Supplementary Material for Enhancing Robustness of Multi-Object Trackers with Temporal Feature Mix

Kyujin Shim, Junyoung Byun, Kangwook Ko, Jubi Hwang, Changick Kim

1 Supplementary Ablation Studies

In this section, we show the experimental results of ablation studies about p and r_{max} with five trackers on our validation split through Table 1 - 5. During the ablative studies for fraction p and maximum mixing ratio r_{max}, maximum temporal distance d_{max} is set to 20 for every case. Also, with the results of the ablative studies, we set p and r_{max} for each tracker by considering the overall metric scores.

2 Quantitative Results on DanceTrack

In this section, we show the additional quantitative results of our proposed method, Temporal Feature Mix (TFM), compared to two baselines, Manifold Mixup (MM) [1] and Noisy Feature Mixup (NFM) [2], on the DanceTrack [3] validation and test sets with three MOT trackers, SORT [4], OC-SORT [5], and BoT-SORT [6]. As in Table 6 and 7, integrating TFM led to notable enhancement in most cases compared to the baselines and those that adopted Manifold Mixup (MM) or Noisy Feature Mixup (NFM) on both the validation and test sets. Although BoT-SORT with TFM shows slightly degraded performance on the test set, it still highly surpasses two comparable methods (MM and NFM), and we expect a more detailed hyper-parameters setting would smoothly solve the issue.

3 Corruption Robustness

In this section, we show the results of the robustness enhancement for each corruption type with each tracker, where the HOTA values of five severity levels are averaged for each case, through Fig. 1 - 5. The results show that our TFM better improves performance in most cases compared to the other methods. As we can see, Manifold Mixup (MM) and Noisy Feature Mixup (NFM) rather degrade the averaged HOTA scores in many cases, although the NFM method is effective for noise-type corruptions. On the contrary, our TFM consistently improves each tracking algorithm in most cases.

Table 1: An ablation study about p and r_{max} with SORT [4] and our validation split.

SORT									
p	r_{max}	HOTA↑	MOTA↑	IDF1↑	DetA↑	rHOTA↑	rMOTA↑	rIDF1↑	rDetA↑
0.05	0.05	65.8	**82.2**	85.7	65.3	51.9	50.2	54.5	53.1
0.05	0.10	65.5	81.8	85.1	65.1	**52.1**	**50.4**	**54.7**	**53.4**
0.05	0.15	**65.9**	**82.2**	85.9	**65.4**	51.9	50.2	54.5	53.1
0.05	0.20	**65.9**	82.0	**86.0**	65.3	51.7	50.1	54.3	53.0
0.10	0.05	65.7	**82.2**	85.5	65.2	52.0	50.2	54.6	53.2
0.10	0.10	65.8	82.1	85.6	**65.4**	51.8	50.2	54.4	53.1
0.10	0.15	65.5	81.9	85.1	65.3	51.4	49.7	54.1	52.6
0.10	0.20	65.6	82.0	85.4	65.3	51.7	50.0	54.4	53.0
0.15	0.05	65.6	81.9	85.4	65.2	51.8	50.0	54.4	52.9
0.15	0.10	65.6	81.8	85.4	65.2	52.0	50.3	54.6	53.2
0.15	0.15	65.8	82.1	85.7	65.3	51.7	50.0	54.4	52.9
0.15	0.20	65.8	82.0	85.6	65.3	51.1	49.4	53.8	52.3
0.20	0.05	65.6	82.0	85.3	65.2	52.0	50.3	54.5	53.2
0.20	0.10	65.5	82.1	85.4	65.2	52.0	**50.4**	54.5	53.3
0.20	0.15	65.8	82.0	85.5	65.3	51.3	49.5	54.0	52.4
0.20	0.20	65.4	81.8	85.3	65.1	51.0	49.3	53.7	52.2

Table 2: An ablation study about p and r_{max} with DeepSORT [7] and our validation split.

