UDC 004.932:004.896:623.746-519 **DOI:** 10.31673/2412-9070.2026.054913

T. SEMENCHENKO, PhD Student,
ORCID: 0009-0007-3259-7007
Taras Shevchenko National University of Kyiv

BENCHMARKING MOTION-BASED TRACKING ALGORITHMS FOR SMALL UAV OBJECT TRACKING

The growing use of small unmanned aerial vehicles (UAVs) creates serious security challenges for airports, critical infrastructure, and military sites, driving demand for automated counter-drone systems that can detect and track aerial targets in real time. Tracking small UAVs remains hard because of low-resolution appearance, irregular flight dynamics, frequent occlusions, and the high sensitivity of the IoU metric to pixel-level localization errors, which together make reliable temporal association difficult. Most multi-object tracking (MOT) benchmarks focus on pedestrians rather than small aerial targets, leaving practitioners without clear guidance on tracker choice for the UAV domain. This paper addresses that gap with an empirical evaluation of five SORT-family multi-object tracking algorithms, namely ByteTrack, BoT-SORT, DeepOC-SORT, Hybrid-SORT, and StrongSORT, on the Anti-UAV-RGBT dataset. All trackers share an RT-DETR-L detector fine-tuned on the Anti-UAV-RGBT training split, and appearance re-identification is disabled so that the comparison reflects only motion-based association strategies. The test set contains 92 sequences with 79,699 annotations, grouped into three size categories (small ≤48 px, medium 48–128 px, large >128 px) to enable a size-stratified breakdown of tracking quality. Experiments run on a single NVIDIA RTX 5090 GPU. BoT-SORT achieves the highest overall MOTA of 0.7802 and also leads on small objects with a MOTA of 0.7479; its Global Motion Compensation module cancels camera-induced displacement before Kalman prediction and preserves IoU overlap when targets span only a few dozen pixels. ByteTrack shows the largest drop on small objects, losing 10.8 percentage points relative to its overall score. DeepOC-SORT and Hybrid-SORT run 38% faster (over 33 FPS) while giving up less than 1.3 percentage points of overall tracking accuracy. Based on these findings, the paper provides practical deployment guidelines that help practitioners pick a tracker according to target scale, accuracy expectations, and available compute budget.

Keywords: UAV tracking, multi-object tracking, SORT, small object tracking, RT-DETR, benchmark, Anti-UAV.

Introduction

The increasing availability of small commercial and consumer unmanned aerial vehicles (UAVs) has raised security concerns across many domains. Unauthorized drone flights near airports, critical infrastructure, military sites, and public events have created strong demand for automated counter-drone systems that can detect and track aerial targets in real time [12], [14], [7]. A key challenge is that UAVs, especially at long range or high altitude, appear as very small objects in camera imagery, often occupying fewer than 48 pixels along their largest dimension. At such scales, targets show minimal visual detail, are prone to motion blur, and produce only small overlap between consecutive frames, making reliable temporal association difficult.

Multi-object tracking (MOT) has advanced significantly in recent years, driven largely by the SORT family of algorithms. The original SORT algorithm [3] introduced a framework based on the Kalman filter for motion prediction and the Hungarian algorithm for detection-to-track association. Later methods extended this framework in different ways: ByteTrack [23] introduced a two-stage association strategy that uses low-confidence detections; BoT-SORT [1] added global motion compensation through sparse optical flow; OC-SORT [5] and its extension DeepOC-SORT [13] proposed observation-centric momentum to improve association under non-linear motion; Hybrid-SORT [20]

© Semenchenko T., 2026

combined height-modulated IoU with tracklet confidence scoring; and StrongSORT [9] introduced noise-scale-adaptive Kalman filtering with motion-appearance fusion. While each of these methods has shown strong performance on pedestrian-focused benchmarks such as MOT17 and MOT20, their effectiveness on small aerial targets has not been systematically tested under controlled conditions.

Statement of the problem

Within multi-sensor counter-drone pipelines that combine visual cameras with radar and acoustic or radio-frequency direction-of-arrival (DOA) estimation [4], the vision subsystem must not only detect UAV targets in each frame but also maintain consistent temporal tracking across frames. Downstream mission logic depends directly on this tracking output: prior work has shown how tra-cked detections from a single optical sensor can be transformed into geospatial coordinates for 3D trajectory prediction [8], and learning-based swarm controllers in cooperative multi-UAV systems rely on continuous visual lock to sustain target-following modes [16]. Without a working tracking component, a vision-based counter-drone pipeline can detect but not follow a target, which limits its practical use.

The tracking problem becomes harder when UAV targets appear small in the camera field of view. At surveillance-relevant distances of hundreds of meters, a UAV may occupy fewer than 48 pixels along its largest dimension. At such scales, the IoU metric that underpins most modern tracking association algorithms becomes highly sensitive to localization errors: a displacement of just a few pixels on a 20×20 target can drop IoU below the matching threshold, causing lost tracks or spurious identity switches. Small targets also provide minimal visual texture, which limits appearance-based re-identification and shifts the burden of temporal association almost entirely onto the motion model. UAV flight dynamics add a further complication: sudden altitude changes, hovering, and lateral drifts produce non-linear motion that challenges standard linear Kalman filter predictions.

