

GMOT-40: A Benchmark for Generic Multiple Object Tracking

—Supplementary Material—

1. Label Format for GMOT-40

The label format for proposed GMOT-40 is shown in the Table 1. We mainly follow the format of the widely-used MOT15 dataset [3]. The only difference is, MOT15 does not take some challenging targets, like small ones, into consideration for evaluation. It uses an extra flag to indicate that these labeled targets should be ignored. On the contrary, GMOT-40 includes all of them in evaluation, no matter how challenging the targets could be, and the flag is not used here. This is consistent with our motivation, i.e., trackers need to deal with these *real world challenges*.

2. Qualitative Analysis

2.1. One-shot GMOT Protocol

The one-shot GMOT protocol visualization results copied from body part are shown in the top three rows of Figure 1 (Protocol A). Each bounding box with a polygon line in color represents a tracked target and its trajectory. Compared with the protocol used in ablation study where groundtruth detections are available, IOU tracker performance drops a lot. We think this is because IOU tracker can not handle errors, like false positive/negative detection results, induced by imperfect detectors. By contrast, MDP, Deep SORT and FAMNet have extra mechanisms to refine these faulty detection results during tracking process. Furthermore, unlike Deep SORT and FAMNet, MDP did not use any pre-trained CNN, which makes itself more robust against during generalization to unseen categories.

2.2. Protocol of Ablation Study

The protocol of ablation study visualization results are shown in the bottom three rows of Figure 1 (Protocol B). For the 2nd row of fish sequence in Protocol B result, Deep SORT [4] and IOU tracker [1] have tracked more targets than MDP [5] and FAMNet [2]. The reason might be Deep SORT and IOU tracker mainly adopt IOU-based tracking paradigm, and they could perform well with all ground truth detection results available. But MDP and FAMNet have superfluous pre-processing on detection which may be harmful under this protocol. Note the visualization result may not be consistent with the quantitative result in the main

body for each sequence, due to the averaging process of computing metrics.

3. Scores for All Sequences

Figure 2 and Figure 3 present the scores for both protocols (one-shot GMOT protocol and the protocol used in ablation study) and all sequences. We can see that the sequences that are easy to handle in the protocol of ablation study may be challenging in one-shot GMOT protocol. Yet the challenging sequences in the protocol of ablation study are still difficult in one-shot GMOT protocol. Such difference and similarity again stress the importance and necessity of a one-shot framework in Generic MOT.

References

- [1] Erik Bochinski, Volker Eiselein, and Thomas Sikora. High-speed tracking-by-detection without using image information. In *AVSS*, 2017. 1, 2
- [2] Peng Chu and Haibin Ling. Famnet: Joint learning of feature, affinity and multi-dimensional assignment for online multiple object tracking. In *ICCV*, 2019. 1, 2
- [3] Laura Leal-Taixé, Anton Milan, Ian Reid, Stefan Roth, and Konrad Schindler. MOTChallenge 2015: Towards a benchmark for multi-target tracking. *arXiv:1504.01942*, 2015. 1
- [4] Nicolai Wojke, Alex Bewley, and Dietrich Paulus. Simple online and realtime tracking with a deep association metric. In *ICIP*, 2017. 1, 2
- [5] Yu Xiang, Alexandre Alahi, and Silvio Savarese. Learning to track: Online multi-object tracking by decision making. In *ICCV*, 2015. 1, 2

Position	Name	Description
1	Frame number	Starts from 0, indicates which frame the target belongs to
2	Identity number	Each trajectory is identified as an unique ID. For detection, it is set to be -1.
3	Bounding box left	Coordinates of the top-left corner of the bounding box
4	Bounding box top	Coordinates of the top-left corner of the bounding box
5	Bounding box width	Width of bounding box in pixels
6	Bounding box height	Height of bounding box in pixels
7	Confidence score	Predicted probability of the detection being foreground. For groundtruth, it is set to be 1.
8-10	-1	Padding to fit MOTChallenge format

Table 1. Annotation format in GMOT-40 dataset.

