Enhancing the Transferability of Adversarial Attacks with Stealth Preservation

Supplementary Material

1. Model Ensemble Algorithm

The *MMIA* algorithm for an ensemble of models is summarized in Algorithm 1.

Algorithm 1 MMIA Algorithm for an ensemble of models

Input: K Classifiers with loss function $J_0, J_1, \ldots, J_{k-1}$, real example $\boldsymbol{x}$, ground-truth label y, outer iterations t, inner iterations n, transformation probability p, perturbation size β, max perturbation ε, perturbation step size α, decay factor μ;

Output: Adversarial example $\boldsymbol{x^*}$;

Init transformations function $T(\cdot)$;
$Grad_0 = 0$; $\boldsymbol{x^*} = \boldsymbol{x}$;
for $i = 0$ to $t - 1$ **do**
  for $t = 0$ to $k - 1$ **do**
    for $j = 0$ to $n - 1$ **do**
      $\boldsymbol{x^*_{itj}} = T(\boldsymbol{x^*_{it}})$ with the probability p
      Get gradient G_j of $\boldsymbol{x^*_{itj}}$ by Eq **??**.
    end for
    Obtain spatial momentum gradient g^s_{it} by

$$g^s_{it} = \frac{1}{n} \sum G_j$$

    $\boldsymbol{Mask_{it}} = SearchMask(g^s_{it}, \beta)$;
  end for
  Get $\boldsymbol{sumGrad_i}$ by Eq **??**.
  Update $Grad_{i+1}$ by

$$Grad_{i+1} = \mu Grad_i + \frac{sumGrad_i}{\|sumGrad_i\|_1}$$

  Update $\boldsymbol{x^*_{i+1}}$ by

$$\boldsymbol{x}^*_{i+1} = \boldsymbol{x}^*_i + \alpha \cdot sign(\boldsymbol{Grad}_{i+1})$$

  Clip $\boldsymbol{x^*_{i+1}}$ according to ε;
end for
return $\boldsymbol{x^*} = \boldsymbol{x^*_t}$

2. Complete Results

Table 2.1 and Table 2.2 show the complete results on the CIFAR-10 dataset and ImageNet dataset, respectively

Table 2.1. The attack success and stealth score (AS^3) (%) of non-targeted adversarial attacks against 6 models on CIFAR-10 dataset. * represent the white-box attacks.

	Attack	Inception v3	ResNet50	VGG16	MobileNet	DenseNet	GoogleNet
Inception v3	FGSM	58.72*	43.84	57.74	64.01	52.39	59.08
	MI-FGSM	76.55*	55.03	66.72	67.56	57.79	67.56
	SMI-FGSM	75.66*	54.58	68.12	66.41	55.78	65.42
	MMIA	**78.73***	**65.01**	**68.13**	**69.15**	**64.30**	**70.20**
ResNet50	FGSM	58.09	51.37*	54.74	62.49	50.13	56.24
	MI-FGSM	75.62	85.63*	70.33	71.57	77.23	70.10
	SMI-FGSM	76.05	86.10*	72.36	73.31	78.39	69.67
	MMIA	**80.98**	**87.10***	**80.60**	**77.52**	**83.72**	**75.94**
VGG16	FGSM	57.48	42.03	48.18*	61.05	46.47	53.17
	MI-FGSM	69.93	57.44	**81.63***	67.43	58.10	**63.25**
	SMI-FGSM	62.72	53.15	76.46*	60.18	55.20	57.56
	MMIA	**70.90**	**68.41**	80.58*	**67.83**	**68.53**	62.94
MobileNet	FGSM	70.18	60.39	66.77	71.96*	62.70	67.55
	MI-FGSM	80.34	63.02	74.95	86.13*	64.80	73.87
	SMI-FGSM	82.60	65.95	75.05	87.81*	67.92	75.14
	MMIA	**84.79**	**76.21**	**80.09**	**87.92***	**75.94**	**77.91**
DenseNet	FGSM	58.38	48.18	52.33	63.92	50.26*	55.55
	MI-FGSM	76.36	76.08	73.22	70.79	85.20*	68.93
	SMI-FGSM	72.75	76.49	71.48	72.60	83.80*	67.67
	MMIA	**79.37**	**82.30**	**77.15**	**75.21**	**87.09***	**74.28**
GoogleNet	FGSM	67.53	51.54	64.02	67.87	56.86	68.46*
	MI-FGSM	75.31	53.55	63.47	73.46	55.56	86.41*
	SMI-FGSM	78.50	53.45	67.88	73.28	56.54	**88.28***
	MMIA	**81.41**	**65.73**	**71.80**	**77.08**	**67.73**	88.21*

Table 2.2. The attack success and stealth score (AS^3) (%) of non-targeted adversarial attacks against 10 models on ImageNet dataset. * represent the white-box attacks.

