Response to the review comments of reviewer 2

Comments and Suggestions for Authors

Comment 1: Better to make the best metrics in bold or colors in Table 1 and 2, so we can find them more easily.

Response:

Dear reviewer,

Thank you for your suggestion, especially regarding highlighting the best indicators in Tables 1 and 2 to make them more easily identifiable for readers. This is a helpful suggestion that will improve the readability of the tables.

Following your advice, we have highlighted the best indicators in Tables 1 and 2 using bold formatting to ensure readers can quickly identify them. This will help readers better understand our research findings. The optimized Tables 1 and 2 are shown below:

Table 1. Comparison between DB-Tracker and the latest multiple trackers tested on VisDrone dataset.

Tracker	MOTA	MOTP	IDF1	IDSW	MT%	ML	FP	FN
DeepSORT	19.4	69.8	33.1	6387	127	592	18635	73853
ByteTrack	25.1	72.6	40.8	4590	189	503	14154	64329
Deep OC-SORT	21.0	69.5	37.8	6489	133	663	17599	50485
BoT-SORT	23.0	71.6	41.4	7014	257	736	13952	69402
UAVMOT	25.0	72.3	40.5	6644	244	496	12804	80476
Strong SORT	34.8	74.3	42.6	3091	201	364	11257	54723
SimpleTrack	33.9	75.1	53.6	2238	**317**	**208**	**10528**	**47239**
DB-Tracker	**37.4**	**75.9**	**54.8**	**2043**	311	247	12903	73425

Table 2. Comparison between DB-Tracker and the latest multiple trackers tested on UAVDT dataset.

Tracker	MOTA	MOTP	IDF1	IDSW	MT%	ML%	FP	FN
DeepSORT	35.9	71.5	58.3	698	43.4	25.7	69091	144760
ByteTrack	39.1	74.3	44.7	2341	43.8	28.1	33129	173228
Deep OC-SORT	36.4	73.7	48.3	1993	35.9	30.6	44612	180921
BoT-SORT	37.2	72.1	53.1	1692	40.8	27.3	52451	146420
UAVMOT	43.0	73.5	61.5	**641**	45.3	22.7	46151	147735
Strong SORT	44.2	76.4	54.3	2024	38.3	24.8	53423	139875
SimpleTrack	45.3	77.9	57.1	1404	43.6	22.5	39452	**138457**
DB-Tracker	**46.2**	**78.3**	**58.2**	1436	**45.9**	**22.1**	**34531**	143625

Thank you once again for your suggestion. We will continue to work hard to improve the quality and comprehensibility of our paper. If you have any further suggestions or require additional modifications, please feel free to let us know, and we will do our best to accommodate your requests.

Comment 2: According to the papers of some existing methods like UAVMOT, Strong SORT, etc, their results are better than the those shown in Table 1 and 2, especially UAVMOT which was also tested on the same dataset VisDrone and UAVDT. Any reasons?

Response:

Dear Reviewer:

We appreciate your insightful observation regarding the comparative results of our proposed DB-Tracker algorithm with existing methods like UAVMOT and Strong SORT, particularly on datasets such as VisDrone and UAVDT. We would like to provide a detailed response to address this concern:

1. **Different experimental setups:** Each study used different experimental setups, including hyperparameters, data augmentation, training set, and test set division, etc. These settings may have a significant impact on the results, but complete details and training tricks are not openly disclosed in the published papers, making it difficult to achieve the performance described in the papers when reproducing the results.
2. **Model complexity:** The complexity and capacity of the model may also be a factor. More complex models may perform better in certain device conditions, but they may be more prone to overfitting or require higher computational resources in other device conditions.

Considering these factors, we encountered issues when reproducing the results on our own devices, where the actual results did not match the results described in the original paper.

Thank you once again for your suggestion. If you have any further suggestions or require additional modifications, please feel free to let us know, and we will do our best to accommodate your requests.

Comment 3: Follow the previus question, this paper misses some detailed comparison with existing compared methods, like when and why these methods don't work well and why the proposed method is better than them.

