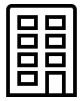
2022 하계 세미나

Face Anti-Spoofing



Sogang University
Vision & Display Systems Lab, Dept. of Electronic Engineering

Presented by
전창렬

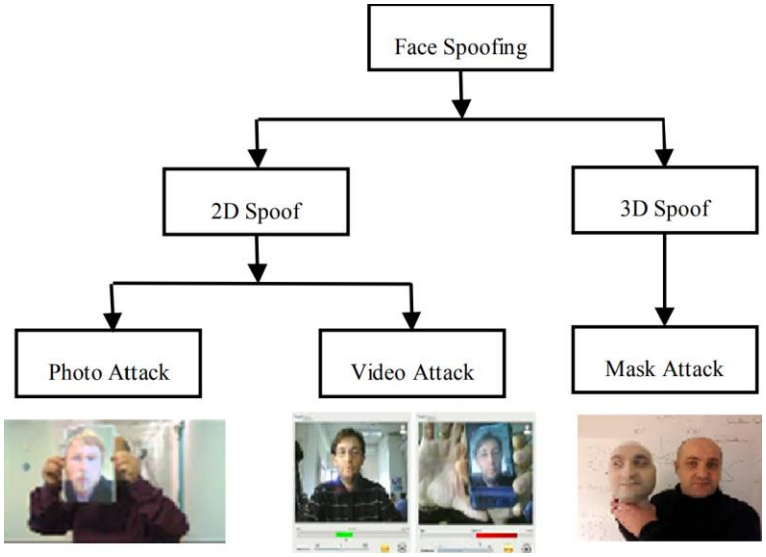
Outline

- Background
  - What is face anti-spoofing?
  - Learn Convolutional Neural Network for Face Anti-Spoofing

- Face anti-spoofing
  - Domain generalization
    - Domain Generalization via Shuffled Style Assembly for Face Anti-Spoofing (CVPR 2022)
  - Fine-grained recognition
    - PatchNet: A Simple Face Anti-Spoofing Framework via Fine-Grained Patch Recognition (CVPR 2022)

Background

- What is face anti-spoofing?
  - 안면 인식은 얼굴을 통해 사람을 식별하거나 특정인의 신원을 인증하는 방법
  - 안면 인식 알고리즘 기술은 AI 기반으로 학습 되기 때문에 이미징 프로세스와 피사체에 따라 상당한 비율의 오류가 발생
  - 이를 활용하여 악의적인 목적을 가지고 타인의 신분으로 위장하는 범죄가 발생
  - Face anti-spoofing은 이러한 spoof attack을 인식하고 실제 얼굴과 구별해내어 안전성을 확보하기 위한 task

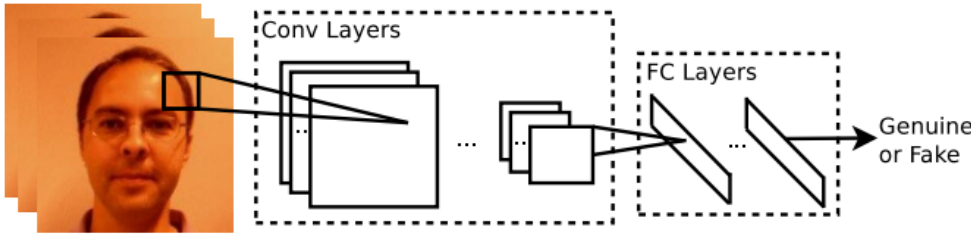


<face spoofing 개요>

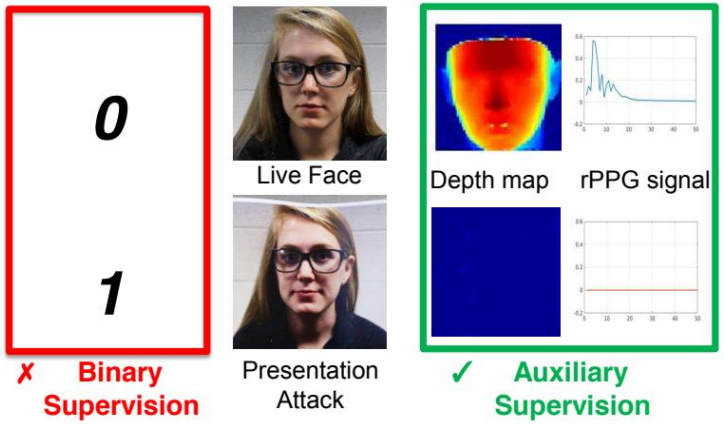
서강대학교 SOGANG UNIVERSITY
VDS LAB

Background

- Binary classification-based face anti-spoofing
  - Face anti spoofing은 기본적으로 live face와 presentation attack 2가지 경우를 구분하는 binary classification 관점
  - 이와 같은 binary classification 관점은 2가지 문제가 존재
    - Spoofing pattern에 따라 이미지 degradation 정도가 달라지는데 학습에 사용된 data로 모든 spoofing pattern에 대한 generalize가 어려움
    - 단순히 image를 binary supervision을 통해 분류하는 것이기 때문에 결과에 대한 해석이나 이유가 따르지 않음
  - 따라서 해당 분야의 연구에서는 Depth map 등의 추가 정보가 활용되기도 함