Despite rapid progress in multi-object tracking, existing benchmarks have not systematically addressed these challenges in the UAV domain. Most tracker evaluations run on pedestrian datasets such as MOT17 and MOT20, where targets are relatively large and move predictably. UAV-focused studies, in turn, have concentrated on detection architectures or single-object trackers rather than isolating the association component of a multi-object pipeline. Aggregate metrics are typically reported without examining how performance varies with object size — a critical dimension for UAV applications, where scale changes significantly with distance. This leaves practitioners without clear guidance on which tracker suits small, medium, or large targets, and how to balance accuracy against processing speed in real-time deployment.

Analysis of recent studies and publications

Interest in counter-drone technologies has led to the development of several dedicated datasets and benchmarks. The Anti-UAV challenge [11] introduced a single-object tracking benchmark with diverse UAV targets under varying lighting and weather conditions. Its extension, Anti-UAV-RGBT [11], provides paired visible and infrared sequences for multi-modal evaluation across 100+ annotated test sequences. The VisDrone dataset [25] covers a broader range of aerial objects captured from drone-mounted cameras, including pedestrians, vehicles, and other UAVs, and has been widely used for detection and tracking evaluations. More recently, several works have applied transformer-based detection architectures to UAV applications. RT-DETR [24], building on the original DETR end-to-end detection framework [6], combines the accuracy of detection transformers with real-time inference, making it a good fit as a detection backbone for tracking pipelines. Recent work has benchmarked RT-DETR and YOLO-family detectors for UAV detection on edge devices [22] and in a follow-up study applied an RT-DETR + ByteTrack pipeline with motion-consistent temporal fusion that improved MOTA and reduced identity switches [21]. The CST Anti-UAV benchmark [19] further extends the Anti-UAV dataset family with thermal infrared sequences focused on tiny UAV tracking in complex scenes, confirming continued interest in this evaluation domain. Despite this progress, most UAV tracking studies focus on detector performance or use single-object tracking setups, and the multi-object tracking association component remains largely unstudied in the UAV domain.

Small Object Tracking

Small object tracking presents challenges that set it apart from general MOT scenarios. When targets occupy only a small number of pixels, the IoU metric becomes very sensitive to even minor localization errors. This sensitivity directly affects IoU-based association, increasing the chance of missed matches and identity switches. In addition, small objects provide limited appearance information, making re-identification features less effective and increasing the reliance on motion models for temporal association, which mirrors the broader challenges identified in CNN-based video analysis for moving objects under complex backgrounds and varying scales [15]. Several studies have looked at small object detection through techniques such as image tiling, super-resolution, and multi-scale feature fusion [10], and Wang et al. [17] recently proposed a detector-tracker combination specifically designed for small UAV objects. Still, little attention has been given to how existing tracking association algorithms handle small target sizes in a controlled comparison. Existing MOT benchmarks such as MOT17 and MOT20 contain mostly medium-to-large pedestrian targets and do not include size-stratified evaluation protocols. This gap motivates the size-aware analysis in this work, which examines tracker performance across small, medium, and large UAV targets.

Research objectives and tasks

The objective of this study is to provide a controlled comparative evaluation of five state-of-the-art SORT-family multi-object trackers for UAV tracking on the Anti-UAV-RGBT dataset, with particular focus on small object performance and the speed-accuracy trade-off.

This objective is broken down into the following tasks:

1. Establish a fair and reproducible benchmark: where all five trackers operate on identical detections from a shared RT-DETR-L detector, with re-identification features disabled to isolate differrences in motion-based association strategies.

2. Conduct a size-stratified performance analysis: that groups ground truth annotations into small (max side ≤ 48 px), medium (max side ≤ 128 px), and large (max side > 128 px) categories, revealing how each tracker responds to varying object scales.

3. Evaluate the speed-accuracy trade-off: by measuring processing speed alongside tracking accuracy for each algorithm.

4. Formulate practical deployment guidelines: that connect tracker choice to specific application needs, giving concrete recommendations for counter-drone system design.

Research results

This section first describes the experimental setup (dataset, detector, tracker configurations, evaluation metrics, and hardware) and then reports the tracking results: overall performance, size-stratified performance, error decomposition, and speed-accuracy trade-offs. As shown in fig. 1, all trackers share a common detection stage based on RT-DETR-L, so performance differences reflect only the tracking association component. The pipeline processes input video sequences from the Anti-UAV-RGBT dataset through the detector, which produces per-frame bounding boxes with confidence scores. These detections are passed to each tracker, which applies its prediction and association strategy to output tracked bounding boxes with persistent identity labels.

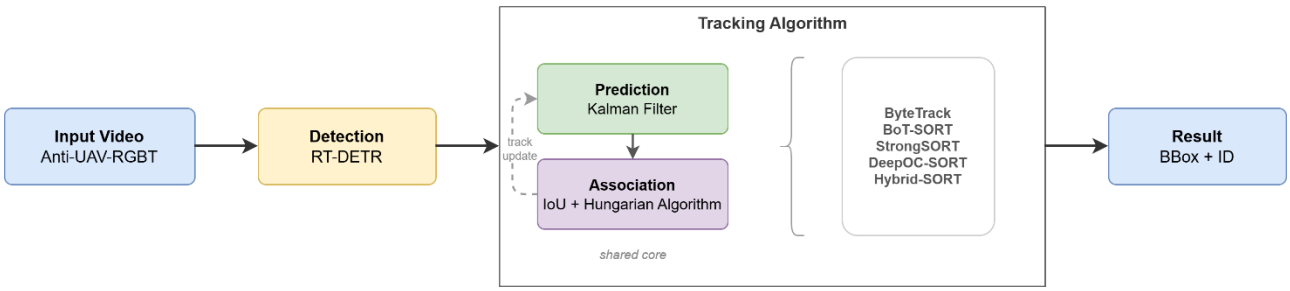


Fig. 1. Experimental pipeline with a shared RT-DETR detector and five tracking association strategies

