Supplementary Material

6. Sensitivity Analysis of Proposed Metrics

To validate the robustness and behavior of our proposed evaluation metrics, Center-Scale Consistency (CSC) and Geometric Alignment Score (GAS), we conducted a sensitivity analysis of parameters. We systematically varied the key parameters of each metric and re-evaluated the performance of our tracker on the UOT32 [4] dataset. The results, visualized as heatmaps, demonstrate the stability of the metrics and provide justification for our selected default parameter values.

6.1. Analysis of Center-Scale Consistency (CSC)

The CSC metric is designed to provide a strict, binary measure of geometric similarity by thresholding both normalized center error (τ_c) and relative scale error (τ_s). Figure 4 illustrates how the mean CSC score varies as these thresholds are adjusted.

Observations: The heatmap shows a clear and expected monotonic relationship: as both τ_c and τ_s increase, the overall CSC score rises. This is because larger thresholds represent a more lenient version of a correct prediction, making the success condition easier to satisfy in the metric computation. We observe that the metric is particularly sensitive to the scale threshold τ_s at lower values (e.g., between 0.05 and 0.20), where a small increase in tolerance leads to a significant jump in the score. In contrast, the sensitivity to the center threshold τ_c is more gradual across the tested range.

Justification of Parameters ($\tau_c = 0.2, \tau_s = 0.2$): Our chosen default values are located in a stable knee region of the sensitivity curve. At this point, the metric is sufficiently strict to penalize trackers with poorer geometric control, as validated by the sharp drop in scores for lower thresholds. However, it is not overly penalizing, as the rate of score increase begins to plateau beyond this point. A 20% tolerance for both normalized center error and relative scale error represents a practically reasonable alignment for most tracking applications, making these values both informative and fair for comparison.

6.2. Analysis of Geometric Alignment Score (GAS)

The GAS metric offers a continuous, soft penalty for geometric misalignment, controlled by tolerance factors for center error (σ_c) and scale error (σ_s). These factors determine the standard deviation of the Gaussian penalty functions. The sensitivity of GAS to these parameters is shown in Figure 5.

Observations: Similar to CSC, the GAS score increases monotonically as the tolerance factors σ_c and σ_s are increased. A larger σ value corresponds to a wider, flatter Gaussian curve, which applies a smaller penalty (more lenient) for the same magnitude of error. The heatmap shows a smooth and continuous gradient, confirming that the metric is well-behaved and does not exhibit erratic responses to parameter changes.

Justification of Parameters ($\sigma_c = 0.5, \sigma_s = 0.5$): We selected $\sigma_c = 0.5$ and $\sigma_s = 0.5$ as our default values because they strike an effective balance. These parameters define a Gaussian penalty that is lenient for minor, almost insignificant errors but becomes progressively stricter for more significant deviations. This ensures that trackers are rewarded for precise alignment while still allowing for small, realistic variations in bounding box prediction. Using a factor of 0.5 of the ground truth dimensions (diagonal for center, width + height for scale) as the standard deviation provides a robust and scale-normalized tolerance that is applicable across objects of varying sizes.

In summary, our analysis confirms that both CSC and GAS are stable and interpretable metrics, and our chosen default parameters provide a challenging yet fair basis for evaluating the geometric accuracy of underwater object trackers (which could also be used in general tracking applications).

7. Correlation Analysis of Proposed Metrics

To validate our proposed geometric metrics and understand their relationship with standard tracking metrics, we perform a correlation analysis. We gather per-video evaluation scores from nine different trackers including MANTA on the UOT32 dataset for comparison. We then compute the Pearson correlation coefficient (r) between our proposed metrics Center-Scale Consistency (CSC) and Geometric Alignment Score (GAS) and the standard metrics of Success (AUC) and Precision (at 20px). The results provide insight into the evaluation properties of CSC and GAS.

