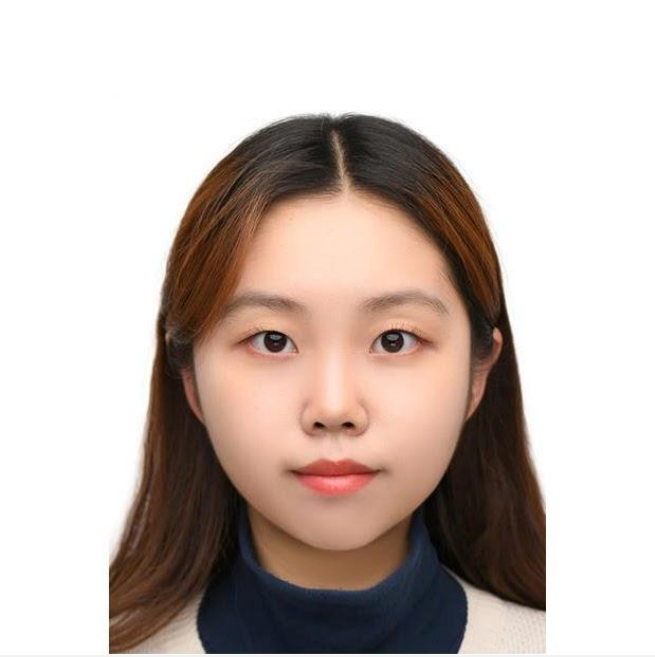
캡스톤 디자인 최종발표

Developing a Model for Face Anti-Spoofing
위조 얼굴 탐지를 위한 모델 개발

컴퓨터공학과 김동수, 오서연

목차

팀원 소개



오서연

FAS 논문 조사 및 재현 실험

Domain Generalization 논문 조사 및 재현 실험

김동수

FAS 논문 조사 및 재현 실험

모델 개발을 위한 Baseline 코드 작성

연구 목표

Face Anti-Spoofing(FAS)란?

- ✓ 얼굴 인식 시스템에서 위조 얼굴 이미지로 인한 보안 위협을 방지하는 기술
- ✓ 위변조 된 얼굴 데이터로 인식 시스템을 속이는 스푸핑 공격을 방지, 발생 피해 최소화 필요

2D Photo Mask | 2D Cardstock Mask | 2D Paper Mask | 3D Layered Mask | Antispoofing Movie File

연구 목표

Attack Type의 한계

- ✓ 다양한 Attack Type 존재 → 모든 Attack Type Data를 확보하는 것은 현실적으로 불가능
- ✓ 국내에서 Face data의 신원정보 문제 → Benchmark Data 사용해 성능 측정


(a)


(b)


(c)


(d)


(e)


(f)


(g)

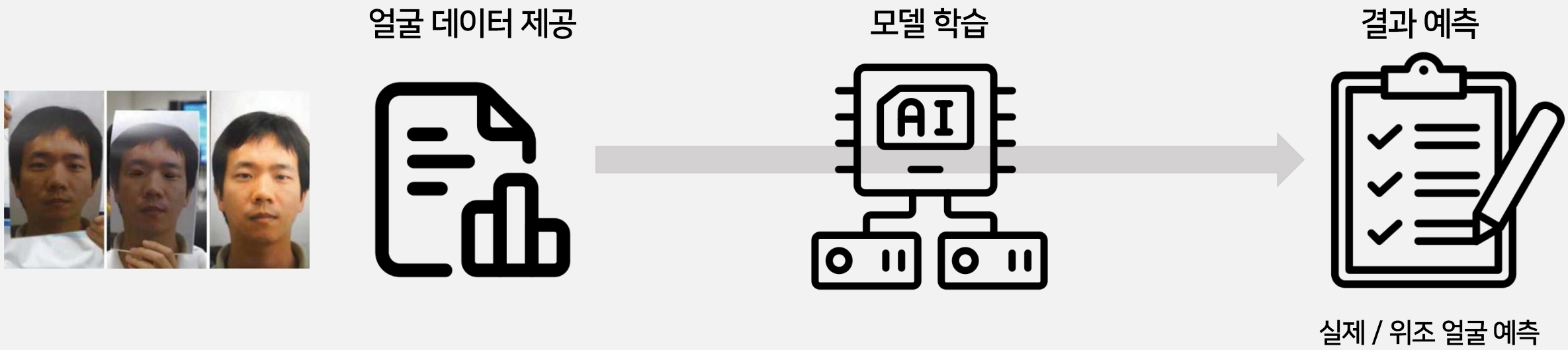
기호	Attack 종류
(a)	Paper glasses
(b)	Print (2D)
(c)	Replay
(d)	Fake head
(e)	Rigid mask
(f)	Flexible mask
(g)	Paper mask

* Wide Multi-Channel presentation Attack(WMCA) dataset

연구 목표

Face Anti-Spoofing(FAS) 모델 개발

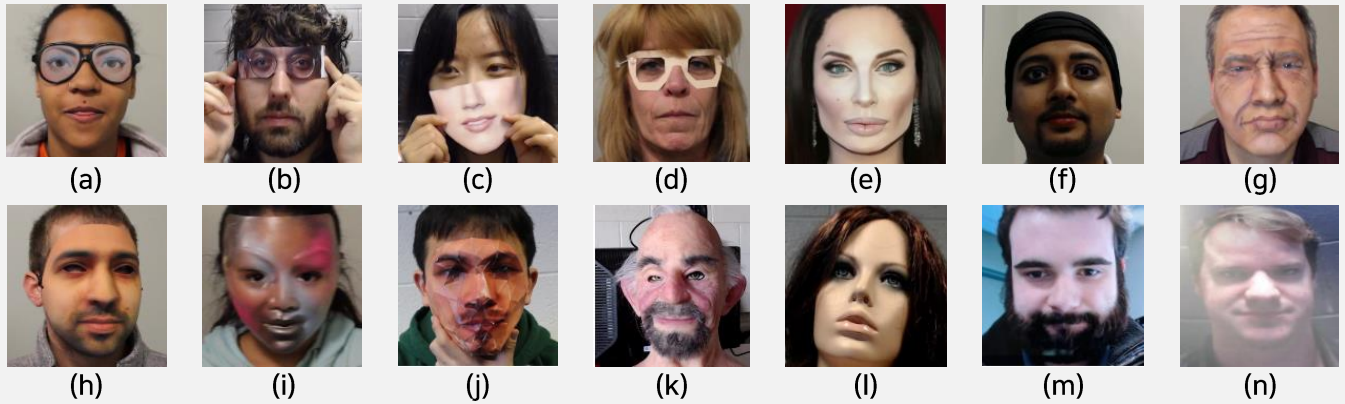
✓ Domain Generalization을 통해 Unknown Attack을 탐지하는 FAS 모델 개발



