Supplementary Material for MOTRv2: Bootstrapping End-to-End Multi-Object Tracking by Pretrained Object Detectors

Yuang Zhang[1], Tiancai Wang[2], Xiangyu Zhang[2,3]
[1]Shanghai Jiao Tong University [2]MEGVII Technology [3]Beijing Academy of Artificial Intelligence

A. Appendix

A.1. Significance Analysis of Query Denoising

To determine the effectiveness of query denoising, we conducted the corresponding ablation study (rows 1 and 3 in Table 9) four times. The results are shown in A-Table 1. The T-test result[1] indicates that the means are significantly different at $p < 0.05$. Furthermore, query denoising also helps to reduce the variance of the HOTA metric.

A-Table 1. Comparison of multiple repeated HOTA results on DanceTrack validation set with and without Query Denoising.

QD	Run1	Run2	Run3	Run4	Mean	STD
	63.51	64.13	62.28	63.70	63.41	0.69
✓	64.50	64.64	63.68	64.50	64.33	0.38

A.2. Effect of the Number of Extra Detect Queries

We concatenate a small set of learnable detect queries to YOLOX proposal queries and track queries to recall those objects missed by the YOLOX detector and avoid the boundary case of having no query at all. We find that adding 10 additional learnable anchors maximizes detection and tracking performance (see A-Table 2). Among all MOTR predictions, the extra detect queries account for the smallest fraction. Most of the predictions are from track queries since most objects are tracked throughout the video.

A-Table 2. Effect of the number of extra detect queries (#Q).

#Q	HOTA	DetA	AssA
1	63.2	77.2	51.9
5	64.1	77.9	53.0
10	**64.5**	**78.7**	**53.0**
50	63.6	77.0	52.8
100	62.9	78.1	50.9

[1] We used the tool in https://www.mathportal.org/calculators/statistics-calculator/t-test-calculator.php to calculate it.