Supplementary Material: DRT: Detection Refinement for Multiple Object Tracking

Bisheng Wang[1, 3]
316106002478@njust.edu.cn
Christian Fruhwirth-Reisinger[2, 3]
christian.reisinger@icg.tugraz.at
Horst Possegger[3]
possegger@icg.tugraz.at
Horst Bischof[2, 3]
bischof@icg.tugraz.at
Guo Cao[1]
caoguo@njust.edu.cn

[1] School of Computer Science and Engineering
Nanjing University of Science and Technology

[2] Christian Doppler Laboratory for Embedded Machine Learning

[3] Institute of Computer Graphics and Vision
Graz University of Technology

In the following, we provide additional experimental results to demonstrate the performance of our method, Detection Refinement for Tracking (DRT). This supplementary material includes two parts. First, we show the tracking results on the recent MOT20 [4] dataset and compare our method with the baseline method, FAIR [11]. Second, we compare the detection results of our DRT on MOT17Det detection dataset [5] with other methods.

1 Results on MOT20 dataset

MOT20 [4] is a new benchmark for multi-object tracking which consists of 8 challenging sequences. Compared with MOT16 & 17 [5], MOT20 has much more crowded scenes and complex backgrounds. Furthermore, the test sets of MOT20 include another two unknown scenes, which are different from the scenes in the training dataset, to measure the generalization capabilities of detectors and trackers.

Our training settings and procedure are the same as described in the main paper. We use the detections of FAIR as input for our DRT-net for this comparison. The MOT20 results are shown in Table 1.

Our DRT achieves 2.5% improvement of MOTA compared to FAIR [11]. The improvements arise from that our DRT can perform well in reducing false positives. The ID switches are slightly higher than FAIR because the trackers may drift between the occluded targets and foreground people in occlusion areas.

Method	MOTA↑	IDF1↑	MOTP↑	MT↑	ML↓	FP↓	FN↓	IDS↓	Hz↑
SORT [1]	42.7	45.1	78.5	16.7	26.2	**27521**	264694	**4470**	**57.3**
TransCenter [8]	58.3	46.8	**79.7**	35.7	18.6	<u>35959</u>	174893	<u>4947</u>	1
FAIR [11] ∗	<u>61.8</u>	**67.3**	78.6	**68.8**	**7.6**	103440	**88901**	5243	<u>13.2</u>
DRT (Ours) ∗	**64.3**	<u>58.7</u>	<u>79.1</u>	<u>60.2</u>	<u>8.9</u>	70794	<u>108316</u>	5389	0.5

Table 1: Comparison with other state-of-the-art trackers on MOT20. All methods are under the "private detector" protocol. ↑/↓ indicates that higher/lower scores correspond to better performance. Methods marked with the same sign use the same detector.

Method	AP↑	MODA↑	MODP↑	FAF↓	TP↑	FP↓	FN↓	Recall↑	Precision↑	F1↑
FRCNN [6]	0.72	68.5	78.0	1.7	88601	100881	25963	77.3	89.8	83.1
SDP [9]	0.81	76.9	78.0	**1.3**	95699	**7599**	18865	83.5	**92.6**	87.9
YTLAB [2]	**0.89**	76.7	<u>80.2</u>	2.8	104555	16685	10009	91.3	86.2	88.7
KDNT [10]	**0.89**	67.1	80.1	4.8	<u>105473</u>	28623	<u>9091</u>	92.1	78.7	84.8
F_ViPeD_B [3]	**0.89**	-14.4	77.4	20.8	**106698**	123194	**7831**	**93.2**	46.4	62.0
GNN_SDT [7]	**0.89**	<u>78.1</u>	**81.3**	2.4	103895	14397	10669	90.7	87.8	89.2
DRT (Ours)	<u>0.88</u>	**83.2**	**81.3**	<u>1.4</u>	103909	<u>8544</u>	10655	90.7	<u>92.4</u>	**91.5**

Table 2: Comparison with state-of-the-art object detection approaches on MOT17Det. ↑/↓ indicates that higher/lower scores correspond to better performance.

2 Results on MOT17Det test set

Since our method is based on detection refinement, we also test our results on the MOT17 detection dataset, which uses the same sequences as MOT 16/17 tracking datasets. The detection results are shown in Table 2.

Compared with other state-of-the-art detectors, our DRT achieves impressive detection performance. The average precision (AP) of 0.88 is much higher than that of traditional detection methods, SDP [9] and Faster RCNN [6]. The performance is also competitive to recent detection methods in two aspects. On the one hand, it achieves a good balance between recalling as many targets as possible and introducing less false postives. On the other hand, the detection accuracy and precision are much better than other detection methods, indicating the capabilities of DRT in refining detections.

References

[1] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft. Simple Online and Realtime Tracking. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 2016.

[2] Zhaowei Cai, Quanfu Fan, Rogerio S Feris, and Nuno Vasconcelos. A Unified Multi-Scale Deep Convolutional Neural Network for Fast Object Detection. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2016.

[3] Luca Ciampi, Nicola Messina, Fabrizio Falchi, Claudio Gennaro, and Giuseppe Amato. Virtual to Real Adaptation of Pedestrian Detectors. *Sensors*, 20(18):5250, 2020.

[4] Patrick Dendorfer, Hamid Rezatofighi, Anton Milan, Javen Shi, Daniel Cremers, Ian Reid, Stefan Roth, Konrad Schindler, and Laura Leal-Taixé. MOT20: A Benchmark for Multi Object Tracking in Crowded Scenes. *arXiv preprint arXiv:2003.09003*, 2020.

[5] 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.

[6] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-cnn: Towards Real-Time Object Detection with Region Proposal Networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2015.

[7] Yongxin Wang, Xinshuo Weng, and Kris Kitani. Joint Detection and Multi-Object Tracking with Graph Neural Networks. *arXiv preprint arXiv:2006.13164*, 2020.

[8] Yihong Xu, Yutong Ban, Guillaume Delorme, Chuang Gan, Daniela Rus, and Xavier Alameda-Pineda. Transcenter: Transformers with Dense Queries for Multiple-Object Tracking. *arXiv preprint arXiv:2103.15145*, 2021.

[9] Fan Yang, Wongun Choi, and Yuanqing Lin. Exploit All the Layers: Fast and Accurate Cnn Object Detector with Scale dependent Pooling and Cascaded Rejection Classifiers. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[10] Fengwei Yu, Wenbo Li, Quanquan Li, Yu Liu, Xiaohua Shi, and Junjie Yan. Poi: Multiple Object Tracking with High Performance Detection and Appearance Feature. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2016.

[11] Yifu Zhan, Chunyu Wang, Xinggang Wang, Wenjun Zeng, and Wenyu Liu. FairMOT: On the Fairness of Detection and Re-Identification in Multiple Object Tracking. *arXiv preprint arXiv:2004.01888*, 2020.