進度報告

Anti-spoofing

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
- Type 1:
  - Attack_success: Logit_real >= Logit_spoofing
- Type 2:
  - Attack_success: Logit_real >= Logit_spoofing

FGSM

OULU

- Model

acc_mean	apcer	bpcer	acer
0.9591	0.1512	0.0128	0.0820

- Attack

epsilon	type	ASR(%)	Attack Success	Attack Fail	Test Real
0.3	1	0.9202	415	36	17
0.3	2	0	0	451	17
0.4	1	0.9977	450	1	17
0.4	2	0	0	451	17
0.5	1	0.9977	450	1	17
0.5	2	0	0	451	17

Replay Attack

- Model

acc_mean	apcer	bpcer	acer
0.9479	0.1625	0.03	0.09625

- Attack

epsilon	type	ASR(%)	Attack Success	Attack Fail	Test Real
0.3	1	0.172	62	298	40
0.3	2	0.172	62	298	40
0.5	1	0.2917	105	255	40
0.5	2	0.2917	105	255	40

iFGSM

OULU

- Model

acc_mean	apcer	bpcer	acer
0.9591	0.1512	0.0128	0.0820

- Attack

epsilon	type	ASR(%)	Attack Success	Attack Fail	Test Real
0.3	1	0.9091	410	41	17
0.3	2	0.5964	269	182	17

image

OULU

FGSM eps = 0.3
ASR = 0.9202

iFGSM eps = 0.3
ASR = 0.9091

original

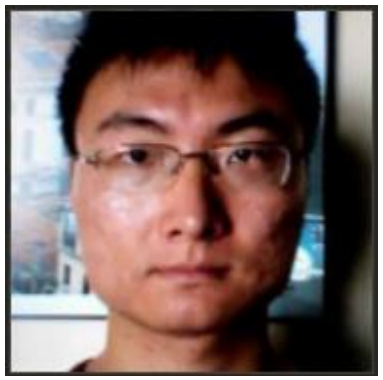
FGSM eps = 0.4
ASR = 0.9977

FGSM eps = 0.5
ASR = 0.9977

Replay Attack

FGSM eps = 0.3
ASR = 0.172

original



FGSM eps = 0.5
ASR = 0.2917

困難

- FGSM
  - 一堆[0.5 0.5]
  - Type1 : FGSM > iFGSM
  - Type2 : FGSM < iFGSM
- Dataset
  - Replay attack難攻