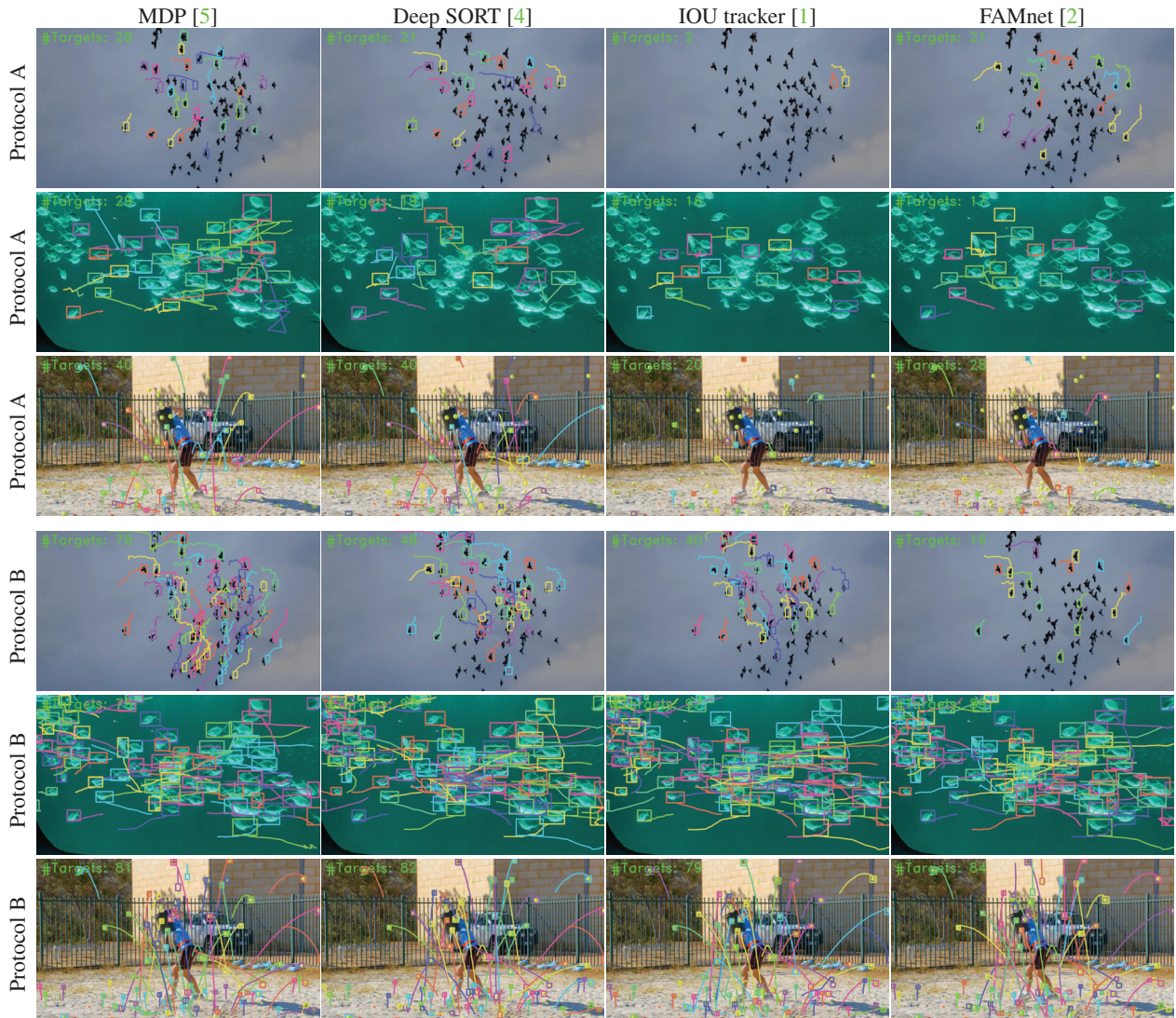


Figure 1. Results visualization of four trackers on several sequences using different protocols.