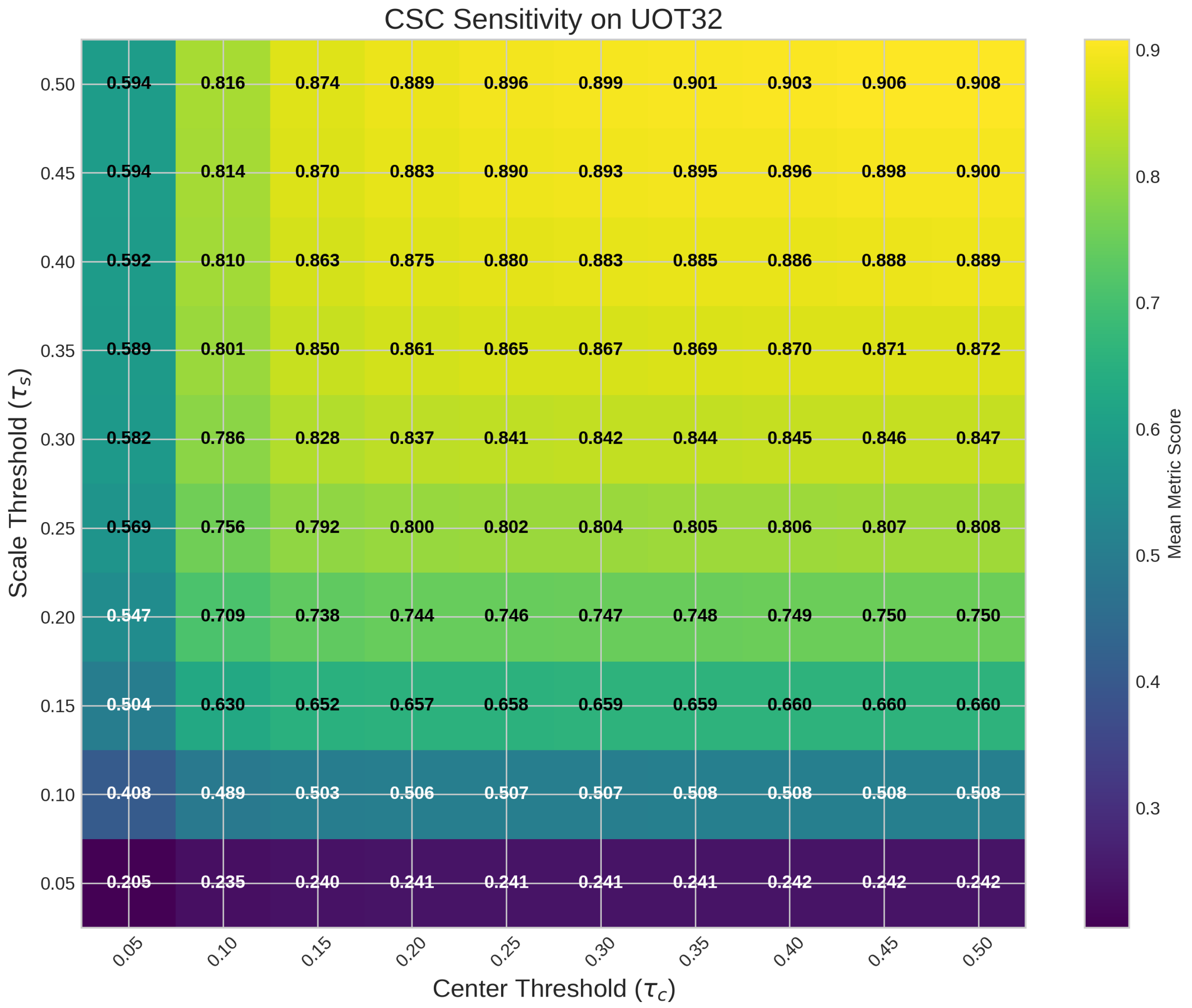


Figure 4. Sensitivity analysis of the CSC metric on the UOT32 dataset using MANTA. The heatmap shows the mean CSC score as a function of the center threshold (τ_c) and the scale threshold (τ_s). The score increases as the thresholds are relaxed, indicating a well-behaved metric.

Metric	Success AUC	Precision Score	CSC Score	GAS Score
Success AUC	1.000	0.794	0.883	0.962
Precision Score	0.794	1.000	0.677	0.824
CSC Score	0.883	0.677	1.000	0.767
GAS Score	0.962	0.824	0.767	1.000

Table 4. Pearson correlation matrix for standard and proposed evaluation metrics on the UOT32 dataset.