Response:

Dear Reviewer：

Thank you for your review and valuable feedback. Based on your suggestion, as well as considering the issue raised in comment 4, we have added additional details comparing our proposed method with existing contrast methods to highlight the contributions and advantages of our approach more clearly. The following is the added content:

> However, according to Tables 1 and 2, it is found that these existing methods have some problems that limit their performance in multi-target tracking tasks in drone aerial video. Although DeepSORT integrates deep learning and SORT algorithms, its high computing resource requirements make it less efficient in large-scale scenarios. ByteTrack is less robust in some complex situations, especially for target occlusion and appearance changes. Although BoT-SORT integrates a variety of tracking modules, the complex model structure requires more computing resources. In addition, although UAVMOT is

optimized for multi-target tracking in UAV aerial video, it still needs further improvement in some complex scenes. Although Deep OC-SORT combines appearance and motion information, its computational overhead is relatively large. Although Strong SORT has a certain degree of robustness, it is less efficient in large-scale scenes of drone aerial videos. Although SimpleTrack is simple and efficient, its robustness in complex scenarios is limited. Compared with these methods, the DB-tracker algorithm combines detection and tracking and comprehensively considers the location and appearance information of the target. Its outstanding robustness enables it to handle challenges such as complex scenes and target occlusions. DB-tracker can efficiently handle the diverse scenarios of drone aerial video, making it a promising solution in this field.

We believe that these additions will address the concerns raised and provide a more comprehensive understanding of the novelty and effectiveness of our proposed method.

Thank you once again for your input. If you have any further suggestions or require additional modifications, please feel free to let us know, and we will do our best to accommodate your requests.

Comment 4: Miss some qualitative/visual result of the comparison between the proposed DB-Tracker and other existing methods, especially some cases where other methods are worse than the proposed methods (together with analysis of the reason). Better to be on both public dataset and self-collected dataset.

Response:

Dear reviewer,

Thank you for your valuable suggestions and feedback. We agree that adding qualitative aspects would better support our research work and explain the performance differences between DB-Tracker and other existing methods. We have included more detailed textual descriptions to explain the performance differences between DB-Tracker and other existing methods, especially highlighting cases where other methods perform poorly in certain scenarios. We emphasize the advantages of DB-Tracker in diverse settings.

The following is the content we have added, as suggested in comment 3:

However, according to Tables 1 and 2, it is found that these existing methods have some problems that limit their performance in multi-target tracking tasks in drone aerial video. Although DeepSORT integrates deep learning and SORT algorithms, its high computing resource requirements make it less efficient in large-scale scenarios. ByteTrack is less robust in some complex situations, especially for target occlusion and appearance changes. Although BoT-SORT integrates a variety of tracking modules, the complex model structure requires more computing resources. In addition, although UAVMOT is optimized for multi-target tracking in UAV aerial video, it still needs further improvement in some complex scenes. Although Deep OC-SORT combines appearance and motion information, its computational overhead is relatively large. Although Strong SORT has a certain degree of robustness, it is less efficient in large-scale scenes of drone aerial videos. Although SimpleTrack is simple and efficient, its robustness in complex scenarios is limited. Compared with these methods, the DB-tracker algorithm combines detection and tracking and comprehensively considers the location and appearance information of the target. Its outstanding robustness enables it to handle challenges such

as complex scenes and target occlusions. DB-tracker can efficiently handle the diverse scenarios of drone aerial video, making it a promising solution in this field.

We believe that these additions will strengthen our research and provide a more comprehensive understanding of the capabilities and advantages of DB-Tracker.

Thank you once again for your input. If you have any further suggestions or require additional modifications, please feel free to let us know, and we will do our best to accommodate your requests.

Comment 5: Would be better and necessary for authors to provide some supplemental video results to prove the continuous consistency and efficiency of the proposed method.

Response:

Dear reviewer,

Thank you very much for your suggestion, especially regarding providing supplementary video results to demonstrate the continuity and efficiency of our proposed method. We greatly appreciate your point of view, as supplementary video results can provide readers with a more comprehensive and intuitive showcase of the method's performance.

