進度報告

Anti-spoofing

進度

- 增加新的dataset MSU-MFSD
- Attack
- 觀察
- 下次進度

Attack 前情提要

- Attack對象設定
  - 只跑label是spoof且被test判別為spoof的frame
  - real被test判別為spoof的不會進行攻擊

OULU-1

- Model

acc_mean	apcer	bpcer	acer
0.9591	0.1512	0.0128	0.0820

- Attack

epsilon	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
FGSM	15.66	1.11	0	0	0.89	7.61	14.09	19.02	21.92	26.17
iFGSM	36.91	30.20		46.98	49.44					

OULU-2

- Model

acc_mean	apcer	bpcer	acer
0.9430	0.0279	0.0715	0.04973

- Attack

epsilon	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
FGSM	0	0	0	0	0	0	32.78	38.97	44.86	50
iFGSM	34.74	29.31		69.03						

OULU-3

- Model

acc_mean	apcer	bpcer	acer
0.9274	0.0826	0.0339	0.0583

- Attack

epsilon	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
FGSM	0	0	0	0	0.47	0.47	1.89	9.85	16.11	18.96
iFGSM	23.22	21.32								

OULU-4

- Model

acc_mean	apcer	bpcer	acer
0.9815	0	0.0526	0.0263

- Attack

epsilon	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
FGSM	2.85	0	0	0	5.71	11.42	11.42	11.42	11.42	11.42
iFGSM	2.85	5.71	42.86	51.43	57.14					

Replay Attack

- Model

acc_mean	apcer	bpcer	acer
0.9125	0.025	0.1	0.0625

- Attack

epsilon	0.3	0.4	0.5	0.6
FGSM	17.22	22.22	29.17	30.55
iFGSM			98.06	

CASIA

- Model

acc_mean	apcer	bpcer	acer
0.9175	0.3016	0.0185	0.1601

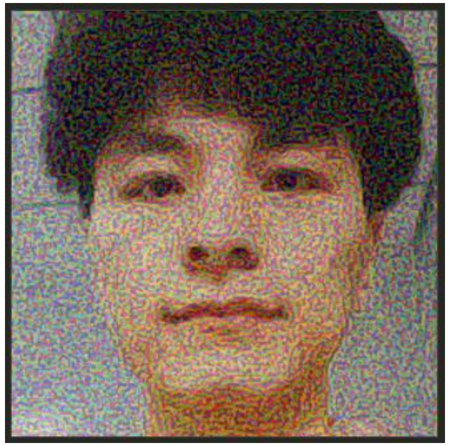
- Attack

epsilon	0.3	0.4	0.5	0.6
FGSM	0	45.75	93.40	95.28

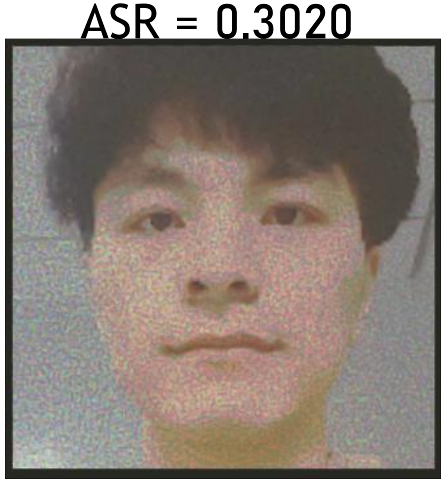
image

OULU

FGSM eps = 0.3
ASR = 0.0111

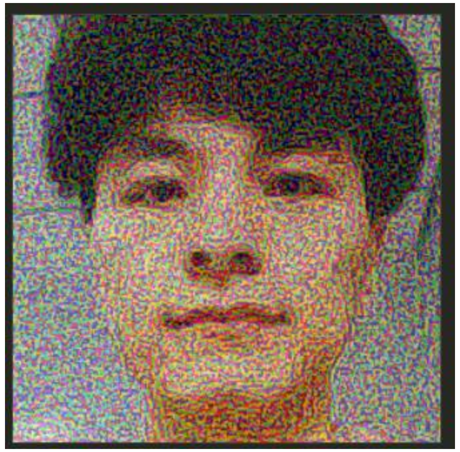


iFGSM eps = 0.3
ASR = 0.3020

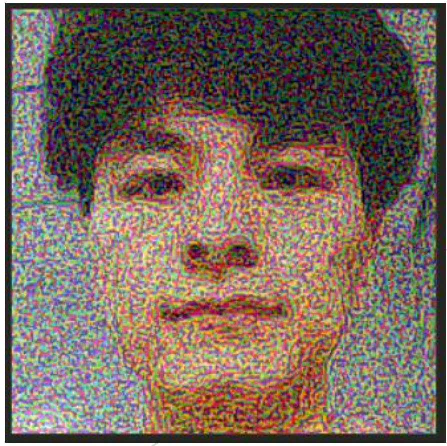


original

FGSM eps = 0.4
ASR = 0.0089

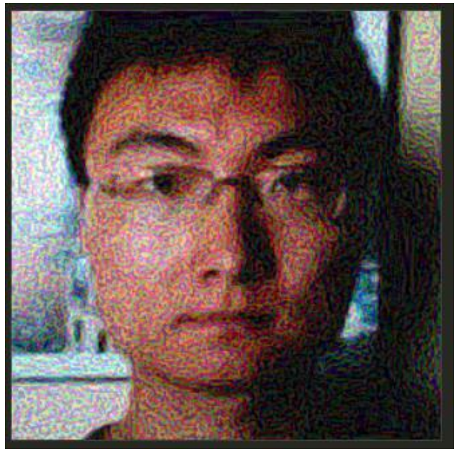


FGSM eps = 0.5
ASR = 0.0761



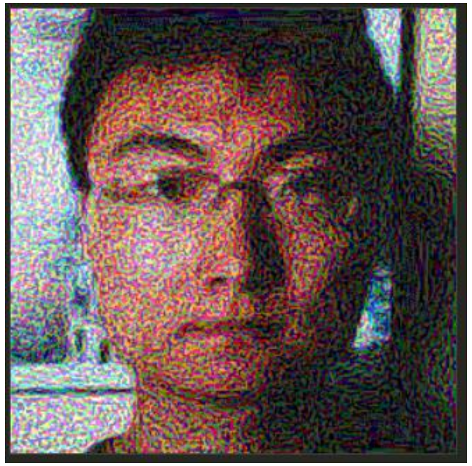
Replay Attack

FGSM eps = 0.3
ASR = 0.172



original

FGSM eps = 0.5
ASR = 0.2917



觀察

- 加夠多noise才攻擊得了
- Eps極小時的異常現象
- Performance iFGSM >> FGSM

下次進度

- Attack new dataset MSU-MFSD
- 跑完實驗
- 在perturbation加各種filter(Gaussian)
- Attack完後加各種filter(Gaussian)
- 改成其他用來攻擊的loss