모델 제안

데이터셋

- ✓ Spoof in Wild with Multiple Attacks Version 2 (SiW-Mv2)
- ✓ 총 14가지의 Attack Type

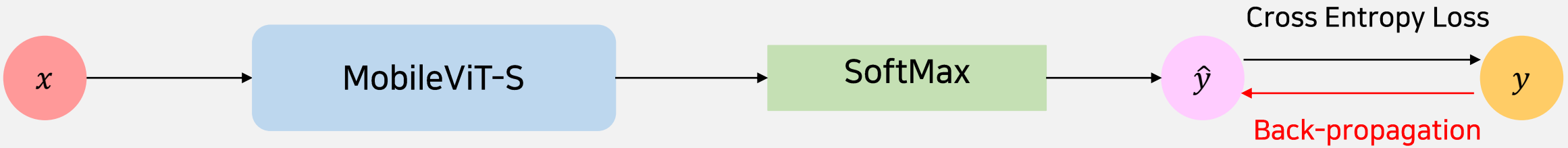


기호	Attack Type
a	Funny Eyes
b	Partial Eyes
c	Parital Mouths
d	Paperglass
e	Impersonate Makeup
f	Obfuscation Makeup
g	Cosmetic Makeup
h	Full Mask
i	Transparent Mask
j	Paper Mask
k	Silicone Head
l	Mannequin
m	Print
n	Replay

모델 제안

FAS Baseline

✓ Backbone으로 MobileViT-S를 활용한 기본적인 FAS 모델



모델 제안

Baseline 개선점

- ✓ **정보 손실**
  - ✓ MobileViT-S는 연산 효율성을 위해 모델의 첫 번째 Convolution Layer에서 Stride를 2로 설정하여 입력 이미지의 세부 정보 손실 → Stride를 1로 변경해서 정보 손실 감소

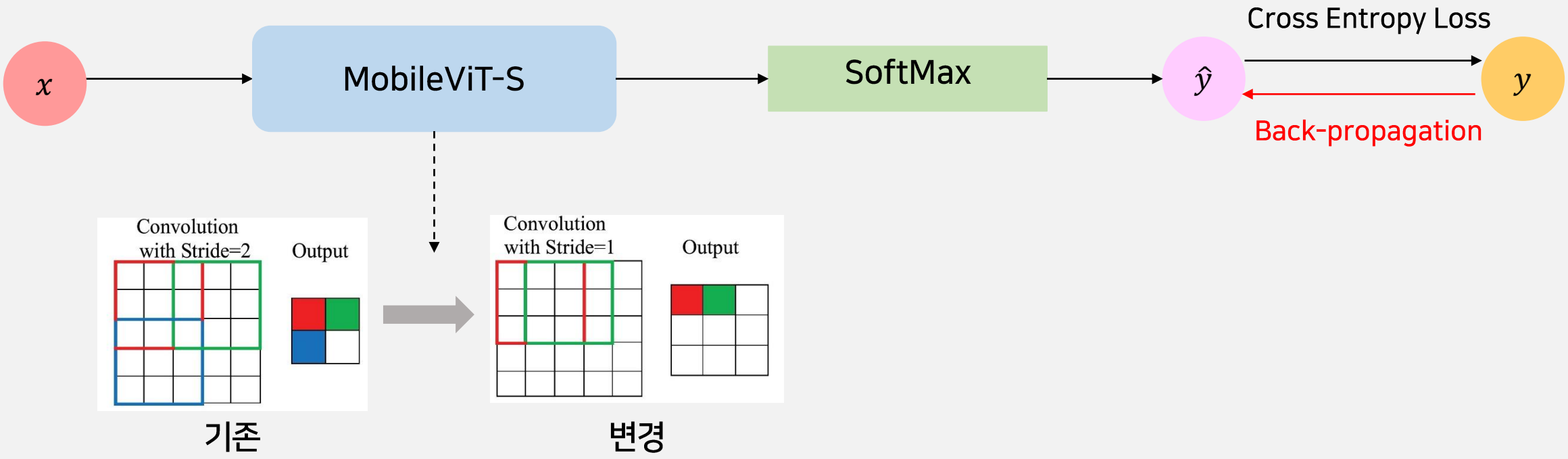
- ✓ **Feature 간 분리도**
  - ✓ Cross Entropy는 단순히 클래스를 분리하고 Feature 간의 분리도는 고려 X → SupCon Loss를 사용해 Feature 간 분리도 강화

