Hard to Track Objects with Irregular Motions and Similar Appearances? Make It Easier by Buffering the Matching Space —Supplemental Material—

Fan Yang, Shigeyuki Odashima, Shoichi Masui, Shan Jiang
Fujitsu Research, Japan
contact: fan.yang@fujitsu.com

In this supplementary material, we present additional experimental results to support our main paper. First, we perform ablation studies to verify the effect of max_age and interpolation in our C-BIoU tracker (Sec. 1). Then, we analyze the tracking evaluation details on each video of the MOT17 and DanceTrack test sets (Sec. 2). Next, we compare tracking results on the validation sets of the MOT17 and DanceTrack when oracle detections are given (Sec. 3). Finally, we illustrate more visualization results on our experimental datasets (Sec. 4). In addition to this file, we provide an extra video for viewing.

1. Additional Ablation Studies

1.1. The Effect of max_age

Table 1 shows the results of the ablation study using different values of max_age. Although the best setting of max_age is different for each data set, the overall performance is similar by setting max_age between 20 and 100. Our C-BIoU tracker may not be sensitive to the max_age value on the DanceTrack [8] validation set, SoccerNet [4, 5] test set, and GMOT-40 [1] test set.

Meanwhile, since the temporal spans of objects are different in each dataset, the high value of max_age could generate better tracking results for long-term objects. In Fig. 1, the distribution of tracklet length is consistent with the optimized value of max_age, as the long tracklet length corresponds to a large max_age.

Table 1: **Ablation for max_age on DanceTrack validation set [8], SoccerNet test set [4, 5], and GMOT-40 test set [1].** We select three levels of max_age, as 20, 60, and 100 for the investigation.

Tracker	HOTA↑	DetA↑	AssA↑	MOTA↑	IDF1↑
DanceTrack Validation Set [8]. Using Oracle Detections.					
C-BIoU Tracker ($max\_age = 20$)	80.8	**97.6**	66.8	**99.3**	79.0
C-BIoU Tracker ($max\_age = 60$)	81.3	**97.6**	67.7	**99.3**	79.9
C-BIoU Tracker ($max\_age = 100$)	**81.7**	**97.6**	**68.4**	**99.3**	**80.5**
SoccerNet Test Set [4, 5]. Using Oracle Detections.					
C-BIoU Tracker ($max\_age = 20$)	88.5	**99.5**	78.7	**99.4**	84.1
C-BIoU Tracker ($max\_age = 60$)	**89.2**	99.4	**80.0**	**99.4**	**86.1**
C-BIoU Tracker ($max\_age = 100$)	89.0	99.4	79.7	**99.4**	**86.1**
GMOT-40 Test Set [1]. Using Oracle Detections.					
C-BIoU Tracker ($max\_age = 20$)	**96.4**	**99.7**	**93.2**	**99.6**	**95.6**
C-BIoU Tracker ($max\_age = 60$)	96.2	**99.7**	92.9	**99.6**	95.4
C-BIoU Tracker ($max\_age = 100$)	96.1	**99.7**	92.7	**99.6**	95.2

1.2. The Effect of Linear Interpolation

In order to recover missing detections, linear interpolation is commonly applied in MOT methods. Using the same interpolation approach as OC-SORT [3], we investigate the effect of the linear interpolation on our C-BIoU tracker. As shown in

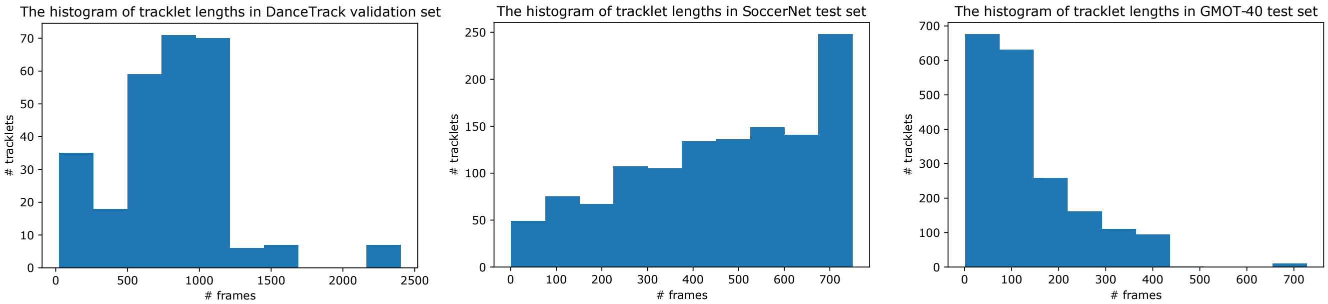

Figure 1: **The histogram of tracklet length on DanceTrack validation set [8], SoccerNet test set [4, 5], and GMOT-40 test set [1]**. The distribution of tracklet length is consistent with the optimized value of max_age, as the long tracklet length corresponds to a large max_age.

