AI 기반 사용자 인증

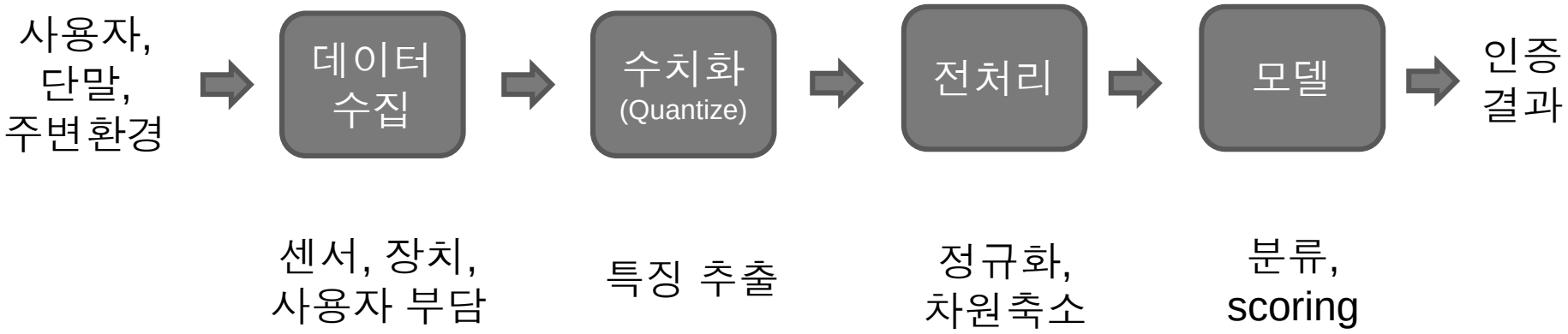
공주대학교
지능보안연구실
최대선

내용

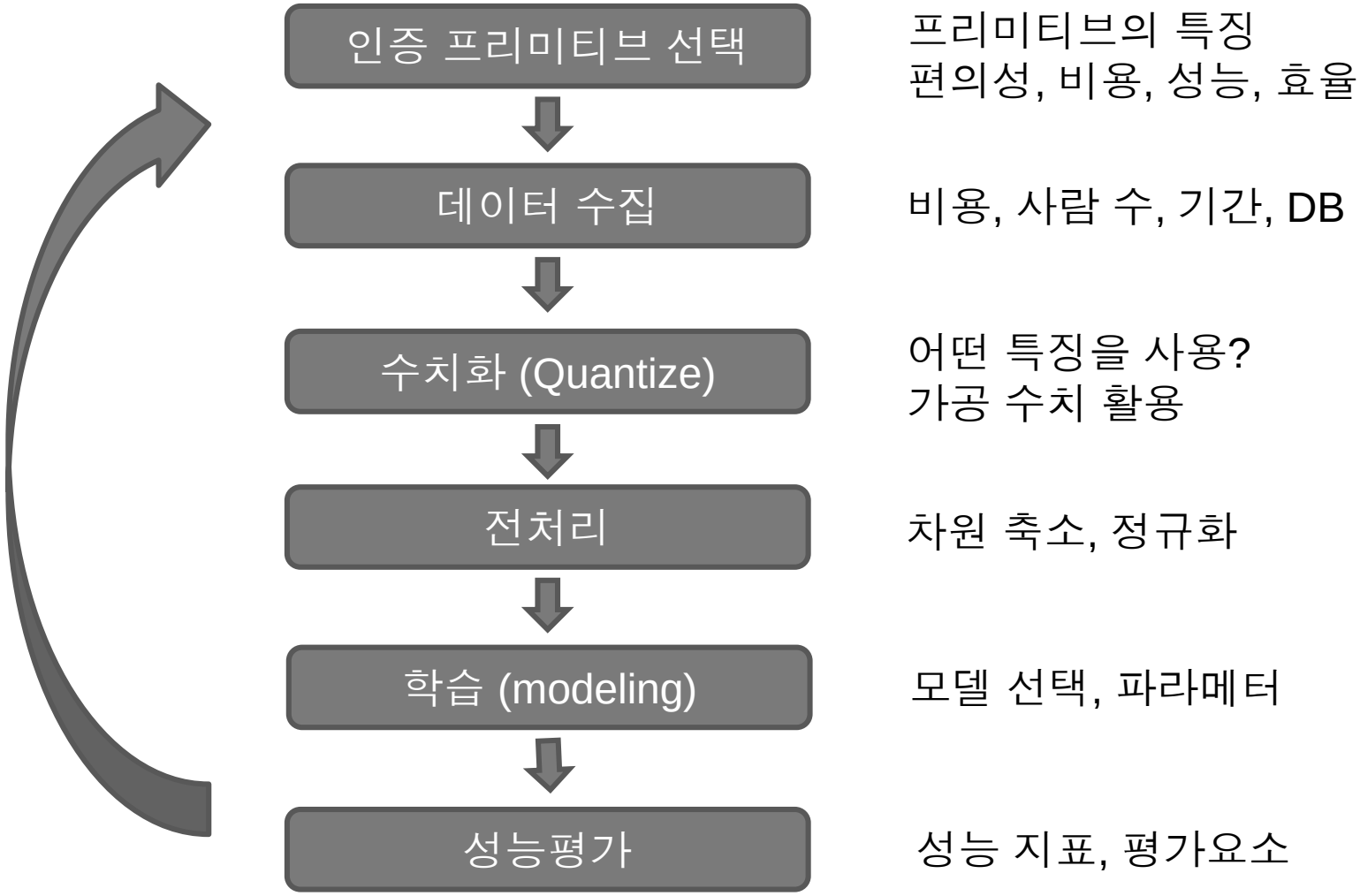
- AI 기반 사용자 인증
- 특징 데이터 수집
- 데이터 가공
- 모델 학습
- 성능 평가

AI 기반 사용자 인증

AI기반 사용자 인증 구조



개발 흐름



프리미티브의 특징
편의성, 비용, 성능, 효율

비용, 사람 수, 기간, DB

어떤 특징을 사용?
가공 수치 활용

차원 축소, 정규화

모델 선택, 파라메터

성능 지표, 평가요소

인증 프리미티브

사용자 인증 프리미티브

지식기반

패스워드
↓
그림, 패턴, 기억

생체기반

생체정보(지문, 홍채..)
↓
생체신호(뇌파, 심전도..)

소지기반

USB토큰, IC카드
↓
웨어러블, 알약, 임플란트

행위기반

키보드/마우스 이용패턴
웹 서핑 패턴
문체, **동적서명**

환경기반

사용자 단말
위치, 시간
주변영상, 주변 **전파신호**

인증 프리미티브

- 특성 비교, 고려요소

Factor	Universality	Uniqueness	Collectability	Performance	Acceptability	Spoofing
Password	n/a	L	H	H	H	H
Token	n/a	M	H	H	H	H
Voice	M	L	M	L	H	H
Facial	H	L	M	L	H	M
Ocular-based	H	H	M	M	L	H
Fingerprint	M	H	M	H	M	H
Hand geometry	M	M	M	M	M	M
Location	n/a	L	M	H	M	H
Vein	M	M	M	M	M	M
Thermal image	H	H	L	M	H	H
Behavior	H	H	L	L	L	L
Beam-forming	n/a	M	L	L	L	H
OCS	n/a	L	L	L	L	M
ECG	L	H	L	M	M	L
EEG	L	H	L	M	L	L
DNA	H	H	L	H	L	L

개발해본 PRIMITIVE

프리미티브	응용	스폰서
얼굴	얼굴인식공격	암호연구회
얼굴	온오프라인 티케팅	IITP
동적서명	간편결제 인증	IITP
문자확인동작	핀테크 보안	IITP
단말 핑거프린트	간편결제 인증	SKP
오픈웹 핑거프린트	플러그인 제거	금융보안원
뇌파	차세대인증	IITP
음성(음악)	히든싱어 찾기	-
Wifi	삼성페이 보안	IITP
주소	비대면 본인확인	IITP
웨어러블/주변환경	간편결제 인증	와이키키
얼굴, 자세, 환경	무자각인증	ETRI

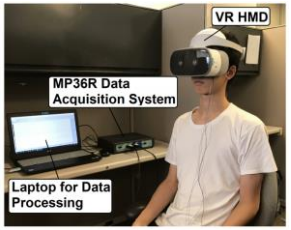
센서, 장치

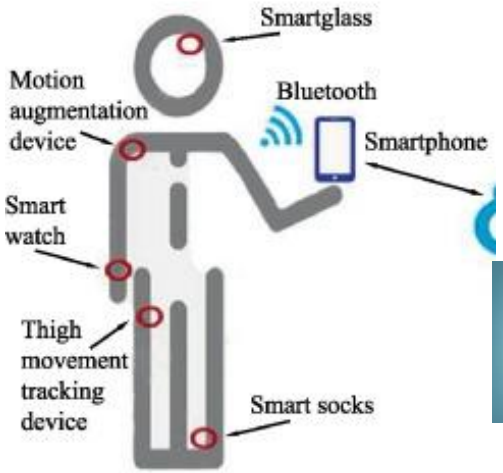
스마트폰 센서

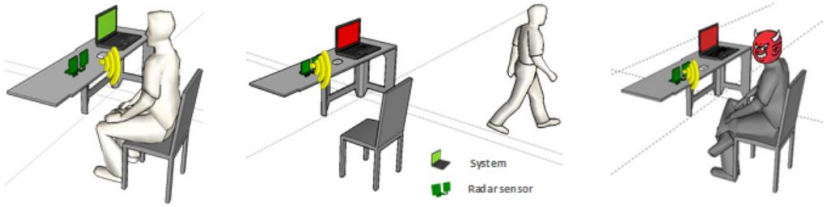

Proximity sensor
Ambient Light Sensor
CMOS Imager sensors
GPS sensor
Accelerometer
Humidity sensor
Microphone
Temperature sensor
Pressure sensor
Fingerprint sensor
Magnetometer
Gyroscope
Touch sensor

