1. 实验准备:

Target Model	Substitute1	Substitute2	Substitute3	Substitute4	Substitute5
Conv(128,3,3)+Relu Conv(64,3,3)+Relu Dropout(0.25) FC(128)+Relu Dropout(0.5) FC+Softmax	Conv(64,8,8)+Relu Conv(128,6,6)+Relu Conv(128,5,5)+Relu Dropout(0.5) FC+Softmax	FC(300)+Relu Dropout(0.5) FC(300)+Relu Dropout(0.5) FC(300)+Relu Dropout(0.5) FC(300)+Relu Dropout(0.5) FC+Softmax	Conv(32,3,3)+Relu Conv(32,3,3)+Relu Conv(64,3,3)+Relu Conv(64,3,3)+Relu FC(200)+Relu Dropout(0.5) FC+Softmax	Conv(32,3,3)+Relu Conv(32,3,3)+Relu Conv(64,3,3)+Relu Conv(64,3,3)+Relu Conv(128,3,3)+Relu Conv(128,3,3)+Relu Drop(0.2) FC(512)+Relu Dropout(0.5) FC+Softmax	Conv(64,3,3)+Relu Conv(64,3,3)+Relu Conv(128,3,3)+Relu Conv(128,3,3)+Relu FC(256)+Relu FC(256)+Relu FC+Softmax

Name	Training data	Test data	Features	Labels	Task
MNIST	50000	10000	28×28×1	10	digit recognition
GTSRB	39209	12630	32×32×3	43	traffic sign recognition

2. MNIST数据集实验结果:

FGSM(α=0.1)	*accuracy*	*Success rate*	*Transfer rate*
Black-Box Model	99.19%	--------	--------
Sub1	81.07%	75.53%	2.64%
Sub2	79.21%	74.91%	2.01%
Sub3	86.13%	39.85%	1.51%
Sub4	86.53%	35.83%	1.62%
Sub5	83.47%	48.85%	1.52%
Iter_Casc(k=3)	----------	**77.32%**	**3.69%**
Stack_Paral(k=3)	----------	**36.73%**	**2.47%**

I-FGSM(α=0.1)	*accuracy*	*Success rate*	*Transfer rate*
Black-Box Model	99.19%	--------	--------
Sub1	81.07%	91.16%	2.34%
Sub2	79.21%	82.04%	1.68%
Sub3	86.13%	59.20%	1.37%
Sub4	86.53%	43.24%	1.54%
Sub5	83.47%	69.03%	1.33%
Iter_Casc(k=3)	----------	**77.32%**	**3.69%**
Stack_Paral(k=3)	----------	**36.73%**	**2.47%**

R+FGSM(α=0.1)	*accuracy*	*Success rate*	*Transfer rate*
Black-Box Model	99.19%	--------	--------
Sub1	81.07%	77.14%	1.75%
Sub2	79.21%	60.51%	1.47%
Sub3	86.13%	40.32%	1.18%
Sub4	86.53%	29.43%	1.31%
Sub5	83.47%	43.32%	1.16%
Iter_Casc(k=3)	----------	**77.32%**	**3.69%**
Stack_Paral(k=3)	**----------**	**36.73%**	**2.47%**

FGSM(α=0.2)	*accuracy*	*Success rate*	*Transfer rate*
Black-Box Model	99.19%	--------	--------
Sub1	81.07%	90.03%	11.17%
Sub2	79.21%	92.88%	11.81%
Sub3	86.13%	51.51%	4.99%
Sub4	86.53%	51.45%	5.92%
Sub5	83.47%	64.05%	5.23%
Iter_Casc(k=3)	----------	**97.83%**	**20.54%**
Stack_Paral(k=3)	**----------**	**60.07%**	**16.25%**

I-FGSM(α=0.2)	*accuracy*	*Success rate*	*Transfer rate*
Black-Box Model	99.19%	--------	--------
Sub1	81.07%	99.36%	7.59%
Sub2	79.21%	98.33%	5.28%
Sub3	86.13%	81.87%	2.49%
Sub4	86.53%	74.36%	3.53%
Sub5	83.47%	86.12%	3.36%
Iter_Casc(k=3)	----------	**97.83%**	**20.54%**
Stack_Paral(k=3)	**----------**	**60.07%**	**16.25%**

R+FGSM(α=0.2)	*accuracy*	*Success rate*	*Transfer rate*
Black-Box Model	99.19%	--------	--------
Sub1	81.07%	94.84%	4.94%
Sub2	79.21%	87.96%	4.17%
Sub3	86.13%	57.43%	2.20%
Sub4	86.53%	47.20%	2.73%
Sub5	83.47%	60.78%	2.40%
Iter_Casc(k=3)	----------	**97.83%**	**20.54%**
Stack_Paral(k=3)	**----------**	**60.07%**	**16.25%**

FGSM(α=0.3)	*accuracy*	*Success rate*	*Transfer rate*
Black-Box Model	99.19%	--------	--------
Sub1	81.07%	94.11%	32.47%
Sub2	79.21%	96.03%	37.00%
Sub3	86.13%	63.12%	18.56%
Sub4	86.53%	65.72%	19.04%
Sub5	83.47%	72.84%	16.60%
Iter_Casc(k=3)	----------	**99.87%**	**46.12%**
Stack_Paral(k=3)	**----------**	**78.81%**	**48.11%**

I-FGSM(α=0.3)	*accuracy*	*Success rate*	*Transfer rate*
Black-Box Model	99.19%	--------	--------
Sub1	81.07%	99.93%	18.95%
Sub2	79.21%	99.27%	16.90%
Sub3	86.13%	92.32%	6.50%
Sub4	86.53%	92.87%	9.89%
Sub5	83.47%	94.42%	9.57%
Iter_Casc(k=3)	----------	**99.87%**	**46.12%**
Stack_Paral(k=3)	**----------**	**78.81%**	**48.11%**

R+FGSM(α=0.3)	*accuracy*	*Success rate*	*Transfer rate*
Black-Box Model	99.19%	--------	--------
Sub1	81.07%	98.80%	16.54%
Sub2	79.21%	95.90%	14.69%
Sub3	86.13%	70.00%	6.13%
Sub4	86.53%	65.40%	8.85%
Sub5	83.47%	76.67%	7.53%
Iter_Casc(k=3)	----------	**99.87%**	**46.12%**
Stack_Paral(k=3)	----------	**78.81%**	**48.11%**