- ✓ **일반화 성능 향상**
  - ✓ 다양한 데이터 분포에서도 일관된 성능을 발휘하도록 SAM Optimizer를 사용

모델 제안

FAS Baseline + Stride

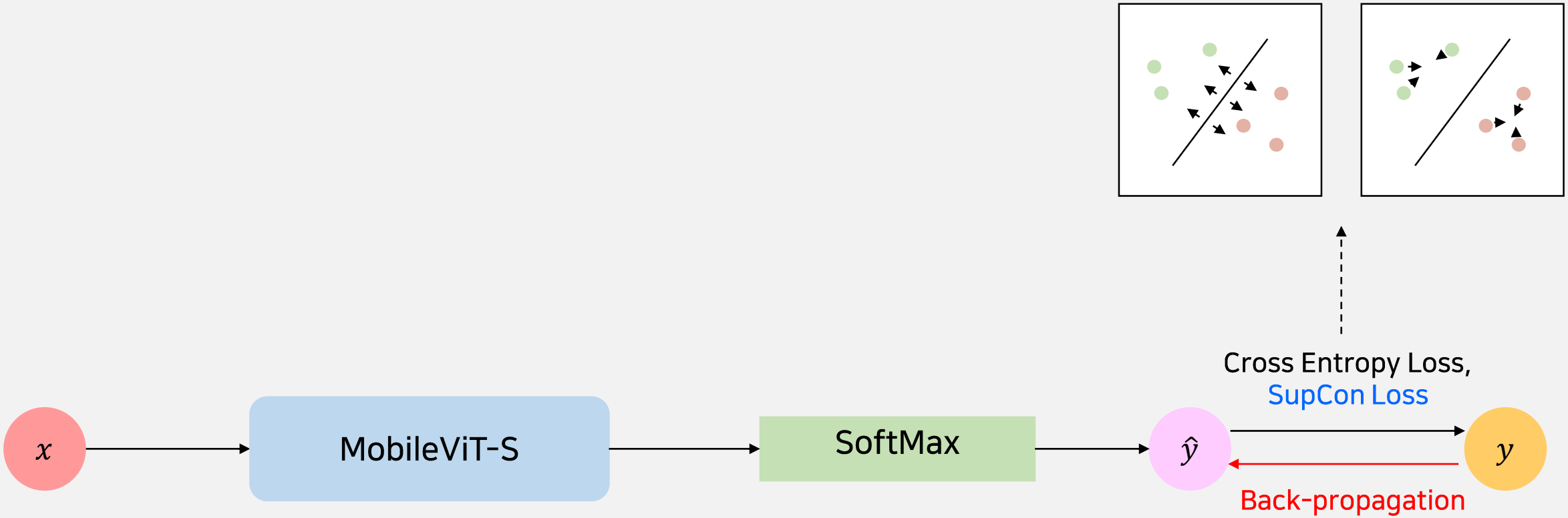
✓ Baseline의 첫 번째 Convolution layer의 stride를 2에서 1로 변경



모델 제안

FAS Baseline + SupCon Loss

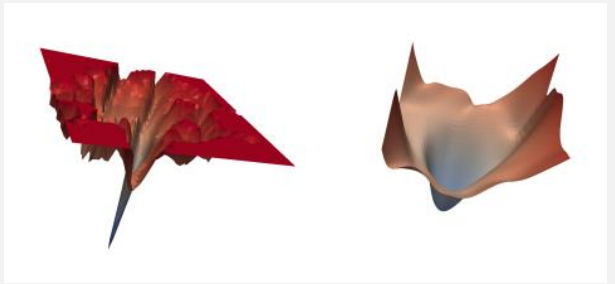
✓ Baseline에 Supervised Contrastive Loss를 추가

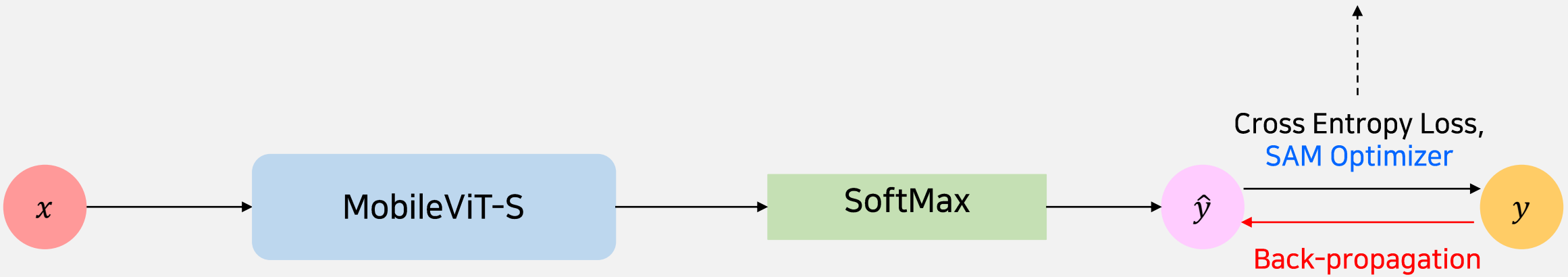


모델 제안

FAS Baseline + SAM

✓ Baseline에 Sharpness-Aware Minimization(SAM) Opimizer를 추가





실험 결과

ACER (Average Classification Error Rate)

- ✓ FAS 성능을 평가하기 위해 사용되는 지표로 APCER와 BPCER의 평균을 사용해 모델이 한 쪽 오류율에만 치우치지 않도록 함.
- ✓ **APCER(Attack Presentation Classification Error Rate)** : 스푸핑 공격을 실제 얼굴로 잘못 분류하는 오류율
- ✓ **BPCER(Bonafide Presentation Classfication Error Rate)** : 실제 얼굴을 스푸핑 공격으로 잘못 분류하는 오류율