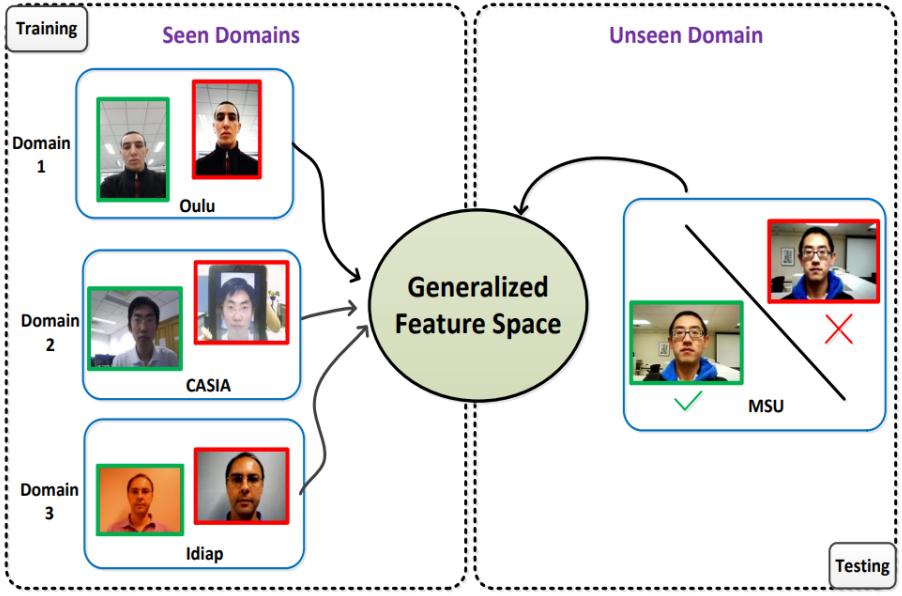
<binary classification을 통한 face anti-spoofing>



<Auxiliary supervision을 활용한 face anti-spoofing>

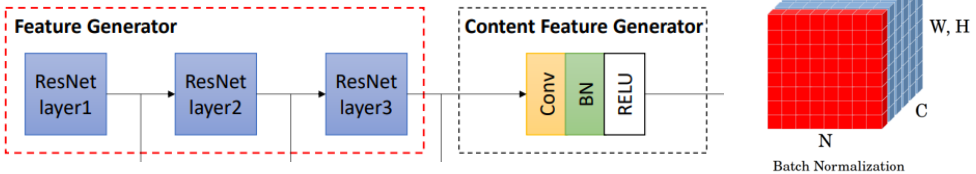
Face anti-spoofing (FAS)

- Domain Generalization via Shuffled Style Assembly for Face Anti-Spoofing
  - 기존의 FAS 기법들은 intra-domain 환경에서 높은 성능을 얻었으나 cross-domain setting에서 성능 저하가 발생
    - 학습 데이터 수와 네트워크의 capability의 한계로 인한 문제
    - 이로 인해 모델이 dataset bias의 영향을 받으며, 새로운 domain에 대한 generalization이 잘 되지 못함
  - 이러한 문제를 해소하기 위해 domain generalization 기법을 도입하는 연구가 진행

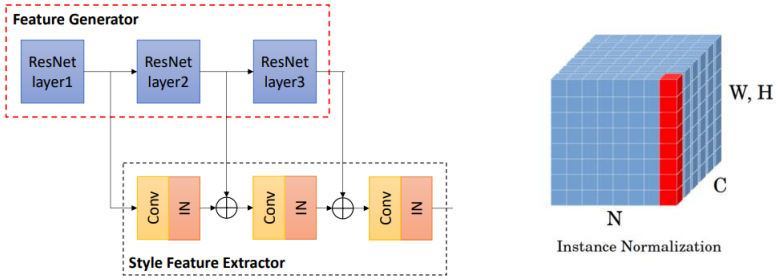


Domain Generalization via Shuffled Style Assembly for Face Anti-Spoofing

- Content and style information aggregation
  - 이미지는 이미지의 외형을 표현하는 contents와 texture등을 표현하는 style로 구성
    - FAS task에서는 이미지의 content information은 common factor로 존재
    - 이미지의 style information은 FAS에서 discriminative cue를 제공하는 특성
    - 따라서 해당 논문에서는 content와 style에 따로 각각의 information aggregation을 진행
      - Content information은 BN을 통하여 global image statistics를 획득함으로써 얻음
      - Style information은 IN을 통하여 개별 sample의 statistics를 획득함으로써 얻음



<content feature aggregation by BN>



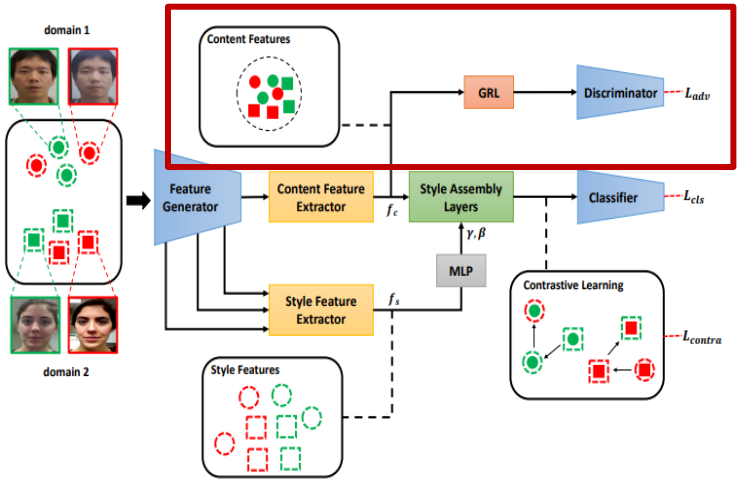
<style feature aggregation by IN>

Domain Generalization via Shuffled Style Assembly for Face Anti-Spoofing

- Content and style information aggregation
  - Content information aggregation
    - Content information에서는 domain간의 distribution discrepancies가 작다고 추측
      - Domain에 무관하게 데이터들은 모두 facial area를 다루고 있기 때문에 유사한 content feature를 가질 것
      - Bona fide or attack presentation에 무관하게 shape나 size와 같은 attribute는 유사할 것
    - 따라서 content 간의 차이가 네트워크에 영향을 미치지 않도록 adversarial 학습을 진행

$$\min_{D}\max_{G} L_{adv}(G,D) = -\mathbb{E}_{(x,y)\sim(X,Y_D)}\sum_{i=1}^{M}\mathbb{1}[i=y]\,logD(G(x))$$