7.1. Quantitative Correlation Matrix

The Pearson correlation matrix, presented in Table 4, summarizes the linear relationships between the four metrics. A value close to 1.0 indicates a strong positive correlation, while a value closer to 0 indicates a weaker relationship.

7.2. Analysis of Geometric Alignment Score (GAS)

The GAS metric demonstrates a very strong positive correlation with Success AUC ($r = 0.962$) and a strong correlation with the Precision Score ($r = 0.824$), as visualized in Figure 6. The extremely high correlation with Success AUC suggests that GAS serves as an excellent proxy for overall tracking quality as measured by IoU. However, unlike the binary success/failure determination at a given IoU threshold, GAS provides a continuous, non-linear Gaussian penalty. This allows it to capture fine-grained differences in performance, especially among high-performing trackers, where it can distinguish between a good prediction and a near-perfect one more smoothly than the IoU metric.

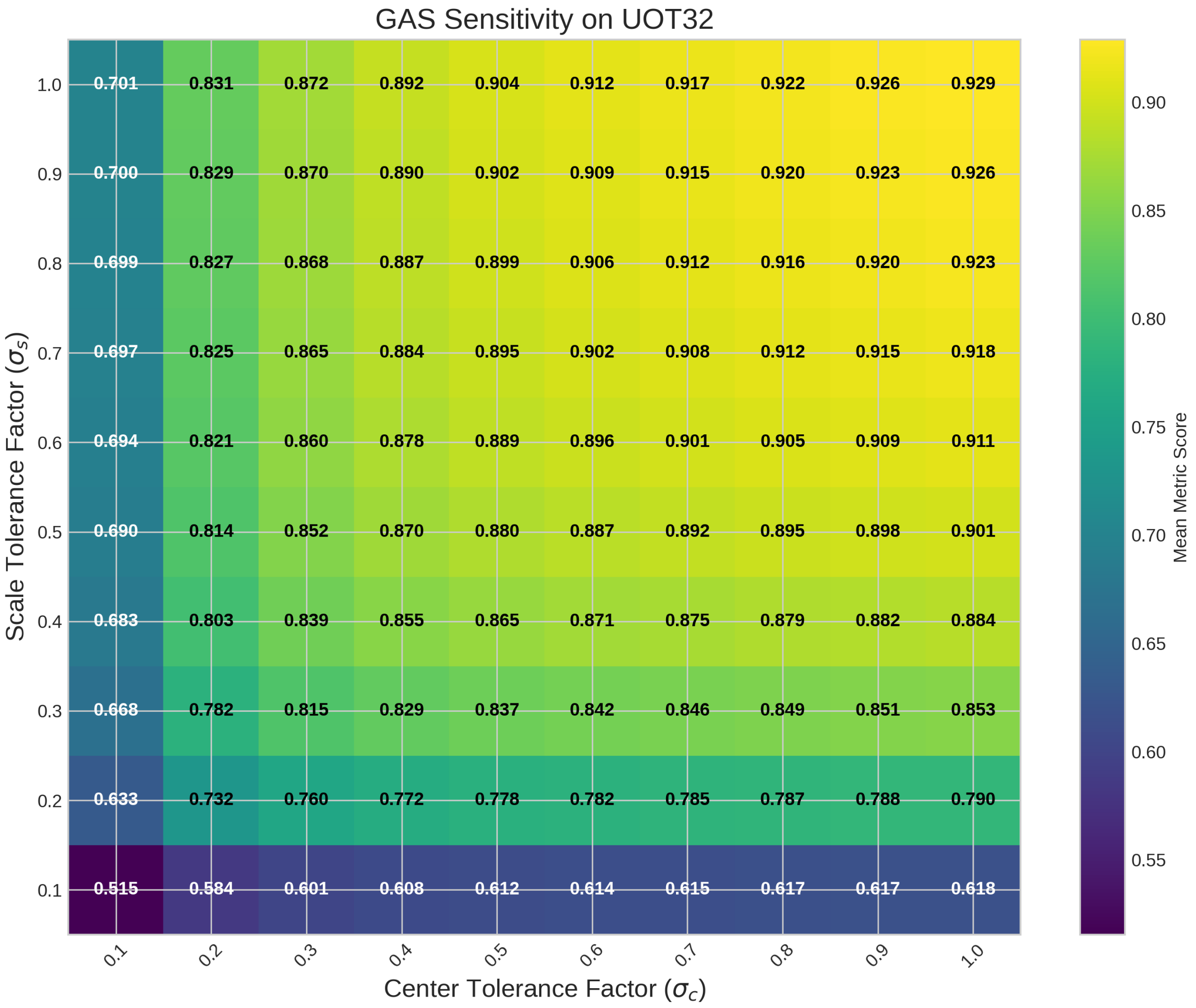


