Breaking Smooth-Motion Assumptions: A UAV Benchmark for Multi-Object Tracking in Complex and Adverse Conditions

Supplementary Material

S1. Additional Dataset Details

This section provides a granular, sequence-level statistical overview of the DynUAV benchmark, along with qualitative visualizations showcasing its diversity and inherent challenges. These details complement the aggregated statistics presented in the main paper.

S1.1. Per-Sequence Statistics

To facilitate fine-grained analysis and allow researchers to target specific challenges, we provide detailed statistics for each sequence in our validation and test sets in Table S1. The table includes metrics for sequence length, object count, and motion characteristics. The *mean inter-frame IoU* serves as a proxy for motion intensity; a lower value typically indicates more severe camera ego-motion or faster object movement, leading to greater temporal inconsistency. The final column provides qualitative labels for the primary challenges inherent in each sequence, linking the quantitative data to real-world tracking difficulties.

S1.2. Qualitative Analysis of IDSW Failures

To visually diagnose the root causes of association failures on DynUAV, Figure S1 presents a gallery of typical IDSW events from several challenging test sequences. Each row showcases a sequence, highlighting specific moments where a single ground-truth object is erroneously assigned a new ID. Our analysis reveals that these failures are not random but are systematically triggered by the complex observational dynamics introduced by our benchmark's design.

Sequence 004 (campus): This sequence illustrates how common tracking challenges combine to induce IDSW.

- *Re-appearance after occlusion (frames 156-214):* The first example shows a classic failure mode. A target (initially ID:8) is briefly occluded by environmental structures. Upon re-entry, the tracker fails to associate it with the original trajectory due to a slight change in appearance or position and incorrectly initializes it as a new object (ID:12).
- *Perspective shift (frames 840-947):* The second case highlights the impact of ego-motion. As the UAV maneuvers, the viewing angle of the target pedestrian shifts significantly. This change in perspective alters the target's appearance features beyond what the tracker's Re-ID model can handle, leading to a loss of identity (ID:25 $\rightarrow$ ID:36).
- *Detection error (frames 1213-1239):* The third example illustrates how tracking performance is fundamentally dependent on detection quality. The detector erroneously identifies only a small part of the large vehicle (ID:1). In the subsequent frame, a more accurate, larger bounding box is detected. The drastic change in box size and position causes the tracker to fail the association, resulting in an IDSW (ID:61).

Sequence 016 (campus under foliage): This sequence demonstrates how partial and dynamic occlusions lead to tracking failures.

- *Dynamic partial occlusion (frames 132-231):* The target vehicle moves under sparse trees. Due to the UAV's continuous motion, the visible, un-occluded parts of the vehicle are constantly changing. This creates a highly unstable appearance feature representation, confusing the tracker and causing it to fragment the trajectory into multiple IDs (e.g., ID:59 $\rightarrow$ ID:86 $\rightarrow$ ID:59 $\rightarrow$ ID:114).
- *Posture change (frames 277-296):* A pedestrian's posture changes (e.g., from walking to turning). This non-rigid deformation, although subtle, is sufficient to cause a mismatch in the appearance feature space, leading to an IDSW (ID:52 $\rightarrow$ ID:132).

Sequence 027 (road scene): This sequence demonstrates how viewpoint changes persistently challenge tracking, even for rigid objects.

- *Continuous position and perspective change (frames 128-880):* As the UAV follows a curving road, the tracked vehicle is observed from a continuously changing perspective (e.g., from rear-view to side-view). Over this long duration, the accumulated appearance change eventually surpasses the tracker's re-identification threshold, leading to a series of IDSW (ID:4 $\rightarrow$ ID:13 $\rightarrow$ ... $\rightarrow$ ID:50). This illustrates the challenge of long-term identity maintenance under sustained ego-motion.

Sequence 055 (nighttime sports field): This sequence demonstrates the compounded difficulty of adverse lighting and unpredictable human motion.

- *Compounded lighting and posture changes (frames 170-996):* The challenges in this nighttime scene are twofold. First, the harsh, directional stadium lighting creates long shadows and corrupts color features, degrading the quality of appearance information. Second, the athletes exhibit rapid and erratic posture changes—such as running and turning. The combination of unreliable appearance cues and unpredictable motion makes re-identification extremely difficult, resulting in frequent IDSW (e.g., ID:45 $\rightarrow$ ID:61, ID:186 $\rightarrow$ ID:217).