센서, 장치

별도 인증 장치





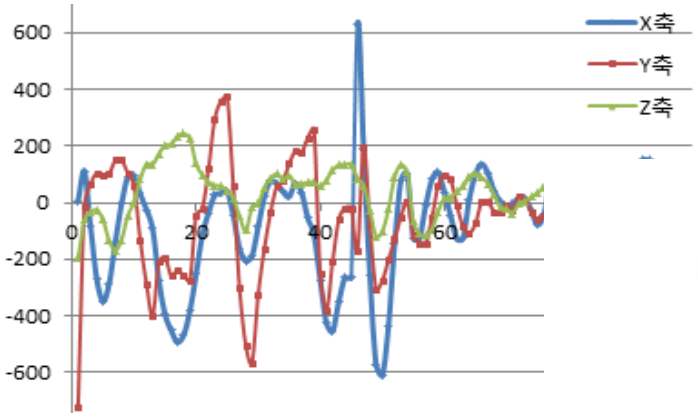


Wearable sensor

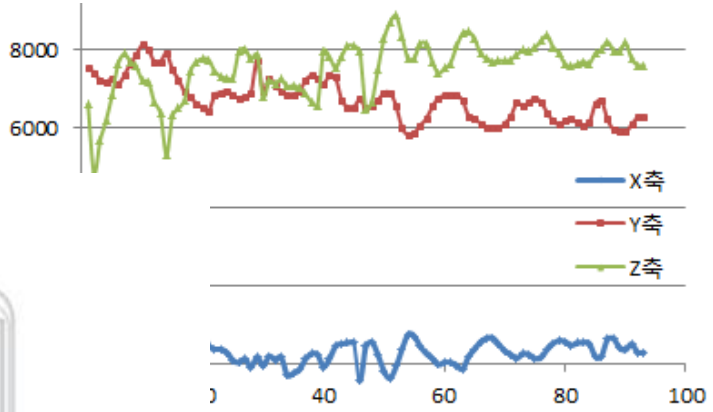
Non wearable sensor

인증 프리미티브

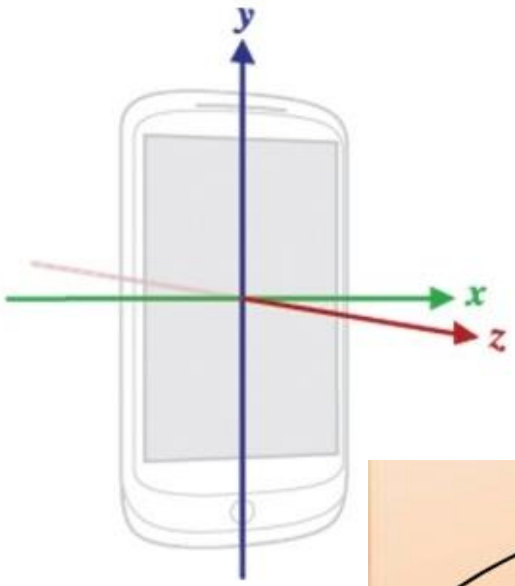
동적 서명

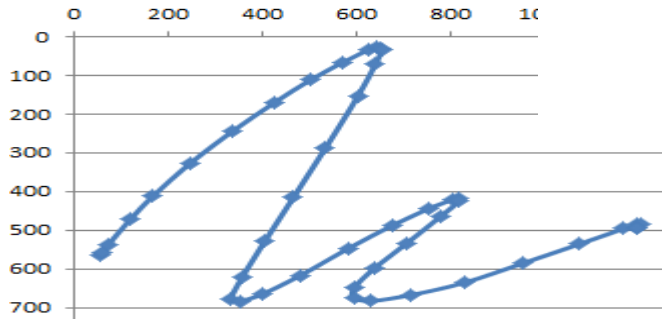


자이로 센서 측정



가속도 센서 측정





좌표 센서 측정

서명 이미지

인증 프리미티브

단말 환경 1

프리미티브	고유성	호환성	불변성	프라이버시
Canvas	상	상	중	상
IP	상	상	중	하
List of Plugins	상	상	하	상
WebRTC	상	중	상	하
HTTP Request Header	중	상	상	상
List of fonts	중	상	상	상
Screen Resolution (Screen size, color depth)	중	상	상	상
Time-Based	-	상	상	상
WebGL 3D	중	상	중	상
Font Metrics	중	상	중	상

인증 프리미티브

단말 환경2

프리미티브	고유성	호환성	불변성	프라이버시
HTTP User agent	중	**상**	하	**상**
AudioContext	중	중	**상**	**상**
Memory	하	**상**	**상**	**상**
Do Not Track	하	**상**	**상**	**상**
Flash enabled	하	**상**	중	**상**
Timezone	하	**상**	중	**상**
Use of local storage	하	**상**	중	**상**
Platform	하	상	중	**상**
Web GL Renderer	하	**상**	중	**상**
Web GL Vendor	하	**상**	중	**상**
Use of Adblock	하	**상**	하	**상**
Cookie enabled	하	**상**	하	**상**
CSS3	-	중	**상**	**상**

인증 프리미티브

단말 환경2

프리미티브	고유성	호환성	불변성	프라이버시
HTTP User agent	중	**상**	하	**상**
AudioContext	중	중	**상**	**상**
Memory	하	**상**	**상**	**상**
Do Not Track	하	**상**	**상**	**상**
Flash enabled	하	**상**	중	**상**
Timezone	하	**상**	중	**상**
Use of local storage	하	**상**	중	**상**
Platform	하	상	중	**상**
Web GL Renderer	하	**상**	중	**상**
Web GL Vendor	하	**상**	중	**상**
Use of Adblock	하	**상**	하	**상**
Cookie enabled	하	**상**	하	**상**
CSS3	-	중	**상**	**상**

데이터 수집

수집 이슈

- 비용과 시간
  - 사람 수
  - 다양한 환경
  - 시차나 기간
- DB / Challenge 활용

LAB FACE DATASET

- Lab Face Datasets
  - 셀피
  - 수집 기간: 2019.04.01 ~ 2019.10.01
  - 사람 당 이미지 수: 440 ~ 568
  - 각도, 장소 변화

얼굴 사진 수집

- 인터넷 유명인 5000명 * 평균 200장
  - ☞ 5000명 명단 구축도 어려움 (나무위키 등등..)
  - 구글링 -> 이미지 다운로드 : 200자 이상 있는 경우도 많지 않음
  - 얼굴 영역 추출 및 150 x 200 사이즈로 가공
- 필터링
  - 20명의 알바생으로 2중 필터링 : 100% 정확하기 어려움
    - 서양인은 같은 사람인지 인식하기도 어려우
  - 가능 : 웃는 사진, 옆 모습이지만 양 쪽 눈이 다 나옴
  - 불가 : 얼굴위로 글자, 얼굴이 망가진 경우, 앞머리가 한쪽 눈을 가림





DB K-FACE

AI 학습용 한국인 안면 이미지 데이터

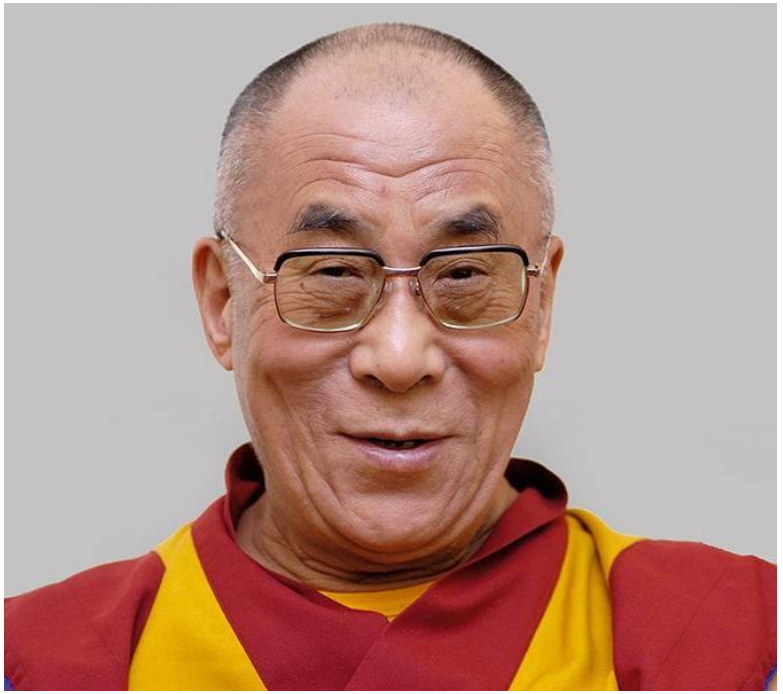
- 100명
- 144,000 [1인 : 1,440장]
- 데이터 특징 : 다양한 환경 반영
  - 표정 [3종], 소품 [6 종], 각도 [5 종], 밝기 [16 종]

4_S001_L02_E01_C06

4_S001_L08_E03_C09

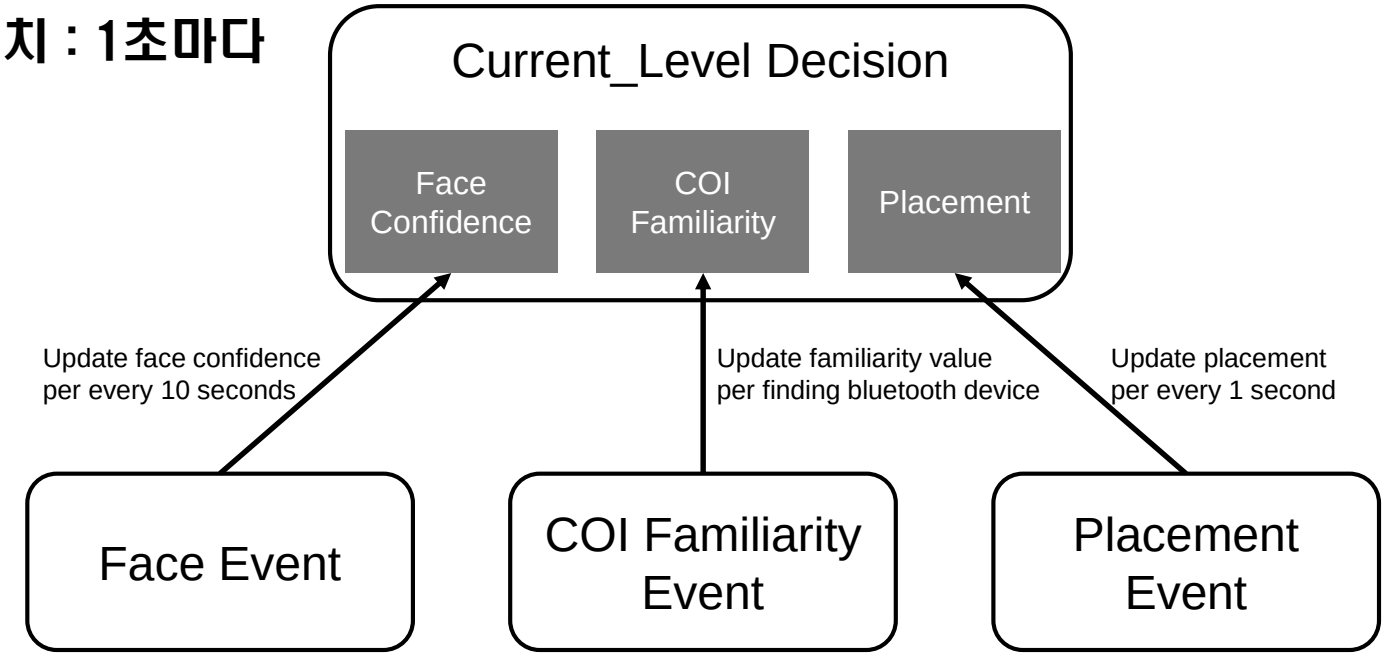
DB VGGFACE

- VGGFace2
  - 인물 수 : 9,131
  - 총 이미지: 3,310,000
  - 인물 당 평균 이미지 수: 362
  - 이미지 크기 → 제 각각



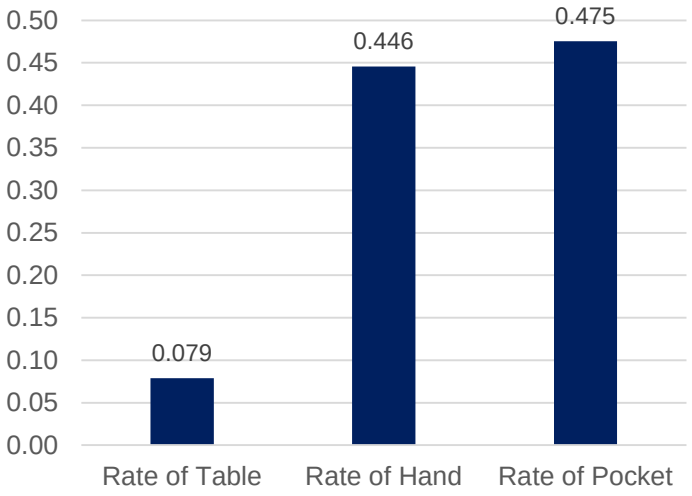
무자각 인증 데이터 수집

- 스마트폰 사용자들 2달 간의 지속적 모니터링
  - 22명(삼성 phone 19명, LG phone 3명) : 가공 데이터 895M
- 주요 구성
  - 얼굴 : 10초마다
  - 주변기기 : BT 기기 발견 마다
  - 폰 위치 : 1초마다