Dataset

The experiments are conducted on the Anti-UAV-RGBT dataset [11], a benchmark designed for UAV detection and tracking in real-world surveillance scenarios. The dataset provides paired visible and infrared video sequences capturing UAV targets under diverse conditions including varying distances, altitudes, lighting, and backgrounds. In this study, only the visible modality is used. The dataset is split into training and test partitions; the training split is used exclusively for detector fine-tuning, while all tracking evaluation is performed on the test split comprising 92 sequences with a total of 79,699 ground truth annotations.

To enable scale-aware evaluation, each ground truth bounding box is categorized by size based on the maximum of its width and height into one of three bins:

– **Small:** max(width, height) ≤ 48 pixels – 5,598 annotations (7.0%)
– **Medium:** 48 < max(width, height) ≤ 128 pixels – 40,127 annotations (50.3%)
– **Large:** max(width, height) > 128 pixels – 33,974 annotations (42.7%)

This distribution shows that medium-sized objects make up the majority of the dataset, while small objects (the most challenging category) represent only 7% of annotations. This imbalance reflects realistic surveillance conditions, where UAVs appear small only at greater distances and occupy more pixels as they approach the camera. Fig. 2 shows example frames from the dataset with ground truth bounding boxes for each size category, illustrating the visual difference between small, medium, and large UAV targets.



Fig. 2. Example frames from Anti-UAV-RGBT showing small, medium, and large UAV targets with ground truth boxes

Detector

All tracking algorithms share a single detection backbone to ensure fair comparison. The detector is RT-DETR-L (Real-Time Detection Transformer, Large variant) [24], a hybrid architecture that combines the representational strength of vision transformers with efficient real-time inference. The model was fine-tuned on the Anti-UAV-RGBT training split using the Ultralytics framework with the following configuration: input resolution of 640x640 pixels, batch size of 16, 10 training epochs starting from pretrained weights, and standard data augmentation including mosaic, random augmentation, and horizontal flipping. During inference, a detection confidence threshold of 0.25 is applied to balance precision and recall.

By using the same detector and the same detection outputs for all five trackers, the experimental design isolates the contribution of the tracking association algorithm. Any differences in tracking performance can therefore be attributed to the tracker's motion model, association strategy, and post-processing rather than to differences in detection quality.

Tracker Configurations

Five SORT-family trackers are evaluated, each taking a different approach to detection-to-track association. All trackers share common hyperparameters to ensure fair comparison: high detection

threshold of 0.5, low detection threshold of 0.1, new track initialization threshold of 0.6, track buffer of 30 frames, and IoU matching threshold of 0.8. Re-identification (ReID) features are disabled across all trackers, restricting them to motion-based association only. This isolates the performance of each tracker's motion model and association strategy, independent of appearance-based cues. Table 1 summarizes the key features of each tracker.

Table 1

Key features of each tracker

Tracker	Two-Stage Assoc.	GMC	OCM	HM-IoU	TCS	NSA Kalman	MC-Lambda
ByteTrack	+	–	–	–	–	–	–
BoT-SORT	+	+	–	–	–	–	–
DeepOC-SORT	+	–	+	–	–	–	–
Hybrid-SORT	+	–	–	+	+	+	–
StrongSORT	+	+	–	–	–	+	(–)

ByteTrack is the baseline. It matches tracks in two passes: first to high-confidence detections (score ≥ 0.5) via IoU, then leftover tracks are matched against low-confidence ones (score ≥ 0.1). This rescues occluded or distant targets that a single-threshold approach would throw away.

BoT-SORT builds on ByteTrack by adding Global Motion Compensation. It estimates camera motion between frames with sparse optical flow and warps track states accordingly before the Kalman prediction step. Helps a lot when the camera itself is moving or shaking.

DeepOC-SORT adds Observation-Centric Momentum (OCM), which rewards associations that line up with a track's recent direction of travel. The score compares the track's velocity to the vector from its last observation to a candidate detection. OCM weight is 0.3, mixed with standard IoU cost. Useful under non-linear or abrupt motion.

Hybrid-SORT brings three things together. Height-Modulated IoU (HM-IoU) penalizes matches between objects of very different heights. A Tracklet Confidence State (TCS) mechanism tags each track as confident or weak based on its last 10 detection scores (threshold 0.4), and applies stricter matching to weak tracks. Finally, a confidence-scaled Kalman filter shrinks measurement noise as detection confidence grows, so strong detections get more trust.

StrongSORT uses a Noise-Scale-Adaptive (NSA) Kalman filter with the same confidence-to-noise logic as Hybrid-SORT, plus GMC via sparse optical flow. Its full version fuses motion and appearance costs with MC-lambda ($\lambda = 0.985$) and an EMA on appearance features ($\alpha = 0.9$). We disable appearance here for a fair comparison, so MC-lambda collapses to pure motion cost and StrongSORT's active contributions reduce to NSA Kalman plus GMC.