	Attack	Inception v3	ResNet50	VGG16	MobileNet	DenseNet	GoogleNet	ResNet50$_{adv}$	MobileNet$_{adv}$	ShuffleNet$_{adv}$	RegNetX$_{adv}$
Inception v3	FGSM	45.95*	21.51	21.55	32.89	24.62	28.69	**6.81**	21.01	**19.52**	**6.24**
	MI-FGSM	63.23*	24.23	23.29	33.57	24.89	27.52	4.29	18.54	15.42	4.39
	SMI-FGSM	49.82*	19.00	16.37	25.80	18.47	21.35	4.30	21.07	9.73	3.65
	MMIA	**71.54***	**38.41**	**31.59**	**41.73**	**39.73**	**40.13**	5.08	**24.38**	11.86	4.72
ResNet50	FGSM	24.14	37.38*	24.52	31.66	27.48	26.77	**6.37**	21.13	**18.54**	**6.19**
	MI-FGSM	34.5	63.60*	38.55	41.93	44.53	35.68	5.04	20.81	14.27	5.19
	SMI-FGSM	22.94	49.95*	21.02	29.19	28.10	23.06	4.31	24.96	11.25	4.23
	MMIA	**57.76**	**71.67***	**56.46**	**58.31**	**65.42**	**55.12**	5.43	**30.78**	10.79	5.27
VGG16	FGSM	22.69	24.37	44.92*	32.31	26.61	27.63	**6.69**	20.35	**19.35**	**6.32**
	MI-FGSM	33.92	37.77	64.20*	45.30	42.68	37.68	5.10	20.73	14.63	4.63
	SMI-FGSM	19.37	21.58	52.67	27.90	24.14	22.68	4.49	23.52	11.72	4.62
	MMIA	**49.08**	**54.22**	**71.20**	**57.84**	**58.82**	**50.16**	5.30	**28.35**	12.81	4.92
MobileNet	FGSM	24.81	25.11	28.15	52.87*	27.74	29.85	**6.86**	23.36	**19.62**	**6.92**
	MI-FGSM	33.53	34.72	39.00	62.96*	37.68	34.09	4.99	26.83	17.17	5.54
	SMI-FGSM	21.98	20.60	23.46	58.52*	25.51	25.12	5.01	28.65	12.52	4.66
	MMIA	**60.56**	**58.88**	**58.30**	**70.90***	**61.63**	**58.22**	5.14	**43.30**	14.30	6.47
DenseNet	FGSM	22.73	23.23	23.60	32.76	38.72*	25.72	**6.51**	21.18	**19.33**	**5.68**
	MI-FGSM	33.44	38.15	34.36	40.55	63.80*	34.75	4.94	21.84	15.04	4.49
	SMI-FGSM	21.25	23.78	20.36	30.42	52.97*	23.03	3.74	25.75	11.35	4.04
	MMIA	**50.98**	**56.87**	**51.17**	**55.66**	**71.85***	**50.12**	5.19	**29.66**	13.06	5.40
GoogleNet	FGSM	26.23	22.69	24.94	35.35	24.84	53.21*	**6.73**	23.16	**19.28**	**6.26**
	MI-FGSM	31.47	26.72	29.98	37.99	30.11	63.67*	5.15	21.26	15.74	4.69
	SMI-FGSM	21.70	17.78	17.69	29.16	19.78	56.44*	4.22	22.73	13.34	4.57
	MMIA	**55.47**	**47.76**	**44.62**	**54.57**	**51.82**	**71.59***	5.05	**31.11**	13.22	5.50
ResNet50$_{adv}$	FGSM	45.26	43.28	40.72	46.94	43.03	48.23	49.27*	50.01	36.03	36.37
	MI-FGSM	59.99	**55.22**	**50.67**	**59.96**	**55.64**	60.26	67.61*	65.65	46.21	50.70
	SMI-FGSM	26.47	21.51	21.72	32.86	22.40	28.78	44.91*	45.58	35.32	27.79
	MMIA	**60.95**	52.34	48.42	57.71	52.88	**62.58**	**71.12***	**68.46**	**52.96**	**56.08**
MobileNet$_{adv}$	FGSM	23.92	19.49	22.65	32.01	22.73	28.73	7.86	51.52*	15.19	6.78
	MI-FGSM	28.30	19.30	20.98	35.05	22.59	32.37	**21.50**	69.09*	**39.56**	**18.17**
	SMI-FGSM	19.00	16.09	17.39	29.97	18.81	21.93	5.32	59.93*	12.63	5.30
	MMIA	**45.86**	**33.39**	**34.93**	**51.19**	**38.84**	**45.05**	10.95	**70.95***	17.85	10.44
ShuffleNet$_{adv}$	FGSM	25.38	23.09	23.16	35.99	24.71	33.06	21.65	40.89	53.88*	20.02
	MI-FGSM	**33.80**	27.11	27.55	**42.41**	**30.46**	37.96	28.27	52.53	70.94*	25.69
	SMI-FGSM	21.25	16.96	18.84	30.80	20.13	25.56	21.31	38.89	54.80*	18.92
	MMIA	30.94	**27.55**	**28.20**	41.09	29.52	**39.96**	**35.47**	**54.71**	**78.75***	**31.25**
RegNetX$_{adv}$	FGSM	46.39	43.65	40.61	46.64	44.30	46.77	35.66	50.65	32.77	51.26*
	MI-FGSM	62.93	57.97	**53.35**	61.36	**59.10**	63.38	50.36	63.32	44.66	70.22*
	SMI-FGSM	29.12	24.15	23.84	33.79	25.61	31.25	30.07	42.72	33.57	42.07*
	MMIA	**66.00**	**58.97**	53.04	**61.66**	57.53	**67.85**	**56.26**	**69.28**	**50.63**	**74.27***