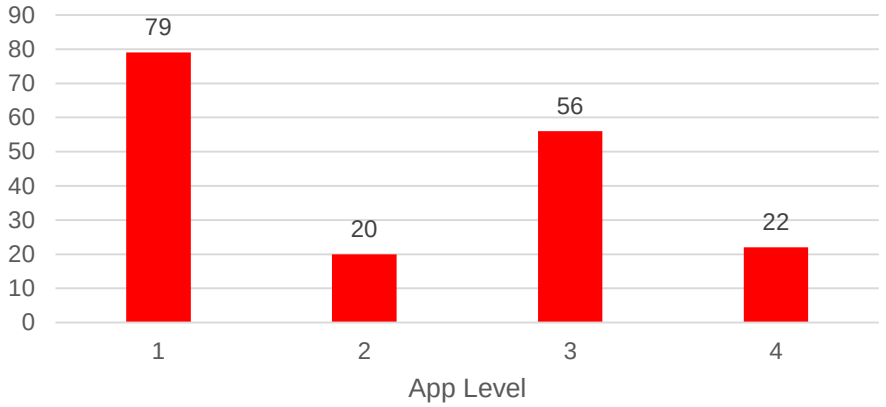


무자각 인증 데이터 통계

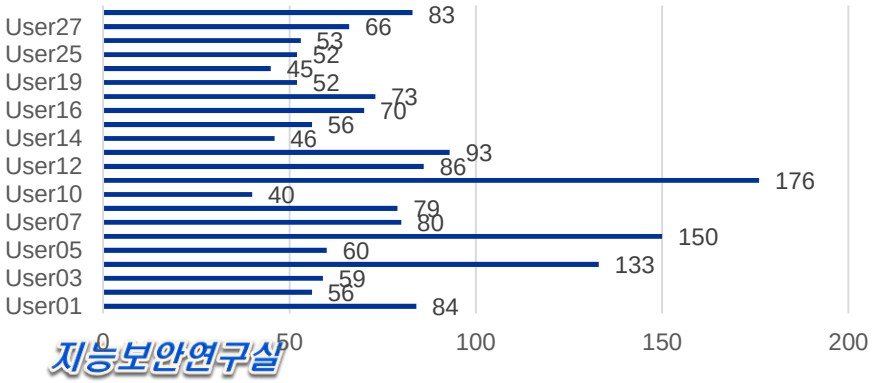
- Placement 각 항목 비율



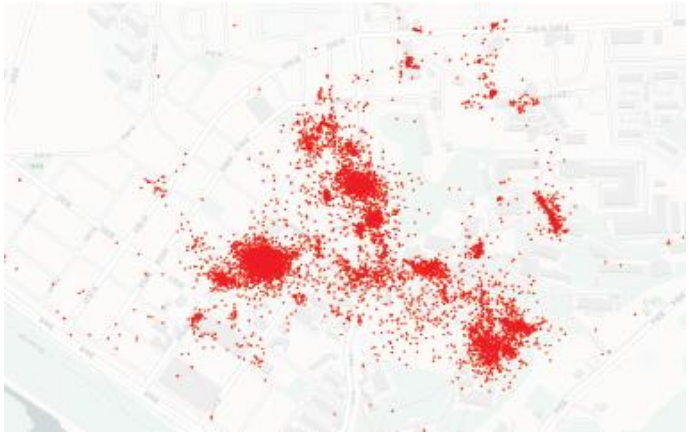
- Screen Off한 후 사용한 App의 인증수준



- 사용자 별 app 사용 갯수



- 주소 인증을 위한 이동 패턴



수집 이슈

- 수집을 끝까지 수행하지 못하고 도중에 포기하는 사람이 많음.
  - 발열이 심해 폰에 이상이 생기지 않을까 걱정
  - 배터리 수명에 대한 걱정
  - 개인정보 수집에 대한 거부감
  - 큰 용량의 데이터를 보내주는 것에 많이 어려워함

환경 기반 인증 실험 데이터

Wifi 기반 위치 인증 실험용

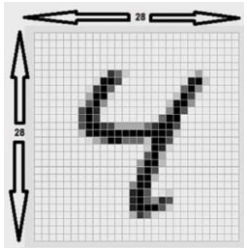
- 7개의 store, store 마다 6곳의 위치
- 7일치 측정

데이터 가공

수집 RAW 데이터

이미지

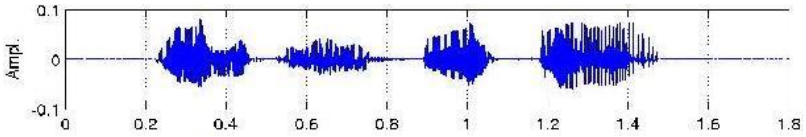
- 얼굴, 서명
- 1-D, 2-D, 3-D



00000000
00010010
00100100
00111100
00001000
00000100

특징 벡터 X

{0, 0, 0... 1, 0,0,1
,0,0,1,1,1,.. }

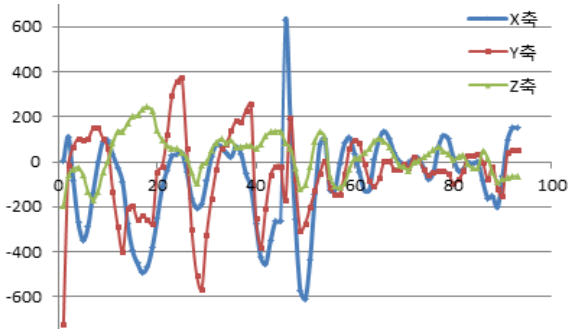


Sampling data

{150, -13, -200... 30, 170, 40, -30, -180 …

시퀀스

- 뇌파, 음성, 서명(행위), 위치
- Sequence of 1-D, 2-D



동적서명의 sequence data

수집 RAW 데이터

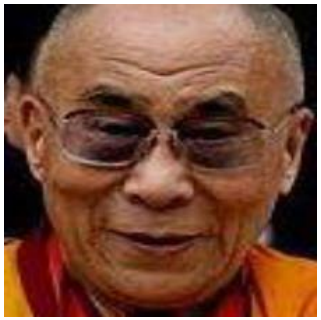
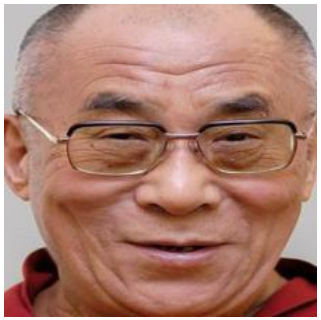
- 트랜잭션
  - 로그, RDB 레코드
  - 필드
  - 수치형 데이터
  - 명목형 데이터

범주형	수치형
Time	Host / Ip
Language	Screen width
Platform	Screen height
Donottrack	Screen depth
Cookies enabled	Canvas hash
List of plugins	
User agent	
Webgl vender	
Webgl renderer	
List of font	

특징 추출

Raw data에서 일부 추출

- 얼굴 영역 Detection 및 Alignment
  - Multi-task CNN(MT CNN) 사용 : 이미지 크기: 224X224X3 (margin: 20 포함)

  - Pix2pix : GAN을 이용, 얼굴영역이 아닌 곳을 검게 칠함



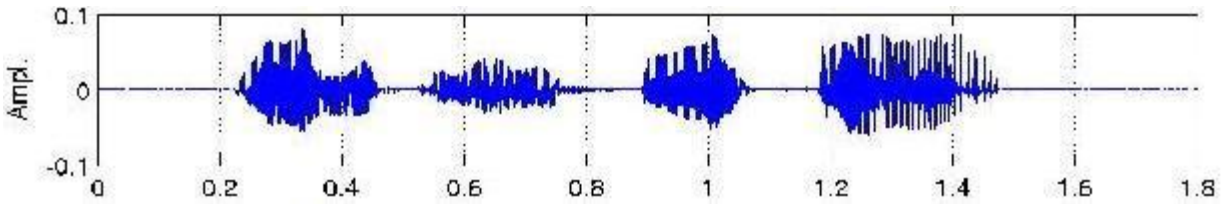

특징 추출

파형

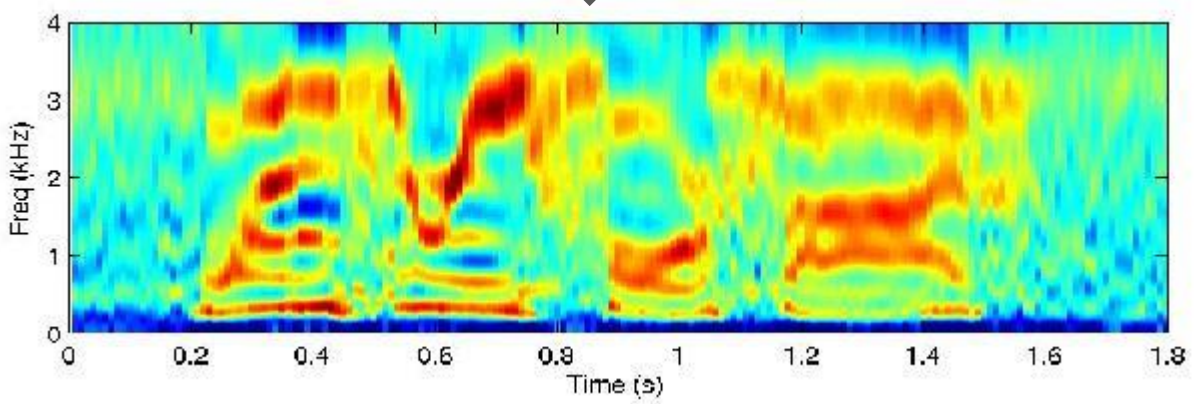
- 주파수 분할
- MFCC

☞ 샘플링 데이터 1-D => 주파수 별 값 2-D

음파



특징추출



특징 추출

- Raw data를 가공해서 만들어 냄
  - Set of data 의 기술 통계값 : max, min, median, freq, time, interval …

Extracted feature	Description
# of transactions	Number of transactions
# of payment_methods	Number of payment methods
# of countries	Number of payment countries
Avg(charge_amount)	Average of charge_amounts
Avg(purchase_amount)	Average of purchase_amounts

코딩

▶ 범주형 데이터의 수치화

- FC1 (Label Encoding) : 필드마다 고유한 모든 값에 0부터 번호 부여

☞ 값의 크고 작고가 의미가 없음

- FC2 (One Hot Encoding) : 값의 개수만큼 차원을 만들어 값에 따라 해당 위치만 1로 세팅
- FC3 (변형 One Hot Encoding) : 사용자 별로 모델이 존재하므로, 사용자가 사용한 값의 개수만큼 차원을 만듦

Field 1	Field 1 Label Encoding	Field 1 One Hot	Field 2	Field 2 Label Encoding	Field 2 One Hot	Features that user used	One Hot	Modified One Hot Encoding
'A1'	0	10000	'B1'	0	1000	Field 1: 0, Field 2: 3	1000000 01	10000
'A2'	1	01000	'B2'	1	0100	Field 1: 2, Field 2: 1	0010001 00	00101
'A3'	2	00100	'B3'	2	0010	Field 1: 1, Field 2: 0	0100010 00	01010
'A4'	3	00010	'B4'	3	0001	Field 1: 4, Field 2: 2	0000100 10	00000
'A4'	4	00001						