Table 2: **The effect of linear interpolation on DanceTrack test set.** Performing linear interpolation may recover part of the missing detections and thus improve the tracking performance.

Tracker	HOTA↑	DetA↑	AssA↑	MOTA↑	IDF1↑
OC-SORT [3]	55.1	80.3	38.0	89.4	54.2
OC-SORT [3] + Interpolation	55.7	**81.7**	38.3	**92.0**	54.6
C-BIoU Tracker	**60.6**	81.3	45.3	91.6	61.4
C-BIoU Tracker + Interpolation	**60.6**	81.3	**45.4**	91.6	**61.6**

Table 2, applying linear interpolation improves the performance of both the C-BIoU tracker and OC-SORT. However, since our C-BIoU tracker performs better tracking before the interpolation, the improvement of applying linear interpolation is less significant for our C-BIoU tracker.

2. Details of Tracking Results on the Test Set of the MOT17 and DanceTrack

Fig. 2 represents the tracking performance of each video on MOT17 test set [7] and DanceTrack test set [8]. We choose HOTA metrics (*i.e.*, HOTA, DetA and AssA) [6] for the evaluation. On the MOT17 test set, the data association performance (*i.e.*, AssA) is proportional to the detection performance (*i.e.*, DetA). On the DanceTrack test set, however, the data association performance may not be proportional to the detection performance. Therefore, for the MOT17 dataset, the bottleneck in tracking performance lies in the detection quality more than the data association ability. In contrast, for the DanceTrack dataset, the data association ability is as important as the detection quality. Our C-BIoU tracker shows its strong data association ability in our experiments. Nonetheless, we also show that our C-BIoU tracker generates poor results on several videos of DanceTrack, which could be caused by the limitations of our C-BIoU tracker. Although we have achieved the best overall performance on DanceTrack, we still have a long way to go to make satisfactory results on each video of DanceTrack.

3. Tracking Results on the Validation Set of the MOT17 and DanceTrack

In the preceding section and our main paper, we have compared tracking results on the test sets of the MOT17 [7] and DanceTrack [8]. To better illustrate the difference between the MOT17 and DanceTrack datasets, as well as to further explore the properties of our C-BIoU tracker, we follow the train-val splitting provided in CenterTrack [10] to construct the validation set of MOT17 and perform experiments on it.

In Table 3, we compare the tracking results on the validation sets of the MOT17 and DanceTrack when oracle detections are given. Our C-BIoU tracker achieves the best performance on both datasets, which is consistent with the results on the GMOT-40 and SoccerNet datasets. The superiority of our C-BIoU tracker is verified. However, although oracle detections are used, we observe that the experimental MOT methods are nearly perfect on the MOT17 validation set, but are far from perfect on the DanceTrack validation set. Compared to the MOT17 dataset, the DanceTrack dataset contains objects with irregular motions in most cases [8], so a more robust data association (*e.g.*, our C-BIoU tracker) is required to achieve good performance. Meanwhile, we once again show the importance of building baselines with oracle detections, which allows us to focus on improving the data association from the initial stage.