Summary. The qualitative gallery presented above provides concrete and insightful evidence for the challenges posed by DynUAV. It clearly reveals that the tracking fail-

Table S1. Detailed statistics and primary challenges for each sequence in the DynUAV validation and test sets.

ID	Subset	Frames	Total BBox	Avg BBox/Frame	Mean IoU	Primary Challenges
			Test Set			
004	Test	1,862	17,574	9.4	0.793	Smooth Motion, Occlusion
009	Test	1,950	80,850	41.5	0.667	**Severe Ego-Motion**, Small Objects
016	Test	1,674	50,357	30.1	0.781	Dense Crowd, Viewpoint Change
027	Test	2,665	15,634	5.9	0.712	High Speed, Ego-Motion
029	Test	1,651	22,578	13.7	**0.568**	**Severe Ego-Motion (Rotation)**
055	Test	1,832	43,498	23.7	0.728	Scale Variation, Occlusion
067	Test	1,808	32,458	17.9	0.859	Smooth Motion, High Density
			Validation Set			
001	Val	2,154	27,795	12.9	0.786	Viewpoint Change
006	Val	1,819	28,924	15.9	0.750	Occlusion, Ego-Motion
025	Val	2,169	16,657	7.7	0.834	Smooth Motion, Low Density
033	Val	2,165	19,176	8.9	0.693	Ego-Motion (Zoom)
040	Val	1,691	38,580	22.8	0.782	Low Light, Motion Blur