We fully agree with your suggestion and will provide the requested supplementary video results. The videos will showcase the continuity and efficiency of our method in different scenarios, further supporting our research. We will ensure the quality and clarity of the video results so that readers can fully understand the performance of our method in practical applications.

The video website is: https://github.com/YubinYuan/DB-Tracker.

Once again, thank you for your valuable suggestion. We will do our best to meet your requirements and improve the quality and comprehensibility of our paper.

Thank you for your attention.

Best regards,

Yubin Yuan

Comments on the Quality of English Language

Comment 1: Some grammar, syntax or typo errors:

Response:

Dear reviewer,

Thank you for the feedback, reviewer. We have carefully reviewed the entire manuscript and made necessary revisions for any grammar, syntax, or spelling errors that were found. We are committed to ensuring the language

quality and accuracy of the paper. If there are any other suggestions or feedback in other aspects, we would be more than happy to listen and make the corresponding improvements. Thank you once again for your valuable suggestions.

Comment 2: I suggest authors to improve the abstract, since it misses the obvious contribution and advantage of the proposed method, but with only technical steps of the pipeline.

Response:

Dear reviewer,

Thank you very much for your valuable suggestions and feedback. We have carefully considered your feedback and made improvements to the abstract based on your suggestions. The revised abstract now better highlights the contributions and advantages of our proposed DB-Tracker algorithm. The improved abstract is as follows:

> **Abstract:** Drone aerial videos are promise in modern digital media and remote sensing applications, but effectively tracking several objects in these recordings is difficult. Drone aerial footage typically includes complicated sceneries with moving objects such as people, vehicles, and animals. Simultaneously, complicated scenarios such as large-scale viewing angle shifts and object crossings may occur. Random finite sets are mixed on a detection-based tracking framework, taking into account the object's location and appearance. It is maintained the detection box information of the detected object and construct the Box-MeMBer object position prediction framework based on the MeMBer random finite set point object tracking. We develop a hierarchical connection structure in the OSNet network, build MB-OSNet to get object appearance information, and connect feature maps of different levels through the hierarchy, such that the network may obtain rich semantic information at different sizes. Similarity measurements are computed and collected for all detections and trajectories in a cost matrix that estimates the likelihood of all possible matches. The cost matrix entries compare the similarity of tracks and detections in terms of position and appearance. The DB-Tracker algorithm performs excellently in multi-target tracking of drone aerial videos, achieving MOTA of 37.4% and 46.2% on the VisDrone and UAVDT data sets respectively. DB-Tracker achieves high robustness by comprehensively considering object position and appearance information, especially in handling complex scenes and target occlusion. This makes DB-Tracker a powerful tool in challenging applications such as drone aerial video.

We believe that these improvements will enhance the clarity and impact of our research. Once again, we sincerely appreciate your valuable suggestions. If you have any further suggestions or require additional modifications, please feel free to let us know, and we will do our best to accommodate your requests.

Comment 3: Ref [39] and [40] are exactly the same.

Response:

Dear reviewer,

Thank you very much for pointing out the issue regarding the duplicate references [39] and [40]. This is an important error that needs to be corrected, and we greatly appreciate your discovery.

We have checked the content of references [39] and [40] and made the necessary corrections. We will ensure that these two references cite the correct sources and have made the corresponding changes in the reference list to rectify this error.

Once again, we sincerely appreciate your thorough review, and your feedback has been extremely helpful in improving the quality of the paper. If you have any further suggestions or require additional modifications, please feel free to let us know, and we will actively address them.

Comment 4: Line 521 [3434]

Response:

Dear reviewer,

Thank you for pointing out the issue with the reference citation in Line 521. This is an error that needs to be corrected, and we greatly appreciate your discovery.

We have verified the reference [3434] cited in Line 521 and ensured the accuracy and consistency of the citation. We have made the necessary corrections to ensure the completeness of the paper's references.

Once again, we sincerely appreciate your thorough review, and your feedback has been extremely helpful in improving the quality of the paper. If you have any further suggestions or require additional modifications, please feel free to let us know, and we will actively make the necessary changes.

Thank you for your attention.

Best regards,

Yubin Yuan