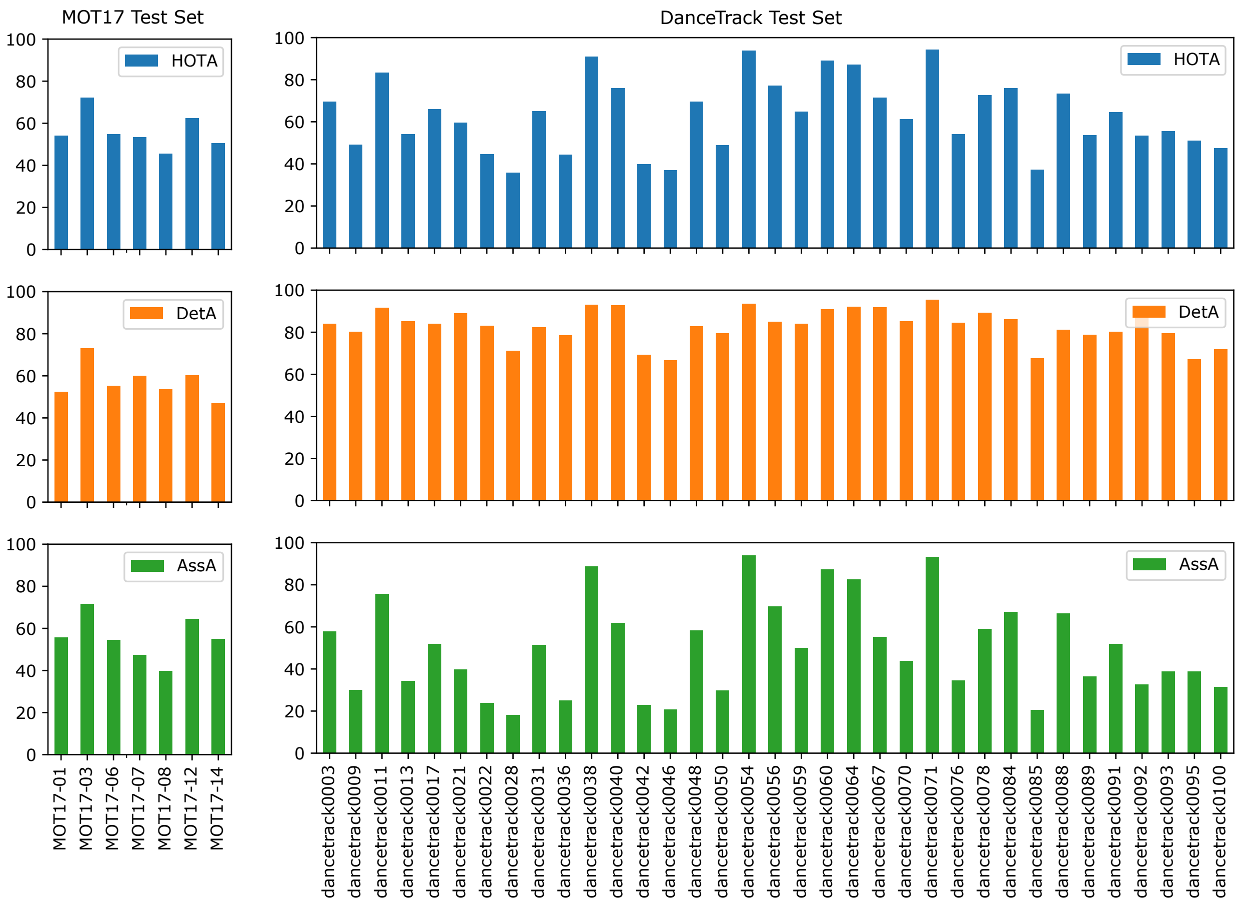

Figure 2: **The tracking performance of each video on MOT17 test set [7] and DanceTrack test set [8].** On MOT17 test set, The data association performance (*i.e.*, AssA) is proportional to the detection performance (*i.e.*, DetA). On DanceTrack test set, for videos with the similar detection performance, the corresponding data association performance may have large differences, and therefore the data association performance may not be proportional to the detection performance.

Table 3: **Results on the validation sets of MOT17 [7] and DanceTrack [8].**

Tracker	MOT17 Validation Set					DanceTrack Validation Set				
	HOTA↑	DetA↑	AssA↑	MOTA↑	IDF1↑	HOTA↑	DetA↑	AssA↑	MOTA↑	IDF1↑
Using Oracle Detections										
DeepSORT [9]	95.0	94.7	95.4	99.3	**98.8**	66.8	86.1	51.8	97.4	68.3
SORT [2]	96.4	97.1	95.8	99.7	98.1	67.6	86.6	52.8	98.1	69.6
OC-SORT [3]	96.9	**99.9**	94.1	**99.9**	95.8	79.1	**97.7**	64.0	**99.6**	76.1
C-BIoU Tracker	**98.6**	**99.9**	**97.3**	**99.9**	98.2	**81.7**	97.6	**68.4**	99.3	**80.5**

4. More Visualizations

In addition to the visualization results in our main paper, we present other results in this section. In Fig. 3, we plot tracking results on GMOT-40 test set [1] to compare our C-BIoU tracker to SORT [2] and OC-SORT [3]. Although tracking on the GMOT-40 test set is relatively easier than on other data sets, traditional methods, such as SORT, may generate more tracking errors than our C-BIoU tracker. Besides, we also give more samples to illustrate our tracking results on the DanceTrack test set and SoccerNet test set in Fig. 4 and Fig. 5, respectively.