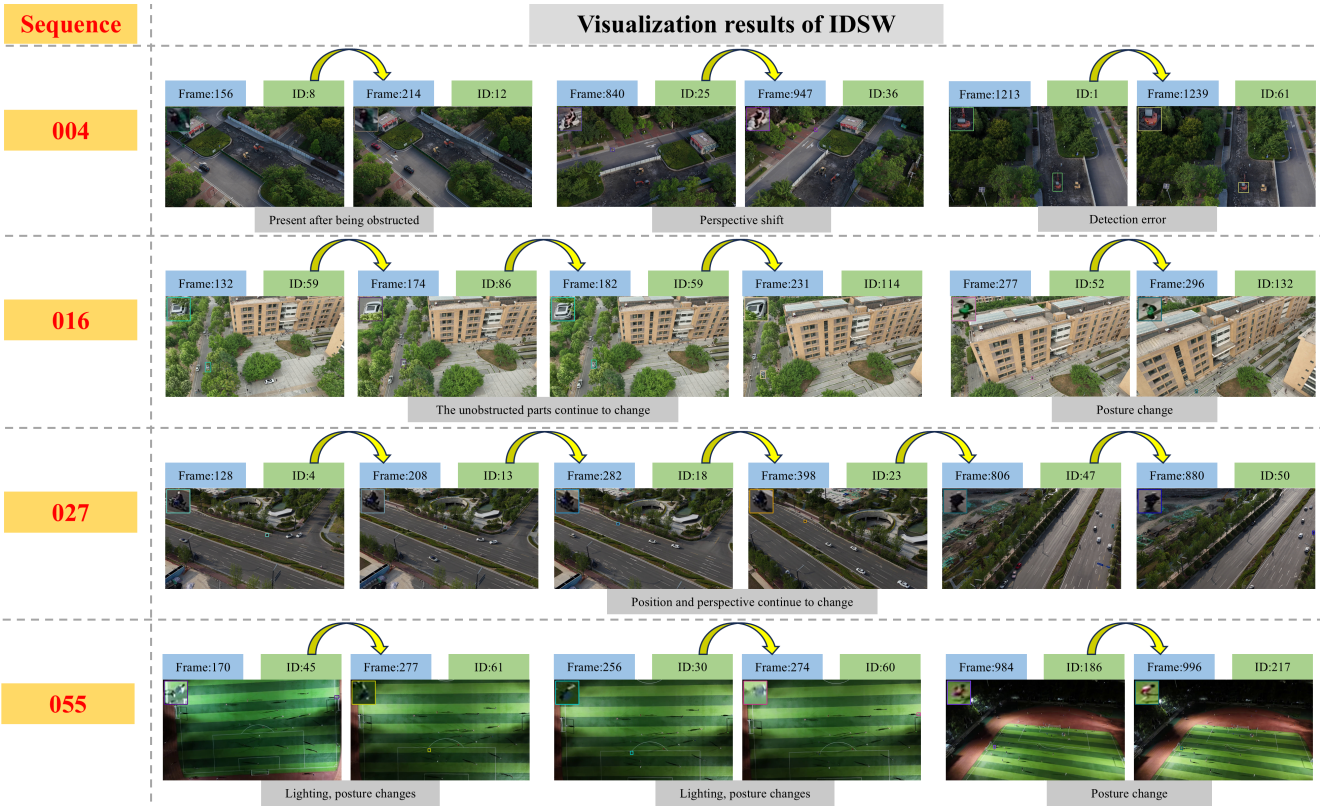

Figure S1. Visualization of IDSW in some sequences

ures observed in our benchmark are not arbitrary events, but rather are systematic consequences of the dynamic scenarios we have deliberately designed. From these typical failure modes, we can distill several core advantages and characteristics of DynUAV:

- **A stress test for motion models:** The "perspective shift" and "continuous position change" cases in Seq. 004 and 027 showcase the highly non-linear apparent trajectories generated by severe camera ego-motion. This directly breaks the linear motion assumptions (e.g., Kalman filters) that many traditional trackers rely on, proving that DynUAV serves as an ideal platform for validating and driving the

development of more advanced motion modeling, such as non-linear prediction or diffusion-based models.

• **A deep probe of appearance feature robustness:** The "posture change", "dynamic partial occlusion", and "adverse lighting" cases in Seq. 016 and 055 all point to a central difficulty: under a dynamic UAV-perspective, a target's appearance features are highly unstable and frequently corrupted by noise. This demands that Re-ID models possess a high degree of viewpoint-invariance, occlusion robustness, and illumination adaptability. DynUAV thus acts as a rigorous testbed for measuring and enhancing the generalization capabilities of existing appearance models under realistic, varied conditions.

• **Exposure of the detection-tracking cascade effect:** The "detection error" case in Seq. 004, along with the prevalent low-confidence detections in nighttime scenes, highlights how minor failures in the upstream detection module can be catastrophically amplified by the downstream tracking module in the tracking-by-detection paradigm. By introducing numerous perceptually challenging scenarios, DynUAV underscores the urgent need for either end-to-end joint optimization or tracking frameworks with stronger robustness to uncertain detection inputs.

The value of DynUAV lies not just in its scene diversity, but in its capacity to *systematically and reproducibly create the compound challenges originating from severe camera motion that are absent in current benchmarks*. It compels algorithms to move beyond the comfort zone of a "smooth world" and to confront the most intractable detection and association problems found in real-world UAV applications.

S2. Detailed Annotation Guidelines

To ensure the highest quality and consistency of our ground truth, we established a comprehensive set of annotation protocols. These guidelines were strictly enforced across all sequences and served as the standard for our three-stage annotation pipeline.

S2.1. Object Category Definitions

DynUAV encompasses eight object categories, with definitions designed for clarity and mutual exclusivity:

Person Any pedestrian that is either walking or stationary.

Car Standard passenger vehicles, including sedans, Sport-Utility Vehicles (SUVs), and hatchbacks.

Bus Public transport vehicles, distinguished from trucks by their elongated shape and capacity for passengers.

Truck Commercial vehicles primarily used for transporting cargo.

Cycle Any parked or unattended two-wheeled vehicle, such as a bicycle, motorcycle, or electric scooter.

Cycler The "Cycler" class denotes a rider on a two-wheeled vehicle, annotated with a single bounding box enclosing both.

Excavator A piece of heavy construction equipment with a digging bucket on an articulated arm.

Crane A type of heavy construction equipment used for lifting and hoisting materials.

S2.2. Occlusion and Truncation Protocols

A precise protocol was defined to handle various levels of occlusion and out-of-view truncation, ensuring temporal consistency and annotation integrity.

Scattered occlusion This pertains to cases where a target is partially obscured by non-solid objects (e.g., tree foliage). The target's full bounding box is estimated and annotated as long as its overall shape and location can be reliably inferred. Annotation is omitted if the target becomes unrecognizable.

Partial occlusion This refers to instances where a part of the target is fully occluded by a solid object. The annotation rule is contingent on the estimated occlusion percentage:

• $Occlusion < 20\%$: The full extent of the object, including both visible and occluded parts, is estimated and enclosed in a single bounding box.

• $20\% \leq Occlusion \leq 50\%$: Only the visible portion of the object is annotated.