DeepSORT									
p	r_{max}	HOTA↑	MOTA↑	IDF1↑	DetA↑	rHOTA↑	rMOTA↑	rIDF1↑	rDetA↑
0.05	0.05	60.9	81.0	80.3	63.2	46.4	50.3	43.5	55.3
0.05	0.10	60.8	80.7	80.2	63.1	46.6	**50.5**	43.7	**55.5**
0.05	0.15	**61.1**	81.0	**80.7**	63.2	46.5	50.3	43.7	55.3
0.05	0.20	60.8	81.0	80.	63.1	46.4	50.3	43.4	55.2
0.10	0.05	60.7	**81.1**	79.9	**63.3**	**46.7**	50.4	**43.8**	55.3
0.10	0.10	60.3	80.7	79.5	63.1	46.4	50.3	43.5	55.3
0.10	0.15	60.3	80.8	79.5	63.1	46.0	49.9	43.0	54.8
0.10	0.20	60.8	80.9	80.3	63.1	46.3	50.2	43.5	55.1
0.15	0.05	60.8	80.9	80.3	63.0	46.5	50.3	43.5	55.1
0.15	0.10	60.2	80.6	79.3	63.0	46.5	50.4	43.6	55.4
0.15	0.15	60.6	80.7	80.0	63.1	46.3	50.2	43.4	55.1
0.15	0.20	60.6	80.7	79.9	63.0	45.7	49.7	43.0	54.5
0.20	0.05	60.5	80.7	79.6	62.9	46.6	**50.5**	43.7	**55.5**
0.20	0.10	60.6	80.9	80.0	63.2	46.5	**50.5**	43.6	**55.5**
0.20	0.15	60.5	80.8	79.8	63.1	45.9	49.8	43.1	54.7
0.20	0.20	60.6	80.5	80.2	63.0	45.6	49.6	42.7	54.5

Table 3: An ablation study about p and r_{max} with ByteTrack [8] and our validation split.

ByteTrack									
p	r_{max}	HOTA↑	MOTA↑	IDF1↑	DetA↑	rHOTA↑	rMOTA↑	rIDF1↑	rDetA↑
0.05	0.05	**64.5**	**82.3**	84.8	**64.8**	51.7	52.2	51.9	56.7
0.05	0.10	64.1	81.8	84.1	64.6	**51.9**	**52.4**	**52.1**	**56.9**
0.05	0.15	64.3	82.1	84.5	64.7	51.7	52.1	52.0	56.6
0.05	0.20	64.2	81.9	84.3	64.7	51.6	52.1	51.8	56.6
0.10	0.05	64.2	**82.3**	84.2	64.7	51.8	52.2	52.0	56.7
0.10	0.10	64.0	81.9	83.8	64.7	51.6	52.1	51.8	56.7
0.10	0.15	64.4	81.9	84.5	64.7	51.2	51.7	51.3	56.2
0.10	0.20	64.2	82.0	84.5	64.7	51.5	52.0	51.9	56.5
0.15	0.05	64.2	82.0	84.3	64.6	51.7	52.1	51.9	56.5
0.15	0.10	64.4	82.0	84.6	64.7	51.7	52.3	51.8	56.8
0.15	0.15	64.2	81.9	84.4	64.5	51.5	52.0	51.8	56.5
0.15	0.20	64.3	81.8	84.5	64.6	50.9	51.4	51.2	55.9
0.20	0.05	64.1	81.7	84.2	64.6	51.8	52.3	51.9	**56.9**
0.20	0.10	**64.5**	82.0	**84.9**	64.5	51.7	52.3	51.8	**56.9**
0.20	0.15	64.3	81.9	84.7	64.6	51.1	51.6	51.4	56.1
0.20	0.20	64.0	81.6	84.1	64.4	50.8	51.4	51.1	55.9

Table 4: An ablation study about p and r_{max} with OC-SORT [5] and our validation split.