$$ACER = \frac{APCER + BPCER}{2}$$

실험 결과

Single-Category-to-Unknown-Attacks

✓ Baseline의 Backbone 모델 탐색을 위한 실험 진행

Methods	From Covering														average ACER
	2D attack		3D attack					Make up			Physical Covering				
	Print	Replay	Half.	Trans.	Paper	Silicone	Mann.	Imp.	Obfu.	Cos.	Fun.	Eye	Mou.	Pap.	
	ACER														
Baseline(MobileViT_S)	24.33	24.54	19.31	17.37	2.13	16.36	9.04	3.36	24.81	27.76	Training				16.90
Stride(1,1)	**19.61**	**19.00**	**20.98**	**17.35**	**0.84**	**17.14**	**11.50**	**7.05**	**14.35**	**24.29**					**15.21**
CE/SupCon(7:3)	28.35	23.99	20.37	26.36	3.11	22.39	10.53	4.48	19.15	33.11					19.18
CE/SupCon(8:2)	28.53	23.20	19.45	25.35	1.88	21.05	8.69	3.70	20.36	33.24					18.55
SAM(rho=0.05)	**22.38**	**24.78**	**22.41**	**11.49**	**0.80**	**16.25**	**7.40**	**2.94**	**24.39**	**29.88**					**16.27**
SAM(rho=0.05), CE/SupCon(7:3)	27.57	23.84	25.32	27.67	1.65	19.13	11.92	3.50	20.25	31.51					19.24

실험 결과

Single-Category-to-Unknown-Attacks

✓ 기존 기법들과의 성능 비교

Methods	From Covering														average ACER
	2D attack		3D attack					Make up			Physical Covering				
	Print	Replay	Half.	Trans.	Paper	Silicone	Mann.	Imp.	Obfu.	Cos.	Fun.	Eye	Mou.	Pap.	
	ACER														
SA-FAS[1]	28.38	33.36	20.33	50.00	2.00	26.58	30.63	11.03	18.38	25.75	Training				24.64
DiVT-M[2]	32.60	35.31	23.48	33.42	5.53	26.83	23.25	9.65	19.06	39.65					24.88
DGUA-FAS[3]	25.57	32.71	16.04	33.78	1.70	19.01	24.34	5.09	25.93	35.22					21.94
Baseline + Stride(1,1) (Ours)	**19.61**	**19.00**	**20.98**	**17.35**	**0.84**	**17.14**	**11.50**	**7.05**	**14.35**	**24.29**					**15.21**

[1] Yiyou Sun, Yaojie Liu, "Rethinking Domain Generalization for Face Anti-spoofing: Separability and Alignment", CVPR, 2023
[2] Chen-Hao Liao, Wen-Cheng Chen, "Domain Invariant Vision Transformer Learning for Face Anti-spoofing", WACV, 2023
[3] Zong-Wei Hong, Yu-Chen Lin, "Domain-Generalized Face Anti-Spoofing with Unknown Attacks", ICIP , 2023

실험 결과

Single-Category-to-Unknown-Attacks

✓ Baseline의 Backbone 모델 탐색을 위한 실험 진행

Methods	From Makeup													average ACER
	2D attack		3D attack					Make up			Physical Covering			
	Print	Replay	Half.	Trans.	Paper	Silicone	Mann.	Imp.	Obfu.	Cos.	Fun.	Eye	Mou.	Pap.
	ACER													
Baseline(MobileViT_S)	15.73	15.86	11.64	4.97	1.55	4.35	1.53	Training			25.00	3.62	5.56	24.09	10.35
Stride(1,1)	15.21	12.19	12.00	5.97	1.62	5.70	2.44				29.82	5.08	9.17	28.13	11.58
CE/SupCon(7:3)	16.03	17.29	7.42	8.90	1.58	4.75	1.62				28.43	2.86	8.76	34.50	12.01
CE/SupCon(8:2)	15.15	16.97	8.14	7.19	2.19	5.87	1.95				28.34	3.62	6.46	26.78	11.15
SAM(rho=0.05)	**12.38**	**12.69**	**5.89**	**5.75**	**1.93**	**2.82**	**0.75**				**29.54**	**3.69**	**7.33**	**24.92**	**9.79**
SAM(rho=0.05), CE/SupCon(7:3)	19.84	16.82	13.51	11.13	4.85	5.83	1.86				31.91	4.61	7.75	33.41	13.77

실험 결과

Single-Category-to-Unknown-Attacks

✓ 기존 기법들과의 성능 비교

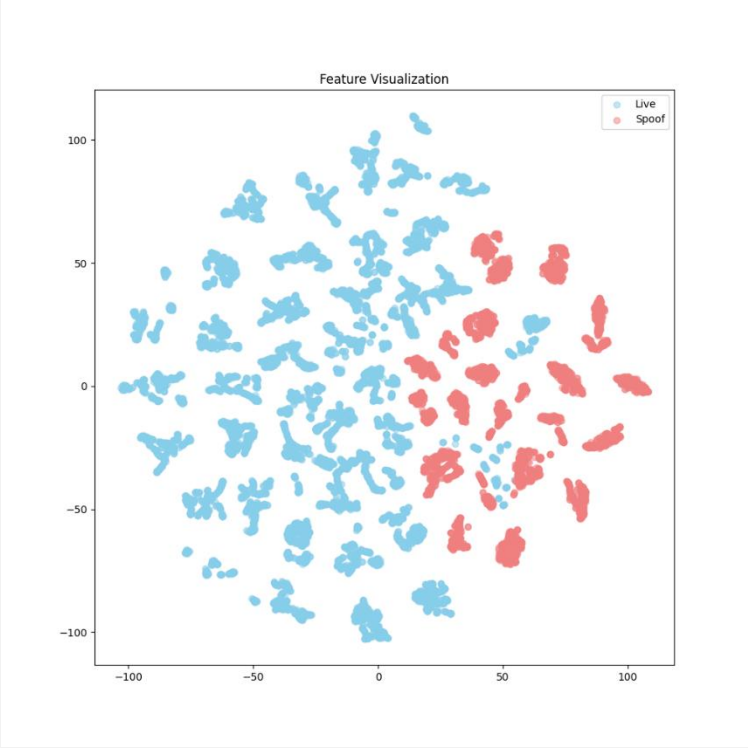
Methods	From Makeup													average ACER
	2D attack		3D attack					Make up			Physical Covering			
	Print	Replay	Half.	Trans.	Paper	Silicone	Mann.	Imp.	Obfu.	Cos.	Fun.	Eye	Mou.	Pap.
	ACER													
SA-FAS	18.20	34.82	21.18	35.28	0.73	13.5	17.47	Training			36.54	12.12	20.97	27.18	20.17
DiVT-M	18.39	29.36	15.07	19.28	0.35	19.37	18.80				38.95	12.21	18.34	19.04	17.23
DGUA-FAS	21.57	33.99	8.78	17.53	0.31	17.53	14.65				17.25	6.98	10.34	15.07	15.77
Baseline + SAM(rho=0.05) (Ours)	**12.38**	**12.69**	**5.89**	**5.75**	**1.93**	**2.82**	**0.75**				**29.54**	**3.69**	**7.33**	**24.92**	**9.79**