• $Occlusion > 50\%$: The object is not annotated in the current frame and is considered temporarily lost.

Truncation (out-of-view) This applies to targets partially outside the camera's field of view. The rule is based on the percentage of the object that is truncated:

• $Truncation < 50\%$: The visible portion of the object within the frame is annotated.

• $Truncation > 50\%$: The object is not annotated and is considered to have exited the scene.

S2.3. Identity (ID) Management Protocol

Maintaining a one-to-one correspondence between a real-world object and its annotated ID is paramount. Our ID management protocol was defined as follows:

ID initialization. A new, unique track ID is assigned to an object upon its first appearance in a sequence, with the condition that it must be at least 80% visible.

ID persistence and termination. The assigned ID is propagated across subsequent frames until the object is considered terminated, either by exiting the scene or by exceeding the occlusion/truncation thresholds defined above.

ID re-association. In cases of re-entry or temporary loss, a target is provisionally assigned a new ID upon reappearance. During the manual review and refinement stages, annotators leveraged the "merge tracks" feature in CVAT to link these fragmented tracklets. This manual re-association, which capitalizes on human cognitive ability to recognize objects across significant appearance shifts, ensures that a single physical object is represented by a single, continuous ID throughout its entire trajectory in the sequence.

S2.4. CMC Parameter Generation Details

To support trackers that leverage Camera Motion Compensation (CMC) and to facilitate research into ego-motion modeling, we provide pre-computed CMC parameters for every sequence in the DynUAV benchmark. Our generation process, inspired by the methodology used in BoT-SORT[1] and implemented via OpenCV, robustly estimates the inter-frame camera motion as a 2x3 affine transformation matrix. Providing this "clean" signal of camera movement is crucial for fair algorithm comparison and for enabling trackers to isolate true object motion. To complement these details, we provide the pseudocode in Algorithm 1, which outlines the core logic.

Image pre-processing. To balance computational efficiency with accuracy, input frames are first downscaled by a factor of 2. This creates a lower-resolution image level of an image pyramid, which significantly accelerates the subsequent feature detection and matching stages while retaining sufficient structural information for reliable motion estimation across the majority of our diverse scenes.

Feature detection. We employ the *Good Features to Track (GFTT)* algorithm, based on the Shi-Tomasi criterion, to identify a dense set of up to 4,000 stable feature points per frame. This method is deliberately chosen for its high speed and its effectiveness in generating features that are spatially well-distributed, making them ideal for the gradient-based optical flow tracking implicitly used in the subsequent estimation step.

Motion estimation and robustification. The core estimation is handled by OpenCV's *videostab* module, which incorporates two critical design choices to ensure robustness in our dynamic scenes:

- *Transformation Model:* We model the camera's ego-motion using a *similarity transformation*. This is a 4-degree-of-freedom model constraining the motion to translation, rotation, and uniform scaling. This choice is optimal for UAV scenarios as it robustly represents common movements while being less susceptible to noise and degenerate geometries than more complex, higher-degree-of-freedom transformations, as its stronger constraints act as a form of regularization.
- *Robust estimation:* The process is wrapped within a RANSAC framework. This is critical for estimating the model from data containing a high percentage of outliers. In our context, outliers are feature points on independently moving objects. RANSAC effectively filters these out by finding a geometric consensus among inliers (points on the static background), ensuring the estimated motion purely reflects the camera's global motion. Without such robust filtering, the estimate would be erroneously biased by foreground object movements.

Matrix post-processing and fallback. The final stage involves two crucial adjustments to ensure accuracy and temporal smoothness. Firstly, the translation components (t_x, t_y) of the computed matrix are scaled to map the motion back to the original, full-resolution coordinate system. Secondly, a robust fallback mechanism is implemented. If a confident motion estimate cannot be found between two frames (e.g., due to severe motion blur or large textureless regions like open sky), the transformation matrix from the previously processed, successful frame pair is used instead. This practice is vital for ensuring a temporally smooth motion sequence, preventing abrupt and unrealistic high-frequency jumps in the motion parameters. Such spikes would introduce spurious high-velocity information to a tracker's internal motion model (e.g., a Kalman filter), corrupting its state estimation and potentially requiring many subsequent frames to recover.