OC-SORT									
p	r_{max}	HOTA↑	MOTA↑	IDF1↑	DetA↑	rHOTA↑	rMOTA↑	rIDF1↑	rDetA↑
0.05	0.05	65.8	**83.0**	85.2	65.9	52.6	52.4	53.5	56.3
0.05	0.10	65.8	82.7	85.1	65.8	**52.9**	**52.7**	**53.8**	**56.6**
0.05	0.15	66.0	**83.0**	85.5	65.8	52.7	52.4	53.7	56.2
0.05	0.20	65.8	82.8	85.3	65.8	52.5	52.4	53.4	56.2
0.10	0.05	65.5	**83.0**	84.9	65.8	52.8	52.5	53.7	56.3
0.10	0.10	**66.1**	82.9	**85.6**	65.9	52.5	52.4	53.4	56.3
0.10	0.15	65.7	82.6	84.9	65.8	52.1	51.9	53.1	55.7
0.10	0.20	65.8	82.8	85.0	65.9	52.4	52.2	53.4	56.0
0.15	0.05	65.9	82.8	85.4	65.9	52.6	52.4	53.5	56.1
0.15	0.10	65.6	82.7	84.7	65.7	52.7	52.5	53.7	56.2
0.15	0.15	65.7	82.9	84.9	65.7	52.4	52.2	53.5	55.9
0.15	0.20	65.5	82.7	84.7	65.8	51.8	51.6	52.9	55.3
0.20	0.05	65.9	82.7	85.2	**66.0**	52.7	52.5	53.6	56.4
0.20	0.10	65.5	82.8	84.7	65.7	52.7	52.5	53.5	56.4
0.20	0.15	65.5	82.6	84.7	65.8	52.0	51.8	53.1	55.5
0.20	0.20	65.2	82.4	84.3	65.4	51.7	51.6	52.7	55.3

Table 5: An ablation study about p and r_{max} with BoT-SORT [6] and our validation split.

BoT-SORT									
p	r_{max}	HOTA↑	MOTA↑	IDF1↑	DetA↑	rHOTA↑	rMOTA↑	rIDF1↑	rDetA↑
0.05	0.05	65.0	83.3	83.9	66.1	52.2	53.2	51.9	57.7
0.05	0.10	65.1	83.0	83.8	65.9	**52.4**	**53.3**	**52.1**	**57.9**
0.05	0.15	65.5	83.3	84.6	**66.2**	52.2	53.1	52.0	57.7
0.05	0.20	**65.6**	83.2	**84.9**	66.1	52.1	53.1	51.8	57.6
0.10	0.05	65.2	**83.4**	84.0	**66.2**	52.3	53.2	52.0	57.7
0.10	0.10	65.2	83.1	83.9	**66.2**	52.0	53.1	51.6	57.7
0.10	0.15	65.1	83.0	83.8	66.0	51.6	52.7	51.3	57.2
0.10	0.20	65.5	83.1	84.6	**66.2**	52.0	53.0	51.8	57.5
0.15	0.05	65.2	83.1	84.0	66.1	52.2	53.1	51.9	57.5
0.15	0.10	65.3	83.1	84.0	66.1	52.2	**53.3**	51.8	57.8
0.15	0.15	65.4	83.2	84.5	66.1	52.0	53.0	51.9	57.5
0.15	0.20	64.9	82.9	83.6	66.0	51.4	52.4	51.3	56.9
0.20	0.05	65.0	82.8	83.7	65.9	52.3	**53.3**	52.0	**57.9**
0.20	0.10	65.3	83.3	84.3	66.0	52.2	**53.3**	51.8	**57.9**
0.20	0.15	65.1	82.9	84.0	66.0	51.6	52.6	51.4	57.0
0.20	0.20	65.0	82.8	83.9	65.9	51.3	52.4	51.2	56.9

Table 6: Evaluation results on the DanceTrack [3] validation set with three trackers. MM and NFM denote Manifold Mixup [1] and Noisy Feature Mixup [2], respectively.

DanceTrack Val					
Tracker	HOTA↑	MOTA↑	IDF1↑	DetA↑	AssA↑
SORT [4]	47.4	87.2	49.2	**72.8**	31.0
SORT [4] + MM [1]	45.7	84.1	47.7	70.3	29.8
SORT [4] + NFM [2]	47.5	86.2	49.4	71.9	31.5
SORT [4] + TFM (ours)	**49.1**	**87.4**	**50.9**	**72.8**	**33.3**
OC-SORT [5]	50.4	85.5	**49.1**	75.3	33.9
OC-SORT [5] + MM [1]	49.2	**86.4**	47.5	**76.0**	32.0
OC-SORT [5] + NFM [2]	49.5	86.2	48.4	75.8	32.4
OC-SORT [5] + TFM (ours)	**50.8**	85.9	**49.1**	75.8	**34.3**
BoT-SORT [6]	53.4	85.9	55.2	74.9	38.3
BoT-SORT [6] + MM [1]	54.0	86.2	56.5	74.7	39.2
BoT-SORT [6] + NFM [2]	53.7	**87.0**	55.7	75.4	38.4
BoT-SORT [6] + TFM (ours)	**55.4**	86.7	**57.2**	**75.5**	**40.8**

Table 7: Evaluation results on the DanceTrack [3] test set with three trackers. MM and NFM denote Manifold Mixup [1] and Noisy Feature Mixup [2], respectively.

