability:** Effectively fusing multi-modal data, especially under conditions of sensor noise, calibration errors, and significant spatial or temporal misalignment, requires further advancement.[26]
- **Real-Time Constraints:** Balancing state-of-the-art accuracy with the computational limitations of onboard or edge platforms remains a key design challenge.[44]
- **Discrimination:** Reliably distinguishing drones from visually similar objects like birds or other background elements requires robust classification capabilities.[10]
- **Data Limitations:** A significant bottleneck is the availability of large-scale, diverse, accurately annotated datasets, particularly for multi-modal fusion (with synchronized and calibrated sensors), complex scenarios (swarms, adversarial tactics, specific payloads), and varied environmental conditions.[11] While numerous datasets exist [2], the persistent nature of the above challenges suggests current benchmarks may not fully capture the complexities needed to drive robust real-world solutions. This continuous need for better data is crucial for validating and advancing methods presented at top conferences.

6.3. Emerging Trends from Recent Conferences

Based on recent publications in the target venues, several trends are shaping the future of drone perception:

- **Advanced and Efficient Transformers:** Research continues to refine Transformer architectures for vision, focusing not only on accuracy but also on computational efficiency (e.g., MixFormerV2 [60], SGLATrack [51]), robustness [55], and unifying different tasks like detection and tracking within a single framework.[18] This suggests a move towards more integrated, end-to-end systems.
- **Sophisticated Fusion:** Moving beyond simple feature concatenation or decision averaging towards more principled fusion methods that explicitly handle misalignment (e.g., ODAF [26]), potentially incorporating cross-modal attention [9] or uncertainty modeling.
- **Event-Based Vision:** Increasing exploration of neuromorphic sensors (DVS) and associated processing techniques (SNNs, specialized algorithms) for scenarios

demanding extremely low latency, high dynamic range, and low power, particularly for fast motion.[9]

- **Multi-Modal Integration:** While this review focuses on RGB/IR, there is broader interest in fusing data from other sensors like LiDAR, radar, RF, and audio for even greater robustness, although this often falls outside the scope of pure CV conferences.[8]
- **Domain Adaptation and Generalization:** Developing techniques to ensure models trained on one dataset or environment perform well in others, crucial for real-world deployment.[61]
- **Reduced Supervision:** Exploring self-supervised, weakly-supervised, or unsupervised learning methods to lessen the dependence on large, meticulously labeled datasets.[50]

6.4. Recommendations for Future Research Avenues

Based on the reviewed literature and identified gaps, promising directions for future research include:

- **Highly Robust Fusion:** Developing fusion algorithms inherently resilient to sensor noise, calibration drift, and significant spatio-temporal misalignment, perhaps by explicitly modeling uncertainties or leveraging learned alignment priors.
- **Explainable AI (XAI):** Incorporating XAI techniques to understand the decision-making process of deep learning models for drone perception, enhancing trust and facilitating debugging in safety-critical systems.[72]
- **Long-Term Tracking and Re-Identification:** Improving the ability to maintain track over extended periods, handle prolonged occlusions, and re-identify drones if track is lost, especially in scenarios with multiple similar drones.[3]
- **Adversarial Robustness:** Designing detection and tracking systems explicitly hardened against adversarial examples or physical modifications aimed at evading perception.[22]
- **Fine-Grained Payload Analysis:** Advancing from simple payload detection to detailed classification (e.g., identifying specific equipment types or materials) and potentially estimating payload properties like weight or state.
- **Comprehensive Dataset Curation:** Creating large-scale, diverse, multi-modal datasets with high-quality, synchronized RGB and IR data, accurate ground truth (including 3D trajectories and payload labels), covering a wider range of drone types, payloads, flight dynamics, environmental conditions, and challenging scenarios (e.g., swarms, GPS-denied, adversarial). Standardized fusion protocols and benchmarks are needed.[22]
- **Efficient and Adaptive Onboard Processing:** Continuing research into

lightweight, hardware-aware models that can run efficiently on resource-constrained UAV platforms, potentially adapting their complexity or focus based on the current context or available resources.