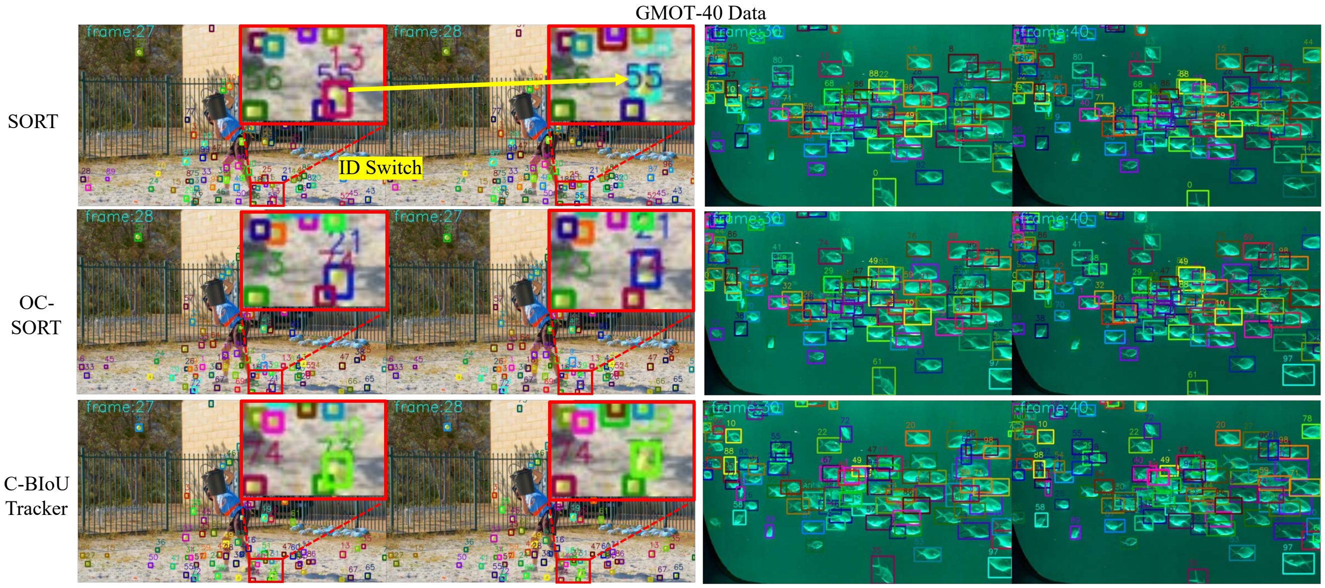

Figure 3: **Comparison of tracking results on GMOT-40 test set [1].** Our C-BIoU tracker generates fewer tracking errors than SORT [2].

References

[1] Hexin Bai, Wensheng Cheng, Peng Chu, Juehuan Liu, Kai Zhang, and Haibin Ling. Gmot-40: A benchmark for generic multiple object tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6719–6728, 2021.

[2] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft. Simple online and realtime tracking. In *IEEE international conference on image processing*, pages 3464–3468, 2016.

[3] Jinkun Cao, Xinshuo Weng, Rawal Khirodkar, Jiangmiao Pang, and Kris Kitani. Observation-centric sort: Rethinking sort for robust multi-object tracking. *arXiv preprint arXiv:2203.14360*, 2022.

[4] Anthony Cioppa, Silvio Giancola, Adrien Deliege, Le Kang, Xin Zhou, Zhiyu Cheng, Bernard Ghanem, and Marc Van Droogenbroeck. Soccernet-tracking: Multiple object tracking dataset and benchmark in soccer videos. *arXiv preprint arXiv:2204.06918*, 2022.

[5] Adrien Deliege, Anthony Cioppa, Silvio Giancola, Meisam J Seikavandi, Jacob V Dueholm, Kamal Nasrollahi, Bernard Ghanem, Thomas B Moeslund, and Marc Van Droogenbroeck. Soccernet-v2: A dataset and benchmarks for holistic understanding of broadcast soccer videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4508–4519, 2021.

[6] Jonathon Luiten, Aljosa Osep, Patrick Dendorfer, Philip Torr, Andreas Geiger, Laura Leal-Taixé, and Bastian Leibe. Hota: A higher order metric for evaluating multi-object tracking. *International Journal of Computer Vision*, pages 1–31, 2020.

[7] Anton Milan, Laura Leal-Taixé, Ian Reid, Stefan Roth, and Konrad Schindler. Mot16: A benchmark for multi-object tracking. *arXiv preprint arXiv:1603.00831*, 2016.

[8] Peize Sun, Jinkun Cao, Yi Jiang, Zehuan Yuan, Song Bai, Kris Kitani, and Ping Luo. Dancetrack: Multi-object tracking in uniform appearance and diverse motion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 0–10, June 2022.