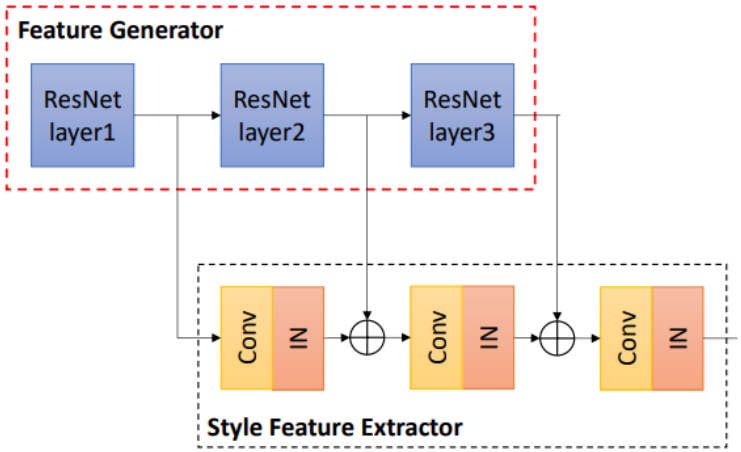
<adversarial loss>



<Content information aggregation>

Domain Generalization via Shuffled Style Assembly for Face Anti-Spoofing

- Content and style information aggregation
  - Style information aggregation
    - Style information은 hierarchical 구조를 통해 multi-layer feature를 모아서 진행
      - Style은 scale 별로 각기 다른 특징을 포함하고 있음
        - ✓밝기 정보는 주로 feature의 broad-scale에 포함
        - ✓Texture와 같은 정보는 이미지의 local-scale에서 담고 있음



<Style information aggregation>

Domain Generalization via Shuffled Style Assembly for Face Anti-Spoofing

- Shuffled style assembly
  - Style feature가 MLP를 통과하여 AdaIN에 적용하기 위한 affine parameter γ, β 추출
  - 3x3 convolution을 적용한 content feature에 AdaIN을 적용하여 intermediate variable z를 얻음
  - z에 추가로 3x3 convolution이 적용되고 AdaIN + content feature를 적용해 style assemble이 적용

$$\mathbf{AdaIN}\left(x, \gamma, \beta\right) = \gamma\left(\frac{x - \mu(x)}{\sigma(x)}\right) + \beta$$

$$\gamma, \beta = \mathbf{MLP}\left[\mathbf{GAP}\left(f_s\right)\right],$$

$$z = \mathbf{ReLU}\left[\mathbf{AdaIN}(K_1 \otimes f_c, \gamma, \beta)\right]$$

$$\mathbf{SAL}\left(f_c, f_s\right) = \mathbf{AdaIN}(K_2 \otimes z, \gamma, \beta) + f_c,$$

$$S\left(x_i, x_i\right) = \mathbf{SAL}\left(f_c\left(x_i\right), f_s\left(x_i\right)\right)$$

$$S\left(x_i, x_{i^*}\right) = \mathbf{SAL}\left(f_c\left(x_i\right), f_s\left(x_{i^*}\right)\right)$$

$$i^* \in random\left\{1, 2, \ldots, N\right\}$$

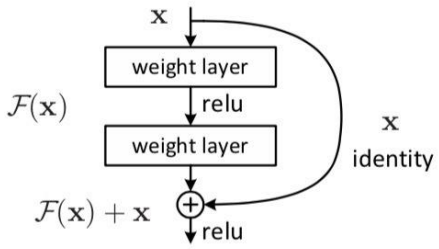
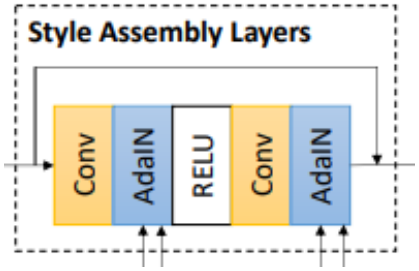
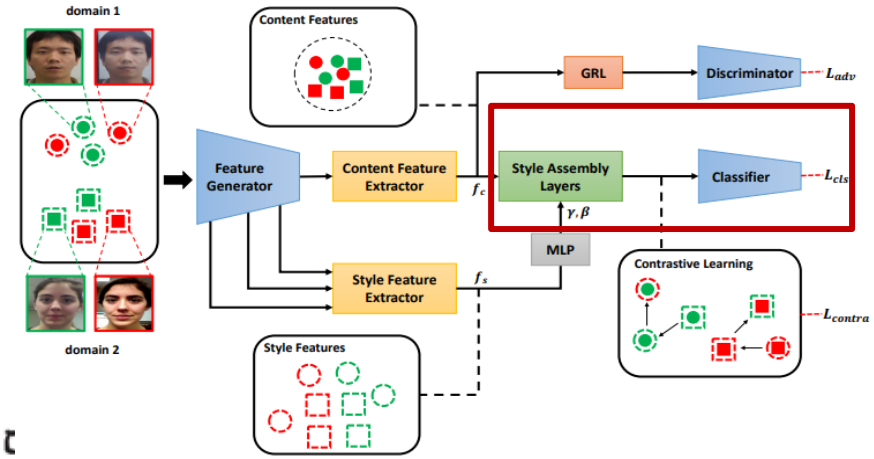


Figure 2. Residual learning: a building block.