7. Conclusion

The task of detecting, tracking, and identifying payloads on drones using RGB and IR sensors is a critical and rapidly evolving area of research, driven by the dual-use nature of UAV technology. This review, focusing on contributions from premier computer vision, robotics, and AI conferences, highlights the dominance of deep learning approaches, particularly CNNs and, more recently, Transformers. Transformers have emerged as the state-of-the-art for high-performance tracking, enabling unified architectures that jointly model spatial and temporal information. RGB-IR fusion is crucial for achieving robustness across diverse operating conditions, with recent advancements in adaptive feature alignment tackling the key challenge of sensor misalignment. While visual trackers provide localization, classical filtering techniques like Kalman and Particle filters remain essential components in robotics-focused literature for refining trajectories and enabling motion analysis, including the inference of approaching or receding behavior. Payload identification currently relies heavily on adapting general object recognition methods. Despite significant progress, challenges related to small object perception, environmental robustness, real-time performance on constrained hardware, and the availability of comprehensive multi-modal datasets persist. Future research directions point towards more robust and explainable fusion techniques, advanced efficient Transformer models, leveraging event-based vision, enhancing long-term tracking and adversarial resilience, developing fine-grained payload analysis capabilities, and curating more challenging benchmark datasets to bridge the gap between academic research and real-world deployment. The convergence of detection and tracking within unified frameworks also represents a significant ongoing trend. Addressing these challenges will be key to fully realizing the potential of vision-based systems for safe and secure drone operations.

Works cited

1. mdpi-res.com, accessed May 3, 2025, https://mdpi-res.com/bookfiles/book/9130/Intelligent_Image_Processing_and_Sensing_for_Drones.pdf?v=1736561179
2. Object Detection in UAV Images via Global Density Fused Convolutional Network - MDPI, accessed May 3, 2025, https://www.mdpi.com/2072-4292/12/19/3140
3. (PDF) Applications, databases and open computer vision research ..., accessed May 3, 2025,

https://www.researchgate.net/publication/349504594_Applications_databases_and_open_computer_vision_research_from_drone_videos_and_images_a_survey
4. Intelligent Recognition and Detection for Unmanned Systems - MDPI, accessed May 3, 2025, https://mdpi-res.com/bookfiles/book/9925/Intelligent_Recognition_and_Detection_for_Unmanned_Systems.pdf?v=1739758103
5. Drone Delivery With Object Detection | PDF | Unmanned Aerial Vehicle - Scribd, accessed May 3, 2025, https://fr.scribd.com/document/600616310/Drone-Delivery-with-Object-Detection
6. Drone High-Rise Aerial Delivery with Vertical Grid Screening - MDPI, accessed May 3, 2025, https://www.mdpi.com/2504-446X/7/5/300
7. Vision-Based Learning for Drones: A Survey - arXiv, accessed May 3, 2025, https://arxiv.org/html/2312.05019v2
8. LiDAR Technology for UAV Detection: From Fundamentals and Operational Principles to Advanced Detection and Classification Techniques - MDPI, accessed May 3, 2025, https://www.mdpi.com/1424-8220/25/9/2757
9. Neuromorphic Drone Detection: an Event-RGB Multimodal Approach - arXiv, accessed May 3, 2025, https://arxiv.org/html/2409.16099v1
10. Real-Time and Accurate Drone Detection in a Video with a Static Background - PMC, accessed May 3, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC7412503/
11. Deep learning-based drone detection in infrared imagery with limited training data, accessed May 3, 2025, https://www.researchgate.net/publication/346822184_Deep_learning-based_drone_detection_in_infrared_imagery_with_limited_training_data
12. Alberto del Bimbo | Papers With Code, accessed May 3, 2025, https://paperswithcode.com/author/alberto-del-bimbo-1
13. Towards Real-Time Fast Unmanned Aerial Vehicle Detection Using Dynamic Vision Sensors - arXiv, accessed May 3, 2025, https://arxiv.org/html/2403.11875v1
14. openaccess.thecvf.com, accessed May 3, 2025, https://openaccess.thecvf.com/content/ICCV2021W/AntiUAV/papers/Isaac-Medina_Unmanned_Aerial_Vehicle_Visual_Detection_and_Tracking_Using_Deep_Neural_ICCVW_2021_paper.pdf
15. Real-Time Detection for Small UAVs: Combining YOLO and Multi-frame Motion Analysis, accessed May 3, 2025, https://arxiv.org/html/2411.02582v1
16. Investigation of UAV Detection in Images With Complex Backgrounds and Rainy Artifacts - CVF Open Access, accessed May 3, 2025, https://openaccess.thecvf.com/content/WACV2024W/RWS/papers/Munir_Investigation_of_UAV_Detection_in_Images_With_Complex_Backgrounds_and_WACVW_2024_paper.pdf
17. Drone Detection using Deep Learning - DiVA portal, accessed May 3, 2025, https://www.diva-portal.org/smash/get/diva2:1738770/FULLTEXT01.pdf
18. A Unified Transformer Based Tracker for Anti-UAV Tracking - CVF Open Access, accessed May 3, 2025, https://openaccess.thecvf.com/content/CVPR2023W/Anti-UAV/papers/Yu_A_Unifi

