Supplementary Material

Yanru Huang[1] Feiyu Zhu[2] Zheni Zeng[1*] Xi Qiu[2] Yuan Shen[1†] Jianan Wu[2]
[1]Tsinghua University [2]Megvii Inc.
{huangyr18, zengzn16}@mails.tsinghua.edu.cn shenyuan_ee@tsinghua.edu.cn
{zhufeiyu, qiuxi, wjn}@megvii.com

In this supplementary material we provide further details and additional results of our self quality evaluation metric SQE on MOT16 Challenge data sets.

To illustrate the effectiveness of our self-evaluation metric upon measuring tracking performance, we compare its score with commonly used ID-based metrics and CLEAR MOT metrics. Take MOT16-02 for example, Table 1 gives the detailed numerical comparison results, and the corresponding visualization results have been shown in the paper. For simplification, the results on the remaining 5 videos are visualized in Figure 1. Although some local ambiguity exists, the overall trends of the left and right images are consistent. The corresponding optimal value comparison results have been summarized in the paper.

Parm	SQE	IDF_1	IDP	IDR	MOTA
0.3	5.8	4.6	86.2	2.4	2.7
0.4	14.0	23.4	95.7	13.4	13.9
0.5	22.2	43.7	91.2	28.7	30.4
0.6	28.2	52.3	84.6	37.8	42.2
0.7	33.3	56.2	81.4	42.9	48.8
0.8	**36.1**	**58.3**	79.3	46.0	51.9
0.9	35.7	57.9	75.5	47.0	53.8
1.0	35.7	56.5	72.7	46.1	51.7
1.1	35.5	55.8	69.5	46.6	51.7
1.2	30.5	49.8	59.2	43.0	48.3
1.3	26.0	37.6	45.0	32.2	43.9
1.4	21.6	31.2	38.7	26.2	39.7
1.5	21.1	31.9	39.7	26.7	38.5
1.6	21.1	31.9	39.7	26.7	38.5

Table 1. Comparison of SQE and other commonly used supervised metrics on MOT16-02 when changing the REID threshold.

Similarly, the detailed numerical comparison results on different tracking algorithms and different parameters are also given in Table 2 and 3 respectively, so as to better explain the visualization results in the paper. The additional results when changing the merging threshold on the remaining 5 videos are visualized in Figure 2. Although the overall similarity is not as promising as the one in Figure 1, the optimal values are more accurate as demonstrated in the paper.

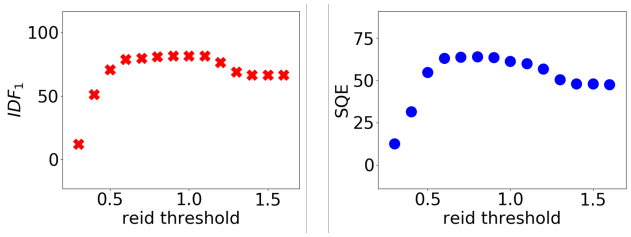
Parm	SQE	IDF_1	IDP	IDR	MOTA
0.3	25.6	44.6	64.3	34.3	46.5
0.325	26.7	44.2	63.2	34	47
0.35	26.5	44.5	63.3	34.3	47.2
0.375	27.0	44.7	63.1	34.6	47.6
0.4	30.1	45.5	63.3	35.5	47.7
0.425	30.1	46.4	64.6	36.2	47.9
0.45	**31.1**	**46.8**	64.5	36.7	47.8
0.475	30.1	45.1	62.1	35.4	47.7
0.5	30.0	45.7	62.4	36	47.9

Table 2. Comparison of SQE and other commonly used supervised metrics on MOT16-02 when changing the matching cosine threshold in Deep SORT algorithm.

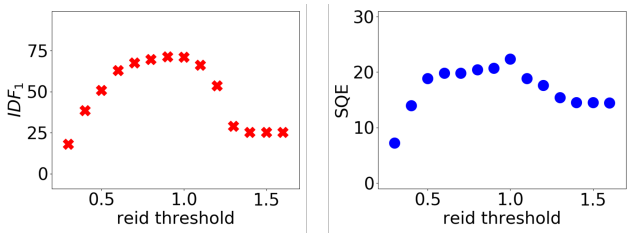
Parm	SQE	IDF_1	IDP	IDR	MOTA
0.5	13.0	49.2	65.3	39.5	49.4
0.6	13.4	49.4	65.6	39.6	49.5
0.7	14.1	50.0	66.7	40.0	49.9
0.8	15.2	51.3	68.6	41.0	50.6
0.9	15.8	55.1	72.7	44.4	52.5
1.0	**16.7**	**57.9**	72.7	46.1	51.7
1.1	16.2	56.8	72.4	46.7	52.5
1.2	14.0	51.2	64.2	42.6	49.7
1.3	12.8	44.9	56.1	37.4	47.5
1.4	13.5	42.2	53.0	35.0	47.3
1.5	13.4	42.2	53.0	35.0	47.3

Table 3. Comparison of SQE and other commonly used supervised metrics on MOT16-02 when changing the merging threshold.