DanceTrack Test					
Tracker	HOTA↑	MOTA↑	IDF1↑	DetA↑	AssA↑
SORT [4]	48.4	89.9	49.9	75.0	31.4
SORT [4] + MM [1]	46.9	88.3	48.6	73.7	29.9
SORT [4] + NFM [2]	46.5	89.0	48.3	73.9	29.4
SORT [4] + TFM (ours)	**48.8**	**90.0**	**50.4**	**75.1**	**31.8**
OC-SORT [5]	51.8	89.4	50.3	78.7	34.3
OC-SORT [5] + MM [1]	50.8	89.3	48.4	78.3	33.0
OC-SORT [5] + NFM [2]	51.6	**89.8**	50.4	**78.8**	33.9
OC-SORT [5] + TFM (ours)	**52.6**	89.1	**50.9**	**78.8**	**35.3**
BoT-SORT [6]	**58.1**	90.0	**60.0**	**79.1**	**42.8**
BoT-SORT [6] + MM [1]	55.0	89.5	56.4	78.0	38.9
BoT-SORT [6] + NFM [2]	57.1	**90.4**	58.6	78.5	41.7
BoT-SORT [6] + TFM (ours)	57.9	90.2	59.5	78.8	42.7

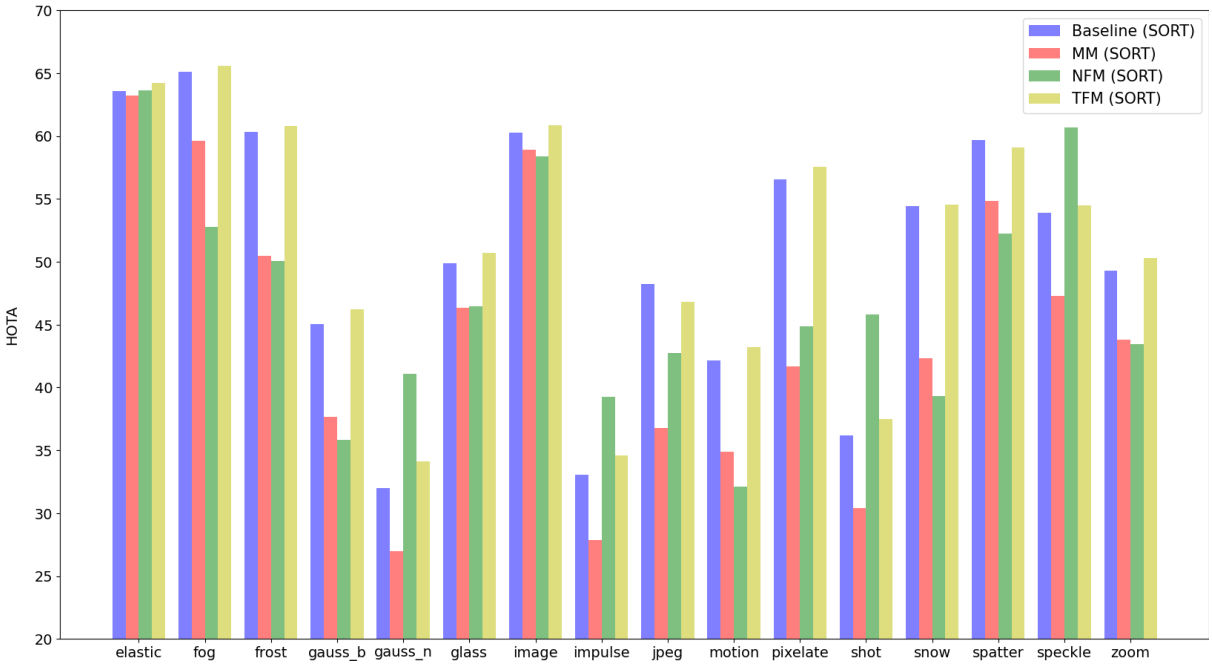

Figure 1: HOTA performances for each corruption type with SORT [4].