ed_Transformer_Based_Tracker_for_Anti-UAV_Tracking_CVPRW_2023_paper.pdf
19. A Unified Transformer Based Tracker for Anti-UAV Tracking - CVPR 2023 Open Access Repository - The Computer Vision Foundation, accessed May 3, 2025, https://openaccess.thecvf.com/content/CVPR2023W/Anti-UAV/html/Yu_A_Unified_Transformer_Based_Tracker_for_Anti-UAV_Tracking_CVPRW_2023_paper.html
20. A Survey of Computer Vision Methods for 2D Object Detection from Unmanned Aerial Vehicles - PubMed Central, accessed May 3, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC8321148/
21. Drone Detection and Tracking in Real-Time by Fusion of Different Sensing Modalities, accessed May 3, 2025, https://www.mdpi.com/2504-446X/6/11/317
22. Securing the Skies: A Comprehensive Survey on Anti-UAV Methods, Benchmarking, and Future Directions - arXiv, accessed May 3, 2025, https://arxiv.org/html/2504.11967v1
23. Vision-Based Drone Detection in Complex Environments: A Survey - MDPI, accessed May 3, 2025, https://www.mdpi.com/2504-446X/8/11/643
24. TractorEye: Vision-based Real-time Detection for Autonomous Vehicles in Agriculture - CORE, accessed May 3, 2025, https://core.ac.uk/download/229217447.pdf
25. Publications - Alexandre Bernardino - Google Sites, accessed May 3, 2025, https://sites.google.com/view/alexbernardino/publications
26. openaccess.thecvf.com, accessed May 3, 2025, https://openaccess.thecvf.com/content/CVPR2024/papers/Chen_Weakly_Misalignment-free_Adaptive_Feature_Alignment_for_UAVs-based_Multimodal_Object_Detection_CVPR_2024_paper.pdf
27. (PDF) Drone Detection and Tracking in Real-Time by Fusion of Different Sensing Modalities, accessed May 3, 2025, https://www.researchgate.net/publication/364766917_Drone_Detection_and_Tracking_in_Real-Time_by_Fusion_of_Different_Sensing_Modalities
28. (PDF) MCFNet: Research on small target detection of RGB-infrared under UAV perspective with multi-scale complementary feature fusion - ResearchGate, accessed May 3, 2025, https://www.researchgate.net/publication/387750223_MCFNet_Research_on_small_target_detection_of_RGB-infrared_under_UAV_perspective_with_multi-scale_complementary_feature_fusion
29. CORE Computer Science Conference Rankings, accessed May 3, 2025, https://people.iiti.ac.in/~artiwari/cseconflist.html
30. Computer Vision & Pattern Recognition - Google Scholar Metrics, accessed May 3, 2025, https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=eng_computervisionpatternrecognition
31. Top Computer Vision Conferences Around The Globe - Deepomatic, accessed May 3, 2025, https://deepomatic.com/blog/top-computer-vision-conferences-around-the-globe
32. Top Computer Vision Conferences in 2025 - AI Superior, accessed May 3, 2025,