결측치 처리

▶ 데이터에 nan값이 많아서

[2019-10-09,nan,nan,……,nan]

[arial,arial black,……,wide latin]

->결측값을 0으로 채우고 , 0을 만나면 건너뛰는 형식으로

```
[12, 'canvas hash', '1032907322', 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
```

[[1, 'time', '2019-11-04 16:34'], [2, 'language', 'ko-KR', 'ko', 'en-US', 'en'],[3, 'Platform', 'Win32'],[4, 'doNotTrack', 'Unspecified'], […]]

NORMALIZE

- 서명의 정규화
  - 시작점, 크기, 길이
- 트랜잭션 data
  - 스트링 : 공백 제거
  - steming

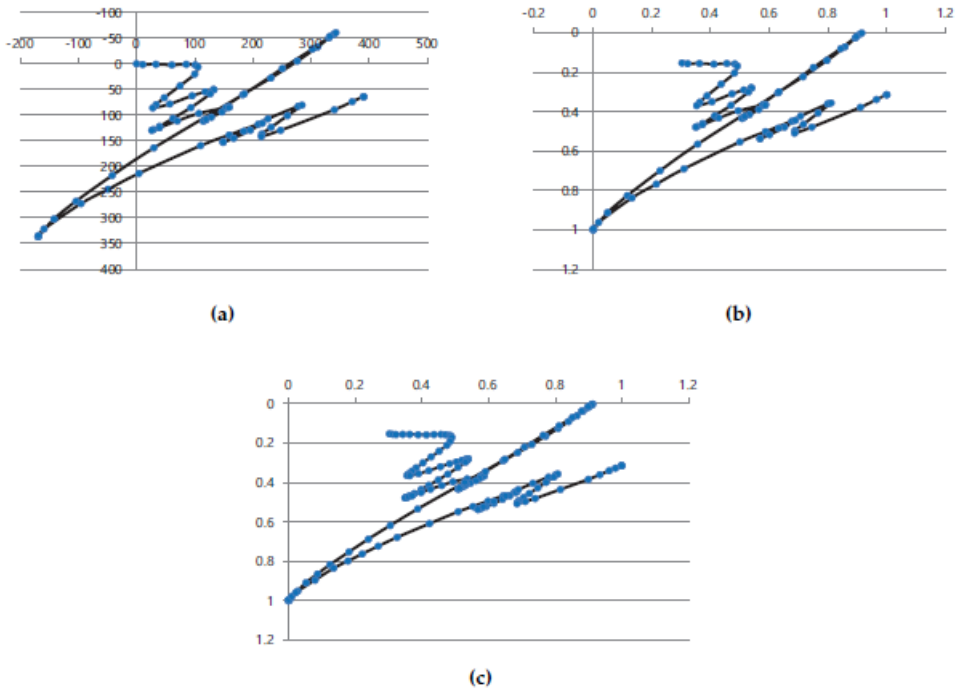
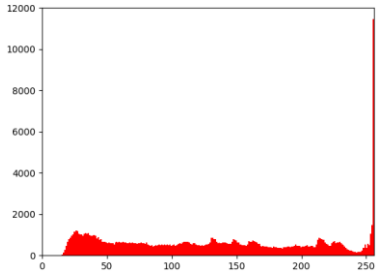
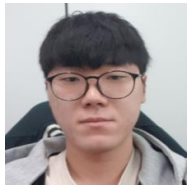


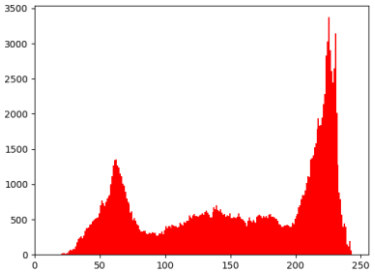
Figure 2. Data normalizations of (**a**) alignment, (**b**) size, and (**c**) length

이미지 정규화

- Alignment
- 밝기 조정
  - 히스토그램을이용한 이미지 평활화
    - 원본 이미지의 히스토그램을 만듦
    - 특정 명암에 대한 빈도수를 계산
    - 빈도수를 기반으로 정규화 시행
    - 정규화된 값을 히스토그램으로 만듦
    - 이미지에 적용