[1] Yiyou Sun, Yaojie Liu, "Rethinking Domain Generalization for Face Anti-spoofing: Separability and Alignment", CVPR, 2023
[2] Chen-Hao Liao, Wen-Cheng Chen, "Domain Invariant Vision Transformer Learning for Face Anti-spoofing", WACV, 2023
[3] Zong-Wei Hong, Yu-Chen Lin, "Domain-Generalized Face Anti-Spoofing with Unknown Attacks", ICIP , 2023

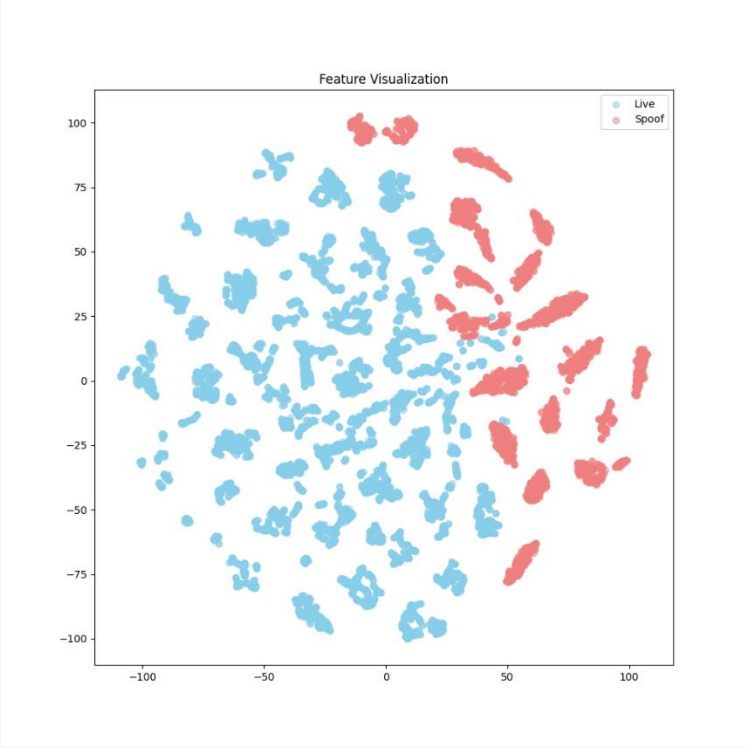
실험 결과 분석

T-SNE 시각화

- ✓ Covering to Obfu
  - ✓ Stride를 변경했을 때 정보 손실이 적어 Baseline 보다 성능 향상



Baseline

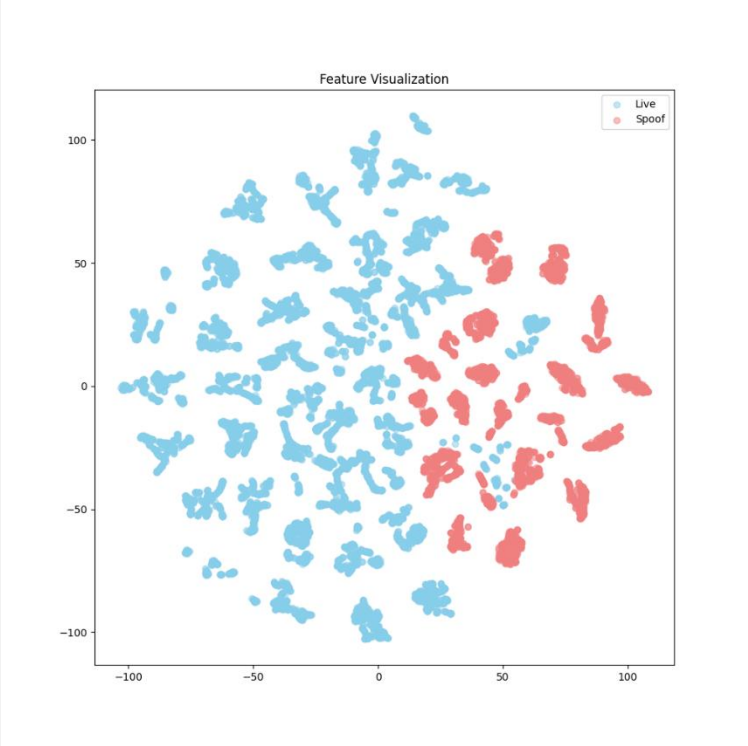


Stride 1

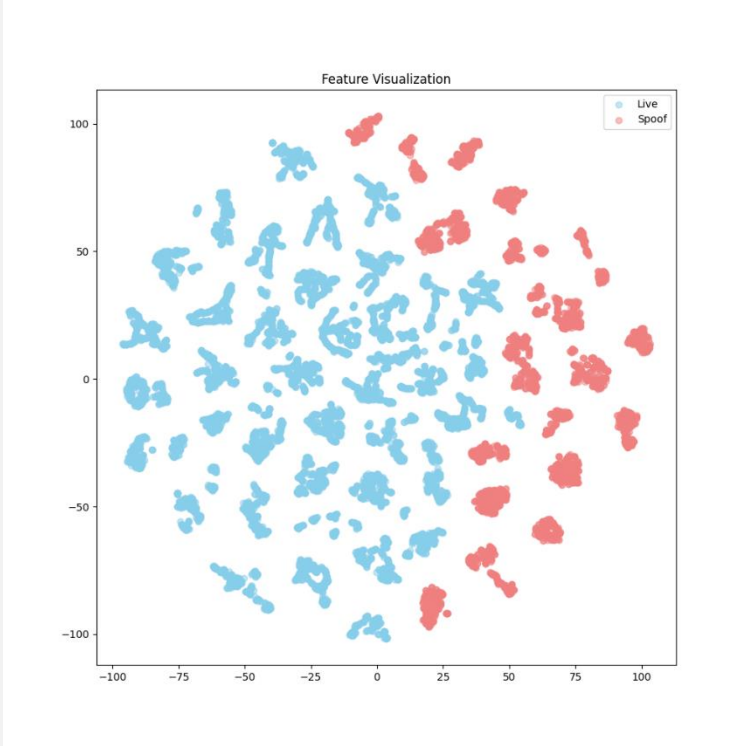
실험 결과 분석

T-SNE 시각화

- ✓ Covering to Obfu
  - ✓ SupCon을 적용하면 성능 자체는 떨어지지만 feature 간 분리도는 높음



Baseline

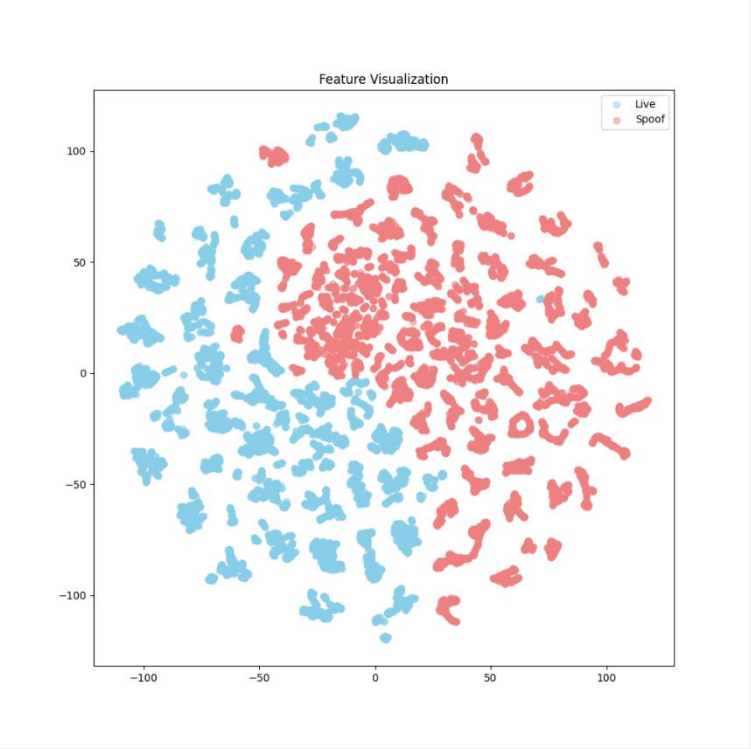


SupCon(8:2)