Figure 5. Sensitivity analysis of the GAS metric on the UOT32 dataset using MANTA. The heatmap shows the mean GAS score as a function of the center tolerance factor (σ_c) and the scale tolerance factor (σ_s). The smooth gradient indicates that GAS is a stable, well-behaved metric.

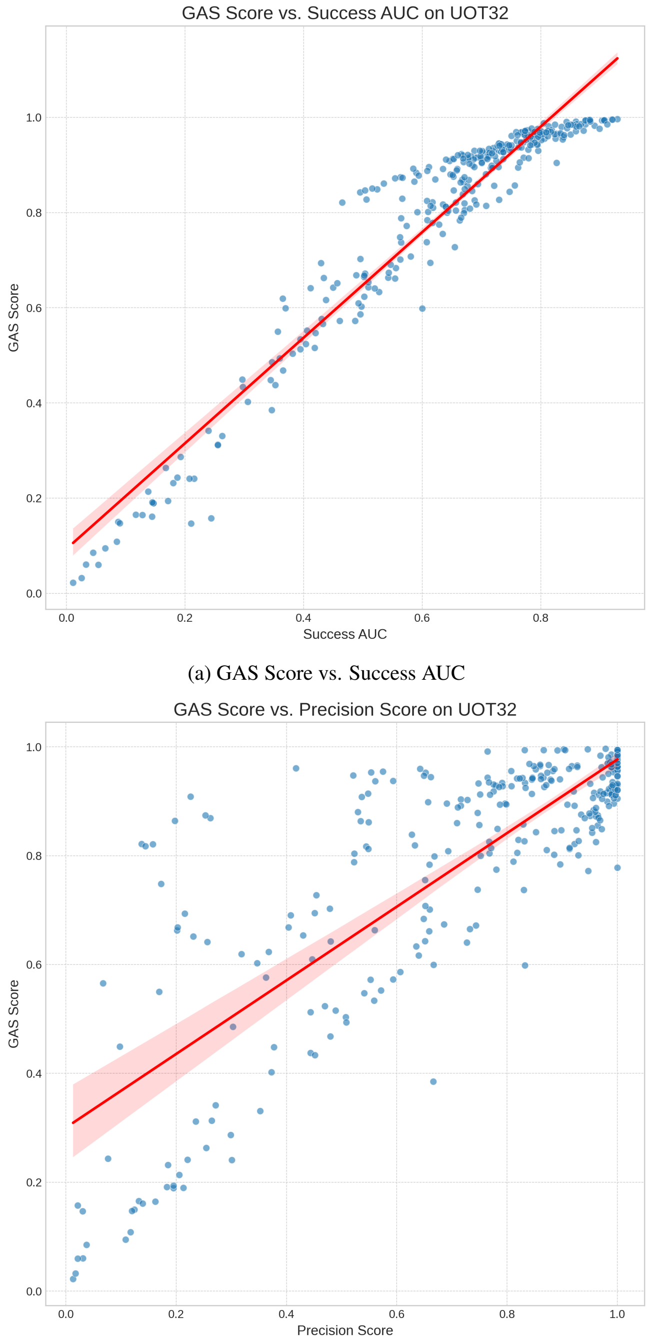
7.3. Analysis of Center-Scale Consistency (CSC)

The CSC metric exhibits a strong correlation with Success AUC ($r = 0.883$) but only a moderate correlation with the Precision Score ($r = 0.677$), as shown in Figure 7. This discrepancy is a key observation and highlights the unique value of the CSC metric. The Precision score solely evaluates the accuracy of the bounding box center, ignoring its dimensions. The moderate correlation ($r = 0.677$) reveals that a tracker can accurately locate the center of a target (while achieving a high Precision score) while simultaneously failing to predict the correct scale (resulting in a low CSC Score). CSC's strict requirement that both center and scale must be accurate makes it a powerful diagnostic tool for identifying trackers that produce geometrically inconsistent or poorly-sized bounding boxes. This is a weakness that is not adequately captured by the Precision metric alone.