https://aisuperior.com/computer-vision-conferences/
33. Best Computer Science Conferences Ranking Image Processing & Computer Vision 2024, accessed May 3, 2025, https://research.com/conference-rankings/computer-science/computer-vision
34. Conference Calendar for Computer Vision, Image Analysis and Related Topics, accessed May 3, 2025, http://conferences.visionbib.com/Iris-Conferences.html
35. Ranked Conference List - University of Oxford Department of Computer Science, accessed May 3, 2025, https://www.cs.ox.ac.uk/people/michael.wooldridge/conferences.html
36. Computer Vision and Pattern Recognition Jun 2023 - arXiv, accessed May 3, 2025, http://arxiv.org/list/cs.CV/2023-06?skip=150&show=2000
37. Vision-Based Online Localization and Trajectory Smoothing for Fixed-Wing UAV Tracking a Moving Target - CVF Open Access, accessed May 3, 2025, https://openaccess.thecvf.com/content_ICCVW_2019/html/VISDrone/Zhou_Vision-Based_Online_Localization_and_Trajectory_Smoothing_for_Fixed-Wing_UAV_Tracking_ICCVW_2019_paper.html
38. List of Accepted Papers - IEEE IGARSS 2024 || Athens, Greece || 7 - 12 July, 2024, accessed May 3, 2025, https://2024.ieeeigarss.org/papers/accepted_papers.php
39. (PDF) Vision-Based Drone Detection in Complex Environments: A Survey - ResearchGate, accessed May 3, 2025, https://www.researchgate.net/publication/385577600_Vision-Based_Drone_Detection_in_Complex_Environments_A_Survey
40. YOLOMG: Vision-based Drone-to-Drone Detection with Appearance and Pixel-Level Motion Fusion - arXiv, accessed May 3, 2025, https://arxiv.org/html/2503.07115v1
41. Real-Time and Accurate Drone Detection in a Video with a Static Background, accessed May 3, 2025, https://www.researchgate.net/publication/342856036_Real-Time_and_Accurate_Drone_Detection_in_a_Video_with_a_Static_Background
42. A Survey on Hough Transform, Theory, Techniques and Applications - ResearchGate, accessed May 3, 2025, https://www.researchgate.net/publication/272195556_A_Survey_on_Hough_Transform_Theory_Techniques_and_Applications
43. Realtime Drone Detection Using YOLOv8 and TensorFlow.JS - Journal of Engineering Sciences, accessed May 3, 2025, https://jespublication.com/uploads/2024-V15I2032.pdf
44. Visual Drone Detection and Tracking for Autonomous Operation from Maritime Vessel, accessed May 3, 2025, https://mecatron.rma.ac.be/pub/2022/ISMCR-Drone_detection_tracking_FullPaper.pdf
45. UAV object tracking for air-ground targets based on status detection and Kalman filter, accessed May 3, 2025, https://www.sciopen.com/article/10.7527/S1000-6893.2024.29834
46. (PDF) Drone Tracking with Drone using Deep Learning - ResearchGate, accessed May 3, 2025,

