Appendix

In the supplementary material, we provides the comparison with gradient attacks on other models, discussion on other possible variants.

A Comparison with Gradient-based Attack

In this section, we report the comparison results on various gradient-based attack methods and our EMI-FGSM method integrated with various input transformations when attacking the other three normally trained models, *i.e*. Inc-v4, IncRes-v2, and Res-101 respectively.

We first report the attack success rates of various gradient-based attacks on the other three normally-trained models. The results are summarized in Table 7. Compared with other advanced attacks, EMI-FGSM also exhibits better white-box attack success rates and higher transferability, which are consistent to the results on the Inc-v3 model in the main text.

The results for the EMI-FGSM integrated with various transformation-based methods under single-model setting, where the adversarial examples are crafted on the other three normally-trained models, are depicted in Table 8. It can be observed that EMI significantly promotes the attack success rates of the baseline attacks with a clear margin, which are consistent to the results on the Inc-v3 model in the main text and further verifies the high effectiveness of the proposed enhanced momentum.

(a) Attack success rates (%) for the adversarial examples crafted on **Inc-v4**.

Attack	Inc-v3	Inc-v4*	IncRes-v2	Res-101	Inc-v3$_{ens3}$	Inc-v3$_{ens4}$	IncRes-v2$_{ens}$
FGSM	27.4	52.0	22.5	22.9	15.7	9.4	5.4
I-FGSM	32.8	**100.0**	20.0	19.9	5.3	6.8	3.1
MI-FGSM	56.2	99.9	46.0	40.7	15.7	15.1	8.3
NI-FGSM	63.0	99.9	52.4	45.6	16.5	14.3	7.5
PI-FGSM (**Ours**)	72.4	99.9	59.7	52.5	18.0	15.7	7.3
EMI-FGSM (**Ours**)	**87.7**	**100.0**	**76.2**	**67.5**	**27.6**	**24.4**	**12.4**

(b) Attack success rates (%) for the adversarial examples crafted on **IncRes-v2**.

Attack	Inc-v3	Inc-v4	IncRes-v2*	Res-101	Inc-v3$_{ens3}$	Inc-v3$_{ens4}$	IncRes-v2$_{ens}$
FGSM	27.2	20.2	41.9	23.6	9.5	9.1	5.7
I-FGSM	33.4	25.2	98.2	20.2	6.8	6.4	4.3
MI-FGSM	57.3	50.4	98.2	44.7	21.2	16.0	11.5
NI-FGSM	63.4	55.9	99.0	45.3	20.2	15.8	10.0
PI-FGSM (**Ours**)	71.6	63.4	98.3	53.4	24.3	18.7	12.5
EMI-FGSM (**Ours**)	**89.1**	**82.4**	**99.4**	**72.3**	**36.7**	**30.8**	**21.8**

(c) Attack success rates (%) for the adversarial examples crafted on **Res-101**.

Attack	Inc-v3	Inc-v4	IncRes-v2	Res-101*	Inc-v3$_{ens3}$	Inc-v3$_{ens4}$	IncRes-v2$_{ens}$
FGSM	36.4	31.2	30.0	78.1	14.9	13.3	6.5
I-FGSM	31.4	25.3	23.1	99.3	8.7	8.5	5.4
MI-FGSM	57.6	51.9	49.8	99.3	23.9	22.1	12.6
NI-FGSM	65.5	58.0	57.5	99.4	24.3	21.5	11.3
PI-FGSM (**Ours**)	72.8	66.8	63.7	99.3	28.3	25.3	14.0
EMI-FGSM (**Ours**)	**82.3**	**76.7**	**76.2**	**100.0**	**35.2**	**30.8**	**19.0**

Table 7: Attack success rates (%) against seven baseline models in single-model setting. The adversaries are crafted on **Inc-v4, IncRes-v2 or Res-101** using various adversarial attacks. * indicates the white-box model being attacked.

(a) Attack success rates (%) for the adversarial examples crafted on **Inc-v4**.

Attack	Inc-v3	Inc-v4*	IncRes-v2	Res-101	Inc-v3$_{ens3}$	Inc-v3$_{ens4}$	IncRes-v2$_{ens}$
DIM	74.1	98.5	66.3	58.0	22.3	21.0	11.6
EMI-DIM (**Ours**)	**89.4**	**99.1**	**83.6**	**75.2**	**33.5**	**30.9**	**16.7**
TIM	58.0	99.5	47.2	42.8	25.9	24.0	16.9
EMI-TIM (**Ours**)	**89.0**	**99.8**	**81.2**	**72.3**	**52.1**	**48.3**	**35.2**
SIM	80.6	99.5	73.6	68.8	47.9	44.9	29.2
EMI-SIM (**Ours**)	**96.4**	**99.9**	**93.7**	**89.0**	**59.7**	**56.1**	**36.9**
DTS	84.7	98.0	80.5	76.3	67.9	66.9	54.3
EMI-DTS (**Ours**)	**95.7**	**99.4**	**94.5**	**90.7**	**81.4**	**77.5**	**68.8**

(b) Attack success rates (%) for the adversarial examples crafted on **IncRes-v2**.