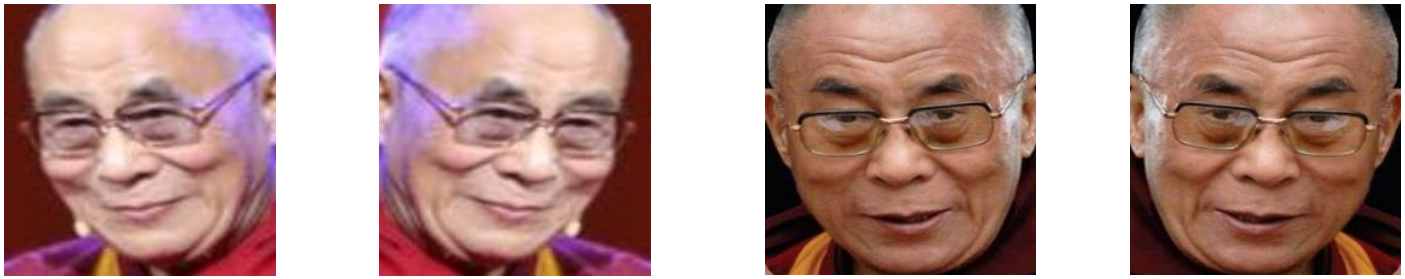



모델링

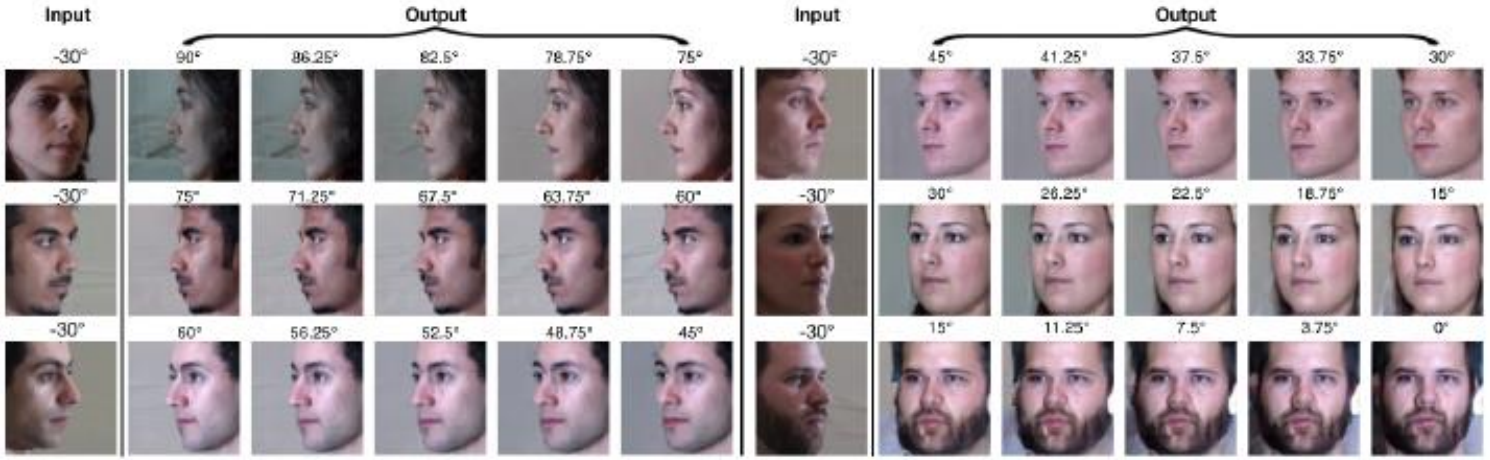
학습데이터 보강

Data Augmentation

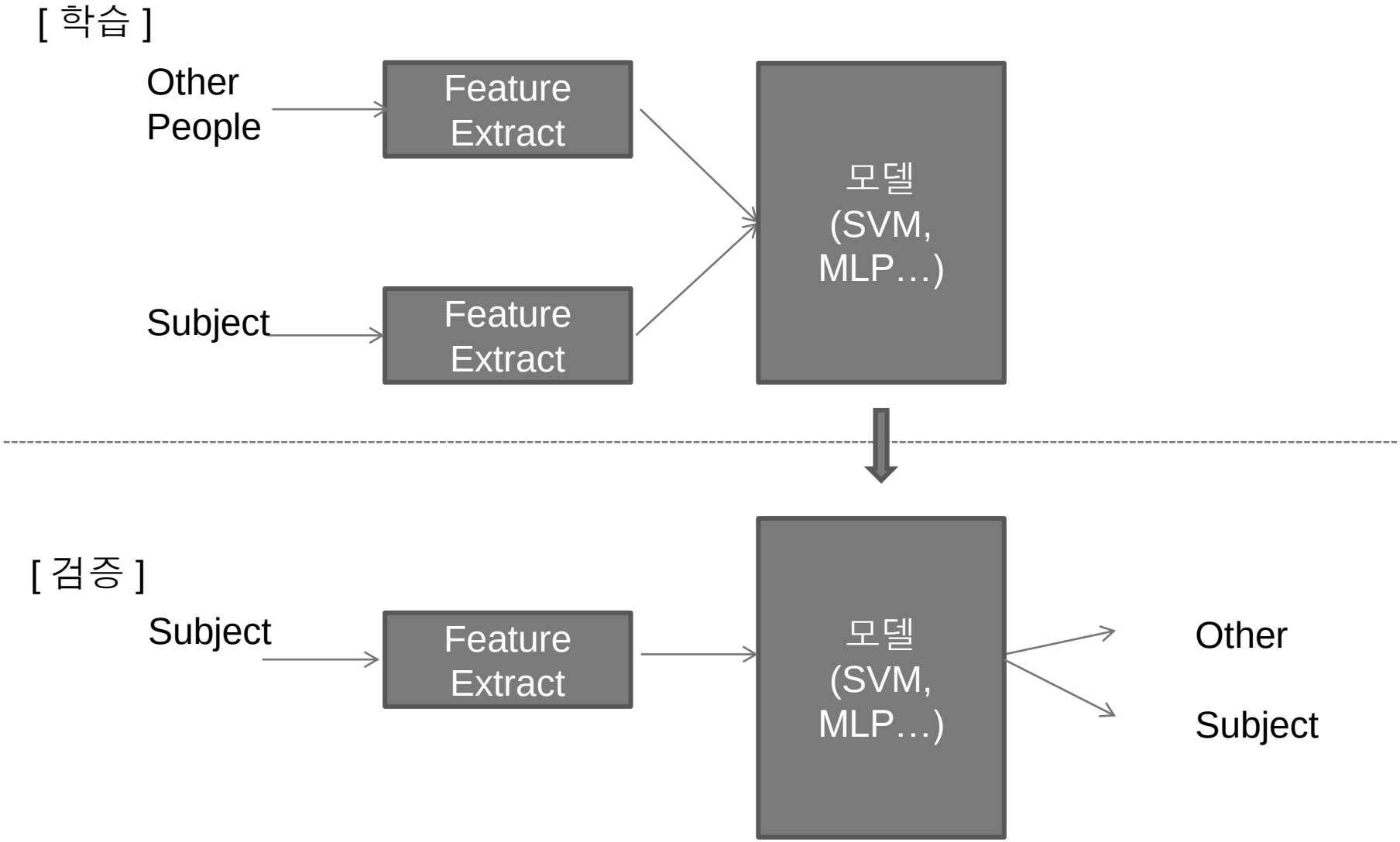
- 좌우 반전 이미지 추가 사용



- GAN으로 각도, 밝기 다른 데이터 생성



인증 모델1 – 2 CLASS 분류기

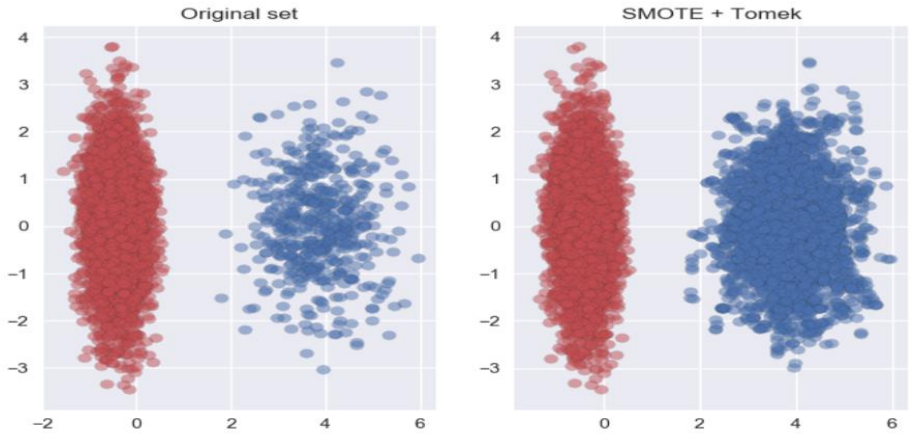


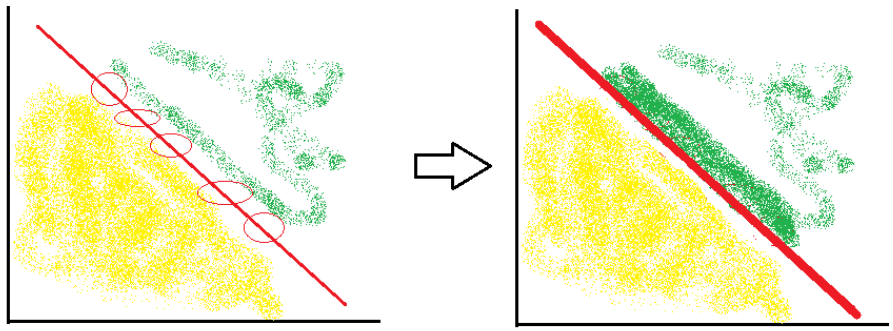
분류 모델

- ML
  - SVM, random forest, KNN, Decision Tree
- DNN
  - MLP
  - RNN
  - CNN
- 파라메터 튜닝
  - Search problem

클래스 불균형

- 서명 Data
  - 1인당 20개의 서명 => 10개 train, 10개 test
  - 20인
- 2 class training data
  - Subject :10개
  - others : 190개
- 해법 : oversampling, undersampling, data 생성 (SMOTE, Balancing GAN…)

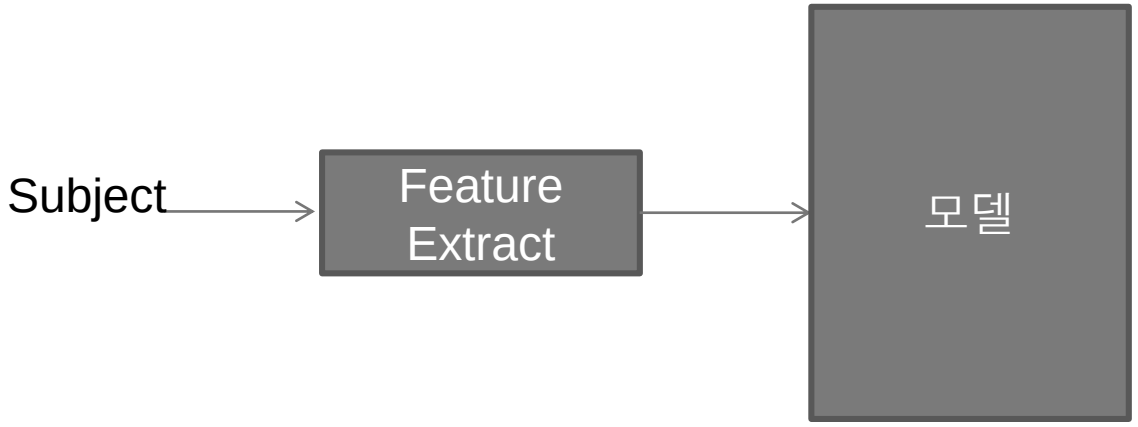




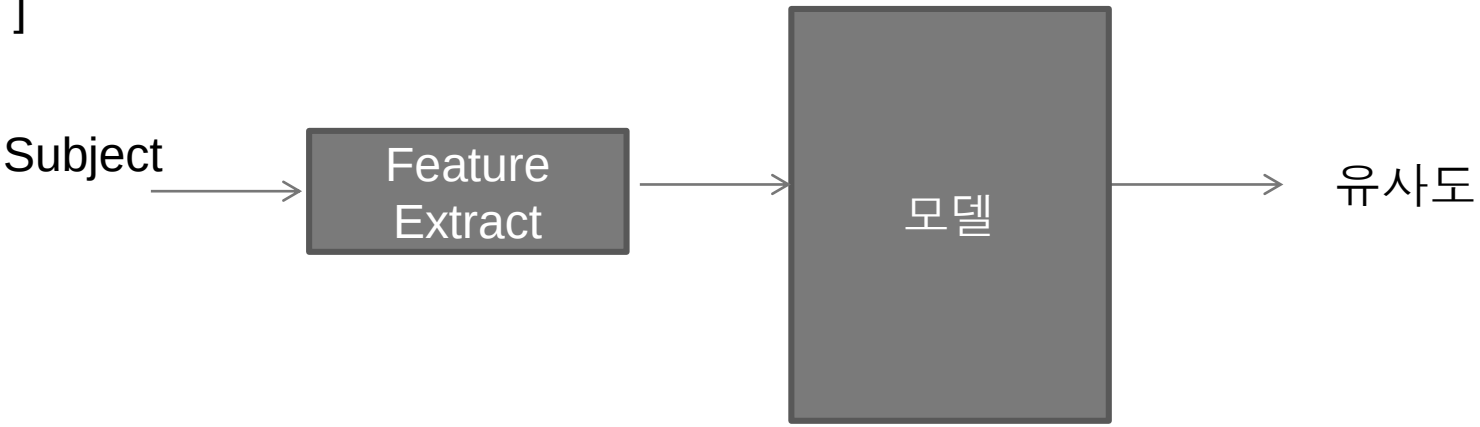
Tomek's-Linked Method + Balancing GAN

인증 모델2 – 1 CLASS SCORER

[Train]



[Verify]



유사도

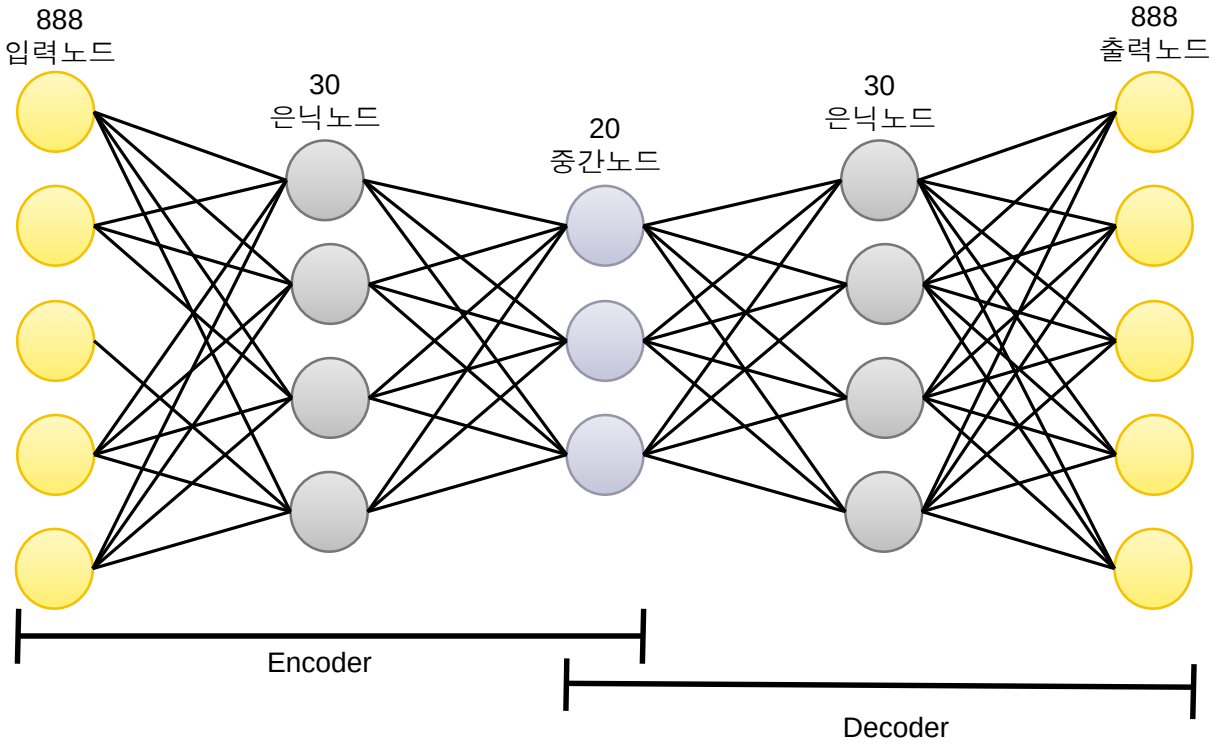
모델과 유사도

- 모델 : 1-train data, 평균, 가우시안믹스쳐
- 유사도 : cosine similarity, Jaccrad, Levenstein

Similarity 방법	핑거프린팅 항목
cosine	platform
	donottrack
	cookise
	canvas
	screen
jaccard	font
	useragent
	webGL
levenshtein	language

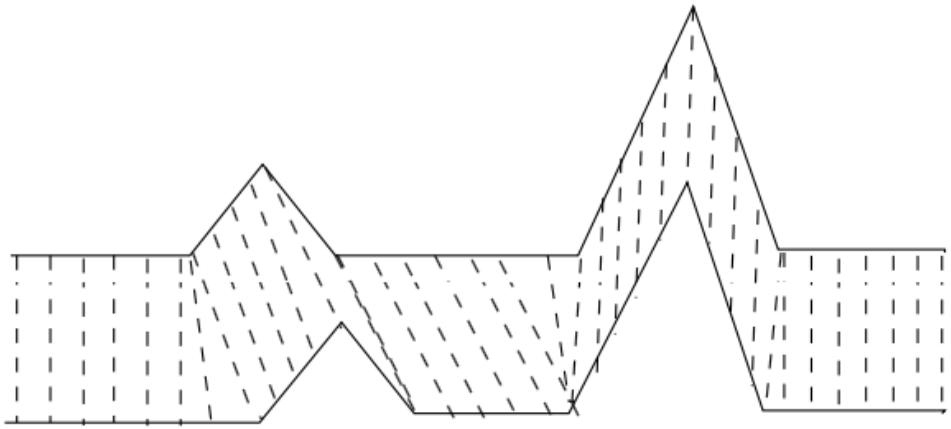
1CLASS SCORER – AUTO ENCODER

- 재 생성 네트워크
  - 학습한 것은 잘 재생성
  - 그렇지 않은 것은 잘못함



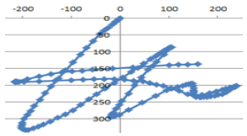
시퀀스 데이터 유사도

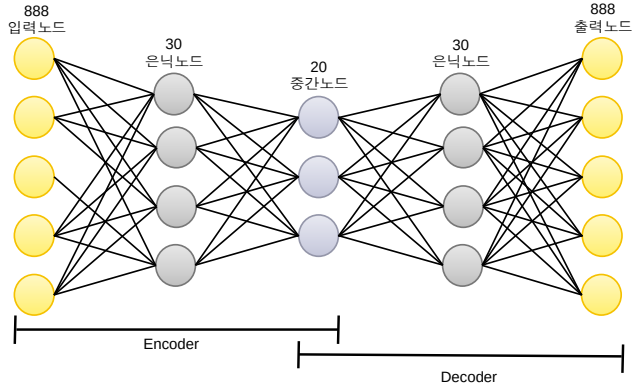
- DTW (Dynamic Time Wraping)
  - 시간적으로 일치하지 않은 2개의 파형의 유사도를 산정