[9] Nicolai Wojke, Alex Bewley, and Dietrich Paulus. Simple online and realtime tracking with a deep association metric. In *IEEE international conference on image processing*, pages 3645–3649, 2017.

[10] Xingyi Zhou, Vladlen Koltun, and Philipp Krähenbühl. Tracking objects as points. In *European Conference on Computer Vision*, pages 474–490. Springer, 2020.

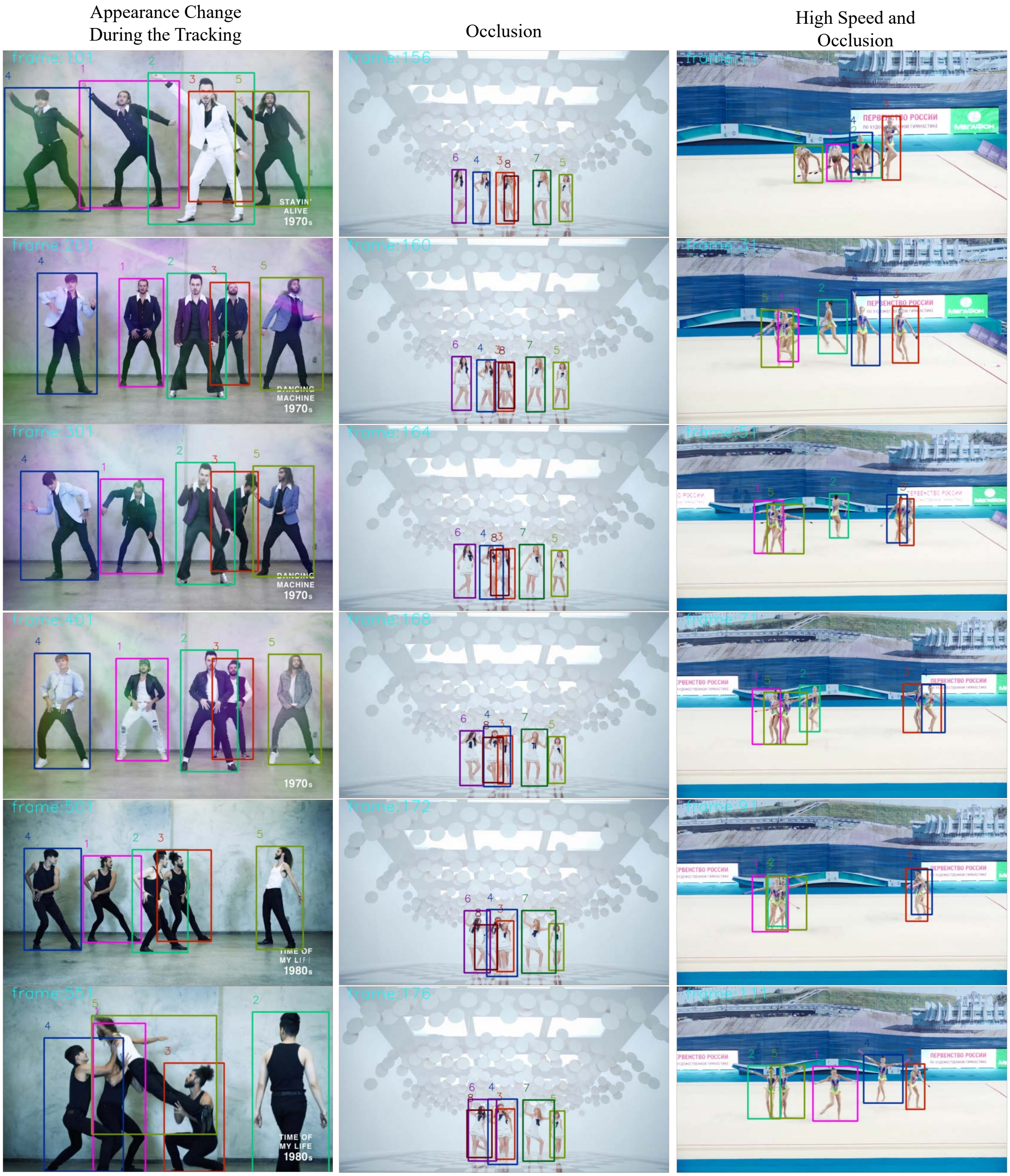

Figure 4: **Tracking results of our C-BIoU tracker on the DanceTrack test set [8].** The detections are generated by YOLOXS.

Figure 5: **Tracking results of our C-BIoU tracker on the SoccerNet test set [4, 5].** The players of the same team wear a similar uniform, and the captured sport field changes, which is nontrivial for the tracking.