Domain Generalization via Shuffled Style Assembly for Face Anti-Spoofing

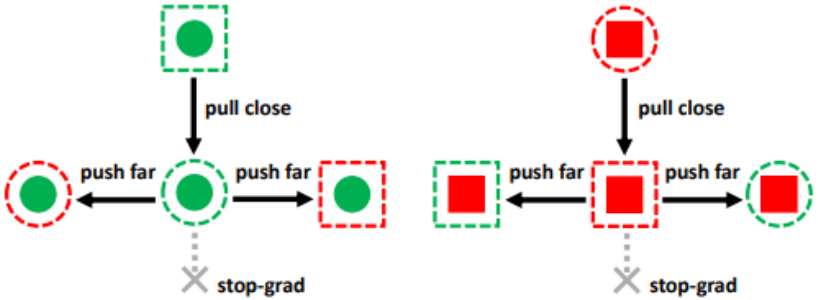
- Contrastive Learning for Stylized Features
  - Shuffled style assembly를 통해 generalization을 진행하는데 이 과정에서 style은 generalize한 style이 domain specific feature인지 liveness-related feature인지 불분명
    - liveness-related style feature를 강조하고 domain-specific feature를 suppress하기 위해 contrastive learning을 적용
  - Content & style feature를 combine함으로써 self-assembly feature $S(x_i, x_i)$, shuffle-assembly features $S(x_i, x_{i*})$를 획득
    - Self-assembly feature는 live / spoofing에 대한 classification 네트워크에 입력
    - Shuffle-assembly feature는 $S(x_i, x_i)$,와 $S(x_i, x_{i*})$ 간의 cosine similarity를 계산



$$Sim\,(a,b) = -\frac{a}{\|a\|_2} \cdot \frac{b}{\|b\|_2}$$

Domain Generalization via Shuffled Style Assembly for Face Anti-Spoofing

- Contrastive Learning for Stylized Features
  - Self-assembly feature는 $S(x_i, x_i)$는 stylized features space의 anchor 역할
    - Anchor의 역할을 하는 feature는 고정되어야 하므로 stop-gradient가 적용
  - shuffle-assembly features $S(x_i, x_{i*})$는 liveness information에 따라 guide됨
    - Live face와 spoofing attack간의 contrastive learning을 통해 live face와 spoofing attack feature들 간에 compact한 feature mapping을 구성하게 됨



$$L_{contra} = \sum_{i=1}^{N} Eq\left(x_i, x_{i^*}\right) \cdot Sim\left(\text{stopgrad}(a), b\right)$$

$$Eq\left(x_i, x_{i^*}\right) = \begin{cases} +1, & \text{label}(x_i) == \text{label}(x_{i^*}) \\ -1, & \text{otherwise.} \end{cases}$$

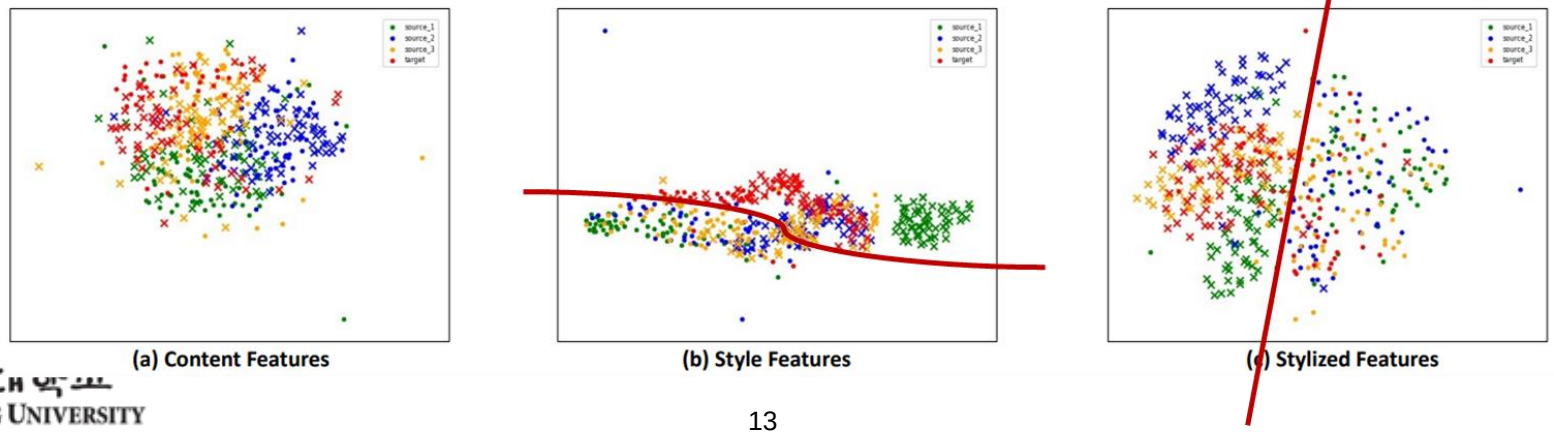
Domain Generalization via Shuffled Style Assembly for Face Anti-Spoofing