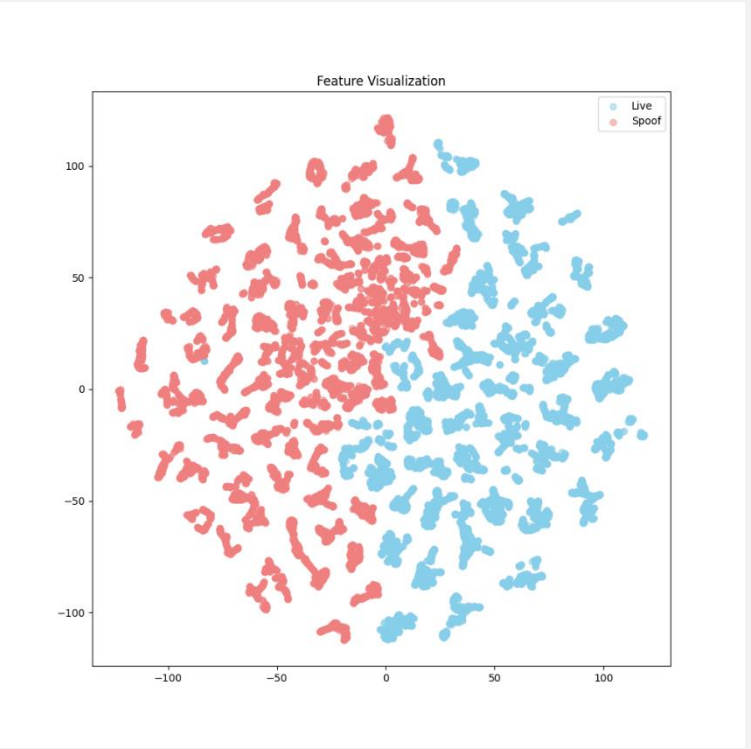
실험 결과 분석

T-SNE 시각화

- ✓ Makeup to HalfMask
  - ✓ SAM을 적용했을 때 feature를 더 명확히 구분하여 Baseline보다 성능 향상



Baseline

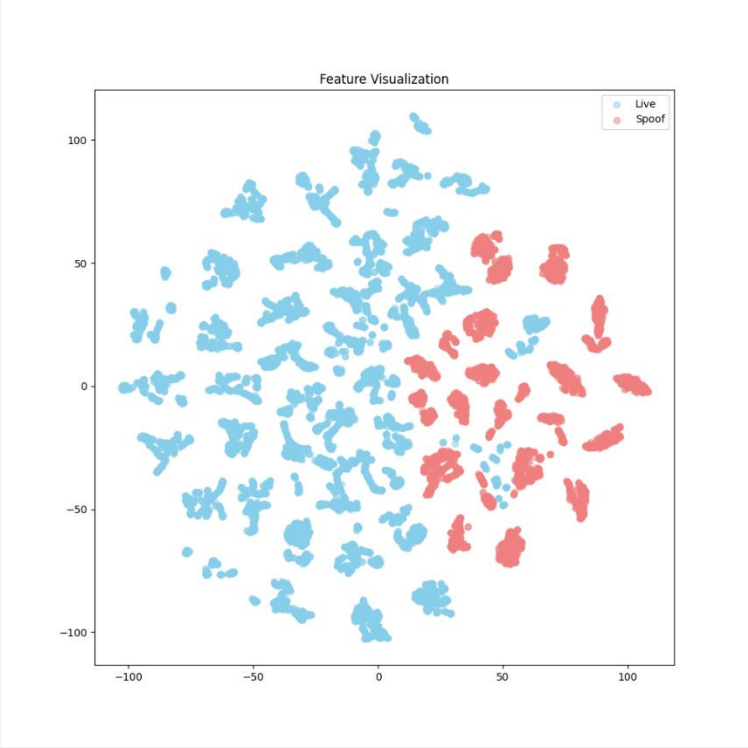


SAM

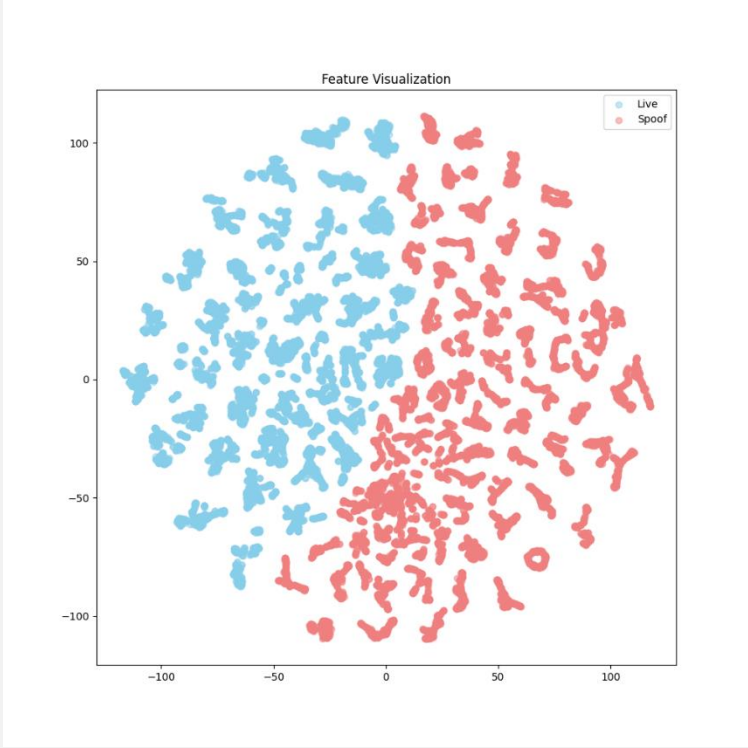
실험 결과 분석

T-SNE 시각화

- ✓ Makeup to HalfMask
  - ✓ SupCon을 적용하면 성능 자체는 떨어지지만 feature 간 분리도는 높음



Baseline



SupCon(7:3)

성과

Competition

Snapshot Spectral Imaging Face Anti-spoofing Challenge

Organized by THU-CVAILAB - Current server time: June 7, 2024, 12:28 p.m. UTC

First phase: Phase 1 — Feb. 29, 2024, midnight UTC

End: Competition Ends — March 31, 2024, midnight UTC

Learn the Details | Phases | Participate | Results

Phase 1 | Phase 2

Phase description

Development phase with result scoring

Max submissions per day: 5

Max submissions total: 100

Download CSV

RESULTS

#	User	Entries	Date of Last Entry	ACER [%] ▲	APCER [%] ▲	BPCER [%] ▲
1	jho-yonsei	9	03/21/24	0.2060 (1)	0.4121 (1)	0.0000 (1)
2	jeu2250	17	03/21/24	0.6868 (2)	0.4121 (1)	0.9615 (3)