The shared hyperparameters (high threshold of 0.5, low threshold of 0.1, track buffer of 30, IoU threshold of 0.8) were not tuned separately for each tracker. This keeps the baseline fair, but some trackers might perform better with algorithm-specific tuning. This is a common trade-off in benchmarking studies and is discussed further in the Conclusions section.

Evaluation Metrics

Tracking performance is measured using standard Multiple Object Tracking (MOT) metrics from the CLEAR MOT protocol [2].

The main metric is Multiple Object Tracking Accuracy (MOTA), which combines three error types into a single score:

$$MOTA = 1 - \frac{FN + FP + ID_{sw}}{N_{GT}}, \quad (1)$$

where FN is the number of false negatives (missed detections), FP is the number of false positives, ID_{sw} is the number of identity switches (cases where a ground truth object gets assigned a different track identity between frames), and N_{GT} is the total number of ground truth annotations.

Localization quality is measured by Multiple Object Tracking Precision (MOTP), which is the average IoU between matched prediction-ground truth pairs:

$$MOTP = \frac{\sum_i IoU_i}{N_{matched}}, \quad (2)$$

where the sum covers all matched pairs across all frames and $N_{matched}$ is the total number of matches.

Predictions are matched to ground truth boxes using Intersection over Union (IoU):

$$IoU(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (3)$$

A prediction counts as a true positive if its IoU with the closest unmatched ground truth box is at least 0.5. Matching is done greedily in order of decreasing IoU.

We also report Precision (TP / (TP + FP)), Recall (TP / N_{GT}), and F1-score (the harmonic mean of Precision and Recall). Processing speed is measured in frames per second (FPS), calculated as the total number of processed frames divided by the total elapsed time, including both detection and tracking.

Hardware and Implementation

All experiments run on a single NVIDIA RTX 5090 GPU. The tracking pipeline is built with the Ultralytics framework in Python, using PyTorch as the deep learning backend. ByteTrack and BoT-SORT use the built-in Ultralytics implementations. DeepOC-SORT, Hybrid-SORT, and StrongSORT are ported from their official repositories into the Ultralytics tracker registry to keep the inference pipeline consistent. The ported versions keep the original algorithm logic but adapt the input/output interface to match the Ultralytics tracking API. Each tracker processes all 92 test sequences one by one, and metrics are collected both globally and per size category.

Overall Tracking Performance

Table 2 summarizes the overall tracking performance of all five trackers across the full test set of 92 sequences. Table 3 provides a detailed breakdown by object size category.

Table 2

Overall tracking results on Anti-UAV-RGBT

Tracker	MOTA	MOTP	Precision	Recall	F1	IDsw	FPS
BoT-SORT	**0.7802**	0.7833	0.9579	0.8404	0.8953	1854	24.39
StrongSORT	**0.7731**	0.7769	0.9568	0.8362	0.8925	2020	24.44
Hybrid-SORT	**0.7698**	0.7776	0.9588	0.8310	0.8903	2027	33.24
DeepOC-SORT	**0.7680**	0.7849	0.9587	0.8299	0.8896	2079	33.73
ByteTrack	**0.7665**	0.7765	0.9513	0.8324	0.8879	1864	23.13

BoT-SORT achieves the highest overall MOTA of 0.7802, followed closely by StrongSORT (0.7731) and Hybrid-SORT (0.7698). The gap between the best and worst tracker is only 1.37 percentage points. Given the high per-sequence variance shown in fig. 3, this gap may not be statistically significant. Still, the consistent ranking across size categories suggests that the ordering reflects real algorithmic differences rather than noise. All trackers show precision above 95%, which confirms that the shared RT-DETR-L detector produces few false positives. The main difference between trackers is recall, which ranges from 0.8299 (DeepOC-SORT) to 0.8404 (BoT-SORT). This means missed detections, not false alarms, are the main source of error. For identity stability, BoT-SORT has

the fewest identity switches (1,854), closely followed by ByteTrack (1,864), while DeepOC-SORT has the most (2,079).

To assess tracker consistency across diverse sequences, fig. 3 presents the distribution of per-sequence MOTA scores.

