MOS-Attack: A Scalable Multi-objective Adversarial Attack Framework

Supplementary Material

Table 7. **Overall Results.** A comparative analysis of attack success rate among MOS-8 attacks with APGD-CE under large magnitude of attack budgets. For MOS-8 Attack, we record its K value, while for others it denoted the number of restarts. The optimal outcome is highlighted in bold and marked with a grey background.

			Attack Success Rate			
CIFAR-10 ($\epsilon = 16/255$)			APGD	**MOS-8**	**MOS-8**	**Diff.(5)**
ID	**Paper**	**Architecture**	(5)	**(1)**	**(5)**	**MOS$\mid$CE**
0	Rade *et al.* (2022) [36] (*ddpm*)	PreActResNet-18	74.30	76.38	**76.80**	+2.50
1	Rade *et al.* (2022) [36] (*extra*)	PreActResNet-18	83.13	82.48	**83.36**	+0.23
2	Sehwag *et al.* (2022) [40]	ResNet-18	**77.51**	76.79	77.37	-0.14
3	Chen *et al.* (2020) [9]	ResNet-50	81.66	82.29	**82.40**	+0.74
4	Gowal *et al.* (2020) [22]	WideResNet-28-10	71.36	72.91	**73.51**	+2.15
5	Wang *et al.* (2023) [47]	WideResNet-28-10	71.12	72.12	**72.67**	+1.55
6	Rebuffi *et al.* (2021) [37]	WideResNet-28-10	70.86	72.39	**72.83**	+1.97
7	Sehwag *et al.* (2022) [40]	WideResNet-34-10	72.45	72.43	**72.95**	+0.5
8	Rade *et al.* (2022) [36]	WideResNet-34-10	75.59	75.76	**76.21**	+0.62
9	Gowal *et al.* (2021) [23]	WideResNet-70-16	66.20	65.98	**66.40**	+0.20
10	Gowal *et al.* (2020) [22]	WideResNet-70-16	70.84	71.22	**71.67**	+0.83
11	Rebuffi *et al.* (2021) [37]	WideResNet-70-16	73.02	73.43	**73.98**	+0.96
	Average Rank		2.58	2.33	1.1	
ImageNet ($\epsilon = 8/255$)						
12	Salman *et al.* (2020) [39]	ResNet-18	89.02	90.74	**90.90**	+1.88
13	Salman *et al.* (2020) [39]	ResNet-50	85.06	85.36	**85.84**	+0.78
14	Wong *et al.* (2020) [49]	ResNet-50	89.64	90.52	**90.64**	+1.00
15	Engstrom *et al.* (2019) [17]	ResNet-50	89.74	90.24	**90.42**	+0.68
16	Salman *et al.* (2020) [39]	WideResNet-50-2	84.80	84.96	**85.16**	+0.36
	Average Rank		3.00	2.00	1.00	