In summary, our correlation analysis confirms that both GAS and CSC are valid and informative metrics. GAS acts as a strong, continuous measure of overall geometric alignment, correlating highly with Success AUC. CSC, on the other hand, provides unique information by enforcing a strict criterion on both location and scale, revealing weaknesses in trackers that are not apparent from the Precision score alone. Together, they form a valuable and complementary addition to the standard suite of tracking evaluation metrics.

7.4. Quantitative Evaluation Results

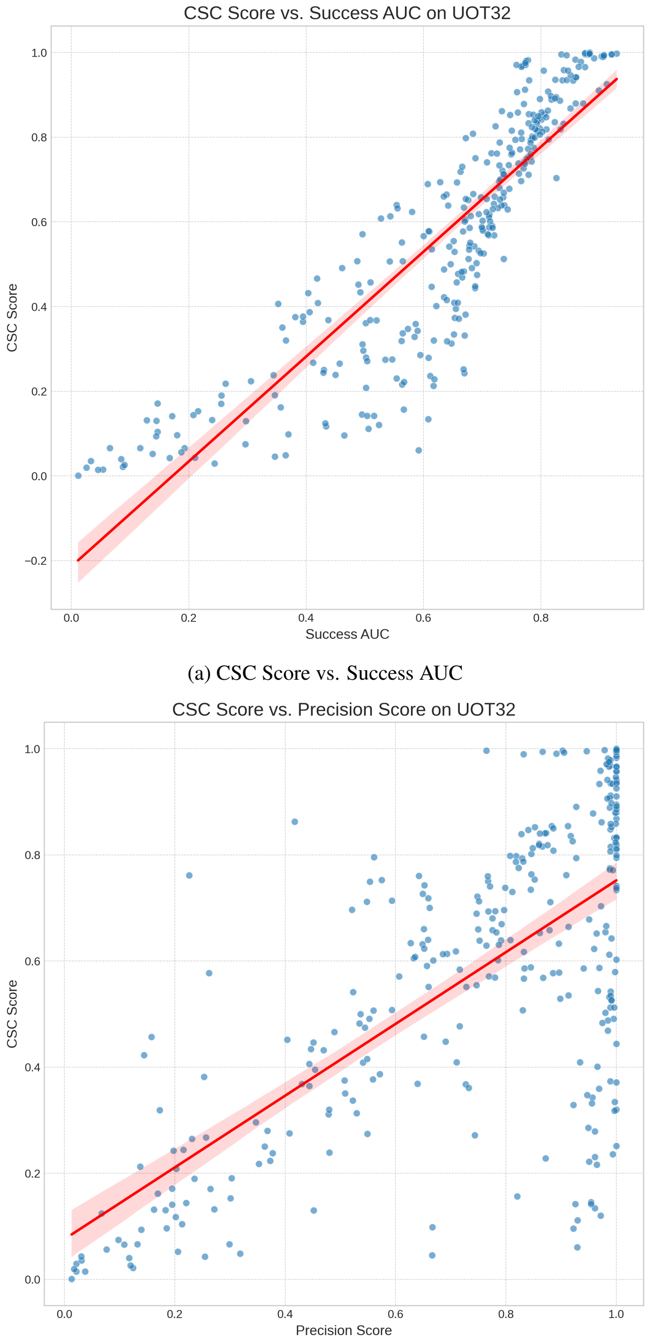
The quantitative evaluation on the UOT32 dataset reveals significant performance differences across the evaluated



(a) GAS Score vs. Success AUC

(b) GAS Score vs. Precision Score

Figure 6. Correlation plots for the Geometric Alignment Score (GAS). The strong linear relationship, particularly with Success AUC, validates GAS as a robust measure of tracking quality.