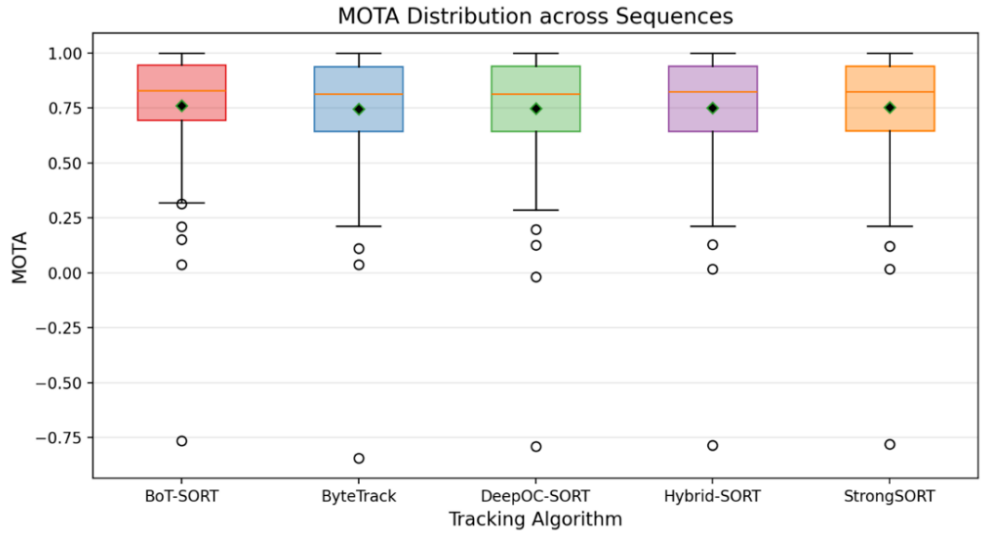


Fig. 3. Distribution of per-sequence MOTA scores across test sequences for each tracker

All five trackers have similar median MOTA values in the 0.75-0.80 range, with comparable interquartile ranges. Some outlier sequences produce MOTA values below zero, meaning the number of errors exceeds the number of ground truth annotations. These hard sequences usually involve long occlusion, extreme distance, or fast maneuvers. No single tracker handles these difficult cases much better than the others. All five algorithms struggle on the same sequences, which suggests that the failures come mainly from detection limitations rather than weaknesses in the tracking algorithms themselves.

Size-Stratified Analysis

The size-stratified breakdown in table 3 shows large performance differences across object scales. This analysis is the central contribution of the study.

Table 3

Tracking performance by object size category

Size Category	Tracker	MOTA	MOTP	Precision	Recall	F1	IDsw
Small (≤48 px)	BoT-SORT	**0.7479**	0.7267	0.9179	0.8705	0.8936	250
	DeepOC-SORT	**0.7345**	0.7271	0.9151	0.8612	0.8874	262
	StrongSORT	**0.7176**	0.7203	0.9098	0.8503	0.8790	271
	Hybrid-SORT	**0.7035**	0.7202	0.9100	0.8326	0.8696	262
	ByteTrack	**0.6584**	0.6997	0.8712	0.8242	0.8471	246
Medium (48–128 px)	BoT-SORT	**0.7156**	0.7776	0.9559	0.7863	0.8629	1382
	StrongSORT	**0.7137**	0.7735	0.9561	0.7854	0.8624	1432
	Hybrid-SORT	**0.7114**	0.7743	0.9589	0.7802	0.8604	1419
	DeepOC-SORT	**0.7044**	0.7785	0.9579	0.7747	0.8566	1453
	ByteTrack	**0.7038**	0.7688	0.9502	0.779	0.8561	1380
Large (>128 px)	BoT-SORT	**0.8640**	0.7982	0.9868	0.8993	0.9410	790
	ByteTrack	**0.8600**	0.7960	0.9853	0.8969	0.9390	798
	StrongSORT	**0.8543**	0.7894	0.9854	0.8938	0.9374	892
	Hybrid-SORT	**0.8513**	0.7900	0.9859	0.8907	0.9359	907
	DeepOC-SORT	**0.8501**	0.8007	0.9862	0.8899	0.9356	927

The size-stratified results lead to several observations:

Small objects are the hardest category for all trackers. Every algorithm shows lower MOTA on small targets compared to its overall score. The biggest drop happens with ByteTrack, which falls from 0.7665 (overall) to 0.6584 (small), a decline of 10.8 percentage points. On the other hand, BoT-SORT handles small objects best, reaching a small-object MOTA of 0.7479, the highest among all trackers and only 3.2 percentage points below its overall score. This comes from BoT-SORT's Global Motion Compensation module: when targets are only a few dozen pixels wide, even small camera motion can shift the predicted bounding box enough to break IoU-based association. By compensating for camera motion before Kalman prediction, BoT-SORT reduces the effective displacement between frames, keeping enough IoU overlap for correct matching.

Large objects are tracked well by all algorithms. All five trackers reach MOTA above 0.85 on large targets, with precision over 98.5%. The gap between the best (BoT-SORT, 0.864) and worst (DeepOC-SORT, 0.8501) tracker is only 1.4 percentage points. This means that when objects are large enough, the choice of tracking algorithm makes little practical difference.

Medium objects produce lower MOTA than small ones for several trackers. Despite accounting for 50.3% of all annotations, medium targets reach MOTA of only 0.70-0.72, below the small-object MOTA for BoT-SORT (0.7156 vs. 0.7479) and StrongSORT (0.7137 vs. 0.7176). This is partly explained by lower recall on medium objects (0.77-0.79 vs. 0.83-0.87 for small). The small-object subset is also much smaller (7% of annotations), so its MOTA may be influenced by a few favorable sequences. Additionally, medium-range sequences appear to contain harder conditions such as cluttered backgrounds and partial occlusions, and medium objects produce the most identity switches, further lowering MOTA. The tracker ranking stays consistent, with BoT-SORT (0.7156) and StrongSORT (0.7137) leading.

Speed-Accuracy Trade-off

Fig. 6 plots small-object MOTA against processing speed for each tracker, showing the trade-off between tracking accuracy and speed.