- Experiments

Method	O&C&I to M		O&M&I to C		O&C&M to I		I&C&M to O	
	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)
MMD-AAE [24]	27.08	83.19	44.59	58.29	31.58	75.18	40.98	63.08
MADDG [41]	17.69	88.06	24.50	84.51	22.19	84.99	27.98	80.02
SSDG-M [21]	16.67	90.47	23.11	85.45	18.21	94.61	25.17	81.83
DR-MD-Net [48]	17.02	90.10	19.68	87.43	20.87	86.72	25.02	81.47
RFMeta [42]	13.89	93.98	20.27	88.16	17.30	90.48	16.45	91.16
NAS-FAS [60]	19.53	88.63	16.54	90.18	14.51	93.84	**13.80**	**93.43**
D^2AM [6]	12.70	95.66	20.98	85.58	15.43	91.22	15.27	90.87
SDA [50]	15.40	91.80	24.50	84.40	15.60	90.10	23.10	84.30
DRDG [31]	12.43	95.81	19.05	88.79	15.56	91.79	15.63	91.75
ANRL [30]	10.83	**96.75**	17.83	89.26	16.03	91.04	15.67	91.90
SSAN-M (Ours)	**10.42**	94.76	**16.47**	**90.81**	**14.00**	**94.58**	19.51	88.17
SSDG-R [21]	7.38	97.17	10.44	95.94	11.71	96.59	15.61	91.54
SSAN-R (Ours)	**6.67**	**98.75**	**10.00**	**96.67**	**8.88**	**96.79**	**13.72**	**93.63**

Domain Generalization via Shuffled Style Assembly for Face Anti-Spoofing

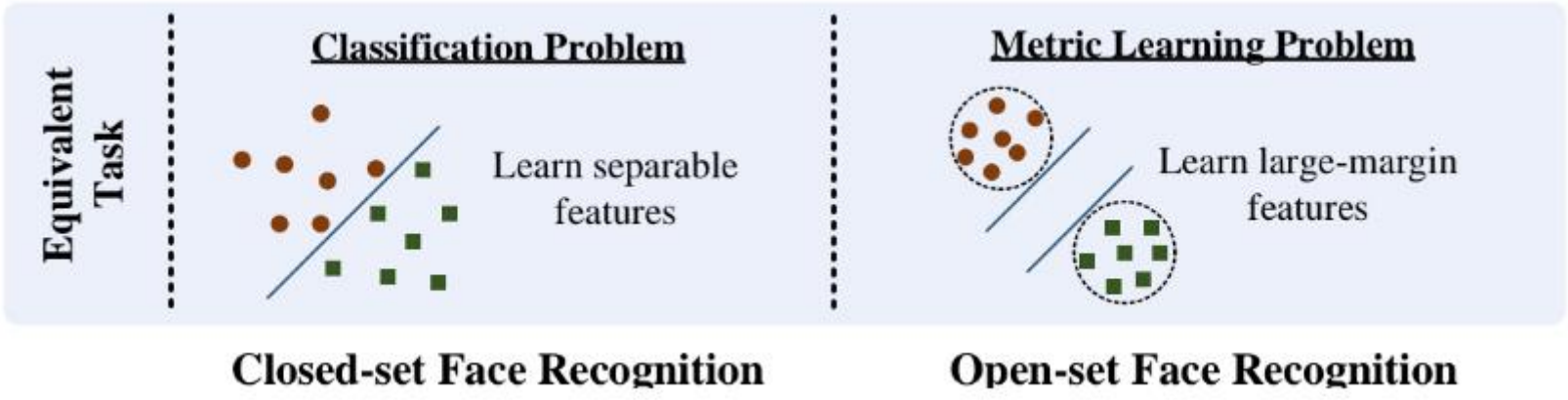
- Experiments
  - Feature visualization
    - Content features의 distribution은 compact하고 mixed되어 있는 형태로 나타남
      - Adversarial learning을 통해 content feature의 기존 distribution 차이가 줄어들었음을 확인 가능
    - Style feature들 간에는 별도의 supervision이 없음에도 coarse boundary가 존재
      - Contrastive learning의 효과로 liveness-related style feature가 emphasize됨
    - Stylized feature는 content와 style information을 combine한 feature로 unseen target domain에 대해서도 잘 generalize되었음을 확인 가능



(a) Content Features (b) Style Features (c) Stylized Features

PatchNet: A Simple Face Anti-Spoofing Framework via Fine-Grained Patch Recognition

- Preliminary
  - 대규모 얼굴 인식을 위해 feature learning을 하는 데에 있어서 주요한 challenge 중 하나는 discriminative power를 강화하는 loss function의 설계임
  - 기존의 softmax기반 loss는 closed-set 분류 문제에 대해서는 feature가 분류 될 수 있지만 open-set 환경에서 충분히 discriminative 하지 못함
  - Angular margin softmax는 마지막 fully connected layer에서의 linear transformation matrix가 angular space에서 class의 중심을 나타내는 표현으로 사용 될 수 있다고 가정