3	dongsukim	30	03/21/24	0.6868 (2)	0.4121 (1)	0.9615 (3)
4	minseok	19	03/22/24	0.8585 (3)	1.2363 (4)	0.4808 (2)
5	ZTT	6	03/21/24	0.9615 (4)	0.9615 (2)	0.9615 (3)
6	whyjlee	13	03/21/24	1.3049 (5)	1.6484 (6)	0.9615 (3)
7	Jimini	25	03/21/24	1.3393 (6)	1.2363 (4)	1.4423 (4)
8	ChenYifan	5	03/20/24	1.5110 (7)	1.0989 (3)	1.9231 (5)
9	yiyao_scb	17	03/17/24	2.0604 (8)	2.1978 (8)	1.9231 (5)
10	stella0831	48	03/23/24	2.1635 (9)	1.9231 (7)	2.4038 (6)
11	SeaRecluse	8	03/08/24	2.4382 (10)	1.5110 (5)	3.3654 (7)
12	sunghun	26	03/19/24	2.4725 (11)	1.0989 (3)	3.8462 (8)
13	yelan.lj	5	03/18/24	2.8846 (12)	5.7692 (12)	0.0000 (1)
14	hexianhua	11	03/20/24	3.6401 (13)	7.2802 (13)	0.0000 (1)
15	Bulbul	11	03/21/24	3.9835 (14)	3.1593 (9)	4.8077 (9)
16	ctyun-ai	11	03/21/24	4.7734 (15)	3.2967 (10)	6.2500 (10)
17	CTEL_AI	4	03/12/24	9.8901 (16)	3.4341 (11)	16.3462 (12)
18	THU-CVAILAB	1	02/29/24	11.0920 (17)	13.0495 (14)	9.1346 (11)
19	Wantongming	3	03/14/24	50.0000 (18)	100.0000 (15)	0.0000 (1)