Algorithm 1 CMC Matrix Generation for an Image Sequence

Input: Image sequence $S = \{I_0, I_1, ..., I_N\}$
Input: Downscale factor d (e.g., 2)
Output: CMC file with affine matrices $M_0, M_1, ..., M_N$
Initialize:
$Estimator \leftarrow$ InitializeRansacSimilarityEstimator()
$FeatureDetector \leftarrow$ InitializeGFTTDetector(4000)
$Estimator$.setDetector($FeatureDetector$)
$M_{prev} \leftarrow$ IdentityMatrix$(2, 3)$
$I_{prev,ds} \leftarrow$ null ▷ Previous downscaled image
for frame index $t = 0$ to N **do**
  $I_t \leftarrow$ ReadImage(S_t)
  $I_{t,ds} \leftarrow$ Resize$(I_t, 1/d)$ ▷ Current downscaled image
  $M_t \leftarrow$ IdentityMatrix$(2, 3)$
  $ok \leftarrow$ false
  if $I_{prev,ds}$ is not null **then**
    ▷ Core estimation step
    $(M_t, ok) \leftarrow Estimator$.estimate$(I_{prev,ds}, I_{t,ds})$
    if not ok **then**
      ▷ Fallback mechanism
      $M_t \leftarrow M_{prev}$
    else
      ▷ Post-process translation components
      $M_t[0, 2] \leftarrow M_t[0, 2] \times d$
      $M_t[1, 2] \leftarrow M_t[1, 2] \times d$
    end if
  end if
  WriteToFile (t, M_t)
  $I_{prev,ds} \leftarrow I_{t,ds}$
  $M_{prev} \leftarrow M_t$
end for