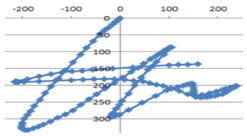
인증 모델2

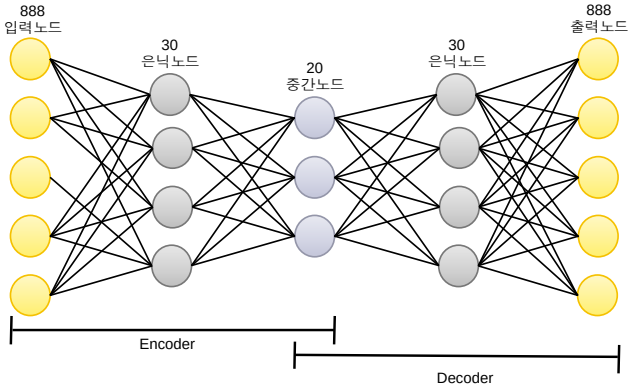
[Train]

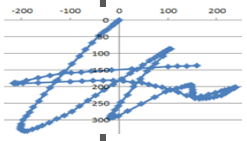




[Verify]



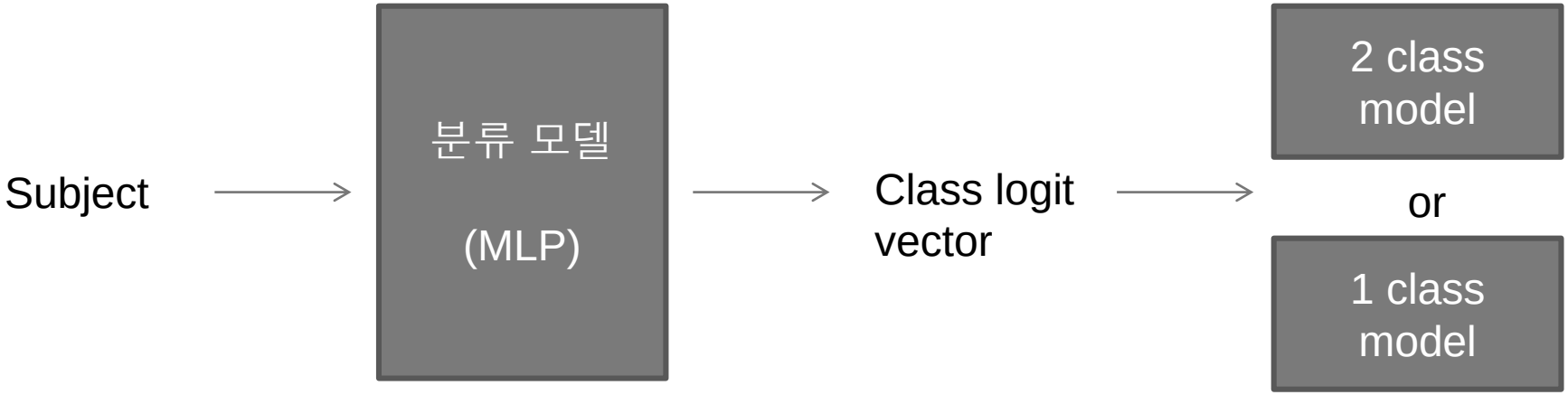




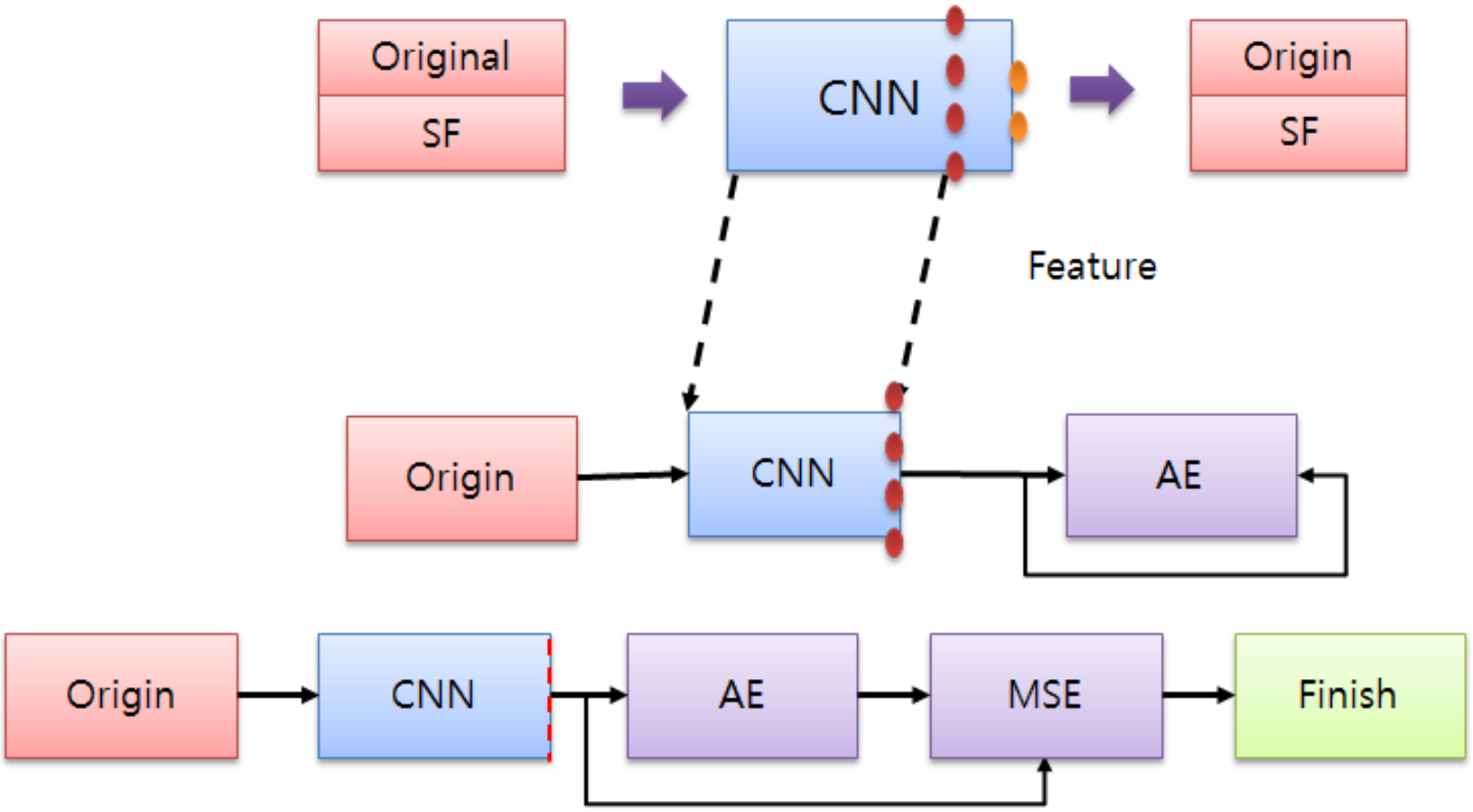
비교기

유사도

CLASSIFIER AS FEATURE EXTRACTOR



특징 추출기로서의 CNN



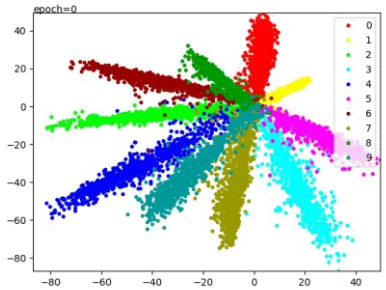
얼굴 인식기 SOTA

ResNetV2-50 사용

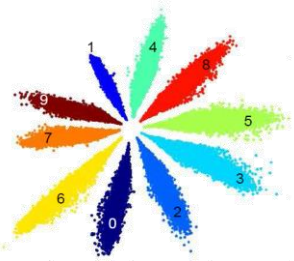
Loss Function:

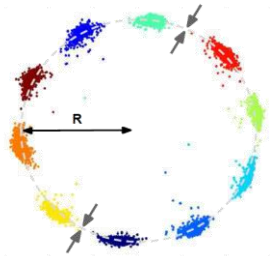
SphereFace(CVPR, 2017)

Ring Loss(CVPR 2018)



SphereFace





Ring Loss

Intra-class variation compactness

동일 인물로부터 추출된 특징의 분산은 작게 함

Inter-class variation separability

다른 인물로부터 추출된 특징의 분산은 크게 함

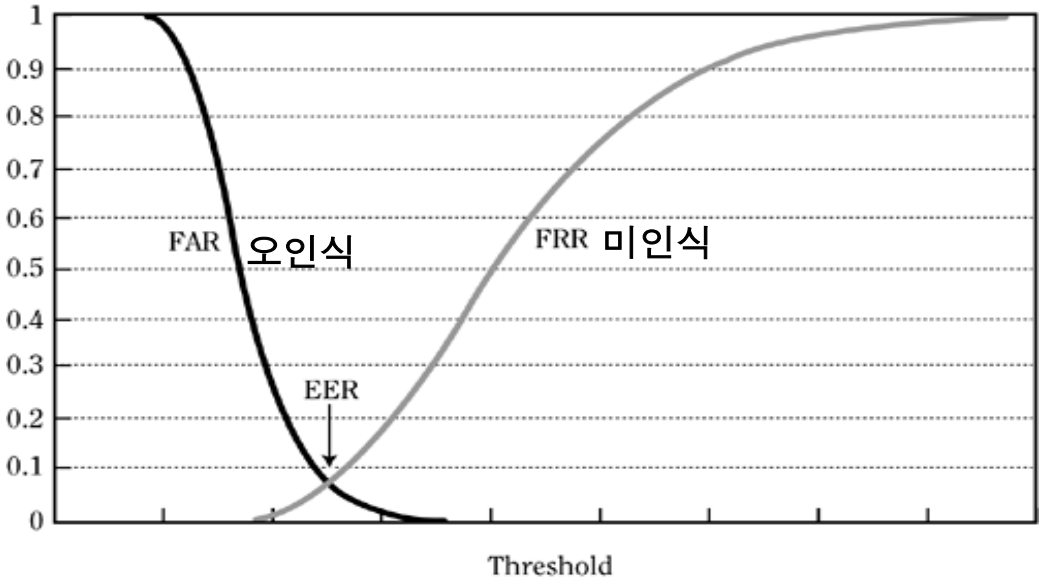
TABLE IV
THE ACCURACY OF DIFFERENT VERIFICATION METHODS ON THE LFW DATASET.