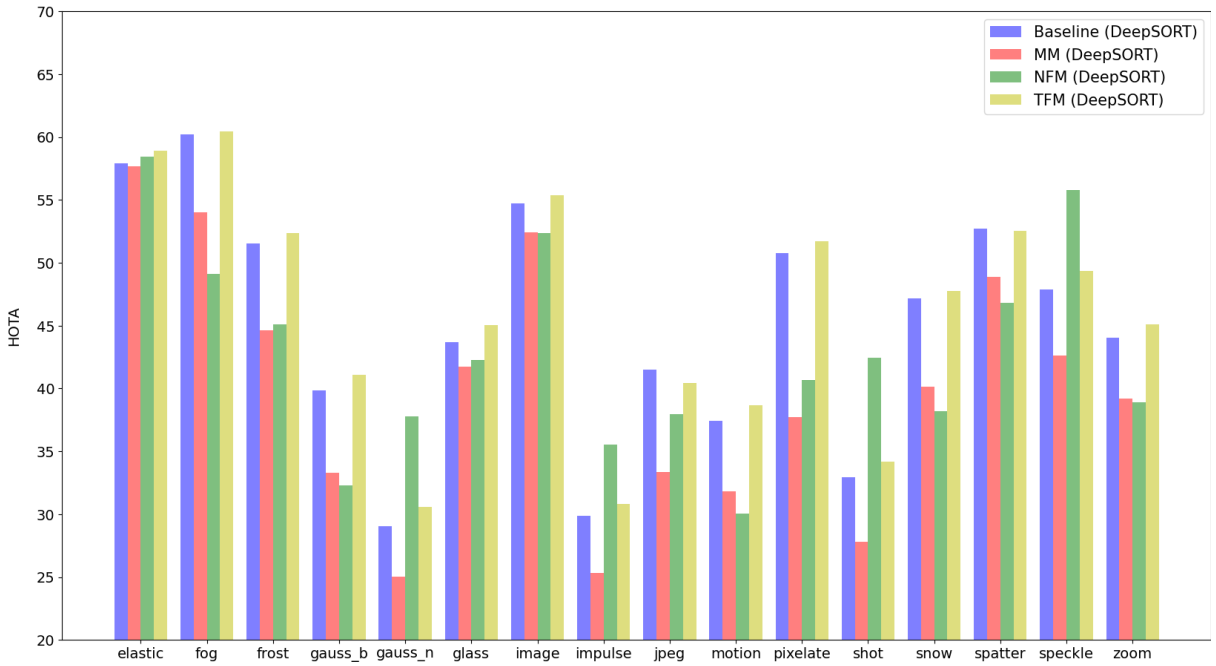

Figure 2: HOTA performances for each corruption type with DeepSORT [7].