(a) CSC Score vs. Success AUC

(b) CSC Score vs. Precision Score

Figure 7. Correlation plots for the Center-Scale Consistency (CSC) Score. The weaker correlation with Precision Score highlights CSC's ability to penalize trackers for poor scale estimation.

tracking methods. Figure 8 presents the success and precision plots, which are the standard evaluation metrics for visual object tracking.

Success Plot Analysis: The success plot measures the percentage of frames where the Intersection over Union (IoU) between predicted and ground truth bounding boxes exceeds various thresholds. MANTA demonstrates superior performance with an Area Under Curve (AUC) of 0.731,

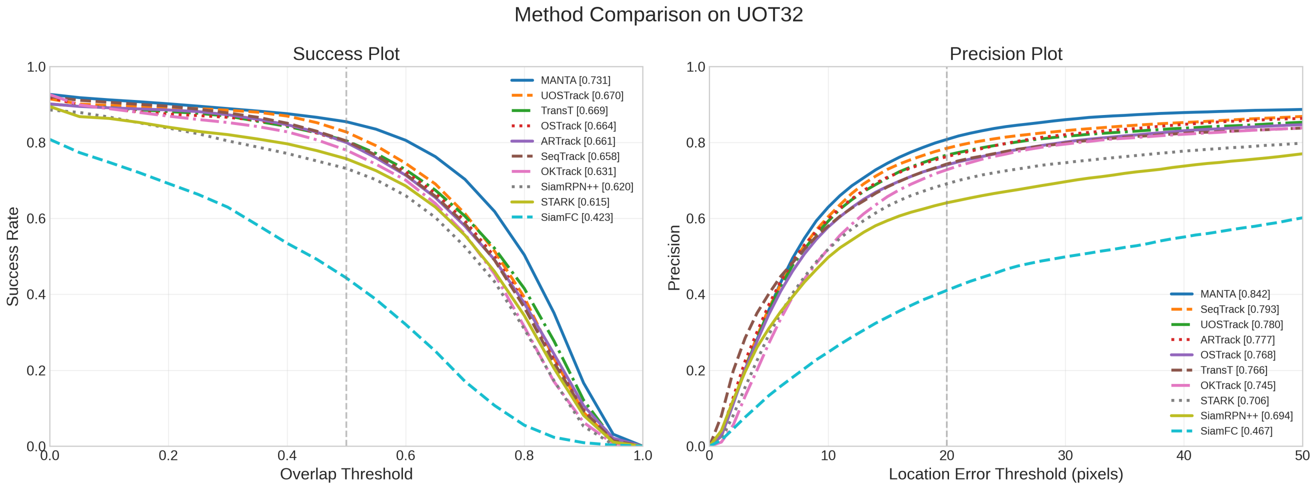


Figure 8. Quantitative comparison of underwater object tracking methods on the UOT32 dataset. **Left:** Success plot showing the percentage of frames with Intersection over Union (IoU) above varying overlap thresholds. Values in brackets indicate Area Under Curve (AUC) scores. **Right:** Precision plot displaying the percentage of frames with center location error below pixel-based thresholds. Values in brackets represent precision scores at 20-pixel threshold. MANTA achieves the highest performance in both metrics (Success AUC: 0.731, Precision@20px: 0.842), followed by UOSTrack and SeqTrack.

followed by UOSTrack (0.670) and TransT (0.669). The performance gap becomes more pronounced at higher overlap thresholds, indicating that MANTA produces more spatially accurate predictions. Notably, SiamFC [1] exhibits the poorest performance (AUC: 0.423), suggesting difficulties in handling the challenging underwater conditions present in UOT32.

Precision Plot Analysis: The precision plot evaluates tracking accuracy by measuring the percentage of frames where the center location error falls below pixel-based thresholds. MANTA again achieves the highest precision at 20 pixels threshold (0.842), demonstrating consistent localization accuracy. SeqTrack (0.793) and UOSTrack (0.780) also show competitive precision scores. The precision curves reveal that most methods plateau around 30-40 pixels, indicating fundamental limitations in precise localization under underwater imaging conditions.