2024 CVPRW FAS 대회 track1 2등

커버 소스 불일치에 대한 이미지 스테그어날리시스 일반화

오서연[O]*, 김동수*, 민지민*, 장한얼†
한밭대학교 컴퓨터공학과
{20211915, 20191766, 30231212}@edu.hanbat.ac.kr, hejang@hanbat.ac.kr

Steganalysis Learning Methods for Resolving Cover Source Mismatch Issues

요 약

모바일 기기 보급으로 인해 스테가노그라피 도구의 활용성이 크게 향상됨에 따라 스테그어날리시스 기술의 중요성이 높아지고 있다. 최근에는 커버를 생성하는 소스 이미지 생성 과정에서 파라미터 변화가 발생하는 경우 스테그어날리시스 성능이 크게 저하되는 커버-소스 불일치 문제가 주목받고 있다. 본 논문에서는 다양한 소스 기기에서 취득한 이미지로 통합 데이터셋을 구축하여 커버-소스 불일치 문제에 효과적으로 대응한다. 또한, 커버와 스테고 이미지 간의 미세한 차이값을 효과적으로 탐지할 수 있도록 CNN 모델의 저수준 특징 추출부를 개선하고 대조 학습을 활용한다. 제안기법은 학습에 사용하지 않은 Galaxy Flip3와 iPhone12 기기에 대하여 베이스라인 모델 대비 평균 6.36% 탐지율 향상을 보였다.

1. 서 론

스테가노그라피(steganography)는 이미지, 비디오, 오디오 등과 같은 다양한 매체에 정보를 비밀리에 삽입하는 기술로, 디지털 매체의 일상화와 함께 정보 보호 분야에서 중요한 방법 중 하나이다. 최근에는 스테가노그라피 모바일 도구 활용성이 크게 향상됨에 따라 스테그어날리시스 기술도 중요해지고 있다. 스테그어날리시스는 스테가노그라피를 통해 숨겨진 정보를 감지하고 분석함으로써 스테가노그라피의 부적절한 사용을 방지하고 보안을 강화하는 데 사용된다. 정보의 디지털화가 지속됨에 따라 스테그어날리시스의 필요성은 더욱 커지고 있다.

스테가노그라피를 탐지하기 위해서는 메시지가 삽입된 스테고(stego) 이미지와 메시지가 삽입되지 않은 커버(cover) 이미지 사이의 미세한 차이를 구별해야 한다. 그림 1을 보면 커버와 스테고 이미지의 차이는 사람의 눈으로 거의 구별할 수 없다. 커버와 스테고 이미지에서 메시지가 삽입된 영역의 차이는 대부분 최하위 비트(least significant bit, LSB)에서 발생하기 때문에 픽셀 차이값은 1이고, 굉장히 미세한 패턴이다. 이미지의 통계적 특성을 주로 활용한 기존 스테그어날리시스 방법들과 달리, 딥러닝을 활용한 스테그어날리시스 방법은 더 복잡한 패턴과 관계를 학습할 수 있으며, 특히 컨볼루션 신경망(convolutional neural networks, CNN)은 고차원 데이터에 내재된 특징들을 추출하여 숨겨진 정보를 효과적으로 식별하는데 효과적이다.

(a) (b) (c)

그림 1. (a) 커버, (b) 스테고, (c) 커버와 스테고 차이 이미지 예시

그러나 딥러닝 기반의 스테그어날리시스 기법은 커버를 생성하는 소스 이미지 생성 과정에서 파라미터 변화가 발생하는 커버-소스 불일치(cover-source mismatch, CSM) 시에 탐지율이 크게 하락하는 제한점이 있다. CSM은 크게 이미지 획득과 이미지 처리 과정에서 발생한다. 이미지 획득 과정에서는 랜즈 종류, ISO, 노출 시간 등에서 파라미터 차이가 발생할 수 있고 이미지 처리 과정에서는 디노이징, 샤프닝, 절삭 등의 알고리즘의 파라미터 차이가 발생할 수 있다. 기존에 공개된 스테그어날리시스 벤치마크의 경우 대부분 같은 기기와 같은 파라미터를 사용하여 소스 이미지를 생성하기 때문에 다른 소스 이미지 생성 과정으로 구축된 데이터셋에 대해서는 성능이 크게 하락하는 문제가 발생한다.

본 논문에서는 CSM 문제에 효과적으로 대응하기 위해 여러 종류의 모바일 기기에서 취득한 다양한 소스 환경의 통합 데이터셋을 구축하였다. 또한, 커버와 스테고 이미지 간의 미세한 차이값을 효과적으로 탐지할 수 있도록 CNN 모델의 저수준 특징 추출부를 개선하고 대조 학습을 활용하는 방법을 제안한다.

2. 관련 연구

2.1 CNN 기반의 스테그어날리시스

최근 딥러닝 기반의 스테그어날리시스 기술이 활발히 연구되고 있다. 딥러닝 모델은 높은 계산 능력과 복잡한 패턴을 학습할 수 있는 능력을 바탕으로, 전통적인 스테가노그라피 탐지 기법보

* 본 연구는 2024년 과학기술정보통신부 및 정보 통신기획평가원의 SW중심대학사업의 연구결과로 수행되었음(2022-0-01068)

2024 KCC

차후 계획

완료

FAS Baseline 구축

FAS Baseline을 기반으로 한 모델 확장

T-SNE 시각화를 통한 정성적 분석

KCC 논문 제출

계획

다양한 데이터셋을 이용한 성능 검증

제안하는 모델 기반 논문 작성

제안하는 모델 기반 어플리케이션 제작

Thank You