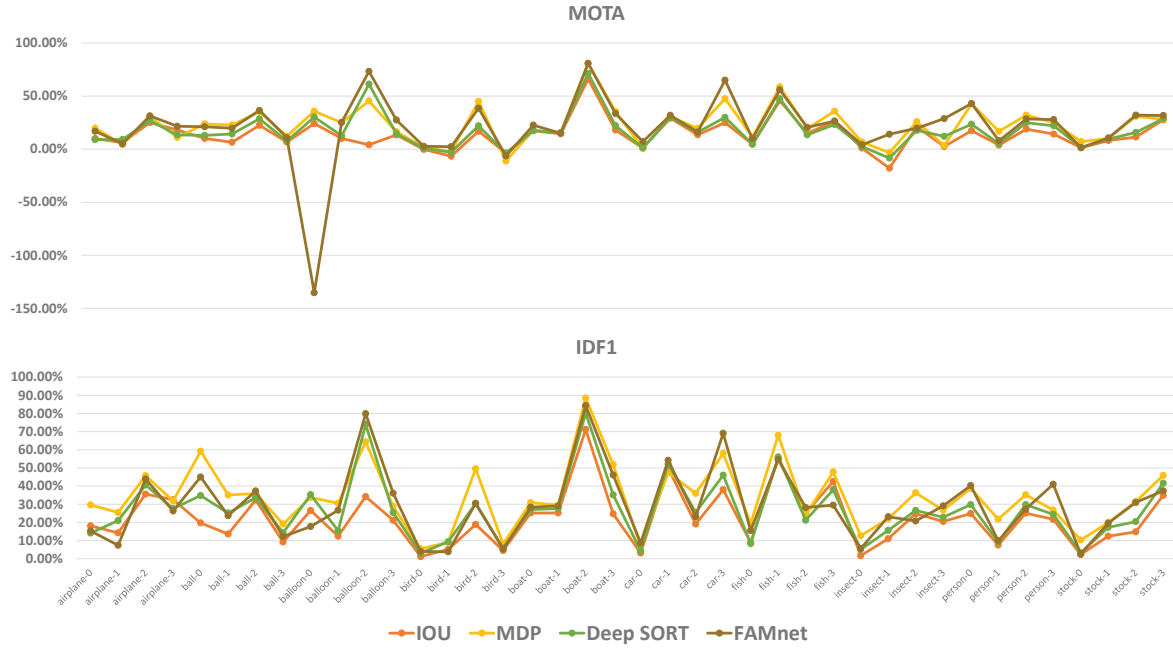


Figure 2. Scores in one-shot GMOT protocol.

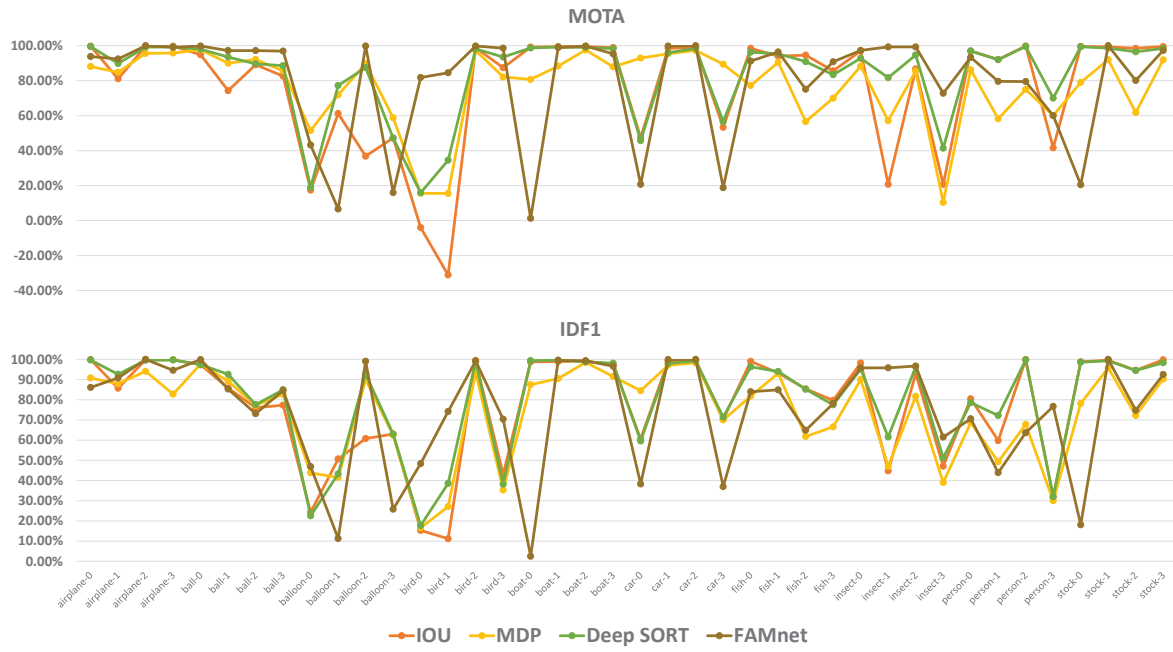


Figure 3. Scores in the protocol of ablation study.