Method	Public. Time	Loss	Architecture	Number of Networks	Training Set	Accuracy±Std(%)
DeepFace [195]	2014	softmax	Alexnet	3	Facebook (4.4M,4K)	97.35±0.25
DeepID2 [187]	2014	contrastive loss	Alexnet	25	CelebFaces+ (0.2M,10K)	99.15±0.13
DeepID3 [188]	2015	contrastive loss	VGGNet-10	50	CelebFaces+ (0.2M,10K)	99.53±0.10
FaceNet [176]	2015	triplet loss	GoogleNet-24	1	Google (500M,10M)	99.63±0.09
Baidu [124]	2015	triplet loss	CNN-9	10	Baidu (1.2M,18K)	99.77
VGGface [149]	2015	triplet loss	VGGNet-16	1	VGGface (2.6M,2.6K)	98.95
light-CNN [225]	2015	softmax	light CNN	1	MS-Celeb-1M (8.4M,100K)	98.8
Center Loss [218]	2016	center loss	Lenet+-7	1	CASIA-WebFace, CACD2000, Celebrity+ (0.7M,17K)	99.28
L-softmax [126]	2016	L-softmax	VGGNet-18	1	CASIA-WebFace (0.49M,10K)	98.71
Range Loss [261]	2016	range loss	VGGNet-16	1	MS-Celeb-1M, CASIA-WebFace (5M,100K)	99.52
L2-softmax [157]	2017	L2-softmax	ResNet-101	1	MS-Celeb-1M (3.7M,58K)	99.78
Normface [206]	2017	contrastive loss	ResNet-28	1	CASIA-WebFace (0.49M,10K)	99.19
CoCo loss [130]	2017	CoCo loss	-	1	MS-Celeb-1M (3M,80K)	99.86
vMF loss [75]	2017	vMF loss	ResNet-27	1	MS-Celeb-1M (4.6M,60K)	99.58
Marginal Loss [43]	2017	marginal loss	ResNet-27	1	MS-Celeb-1M (4M,80K)	99.48
SphereFace [125]	2017	A-softmax	ResNet-64	1	CASIA-WebFace (0.49M,10K)	99.42
CCL [155]	2018	center invariant loss	ResNet-27	1	CASIA-WebFace (0.49M,10K)	99.12
AMS loss [205]	2018	AMS loss	ResNet-20	1	CASIA-WebFace (0.49M,10K)	99.12
Cosface [207]	2018	cosface	ResNet-64	1	CASIA-WebFace (0.49M,10K)	99.33
Arcface [42]	2018	arcface	ResNet-100	1	MS-Celeb-1M (3.8M,85K)	99.83
Ring loss [272]	2018	Ring loss	ResNet-64	1	MS-Celeb-1M (3.5M,31K)	99.50

성능 평가

보안성 = 인식 정밀도 ?

- 인식 정밀도 != 도용(사고) 가능성



- 바이오 정보 복제 가능성
- 인증 체계의 보안성
  - Primitive security : EER, multi-factor, liveness
  - Platform security : Samsung Knox, TPM, TEE
  - Protocol security : SSL 취약성..

다양한 환경 테스트를 위한 데이터

Data set

- Multi-PIE
  - 얼굴 인식에 대해 연구하기위해 사용
  - 20개 조명 변화, $\pm 90^{o}$, 337개의 대상, 6개 포즈
  - 200명 학습, 137명 test사용
- IJB-A
  - 자세변화가 큰 데이터
  - 5,396이미지, 20,412개의 비디오 프레임
- CASIA Web Face
  - 10573개의 대상, 494,414 이미지

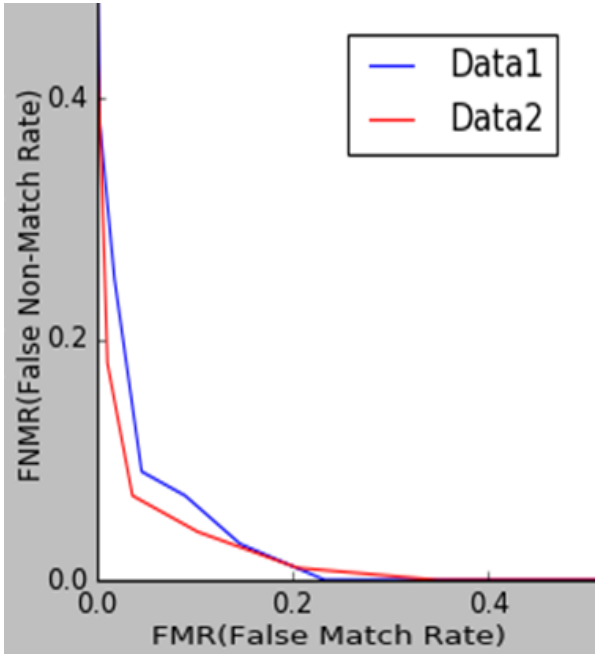
실험 결과

단일 세션 (HTER)

- 같은 날 받은 서명 중 일부를 training
- 일부를 test

	Data 1	Data 2
AE	5%	3.5%

	SVM	MLP	AE
Data2	4.7%	3.8%	3.5%



바이오 인증 이슈

- 등록과 인증 시점의 차이
  - 행위 기반은 특히 더 차이가 많이 남

- 위조 데이터 (보고 따라하기, 가짜 지문/홍채..)
  - 현재, 바이오 인증의 가장 큰 이슈

LIVENESS DETECTION

LivDet 2015 Contest

- 데이터

Dataset	Live Image	Ecoflex	Gelatine	Latex	WoodGlue	Liquid Ecoflex	RTV
Green Bit	1000	250	250	250	250	250	250
Biometrika	1000	250	250	250	250	250	250
Digital Persona	1000	250	250	250	250	250	250
	Live Image	**Body Double**	**Ecoflex**	**Playdoh**	**OOMOO**	**Gelatin**	**-**
Crossmatch	1500	300	270	281	297	300	-

- 결과

Algorithm	1	2	3	4	Overall
nogueira	95.40	94.36	93.72	98.10	95.51
unina	95.80	95.20	85.44	96.00	93.23
jinglian	94.44	94.08	88.16	94.34	92.82
anonym	92.24	92.92	87.56	96.57	92.51
titanz	91.76	92.36	89.04	91.62	91.21
hbirkholz	91.36	93.40	88.00	89.93	90.64
hectorn	90.00	88.20	84.20	86.94	87.32
CSI_MM	86.56	87.84	75.56	89.99	85.20
CSI	82.12	83.20	76.20	88.33	82.71
COPILHA	72.76	75.64	79.96	69.00	74.11
UFPE II	87.68	71.24	75.44	61.16	73.33
UFPE I	82.56	64.32	78.36	59.97	70.82

실험 결과2

시간차

- 등록 후 한달 경과

	Data 1	Data 2
MLP	8.2%	9.1%
AE	9%	7.7%

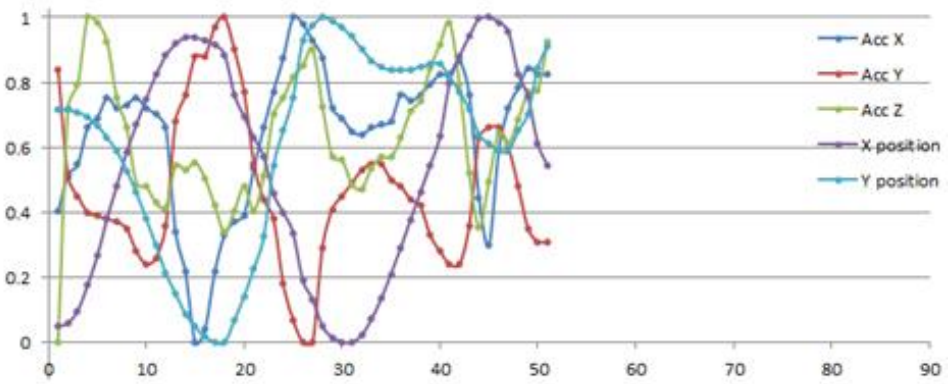
위조서명

- 정상 서명을 보고 그린 것

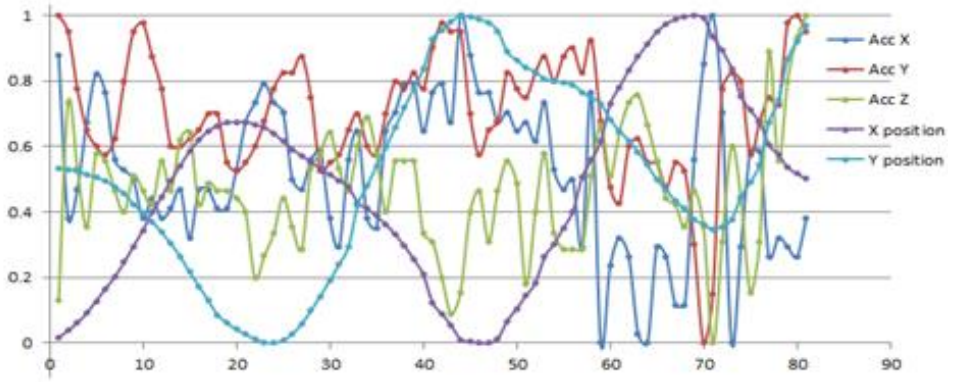
	Data 1	Data 2
MLP	25%	18.1%
AE	19.3%	13.7%

위조서명을 잡자

원본서명

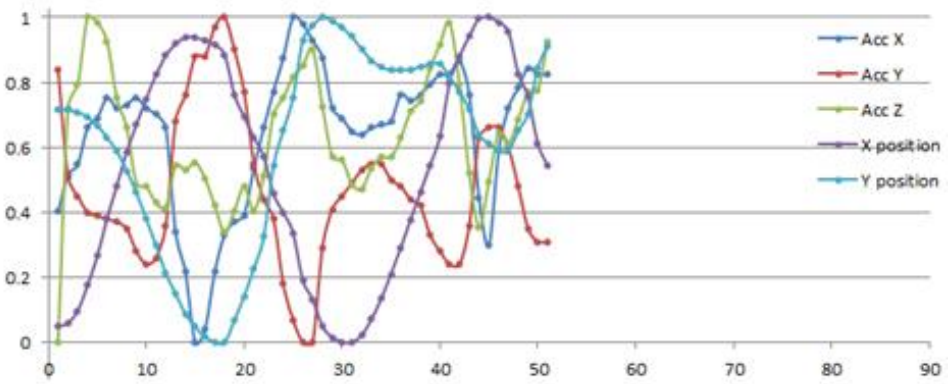


위조서명

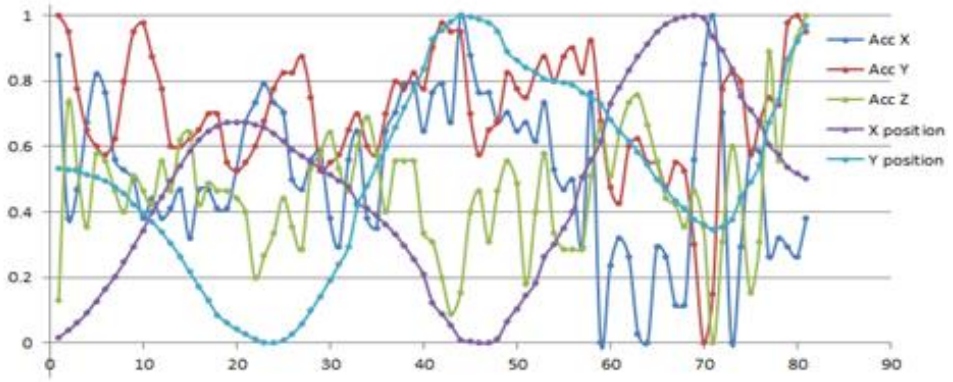


위조서명을 잡자

원본서명



위조서명



EXPERIMENT RESULT

CNN 자체의 분류성능

- A: all subjects vs. all skilled forgery
- B: 20 classes
- C: subject vs. others
- D: subject vs. skilled forgery for the user