The results demonstrate that transformer-based approaches (MANTA, TransT [2]) and recent deep learning methods (UOSTrack [5], SeqTrack [3]) generally outperform traditional correlation filter-based trackers (SiamFC, STARK) on underwater scenarios. This suggests that attention mechanisms and advanced feature representations are particularly beneficial for handling the unique challenges of underwater object tracking, including poor visibility, color distortion, and dynamic lighting conditions.

8. Temporal Performance Analysis on a Challenging Sequence

To assess the robustness of our tracker over long sequences, we conducted a temporal performance analysis on the `UWCOT220_000163` video, a challenging sequence featuring a sea turtle with significant appearance changes and potential occlusions from the UWCOT-220 [8] dataset. We compare the frame-by-frame IoU performance of MANTA against three other leading trackers: OSTrack, OKTrack, and SeqTrack. The results, visualized in Figure 9, highlight the superior stability of our physics-informed approach.

8.1. Discussion of Temporal Stability

The time-series plots provide a clear narrative of tracker stability. As shown in Figure 9, our proposed method, MANTA, achieves an exceptionally high mean IoU of 0.938 and its performance remains remarkably stable across the nearly 1400 frames of the sequence. The few sharp dips in its IoU curve are brief and quickly recovered, suggesting that its re-identification mechanism successfully handles transient challenges like partial occlusion/rapid motion.

In stark contrast, the competitors demonstrate a critical weakness. OSTrack [6], OKTrack [7], and SeqTrack all perform reasonably well for the first half of the sequence. However, they all experience a significant and prolonged drop in performance starting around frame 700, with IoU scores becoming highly erratic and frequently falling to near-zero values. This pattern is indicative of tracker drift, where the model loses the target and fails to reacquire it robustly.