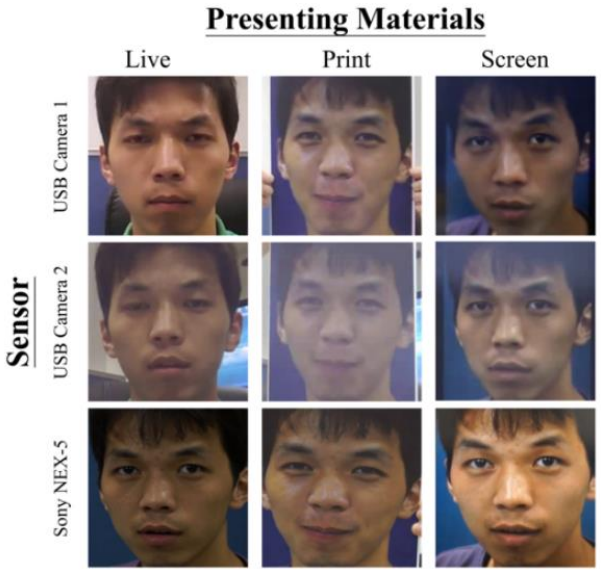


PatchNet: A Simple Face Anti-Spoofing Framework via Fine-Grained Patch Recognition

- Preliminary
  - Example
    - 얼굴 이미지 feature $x = [10, 20, 30]$
    - Feature를 입력으로 받았을 때, 2가지 identity로 분류하는 task를 진행하는 fully connected layer를 구성하는 matrix A = $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$
      - 이때 A의 행은 class의 개수, 열은 feature의 차원을 의미함
      - (1 2 3)은 첫 번째 identity의 중심, (4 5 6)은 두 번째 identity의 중심을 나타냄
    - Fully connected layer를 통한 xA^T 연산 결과는 (10*1 + 20*2+ 30*3 10*4+20*5+30*6)
      - 이러한 과정을 통해 얼굴 이미지 feature 입력 시 각 identity에 대한 prediction score 계산
    - 각 identity의 중심과 prediction score를 normalization한 결과는 각각 $\left(1/\sqrt{14} \quad 2/\sqrt{14} \quad 3/\sqrt{14}\right)$, $\left(4/\sqrt{77} \quad 5/\sqrt{77} \quad 6/\sqrt{77}\right)$, $\left(10/\sqrt{1400} \quad 20/\sqrt{1400} \quad 30/\sqrt{1400}\right)$
    - 이러한 과정을 통해 identity의 중심과 얼굴 이미지 feature 의 prediction이 모두 중심이 (0,0,0)인 구 위에 mapping 됨

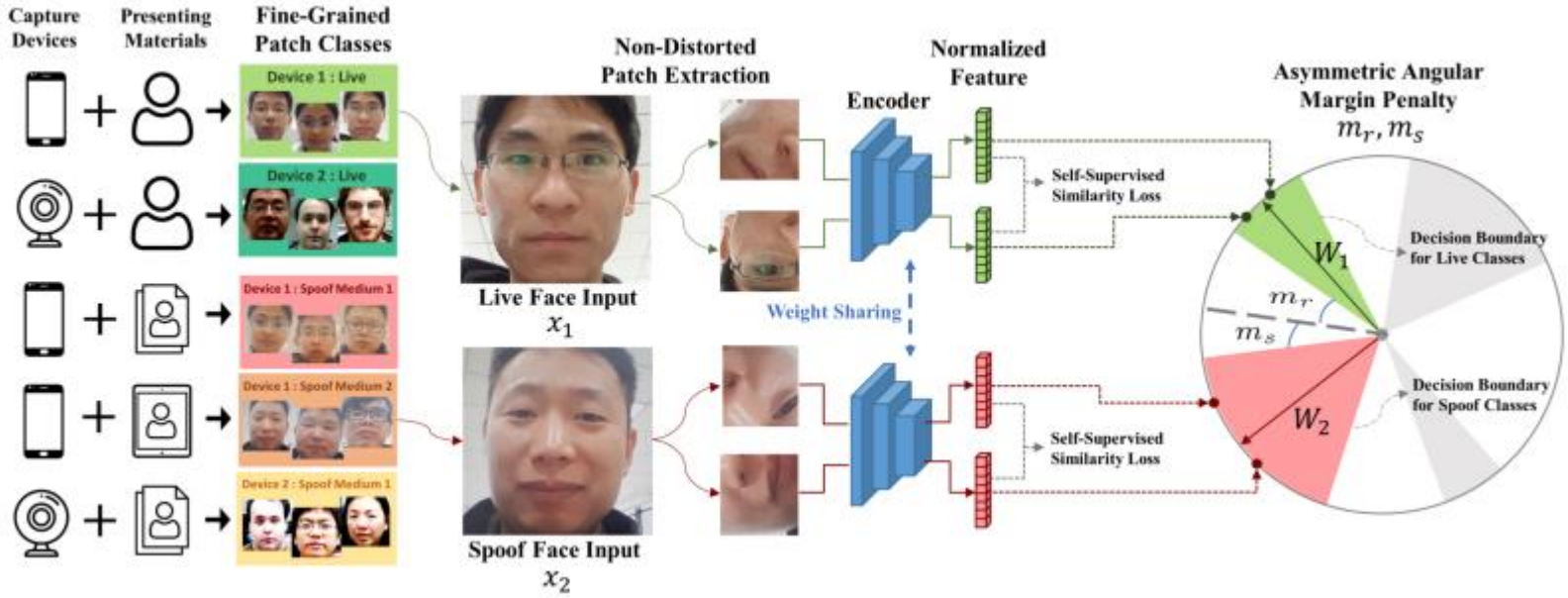
PatchNet: A Simple Face Anti-Spoofing Framework via Fine-Grained Patch Recognition