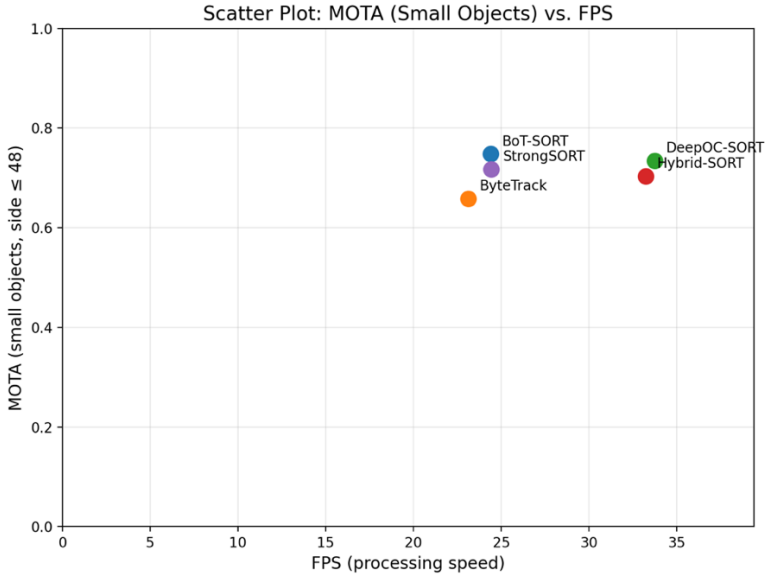


Fig. 4. Scatter plot of small-object MOTA versus processing speed (FPS) for each tracker

The trackers fall into two speed groups. The first group, BoT-SORT (24.39 FPS), StrongSORT (24.44 FPS), and ByteTrack (23.13 FPS), runs at around 24 FPS. The second group, DeepOC-SORT (33.73 FPS) and Hybrid-SORT (33.24 FPS), runs at around 33 FPS, which is 38% faster. The speed difference lines up with Global Motion Compensation: BoT-SORT and StrongSORT both compute sparse optical flow for GMC, which slows them down. DeepOC-SORT and Hybrid-SORT use lightweight association methods (OCM and HM-IoU) that add very little overhead. ByteTrack, though, has no computationally expensive module but still runs the slowest of all five trackers (23.13 FPS). This is likely an implementation artifact rather than an algorithmic property, since ByteTrack and

BoT-SORT use the built-in Ultralytics code path, which may add overhead that the ported implementations do not have. The speed differences should be read with this in mind.

The faster group's speed advantage comes at only a small accuracy cost. DeepOC-SORT reaches a small-object MOTA of 0.7345, just 1.34 percentage points below BoT-SORT's 0.7479. Hybrid-SORT, at 0.7035 on small objects, gives up a larger 4.4 percentage points of small-object MOTA for its speed gain, but stays competitive on overall MOTA (0.7698 vs. 0.7802). ByteTrack sits in the least favorable position: it has neither the accuracy of BoT-SORT nor the speed of DeepOC-SORT, and its performance drops the most on small objects (0.6584).

Conclusions and prospects for further research

This paper presented a controlled, size-stratified evaluation of five SORT-family trackers for UAV tracking on the Anti-UAV-RGBT dataset. By fixing the detector and turning off ReID features, the study isolated the contribution of motion-based association strategies under fair conditions. The key findings are:

1. BoT-SORT's GMC module gives it a clear advantage on small objects (MOTA 0.7479), where even small camera movement can break IoU-based matching. This makes it the best choice when tracking accuracy is the top priority.

2. DeepOC-SORT and Hybrid-SORT run 38% faster (33+ FPS) while losing less than 1.3 percentage points of overall MOTA, making them good candidates for real-time use.

There are several limitations to note. All experiments use a single detector (RT-DETR-L), and different architectures may change the relative tracker rankings. ReID features are turned off to isolate motion-based tracking; turning on appearance models could change the results, especially for StrongSORT and DeepOC-SORT, which were designed with ReID support. The shared hyperparameters were not tuned for each tracker separately, which may put some algorithms at a disadvantage if they are sensitive to specific threshold settings. The evaluation covers only the visible modality of a single dataset, standalone detector performance metrics (AP/AR by size) are not reported, and processing speeds measured on a high-end NVIDIA RTX 5090 GPU may not hold on lower-end hardware. Also, identity-aware metrics like IDF1 and HOTA could give a more complete picture of long-term tracking quality and are left for future work.

These limitations point to several directions for future research: enabling ReID modules to measure the contribution of appearance features; reporting detector-level metrics alongside tracker metrics to separate detection and association errors; extending the evaluation to the infrared modality for nighttime and low-visibility conditions; adding IDF1 and HOTA metrics for a fuller assessment of identity preservation; investigating detector-level improvements for small object recall, such as higher-resolution inference or image tiling; and benchmarking on embedded platforms (e.g., NVIDIA Jetson series) for deployment-focused guidance.

Declaration on artificial intelligence

The author declares that no artificial intelligence (AI) technologies, generative AI tools, or large language models were used in the drafting, writing, or structuring of any element of this article.

Conflict of interest

The author declares that there is no conflict of interest and confirms that during the preparation of this work there were no commercial, financial, or other relationships that could be construed as influencing the results of the study or their interpretation. The work was performed in accordance with the principles of academic integrity, ethical standards for conducting scientific research, and editorial policy requirements for preventing conflicts of interest.