https://www.researchgate.net/publication/362833814_Drone_Tracking_with_Drone_using_Deep_Learning
47. ZAKAUDD/Awesome-Transformer-Attention - GitHub, accessed May 3, 2025, https://github.com/ZAKAUDD/Awesome-Transformer-Attention
48. Algorithms for the Assignment and Transportation Problems | SIAM Journal on Applied Mathematics, accessed May 3, 2025, https://epubs.siam.org/doi/10.1137/0105003
49. CPROS: A Multimodal Decision-Level Fusion Detection Method ..., accessed May 3, 2025, https://www.mdpi.com/2072-4292/16/15/2745
50. chakravarthi589/Event-based-Vision_Resources: Resources Related to Event-based Vision | Event Cameras | DVS - GitHub, accessed May 3, 2025, https://github.com/chakravarthi589/Event-based-Vision_Resources
51. Similarity-Guided Layer-Adaptive Vision Transformer for UAV Tracking - arXiv, accessed May 3, 2025, https://arxiv.org/html/2503.06625v1
52. Pedestrian Tracking in UAV Images With Kalman Filter Motion Estimator and Correlation Filter | Request PDF - ResearchGate, accessed May 3, 2025, https://www.researchgate.net/publication/369077849_Pedestrian_Tracking_in_UAV_Images_With_Kalman_Filter_Motion_Estimator_and_Correlation_Filter
53. [2110.08822] Siamese Transformer Pyramid Networks for Real-Time UAV Tracking - ar5iv, accessed May 3, 2025, https://ar5iv.labs.arxiv.org/html/2110.08822
54. SiamHSFT: A Siamese Network-Based Tracker with Hierarchical Sparse Fusion and Transformer for UAV Tracking, accessed May 3, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC10648809/
55. Little-Podi/Transformer_Tracking: This repository is a paper digest of Transformer-related approaches in visual tracking tasks. - GitHub, accessed May 3, 2025, https://github.com/Little-Podi/Transformer_Tracking
56. Slight Aware Enhancement Transformer and Multiple Matching Network for Real-Time UAV Tracking - MDPI, accessed May 3, 2025, https://www.mdpi.com/2072-4292/15/11/2857
57. Awesome-Transformer-Attention/README_2.md at main - GitHub, accessed May 3, 2025, https://github.com/cmhungsteve/Awesome-Transformer-Attention/blob/main/README_2.md
58. Siamese Transformer Pyramid Networks for Real-Time UAV Tracking - Code Ocean, accessed May 3, 2025, https://codeocean.com/explore/44bcfb34-5375-4c80-b169-6169e66d36b2?query=tag%3Atracking&page=1&filter=all
59. Transformer-in-Computer-Vision/main/tracking.md at main - GitHub, accessed May 3, 2025, https://github.com/Yangzhangcst/Transformer-in-Computer-Vision/blob/main/main/tracking.md
60. MixFormerV2: Efficient Fully Transformer Tracking - NIPS papers, accessed May 3, 2025, https://papers.neurips.cc/paper_files/paper/2023/file/b7870bd43b2d133a1ed95582ae5d82a4-Paper-Conference.pdf

61. Accepted Papers - IROS 2024, accessed May 3, 2025, https://iros2024-abudhabi.org/accepted-papers
62. Tiny Drone Tracking Framework Using Multiple Trackers and Kalman-based Predictor - Journal of Web Engineering, accessed May 3, 2025, https://journals.riverpublishers.com/index.php/JWE/article/download/5455/10209/37913
63. An Integrated Visual System for Unmanned Aerial Vehicles Tracking and Landing on the Ground Vehicles - arXiv, accessed May 3, 2025, https://arxiv.org/pdf/2301.00198
64. UAV Tracking with Lidar as a Camera Sensor in GNSS-Denied Environments - arXiv, accessed May 3, 2025, https://arxiv.org/pdf/2303.00277
65. 3D Model-based estimation for UAV tracking - ResearchGate, accessed May 3, 2025, https://www.researchgate.net/publication/330300778_3D_Model-based_estimation_for_UAV_tracking
66. A Ground-Based Vision System for UAV Tracking - VisLab, accessed May 3, 2025, https://vislab.isr.tecnico.ulisboa.pt/publications/15-Oceans-NSantos.pdf
67. 3D Model-based estimation for UAV tracking - VisLab, accessed May 3, 2025, https://vislab.isr.tecnico.ulisboa.pt/wp-content/uploads/2020/07/nsantos-oceans2018.pdf
68. Drone Based Object Tracking - CatalyzeX, accessed May 3, 2025, https://www.catalyzex.com/s/Drone%20Based%20Object%20Tracking
69. Vision-Based HAR in UAV Videos Using Histograms and Deep Learning Techniques - PMC, accessed May 3, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC10007408/
70. The State of Aerial Surveillance: A Survey - Computer Vision Lab, accessed May 3, 2025, http://cvlab.cse.msu.edu/pdfs/Nguyen_Fookes_Sridharan_Tian_Liu_Liu_Ross_AerialSurveillance.pdf
71. Applications of Knowledge Distillation in Remote Sensing: A Survey - arXiv, accessed May 3, 2025, https://arxiv.org/html/2409.12111v1
72. Alexandre Bernardino - DBLP, accessed May 3, 2025, https://dblp.org/pid/53/5306