Attentively Fabricate and Erase: A Pervasive Black-box Adversarial Attack with Cross-task Transferability

Surrogate	mAP@50↓ Attack Mtd.	FasterRCNN Res50FPN	FasterRCNN Mobv3L320	MaskRCNN Res50FPN
VGG16	DR (N=100, α=4)	6.57	6.8	7.73
	DR (N=40, α=2)	6.47	7.28	7.94
	FIA (N=10, α=1.6)	8.71	8.15	9.74
	FIA (N=40, α=2)	7.01	6.99	7.84
	AF&E(ours)	**5.27**	**5.07**	**6.18**
Res152	DR (N=100, α=4)	10.2	**7.04**	11.69
	DR (N=40, α=2)	12.69	7.22	12.42
	FIA (N=10, α=1.6)	11.59	10.81	12.21
	FIA (N=40, α=2)	**9.66**	9.46	**10.25**
	AF&E(ours)	11.33	7.71	12.45
Mobile$_v$2	DR (N=100, α=4)	11.95	8.15	13.12
	DR (N=40, α=2)	12.69	8.07	13.8
	FIA (N=10, α=1.6)	18.46	17.49	19.14
	FIA (N=40, α=2)	15.26	15.78	16.6
	AF&E(ours)	**10.48**	**6.57**	**12.35**

Table 1. **Object Detection** Results on MS COCO dataset (lower mAP indicates better attack performance).

Surrogate	mIoU% ↓ Attack Mtd.	FCN Res101	Deeplabv3 Mobilev3L	LRASPP Mobilev3L
VGG16	DR (N=100, α=4)	12.92	12.95	10.65
	DR (N=40, α=2)	13.06	13.85	11.44
	FIA (N=10, α=1.6)	13.65	11.09	9.67
	FIA (N=40, α=2)	11.79	9.35	8.22
	AF&E(ours)	**10.5**	**6.54**	**6.57**
Res152	DR (N=100, α=4)	16.35	16.04	12.33
	DR (N=40, α=2)	16.81	16.23	12.34
	FIA (N=10, α=1.6)	17.83	16.34	13.99
	FIA (N=40, α=2)	**15.81**	13.63	11.86
	AF&E(ours)	15.96	**11.83**	**11.26**
Mobile$_v$2	DR (N=100, α=4)	16.46	13.84	10.39
	DR (N=40, α=2)	16.68	15.73	11.74
	FIA (N=10, α=1.6)	22.25	24.82	23.02
	FIA (N=40, α=2)	19.4	21	18.63
	AF&E(ours)	**14.3**	**9.54**	**8.63**

Table 2. **Semantic Segmentation** Results on MS COCO dataset(lower mIoU indicates better attack performance).

Surrogate	Acc% ↓ Attack Mtd.	Inc-v3 ens3	Inc-v3 ens4	IncRes-v2 ens
VGG16	DR (N=100, α=4)	68.7	72.18	76.54
	DR (N=40, α=2)	67.72	71.28	75.26
	FIA (N=10, α=1.6)	60.49	61.33	69.69
	FIA (N=40, α=2)	46.94	46.86	56.31
	AF&E(ours)	**42.20**	**44.82**	**49.84**
Res152	DR (N=100, α=4)	66	68.88	73.18
	DR (N=40, α=2)	65.36	68.58	72.3
	FIA (N=10, α=1.6)	71.69	72.22	79
	FIA (N=40, α=2)	56.9	56.51	63.47
	AF&E(ours)	**33.22**	**35.24**	**38.26**
Mobile$_v$2	DR (N=100, α=4)	72.88	76.5	80.5
	DR (N=40, α=2)	73.5	76.14	80.14
	FIA (N=10, α=1.6)	82.47	84.29	88.18
	FIA (N=40, α=2)	74.33	74.55	79.8
	AF&E(ours)	**37.16**	**40.52**	**43.84**

Table 3. Results on adversarially robust trained classifiers (lower accuracy indicates better attack performance).