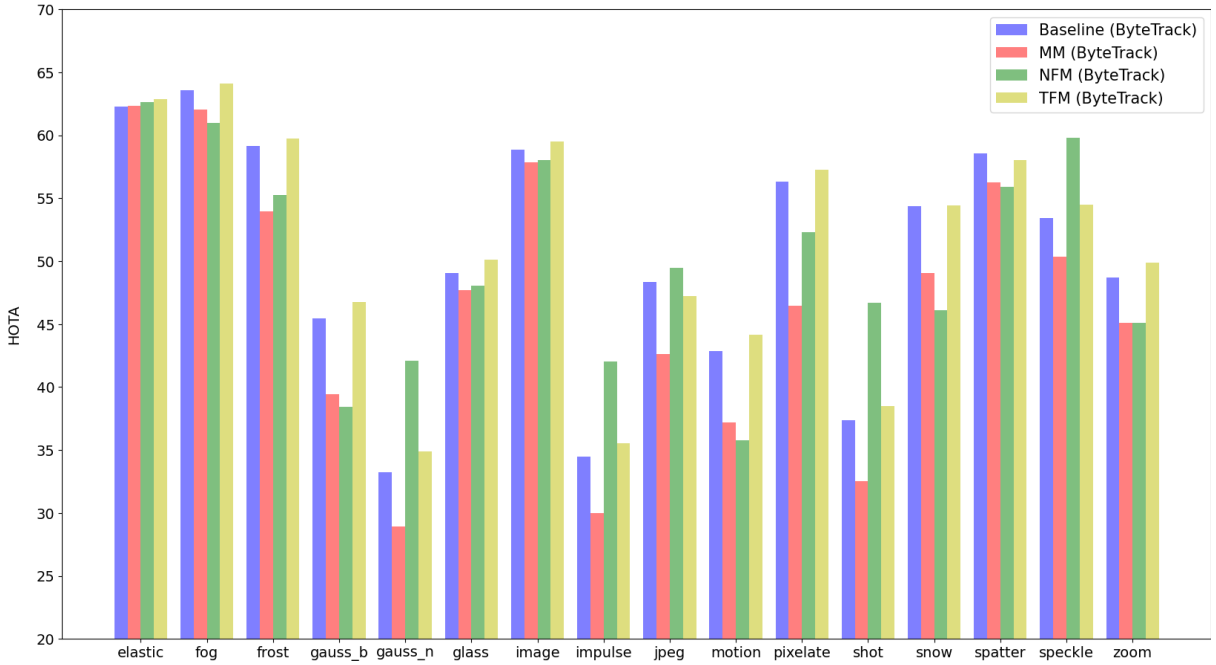

Figure 3: HOTA performances for each corruption type with ByteTrack [8].

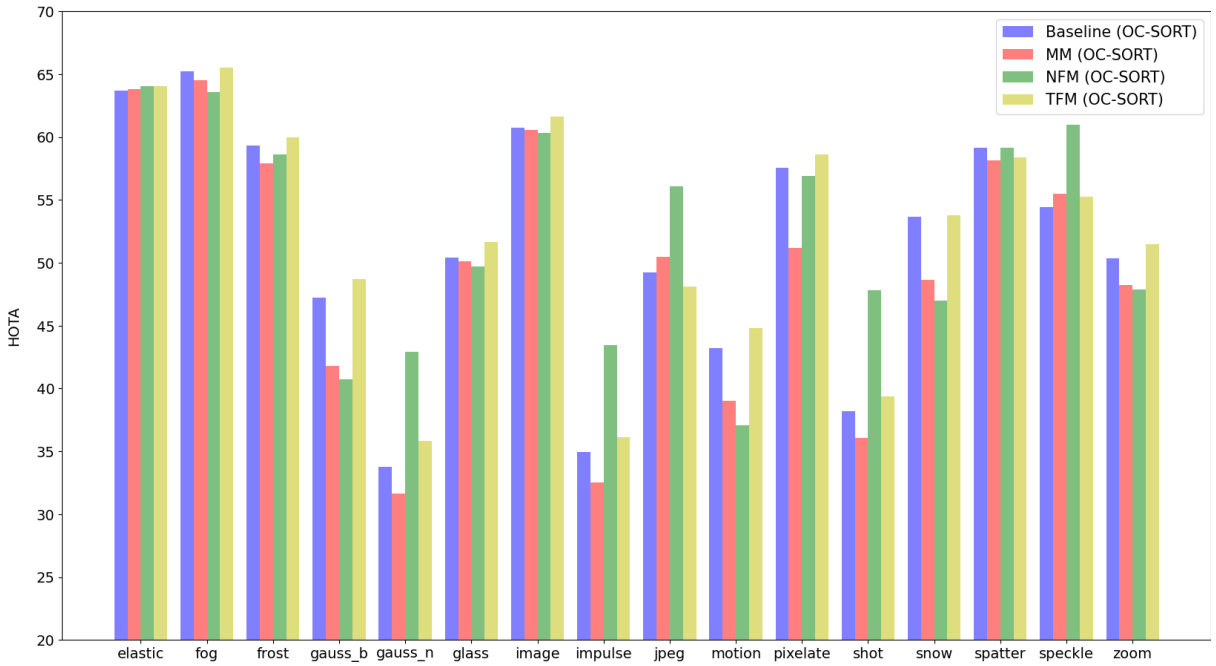

Figure 4: HOTA performances for each corruption type with OC-SORT [5].