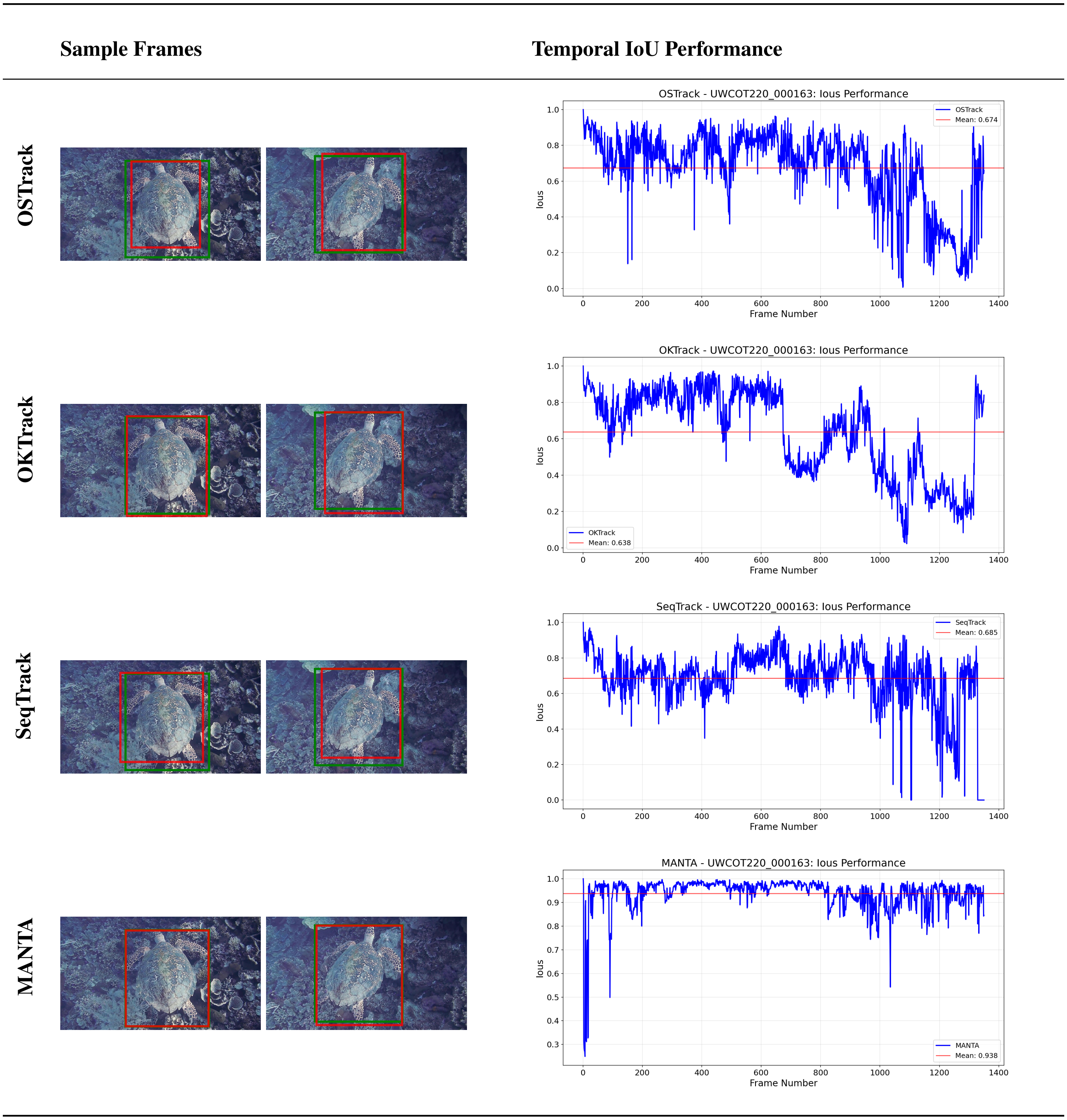


Figure 9. Temporal performance comparison on sequence `UWCOT220_000163`. Each row in the table presents a tracker's qualitative output (sample frames) alongside its quantitative IoU performance over the entire sequence. Each predicted bounding box is shown in red and the corresponding ground truth in green. While MANTA is able to efficiently track the turtle across almost all frames in the sequence,

This analysis strongly suggests that conventional trackers, even powerful transformer-based models, lack the robustness required for challenging underwater conditions. Their learned features are not sufficiently invariant to the severe appearance changes caused by shifting light, water turbidity, and the object's own movement. MANTA, with its physics-informed representations, is explicitly trained to be independent of these distortions, resulting in a far more stable and reliable track throughout the entire sequence. This consistent high performance is crucial for any real-world deployment in marine applications.

References

[1] Luca Bertinetto, Jack Valmadre, Joao F Henriques, Andrea Vedaldi, and Philip HS Torr. Fully-convolutional siamese networks for object tracking. In *European conference on computer vision*, pages 850–865. Springer, 2016. 5

[2] Xin Chen, Bin Yan, Jiawen Zhu, Dong Wang, Xiaoyun Yang, and Huchuan Lu. Transformer tracking. In *CVPR*, 2021. 5

[3] Xin Chen, Houwen Peng, Dong Wang, Huchuan Lu, and Han Hu. Seqtrack: Sequence to sequence learning for visual object tracking. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14572–14581, 2023. 5

[4] Landry Kezebou, Victor Oludare, Karen Panetta, and Sos S Agaian. Underwater object tracking benchmark and dataset. In *2019 IEEE International Symposium on Technologies for Homeland Security (HST)*, pages 1–6. IEEE, 2019. 1

[5] Yunfeng Li, Bo Wang, Ye Li, Zhuoyan Liu, Wei Huo, Yueming Li, and Jian Cao. Underwater object tracker: Uostrack for marine organism grasping of underwater vehicles. *Ocean Engineering*, 285:115449, 2023. 5

[6] Botao Ye, Hong Chang, Bingpeng Ma, Shiguang Shan, and Xilin Chen. Joint feature learning and relation modeling for tracking: A one-stream framework. In *ECCV*, 2022. 5

[7] Chunhui Zhang, Li Liu, Guanjie Huang, Hao Wen, Xi Zhou, and Yanfeng Wang. Webuot-1m: Advancing deep underwater object tracking with a million-scale benchmark. *Advances in Neural Information Processing Systems*, 37:50152–50167, 2024. 5

[8] Chunhui Zhang, Li Liu, Guanjie Huang, Zhipeng Zhang, Hao Wen, Xi Zhou, Shiming Ge, and Yanfeng Wang. Underwater camouflaged object tracking meets vision-language sam2. *arXiv preprint arXiv:2409.16902*, 2024. 5