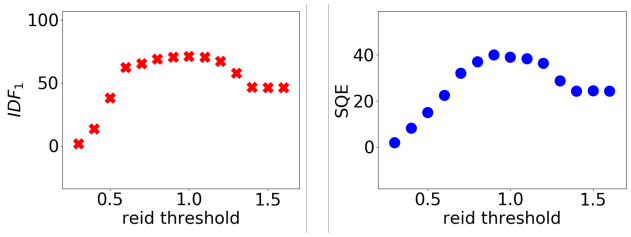
*This work was done during Zheni Zeng was an intern at Megvii Inc.
†Corresponding author.



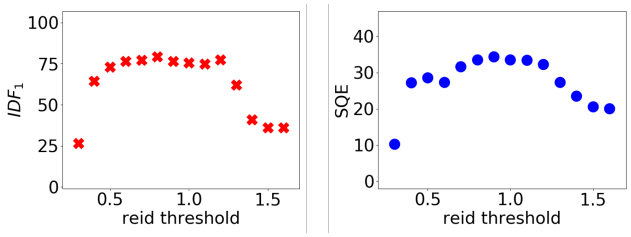
(a) 04



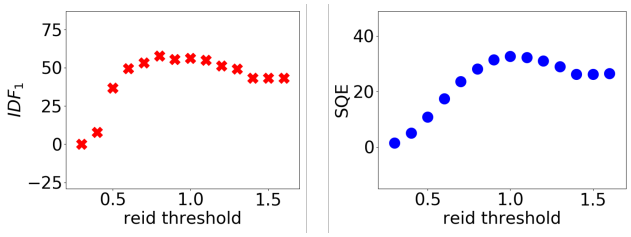
(b) 05



(c) 10

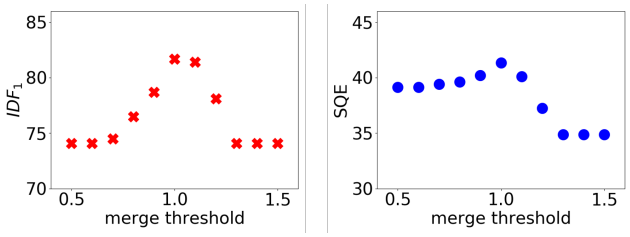


(d) 11

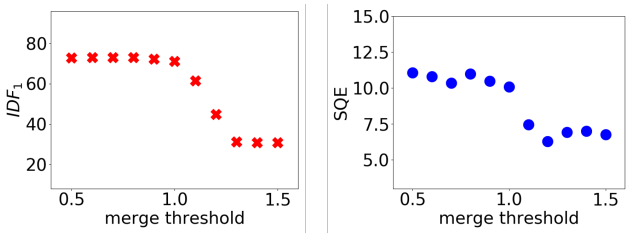


(e) 13

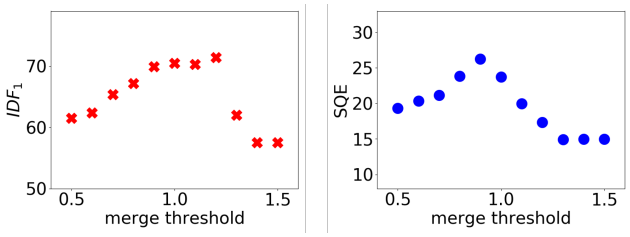
Figure 1. Visualization of IDF$_1$ and SQE on MOT16-{04, 05, 10, 11, 13} when changing the REID threshold.



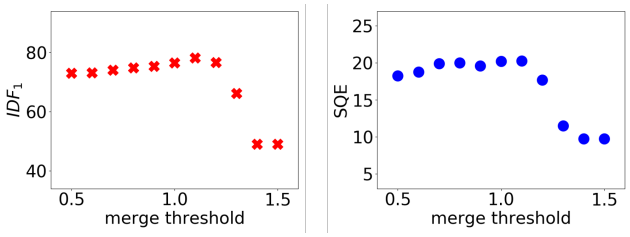
(a) 04



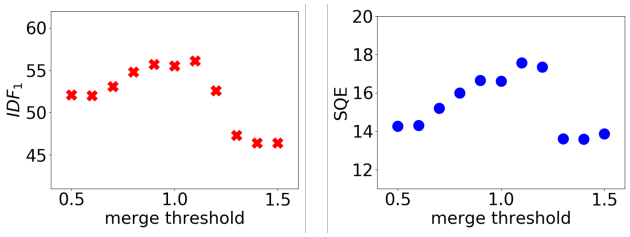
(b) 05



(c) 10



(d) 11



(e) 13

Figure 2. Visualization of IDF$_1$ and SQE on MOT16-{04, 05, 10, 11, 13} when changing the merging threshold.