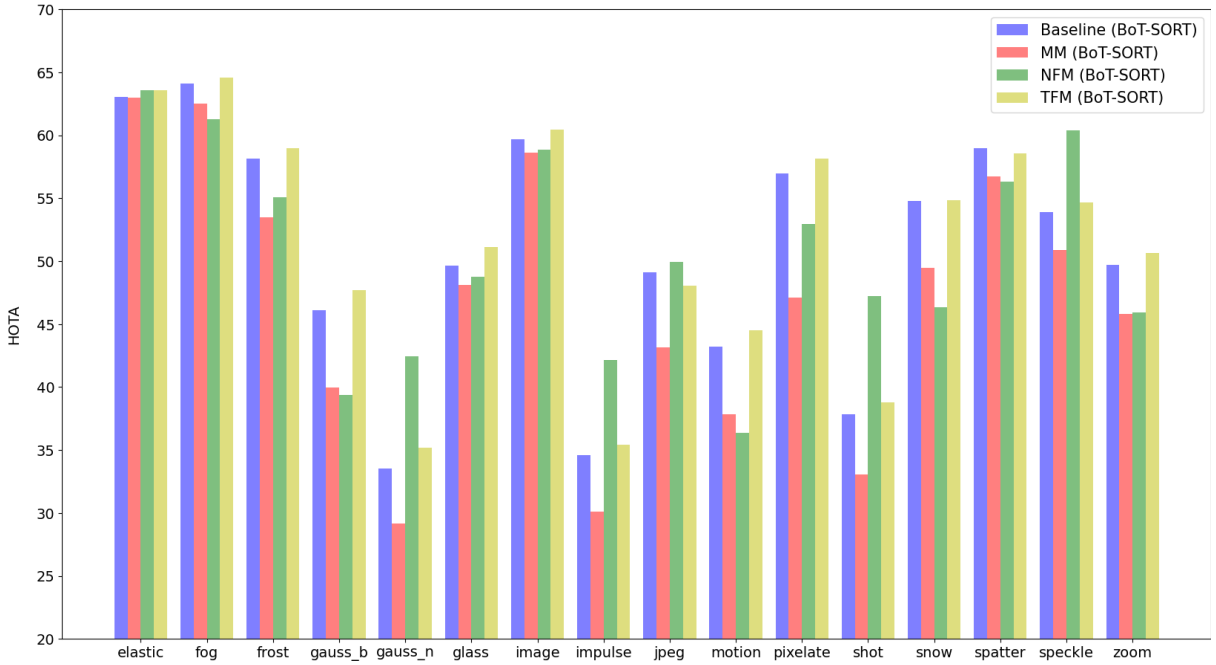

Figure 5: HOTA performances for each corruption type with BoT-SORT [6].

References

[1] V. Verma, A. Lamb, C. Beckham, A. Najafi, I. Mitliagkas, D. Lopez-Paz, and Y. Bengio, "Manifold mixup: Better representations by interpolating hidden states," in *Int. conf. Mach. Learn.*, vol. 97, 2019, pp. 6438–6447.

[2] S. H. Lim, N. B. Erichson, F. Utrera, W. Xu, and M. W. Mahoney, "Noisy feature mixup," in *Int. Conf. Learn. Represent.*, 2022.

[3] P. Sun, J. Cao, Y. Jiang, Z. Yuan, S. Bai, K. Kitani, and P. Luo, "Dancetrack: Multi-object tracking in uniform appearance and diverse motion," in *Conf. Comput. Vis. Pattern Recog.*, 2022.

[4] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *Int. Conf. Image Process.*, 2016, pp. 3464–3468.

[5] J. Cao, J. Pang, X. Weng, R. Khirodkar, and K. Kitani, "Observation-centric sort: Rethinking sort for robust multi-object tracking," in *Conf. Comput. Vis. Pattern Recog.*, 2023, pp. 9686–9696.

[6] N. Aharon, R. Orfaig, and B.-Z. Bobrovsky, "Bot-sort: Robust associations multi-pedestrian tracking," *arXiv preprint*, vol. arXiv:2206.14651, 2022.

[7] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *Int. Conf. Image Process.*, 2017, pp. 3645–3649.

[8] Y. Zhang, P. Sun, Y. Jiang, D. Yu, F. Weng, Z. Yuan, P. Luo, W. Liu, and X. Wang, "Bytetrack: Multi-object tracking by associating every detection box," in *Eur. Conf. Comput. Vis.*, 2022, pp. 1–21.