	A	B	C	D
단일세션	10.1%	35.2%	8.3%	8.4%
시간차	13.7%	27.6%	9.6%	8.8%
위조서명	23.9%	11.6%	10.2%	3.3%

CNN 특징 추출 + AE 사용자모델

	A	B	C	D
단일세션	18.1%	9.1%	4.1%	3.2%
시간차	23.7%	11.3%	4.5%	4.1%
위조서명	4.3%	23.3%	4.6%	3.7%

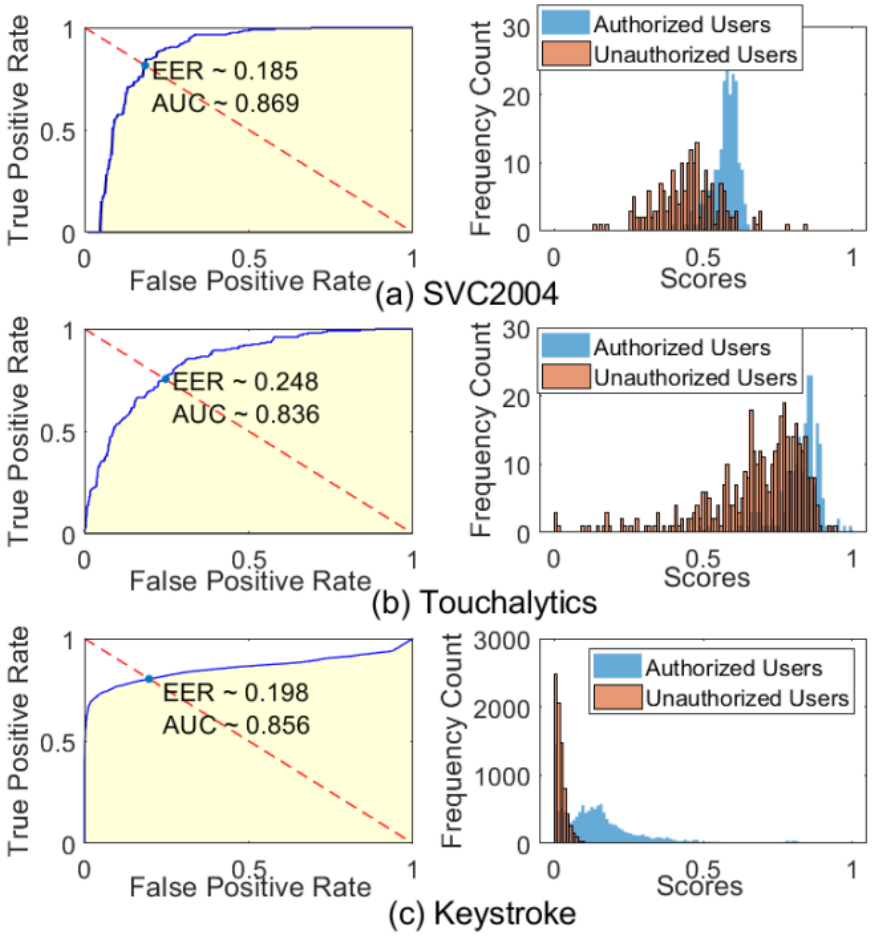
보안성 MEASURE

- 머신러닝 기반 인증 method에 대한 measure
  - Robust Performance Metrics for Authentication System, NDSS2019

- 기존 measure 들은 실험대상자의 쏠림, threshold, trade-off등을 잘 반영하지 못함

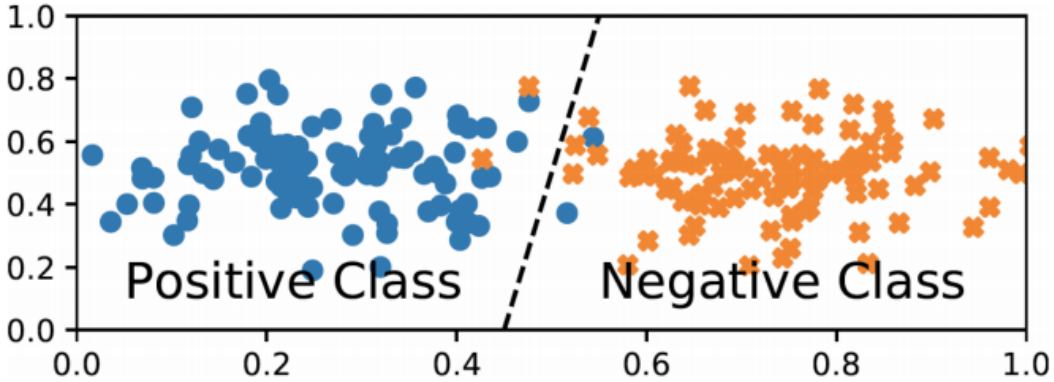
☞ FCS : 진짜 사용자와 공격자의 score 분포 히스토그램

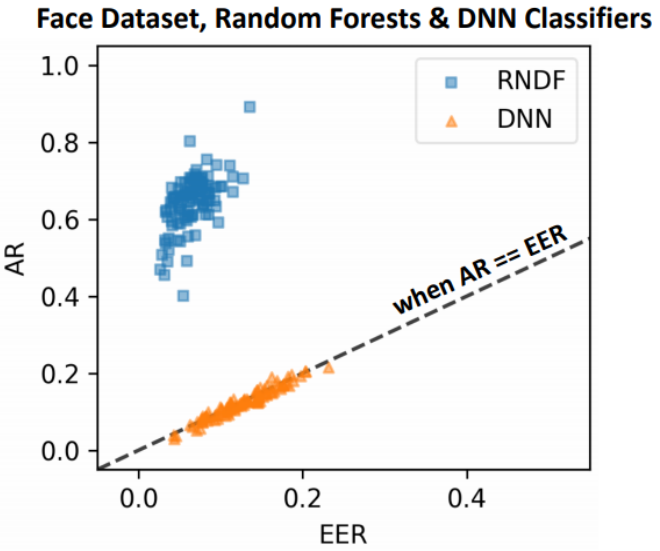
☞ 비슷한 EER, AUC를 갖고 있는 ROC 커브들에 대해서 중첩영역이 가장 작은 score를 만들 수 있는 (a) 방법이 우수

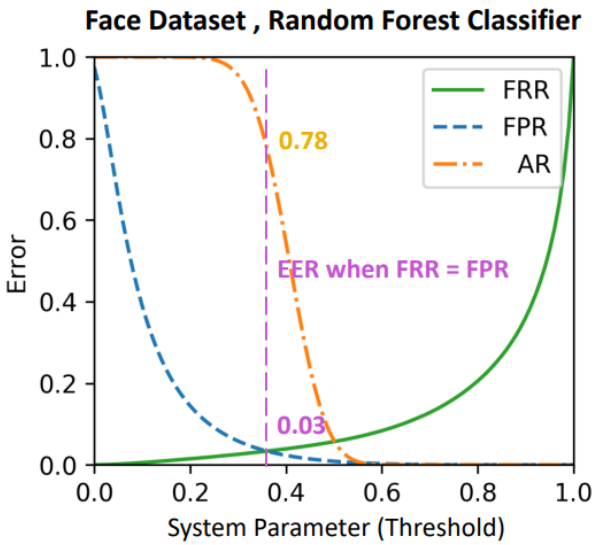


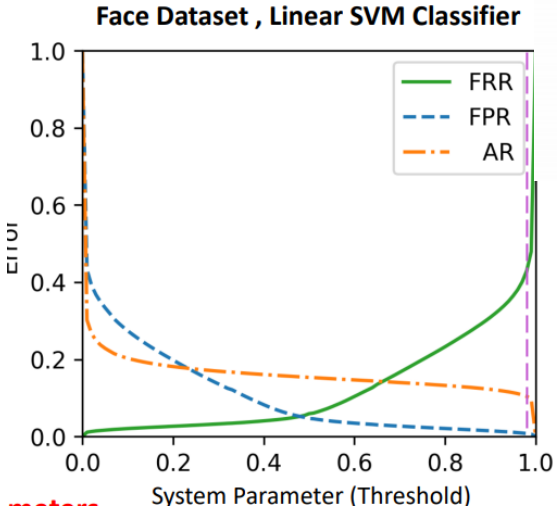
RANDOM INPUT ATTACK

- Positive 면적 만큼의 성공확률



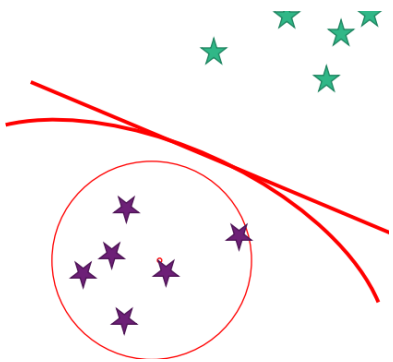


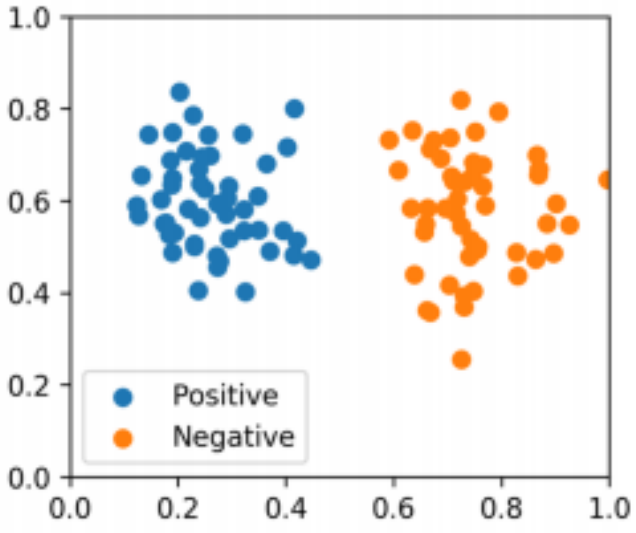


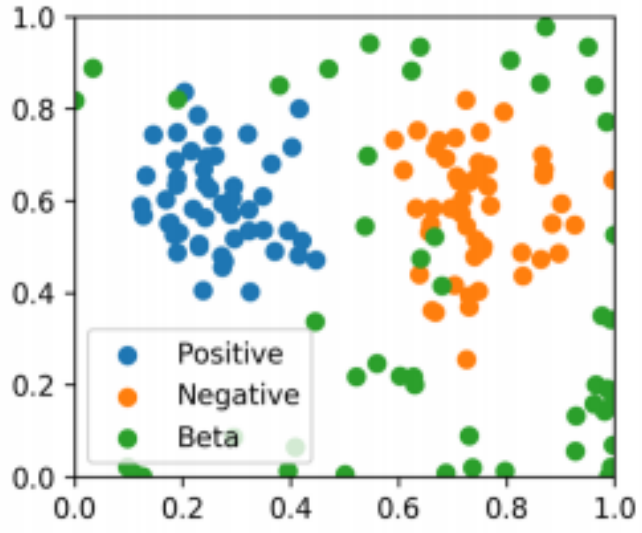


RANDOM INPUT ATTACK

- 경계선을 좁게
  - 3rd class beta noise







Q&A