- Patch Features Extraction
  - 해당 논문에서는 face image를 capturing devices와 presenting materials에 따라 category를 세분화 하여 fine-grained 환경으로 세분화함
    - 예시로 CASIA-FASD 데이터 셋에는 2가지의 spoof medium과 3종류의 capturing device가 존재
      - 3개의 live types과 2*3개의 spoof types으로 총 9개의 class로 나뉘어짐
    - 이를 통해 해당 논문은 기존의 binary 관점의 구분이 아닌 세분화된 구분으로 학습된 robust한 네트워크를 구성함



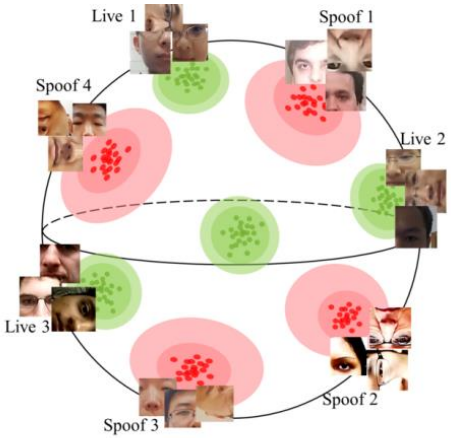
PatchNet: A Simple Face Anti-Spoofing Framework via Fine-Grained Patch Recognition

- Patch Features Extraction
  - 해당 논문에서는 fine-grained patch를 얻는 데에 있어서 이미지의 distortion이나 spoof-related information을 잃는 것을 막기위해 augmentation을 선별적으로 사용
    - Random horizontal flip, random rotation, fixed size cropping을 활용
  - face region x_i에 대하여 2개의 각기 다른 augmentation t_1, t_2를 적용하여 $x_i^{t_1}, x_i^{t_2}$를 얻음



PatchNet: A Simple Face Anti-Spoofing Framework via Fine-Grained Patch Recognition

- Fine-Grained patch Recognition
  - 학습 데이터 셋에 patch type classes가 N개가 있다고 가정 시 k개의 live class, N-k의 spoof classes로 구성
    - 각각의 input patch는 해당 $y_i \in \{L_1, L_2, \ldots, L_k, S_1, S_2, \ldots, S_{N-k}\}$ 중 하나의 class에 포함
  - Patch feature를 regularize하기 위해 AM-Softmax loss를 활용
    - Open-set identity에 최적화된 loss를 활용함으로써 generalization ability를 향상 시킴



Fine-Grained Patch-Type Recognition

PatchNet: A Simple Face Anti-Spoofing Framework via Fine-Grained Patch Recognition

- Asymmetric AM-Softmax Loss
  - AM-Softmax
    - 기존 Softmax는 feature embedding에 대해 intra-class compactness가 얼굴 인식에 적용될 만큼 고려되지 않고 class 구분에만 초점이 맞추어져 있음
    - AM-Softmax loss 를 활용하여 intra-class에서는 compact하면서 generalization 능력을 높임
    - logit $W_j^T f_i = \|W_j\| \|f_i\| \cos\theta_j$ 로 변환
      - θ_j는 W_j와 feature f_i 간의 각도

$$\mathcal{L}_S = -\frac{1}{n}\sum_{i=1}^{n} log\frac{e^{W_{y_i}^T f_i}}{\sum_{j=1}^{c} e^{W_j^T f_i}}$$

$$= -\frac{1}{n}\sum_{i=1}^{n} log\frac{e^{\|W_{y_i}\|\|f_i\|cos(\theta_{y_i})}}{\sum_{j=1}^{c} e^{\|W_j\|\|f_i\|cos(\theta_j)}}$$

PatchNet: A Simple Face Anti-Spoofing Framework via Fine-Grained Patch Recognition

- Asymmetric AM-Softmax Loss
  - AM-Softmax
    - $0 \leq \theta_1 \leq \frac{1}{m}$의 조건 하에 cos함수는 단조 감소의 형태를 나타냄
    - 따라서 함수에 m값을 크게 적용하게 되면 θ_1의 범위가 줄어들고 더 compact한 구성으로써 margin을 크게 가져갈 수 있게 됨
    - Margin을 직관적으로 간편하게 가져가기 위해서 m을 곱하는 형식이 아니라 $\cos\theta_{y_i} - m$으로 처리하여도 의미적으로 같은 뜻을 가짐

$$
\begin{aligned}
\boldsymbol{W}_1^T\boldsymbol{x} &> \boldsymbol{W}_2^T\boldsymbol{x} \\
\|\boldsymbol{W}_1\|\|\boldsymbol{x}\|\cos(\theta_1) &> \|\boldsymbol{W}_2\|\|\boldsymbol{x}\|\cos(\theta_2) \\
\|\boldsymbol{W}_1\|\|\boldsymbol{x}\|\cos(\theta_1) &\geq \|\boldsymbol{W}_1\|\|\boldsymbol{x}\|\cos(m\theta_1) \\
&> \|\boldsymbol{W}_2\|\|\boldsymbol{x}\|\cos(\theta_2). \\
&\quad (0 \leq \theta_1 \leq \tfrac{\pi}{m})
\end{aligned}
$$