Attack	Inc-v3	Inc-v4	IncRes-v2*	Res-101	Inc-v3$_{ens3}$	Inc-v3$_{ens4}$	IncRes-v2$_{ens}$
DIM	68.1	65.1	93.7	58.3	30.2	23.4	17.3
EMI-DIM (**Ours**)	**88.8**	**85.1**	**98.5**	**78.3**	**42.4**	**35.5**	**26.4**
TIM	62.1	55.8	97.2	49.9	31.0	28.3	21.5
EMI-TIM (**Ours**)	**90.6**	**85.0**	**99.4**	**80.1**	**61.5**	**52.2**	**48.2**
SIM	84.6	79.5	98.9	76.1	55.9	49.0	41.7
EMI-SIM (**Ours**)	**97.5**	**95.1**	**99.9**	**90.9**	**69.0**	**60.1**	**51.6**
DTS	87.1	84.3	96.6	81.4	76.4	73.3	69.4
EMI-DTS (**Ours**)	**97.8**	**95.4**	**99.9**	**93.6**	**88.2**	**83.5**	**81.9**

(c) Attack success rates (%) for the adversarial examples crafted on **Res-101**.

Attack	Inc-v3	Inc-v4	IncRes-v2	Res-101*	Inc-v3$_{ens3}$	Inc-v3$_{ens4}$	IncRes-v2$_{ens}$
DIM	73.6	68.5	69.5	97.6	36.2	31.9	20.6
EMI-DIM (**Ours**)	**88.7**	**84.3**	**84.1**	**99.7**	**46.4**	**40.7**	**26.3**
TIM	59.4	54.0	52.3	99.2	35.6	31.8	22.8
EMI-TIM (**Ours**)	**86.0**	**79.1**	**79.8**	**100.0**	**56.4**	**50.3**	**41.7**
SIM	74.4	69.8	68.3	99.7	43.1	39.4	26.0
EMI-SIM (**Ours**)	**92.0**	**88.7**	**88.4**	**100.0**	**57.6**	**50.4**	**35.7**
DTS	84.0	80.0	81.9	98.9	73.3	70.9	59.3
EMI-DTS (**Ours**)	**93.7**	**90.9**	**92.3**	**99.6**	**83.9**	**80.9**	**71.7**

Table 8: Attack success rates (%) of various adversarial attacks against the seven baseline models under single-model setting. The adversarial examples are crafted on **Inc-v4, IncRes-v2 or Res-101** using various adversarial attack methods. * indicates the white-box model being attacked.

B Discussion on Possible Variations

Except for EMI-FGSM, there are also other ways to enhance the momentum. Here we provide two possible implementations of the enhanced momentum, denoted as ENI-FGSM and ERI-FGSM. Specifically, ENI-FGSM samples the data points in the direction of momentum by substituting Eq. 1 in EMI-FGSM with:

$$\tilde{x}_t^{adv}[i] = x_t^{adv} + c_i \cdot g_{t-1} \tag{5}$$

where g_{t-1} is the accumulated momentum of the previous iteration. ERI-FGSM adopts the accumulated gradient of randomly sampled data points by substituting Eq. 1 in EMI-FGSM with:

$$\tilde{x}_t^{adv}[i] = x_t^{adv} + \alpha \cdot U(-1^d, 1^d) \tag{6}$$

where $U(a,b)$ denotes the uniform distribution in $[a,b]$.

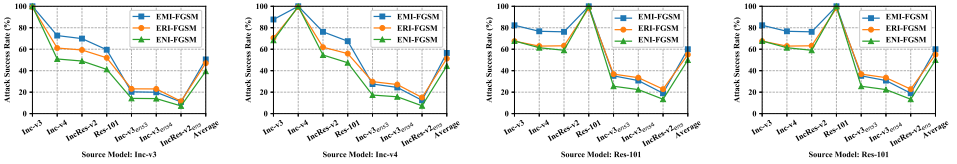

Figure 6: Attack success rates (%) of EMI-FGSM, ERI-FGSM and ENI-FGSM against seven models under single-model setting. (Zoom in for details.)

EMI-FGSM *vs*. ENI-FGSM. The proposed EMI-FGSM accumulates the gradient of the data points in the direction of (t-1)-th gradient at the t-th iteration and exhibits remarkable performance improvement. However, from the perspective of NI-FGSM, *can we accumulate the gradient of the data points in the direction of momentum at the t-th iteration?* To address this concern, we extend NI-FGSM to ENI-FGSM and test the attack performance. As shown in Figure 6, we see that ENI-FGSM is considerably lower than EMI-FGSM. It further supports our hypothesis that the direction of the accumulated momentum cannot provide a precise description of the neighborhood and find proper point for the gradient calculation, as it contains too much accumulated information from the previous iterations.

EMI-FGSM *vs*. ERI-FGSM. The comparison between ENI-FGSM and EMI-FGSM shows that the direction plays a big impact on the performance of the enhanced momentum. Both ENI-FGSM and EMI-FGSM sample the data points in a fixed direction. *What if we accumulate the gradient of the data points in the neighborhood of x_t^{adv} at the t-th iteration?* To address this concern, we test the performance of ERI-FGSM on various models. As shown in Figure 6, ERI-FGSM exhibits considerably lower transferability than EMI-FGSM on normally trained models but achieves slightly better performance on adversarially trained models. A possible reason might be that the data points with noise for the gradient calculation are more similar to the adversaries for the adversarial training.