References

1. Aharon, N., Orfaig, R., & Bobrovsky, B.-Z. (2022). BoT-SORT: Robust associations multi-pedestrian tracking. https://doi.org/10.48550/ARXIV.2206.14651

2. Bernardin, K., & Stiefelhagen, R. (2008). Evaluating multiple object tracking performance: The CLEAR MOT metrics. EURASIP Journal on Image and Video Processing, 2008, 1–10. https://doi.org/10.1155/2008/246309

3. Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016). Simple online and realtime tracking. У 2016 IEEE international conference on image processing (ICIP)(с. 3464–3468). IEEE. https://doi.org/10.1109/ICIP.2016.7533003

4. Bychkov, O., Shatyrko, A., & Ivanenko, A. (2026). Comparative analysis of classical direction finding methods doa. Connectivity, 179(1), 3–17. https://doi.org/10.31673/2412-9070.2026.017402

5. Cao, J., Pang, J., Weng, X., Khirodkar, R., & Kitani, K. (2023). Observation-Centric SORT: Rethinking SORT for Robust Multi-Object Tracking. У 2023 IEEE/CVF conference on computer vision and pattern recognition (CVPR)(с. 9686–9696). IEEE. https://doi.org/10.1109/CVPR52729.2023.00934

6. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). End-to-End object detection with transformers. У A. Vedaldi, H. Bischof, T. Brox & J.-M. Frahm (Ред.), Computer vision – ECCV 2020(Т. 12346, с. 213–229). Springer International Publishing. https://doi.org/10.1007/978-3-030-58452-8_13

7. Coluccia, A., Fascista, A., Schumann, A., Sommer, L., Dimou, A., Zarpalas, D., Akyon, F. C., Eryuksel, O., Ozfuttu, K. A., Altinuc, S. O., Dadboud, F., Patel, V., Mehta, V., Bolic, M., & Mantegh, I. (2021). Drone-vs-Bird Detection Challenge at IEEE AVSS2021. У 2021 17th IEEE international conference on advanced video and signal based surveillance (AVSS)(с. 1–8). IEEE. https://doi.org/10.1109/AVSS52988.2021.9663844

8. Department of Software Systems and Technologies, Taras Shevchenko National University of Kyiv, Kyiv, 01601, Ukraine, Yurchuk, I., & Obertan, D.-M. (2025). Geospatial detection and movement analysis system for unmanned aerial vehicles based on computer vision methods. International Journal of Information Technology and Computer Science, 17(4), 16–27. https://doi.org/10.5815/ijitcs.2025.04.02

9. Du, Y., Zhao, Z., Song, Y., Zhao, Y., Su, F., Gong, T., & Meng, H. (2023). StrongSORT: Make deepsort great again. IEEE Transactions on Multimedia, 25, 8725–8737. https://doi.org/10.1109/TMM.2023.3240881

10. Hua, W., & Chen, Q. (2023, червень). A survey of small object detection based on deep learning in aerial images. https://doi.org/10.21203/rs.3.rs-3074407/v1

11. Jiang, N., Wang, K., Peng, X., Yu, X., Wang, Q., Xing, J., Li, G., Guo, G., Ye, Q., Jiao, J., Zhao, J., & Han, Z. (2023). Anti-UAV: A Large-Scale Benchmark for Vision-Based UAV Tracking. IEEE Transactions on Multimedia, 25, 486–500. https://doi.org/10.1109/TMM.2021.3128047

12. Liu, Z., An, P., Yang, Y., Qiu, S., Liu, Q., & Xu, X. (2024). Vision-Based drone detection in complex environments: A survey. Drones, 8(11), 643. https://doi.org/10.3390/drones8110643

13. Maggiolino, G., Ahmad, A., Cao, J., & Kitani, K. (2023). Deep OC-Sort: Multi-Pedestrian Tracking by Adaptive Re-Identification. У 2023 IEEE international conference on image processing (ICIP)(с. 3025–3029). IEEE. https://doi.org/10.1109/ICIP49359.2023.10222576

14. Svanström, F., Alonso-Fernandez, F., & Englund, C. (2022). Drone detection and tracking in real-time by fusion of different sensing modalities. Drones, 6(11), 317. https://doi.org/10.3390/drones6110317

15. T. Kysil & O. Zinchenko. (2025). Convolutional neural networks for moving object analysis in video streams. Connectivity, 176(4), 48–57. https://doi.org/10.31673/2412-9070.2025.042042

16. Voitovych, M., & Bychkov, O. (2025). Using hybrid multi-agent rl and automata for uav swarm decision-making in uncertain environments. Collection of Scientific Papers «SCIENTIA», (November 28, 2025; Glasgow, Scotland, UK), 232–329. https://previous.scientia.report/index.php/archive/article/view/3268

17. Wang, W., Fu, J., Song, J., Li, K., Qiao, H., Liu, J., Sun, H., & Cao, X. (2025). Dist-Tracker: A small object-aware detector and tracker for UAV tracking. У 2025 IEEE/CVF conference on

computer vision and pattern recognition workshops (CVPRW)(с. 6603–6611). IEEE. https://doi.org/10.1109/CVPRW67362.2025.00657

18. Wojke, N., Bewley, A., & Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. У 2017 IEEE international conference on image processing (ICIP)(с. 3645–3649). IEEE. https://doi.org/10.1109/ICIP.2017.8296962