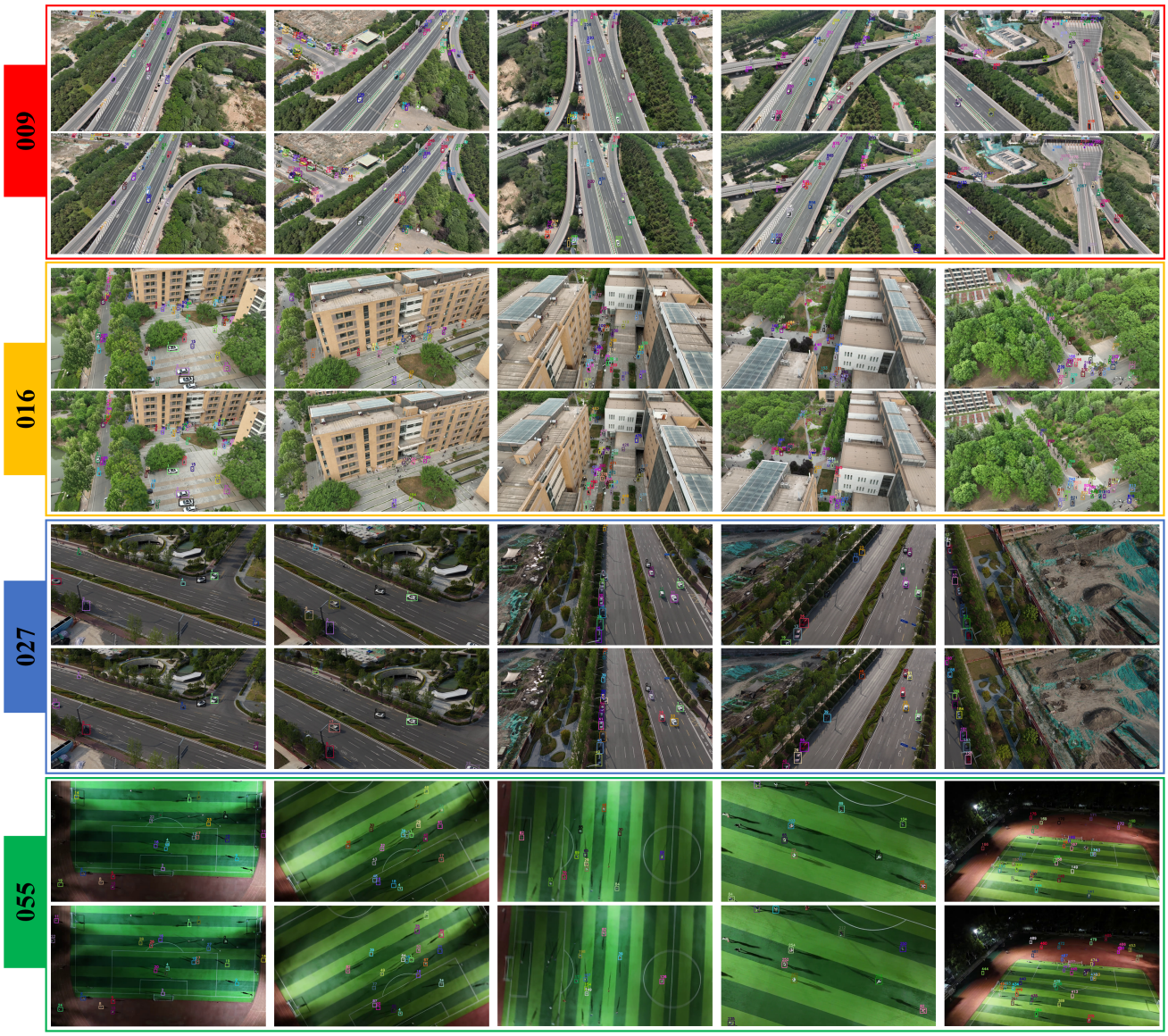

Figure S2. Qualitative comparison between ByteTrack [5] and TrackTrack [3] on challenging sequences from DynUAV.

S3. Extended Experimental Analysis

This section extends our main paper with additional experiments and visual analysis, incorporating the new evaluations provided during the rebuttal phase. We focus on cross-dataset evaluation, long-term continuity metrics, diagnostic CMC analysis, hyperparameter sensitivity, and case studies that highlight the unique challenges in DynUAV.

S3.1. Cross-Dataset Evaluation

To empirically isolate the impact of ego-motion and demonstrate that DynUAV specifically creates a far more challenging data association problem, we conducted a controlled cross-dataset evaluation. We trained a single shared detector on the VisDrone dataset and evaluated representative trackers (DeepOCSORT and ByteTrack) on both UAVDT and DynUAV (using common object classes only).

As shown in Table S2, while the detection accuracy (DetA) is nearly identical on both datasets (e.g., ∼26-27% for ByteTrack), the association accuracy (AssA) on DynUAV is drastically lower. This indicates that even with comparable perceptual input, the severe ego-motion in DynUAV significantly complicates the association stage, validating its unique value as an association stress-test.

S3.2. Long-Term Continuity Metrics (TSR)

To explicitly quantify long-term trajectory continuity, which targets identity persistence across significant tempo-

Table S2. Cross-dataset evaluation results using a VisDrone-trained detector. Note the similar DetA but significantly lower AssA on DynUAV.

Test Set & Tracker	HOTA↑	DetA↑	AssA↑
UAVDT - DeepOCSORT	37.11	26.02	53.82
DynUAV - DeepOCSORT	33.67	26.41	44.23
UAVDT - ByteTrack	40.04	26.99	60.28
DynUAV - ByteTrack	33.14	27.40	41.46

ral gaps (e.g., re-entry after hundreds of frames) rather than merely continuous visibility, we introduce the **Trajectory Survival Rate (TSR)**. TSR is defined as the ratio of the longest uninterrupted tracked segment to the total ground-truth duration of a target. We intentionally induce trajectory breaks via agile maneuvers in DynUAV to stress-test these recovery capabilities.

As shown in Table S3, DiffMOT achieves a high TSR (72.88% at IoU=0.5) on MOT20, where static cameras allow for continuous tracking maintenance. In contrast, the much lower TSR (52.00%) on DynUAV directly evidences that trackers fail to bridge motion-induced gaps. Furthermore, the high sensitivity of TSR to IoU thresholds on DynUAV confirms that severe jitter frequently breaks trajectories, highlighting the difficulty of maintaining identities over a target's entire lifespan under dynamic conditions.

Table S3. Comparison of Trajectory Survival Rate (TSR) across different IoU matching thresholds.

Dataset	IoU=0.4	IoU=0.5	IoU=0.6
MOT20	74.23	72.88	69.55
DynUAV (Ours)	61.52	52.00	39.42

S3.3. CMC as a Diagnostic Probe

We employ Camera Motion Compensation (CMC) as a diagnostic probe to isolate and quantify the two main challenges in DynUAV: severe ego-motion and prolonged trajectory recovery. To isolate each factor, we analyze representative sequences as reported in Table S4.

In Seq 029, enabling CMC yields significant gains (e.g., AssA +5.17, IDSW -48.72%), confirming that **large ego-motion** is the primary cause of failure, as it invalidates standard linear motion models. Conversely, the marginal gain observed in Seq 027 identifies **long-term appearance association** (across occlusions) as the bottleneck, which falls outside the scope of geometric compensation. CMC thus diagnostically reveals that DynUAV's challenges extend beyond motion compensation to demand robust long-term re-identification.

S3.4. Diagnosing ID Inflation

To provide a more granular, qualitative understanding of how different tracking paradigms cope with the challenges

Table S4. Diagnostic analysis of CMC gains across distinct challenges.

Seq (Challenge)	HOTA↑	AssA↑	IDSW↓
029 (Large Motion)	**+4.63**	**+5.17**	**-48.72%**
027 (Long-term)	+0.36	+0.15	-22.2%

of DynUAV, we present a direct visual comparison between two representative algorithms: ByteTrack [5], a highly efficient tracker relying on a simple IoU-based and detection-score-based association, and TrackTrack [3], a state-of-the-art single-stage method with a more robust, unified matching process. Figure S2 visualizes their tracking outputs on several challenging sequences. *ID number inflation* is a key phenomenon observed in ByteTrack's results over time. As the sequences progress, the tracker IDs assigned by ByteTrack [5] become progressively larger than those assigned by TrackTrack [3]. This is a direct visual proxy for the cumulative effect of IDSW, where each new switch forces the tracker to assign a new, higher ID number to a previously seen object.

The drastic inflation of ID numbers in ByteTrack [5] reflects a systemic failure mode induced by core DynUAV challenges, primarily due to three causes:

Brittleness to severe ego-motion (Seq. 009 & 027): The road scenes in Seq. 009 and 027 are characterized by continuous, sweeping UAV maneuvers. This induces highly non-linear apparent trajectories for the vehicles. ByteTrack's association logic, which heavily relies on a linear motion model (Kalman filter) for its primary matching stage, is fundamentally ill-equipped for this challenge. In the visualization, we can see that at nearly every turn or change in UAV speed, ByteTrack's predictions fail, leading to low IoU scores between predicted and detected boxes. This causes a cascade of micro-failures: a correct association is missed, the track is temporarily lost, and upon re-detection, it is incorrectly assigned a new ID. Each of these failures adds to the ever-growing ID count. In contrast, TrackTrack's unified matching process is less reliant on rigid motion priors, allowing it to maintain identity continuity through these transient disturbances, resulting in stable, low-numbered IDs. This aligns with our quantitative results, where TrackTrack [3] achieves a far lower IDSW count (256) compared to ByteTrack [5] (3136).

Failure to re-identify across appearance shifts (Seq. 016): The campus scene in Seq. 016 presents a different challenge: drastic appearance changes. As the UAV orbits the buildings, pedestrians are seen from continuously shifting viewpoints (e.g., from side-profile to top-down). ByteTrack [5], lacking a dedicated, powerful Re-Identification (Re-ID) module, struggles to recognize that these are the same individuals. Its association relies almost entirely on proximity and detection score, cues that become unreliable when appearance changes. The visual-

Table S5. Tracker performance sensitivity to IoU threshold on DynUAV. (Best highlighted; **0.4** is main paper value)

Tracker	$IOU_{threshold}$	HOTA↑	MOTA↑	IDSW↓	DetA↑	AssA↑
Deep OC-SORT[2]	0.25	36.646	13.285	1672	32.229	41.716
	0.4	**61.090**	**66.436**	**567**	**57.579**	**65.493**
	0.5	60.675	66.435	750	57.616	64.602
	0.6	59.802	66.086	963	57.377	63.062
BoostTrack[4]	0.25	53.558	56.558	630	49.387	58.722
	0.4	**53.715**	**56.724**	**609**	**49.519**	**58.977**
	0.5	53.325	56.487	716	49.305	58.322
	0.6	53.130	56.343	766	49.212	58.016
BoT-SORT[1]	0.25	58.978	66.324	1498	57.462	61.172
	0.4	**59.855**	**66.880**	**1274**	**57.836**	**62.648**
	0.5	59.209	66.460	1442	57.560	61.540
	0.6	59.511	66.711	1369	57.715	61.997

ization clearly shows ByteTrack [5] fragmenting the trajectories of pedestrians as they walk through the plaza, assigning new IDs after even minor changes in posture or viewing angle. TrackTrack [3], which incorporates appearance features into its matching cost, demonstrates superior robustness in re-linking these tracklets, thus preventing ID inflation. This directly explains why TrackTrack's AssA (68.89) is significantly higher than ByteTrack's (53.08).