19. Xie, B., Zhang, C., Wang, F., Liu, P., Lu, F., Chen, Z., & Hu, W. (2025). CST anti-uav: A thermal infrared benchmark for tiny UAV tracking in complex scenes. У 2025 IEEE/CVF international conference on computer vision workshops (ICCVW)(с. 6216–6225). IEEE. https://doi.org/10.1109/ICCVW69036.2025.00647

20. Yang, M., Han, G., Yan, B., Zhang, W., Qi, J., Lu, H., & Wang, D. (2024). Hybrid-SORT: Weak cues matter for online multi-object tracking. Proceedings of the AAAI Conference on Artificial Intelligence, 38(7), 6504–6512. https://doi.org/10.1609/aaai.v38i7.28471

21. Yurchuk, I., & Semenchenko, T. (2025a). Motion-consistent temporal fusion for UAV detection and tracking. Advanced Information Technology, (1 (4)), 22–28. https://doi.org/10.17721/AIT.2025.1.03

22. Yurchuk, I., & Semenchenko, T. (2025b). Vision-based UAV detection models for small-edge devices. У CEUR workshop proceedings(с. 81–90). CEUR-WS. https://ceur-ws.org/Vol-4035/Paper7.pdf

23. Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2022). ByteTrack: Multi-object tracking by associating every detection box. У S. Avidan, G. Brostow, M. Cissé, G. M. Farinella & T. Hassner (Ред.), Computer vision – ECCV 2022(Т. 13682, с. 1–21). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-20047-2_1

24. Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., Liu, Y., & Chen, J. (2024). DETRs beat yolos on real-time object detection. У 2024 IEEE/CVF conference on computer vision and pattern recognition (CVPR)(с. 16965–16974). IEEE. https://doi.org/10.1109/CVPR52733.2024.01605

25. Zhu, P., Wen, L., Du, D., Bian, X., Fan, H., Hu, Q., & Ling, H. (2022). Detection and tracking meet drones challenge. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(11), 7380–7399. https://doi.org/10.1109/TPAMI.2021.3119563

Т. О. Семенченко

ОЦІНЮВАННЯ АЛГОРИТМІВ ВІДСТЕЖЕННЯ НА ОСНОВІ РУХУ ДЛЯ ВІДСТЕЖЕННЯ ОБ'ЄКТІВ МАЛИХ БПЛА

Зростаюче використання малих безпілотних літальних апаратів (БПЛА) створює серйозні проблеми з безпекою для аеропортів, об'єктів критичної інфраструктури та військових об'єктів, формуючи попит на автоматизовані системи протидії дронам, здатні виявляти та супроводжувати повітряні цілі в реальному часі. Трекінг малих БПЛА залишається складним через низьку роздільну здатність, нерегулярну динаміку польоту, часті оклюзії та високу чутливість метрики IoU до похибок локалізації на рівні окремих пікселів, що разом ускладнює надійну часову асоціацію. Більшість наявних бенчмарків трекінгу кількох об'єктів (MOT) зосереджені на пішоходах, а не на малих повітряних цілях, через що практики не мають чітких рекомендацій щодо вибору трекера для задач БПЛА. У цій статті представлено емпіричне оцінювання п'яти алгоритмів трекінгу сімейства SORT, а саме ByteTrack, BoT-SORT, DeepOC-SORT, Hybrid-SORT та StrongSORT, на наборі даних Anti-UAV-RGBT. Усі трекери використовують спільний детектор RT-DETR-L, донавчений на тренувальному спліті Anti-UAV-RGBT, а реідентифікацію за зовнішнім виглядом вимкнено, щоб порівняння відображало лише стратегії асоціації на основі руху. Тестовий набір містить 92 послідовності з 79 699 еталонними анотаціями, згрупованими у три категорії розмірів (малі ≤48 пк, середні 48–128 пк, великі >128 пк), що дозволяє зробити розбивку якості трекінгу за масштабом цілі. Експерименти проводилися на одному графічному процесорі NVIDIA RTX 5090. BoT-SORT досягає найвищого загального показника MOTA 0.7802 і лідирує на малих об'єктах зі значенням 0.7479; його мо-

дуль глобальної компенсації руху камери (Global Motion Compensation) усуває зміщення, спричинене рухом камери, до кроку прогнозування Калманом і зберігає перекриття за IoU, коли ціль займає лише декілька десятків пікселів. ByteTrack демонструє найбільше падіння точності на малих об'єктах, втрачаючи 10.8 відсоткового пункту порівняно із загальним показником. DeepOC-SORT та Hybrid-SORT працюють на 38% швидше (понад 33 кадри за секунду), поступаючись менш, ніж на 1.3 відсоткового пункту у загальній точності. На основі цих результатів у статті надано практичні рекомендації щодо розгортання, які допомагають обрати трекер залежно від масштабу цілі, вимог до точності та доступних обчислювальних ресурсів.

Ключові слова: відстеження БПЛА, відстеження кількох об'єктів, SORT, відстеження малих об'єктів, RT-DETR, бенчмарк, Anti-UAV.

Надійшла до редакції: 23.09.2026
Прийнята до друку: 13.10.2026
Опубліковано: 30.10.2026
© 2026 Semenchenko T.
Цей матеріал ліцензовано за умовами CC BY 4.0.https://creativecommons.org/licenses/by/4.0/