Compounded challenges in adverse conditions (Seq. 055): The nighttime scene in Seq. 055 represents a "worst-case scenario" where multiple challenges are compounded. The low-light conditions lead to low-confidence detections, and the artificial lighting creates long, dynamic shadows and corrupts appearance features. The athletes themselves exhibit erratic, non-linear motion. In this environment, ByteTrack's two-stage matching breaks down completely. Low-confidence detections are often discarded, leading to frequent FNs. When targets are detected, their appearance is too noisy for simple IoU-based matching to work reliably. The result, as seen in the visualization, is a chaotic tracking output with rapidly escalating ID numbers. TrackTrack [3], while also challenged, performs significantly better by holistically leveraging all available cues, demonstrating that a robust algorithm must be able to function even when individual cues (motion, appearance, detection score) are unreliable.

This visual analysis of ID number inflation offers an intuitive and powerful illustration of *error accumulation*. The ever-increasing ID numbers in ByteTrack [5] serve as a clear "fever chart", indicating a tracker that is constantly failing and restarting under the sustained pressure of the DynUAV benchmark. The stable, low IDs of TrackTrack [3], in contrast, signify a healthy and robust tracking process capable of long-term identity preservation.

Summary. This comparison empirically validates two central claims of our work: (1) The severe ego-motion and complex scenes in DynUAV systematically violate the core assumptions of simpler, widely-used trackers, leading to high rates of association failure. (2) The benchmark's long duration amplifies these failures over time, making long-term robustness and resistance to error accumulation, not just frame-to-frame accuracy, the critical capabilities for success. DynUAV thus serves as a rigorous diagnostic tool for evaluating the temporal stability of MOT algorithms.

S3.5. The Impact of IoU Threshold

We analyze tracker sensitivity to the IoU threshold parameter under DynUAV's ego-motion to assess their reliance on geometric matching. Three representative methods are tested: Deep OC-SORT [2] (motion + Re-ID fusion), BoT-SORT [1] (alternative motion-appearance fusion), and BoostTrack [4](detection-driven). IoU threshold was varied from 0.25 to 0.6 (see Table S5), revealing key insights into both benchmark characteristics and tracker behavior.

Universal preference for lower IoU threshold. All three trackers achieve peak performance at a lower IoU threshold (0.4) than conventional defaults. Raising it to 0.6 consistently degrades all metrics, particularly increasing IDSW while reducing HOTA and AssA. This trend confirms that *DynUAV's severe ego-motion induces localization jitter, where strict geometric matching penalizes correct associations.* The optimal 0.4 threshold demonstrates that tolerating geometric uncertainty is essential for robust tracking in our benchmark.

Sensitivity analysis and Re-ID's role. Trackers show varying sensitivity to the IoU threshold, revealing differences in their association mechanisms.

• *High sensitivity (Deep OC-SORT[2])*: Deep OC-SORT[2] shows the most volatile performance, collapsing at extremely low IoU thresholds (0.25) with massive IDSW and degrading significantly at high thresholds (0.6). This

confirms its reliance on synergistic motion-appearance fusion: when the initial geometric matching is too loose or strict, its association logic becomes compromised.

• *Moderate sensitivity (BoT-SORT[1])*: BoT-SORT[1] also peaks at the 0.4 IoU threshold, with milder degradation at higher values. This implies greater robustness in its fusion mechanism to initial matching variations, though its consistently high IDSW across all thresholds reveals underlying identity consistency issues on DynUAV.

• *Low sensitivity (BoostTrack[4]):* BoostTrack[4] shows notable insensitivity to the IoU threshold, with stable metrics across all tested values. This reveals a fundamental limitation: as a simpler method lacking strong Re-ID, its performance is already bottlenecked by severe appearance changes and long-term occlusions in our data, making threshold tuning inconsequential.

Summary. This analysis confirms the distinct challenges of DynUAV from three aspects. The consistent preference for a lower IoU threshold reflects widespread motion-induced localization uncertainty. The varying sensitivity among trackers highlights the need for reliable fallback mechanisms like Re-ID. Furthermore, BoostTrack's insensitivity indicates a fundamental mismatch between its design premises and the benchmark's dynamic conditions.