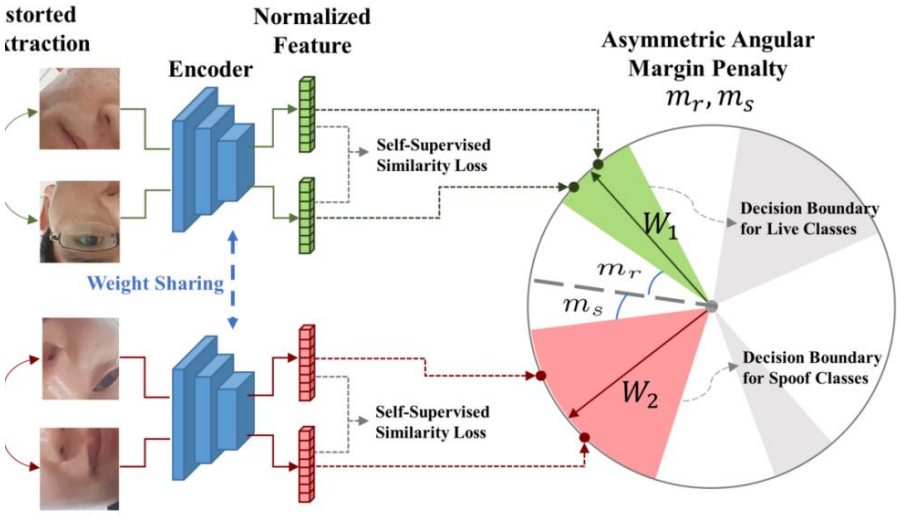
$$
\begin{aligned}
\mathcal{L}_{AMS} &= -\frac{1}{n}\sum_{i=1}^{n} log\frac{e^{s\cdot(cos\theta_{y_i}-m)}}{e^{s\cdot(cos\theta_{y_i}-m)}+\sum_{j=1,j\neq y_i}^{c} e^{s\cdot cos\theta_j}} \\
&= -\frac{1}{n}\sum_{i=1}^{n} log\frac{e^{s\cdot(W_{y_i}^T f_i-m)}}{e^{s\cdot(W_{y_i}^T f_i-m)}+\sum_{j=1,j\neq y_i}^{c} e^{sW_j^T f_i}}
\end{aligned}
$$

PatchNet: A Simple Face Anti-Spoofing Framework via Fine-Grained Patch Recognition

- Asymmetric AM-Softmax Loss
  - Spoofing의 패턴은 live sample에 비해 다양해 질 수 있기 때문에 distribution discrepancies 또한 spoofing sample이 더욱 큼
    - 이러한 특징을 feature space에서 realistic하게 반영하기 위해 AM-Softmax loss를 asymmetrically하게 적용하여 FAS task에 더 적합하게 compact한 cluster를 구성
    - Live type에 lager angular margin을 적용하여 live type을 더 compact하게 구성

$$\mathcal{L}_{AAMS}(f_i) = \begin{cases} -log\dfrac{e^{s\cdot\left(W_{y_i}^T f_i - m_l\right)}}{e^{s\cdot\left(W_{y_i}^T f_i - m_l\right)} + \sum_{j=1, j\neq y_i}^{N} e^{s\cdot W_j^T f_i}} & y_i \in L \\ -log\dfrac{e^{s\cdot\left(W_{y_i}^T f_i - m_s\right)}}{e^{s\cdot\left(W_{y_i}^T f_i - m_s\right)} + \sum_{j=1, j\neq y_i}^{N} e^{s\cdot W_j^T f_i}} & y_i \in S \end{cases}$$

$$\mathcal{L}_{Asym} = -\frac{1}{n}\sum_{i=1}^{n}\left(\mathcal{L}_{AAMS}(f_i^{t_1}) + \mathcal{L}_{AAMS}(f_i^{t_2})\right)$$



PatchNet: A Simple Face Anti-Spoofing Framework via Fine-Grained Patch Recognition

- Self-Supervised Similarity Loss
  - 하나의 face image에서 얻어진 2개의 transformed patch views에 대해 self-supervised similarity loss를 적용
    - Spoof-specific discriminative information이 전체 face region에 대하여 존재
      - 따라서 같은 face capture에서 얻어진 2개의 patch views의 feature가 유사한 특성을 지녀야함
    - 해당 loss를 통해서 같은 이미지에서 얻어진 다른 patch 간의 location과 rotation으로 인한 feature의 정보 차이와 무관한 feature를 구성하게 만듦

$$\mathcal{L}_{Sim}(f_i^{t_1}, f_i^{t_2}) = \frac{1}{n}\sum_{i=1}^{n} \left\| f_i^{t_1} - f_i^{t_2} \right\|_2$$

PatchNet: A Simple Face Anti-Spoofing Framework via Fine-Grained Patch Recognition

- Testing Strategy
  - Test 시에는 face image에서 uniform하게 crop(training에서의 crop size와 같게) 진행
  - 이를 통해 하나의 face image에 대하여 P개의 patch features를 얻음(f^1, f^2,… f^P)
  - 모든 patch features들에서 얻은 live class probabilities를 평균하여 하나의 face image에 대하여 live probability를 얻을 수 있음

$$LiveProb = \frac{1}{P}\sum_{i=1}^{P}\sum_{y\in L} Softmax(s \cdot W_y^T f^i)$$

PatchNet: A Simple Face Anti-Spoofing Framework via Fine-Grained Patch Recognition

- Experiments

Method	O&C&I to M		O&M&I to C		O&C&M to I		I&C&M to O	
	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)
PatchNet w/ coarse cls [12]	10.24	96.45	15.67	92.47	21.65	91.08	16.26	91.33
PatchNet w/o margin	10.0	96.61	18.0	91.57	17.25	90.47	15.04	92.42
PatchNet w/o $\mathcal{L}_{Sim}$	8.9	97.42	13.44	93.99	15.1	92.10	14.24	92.93
PatchNet (Ours)	**7.10**	**98.46**	**11.33**	**94.58**	**14.6**	**92.51**	**11.82**	**95.07**

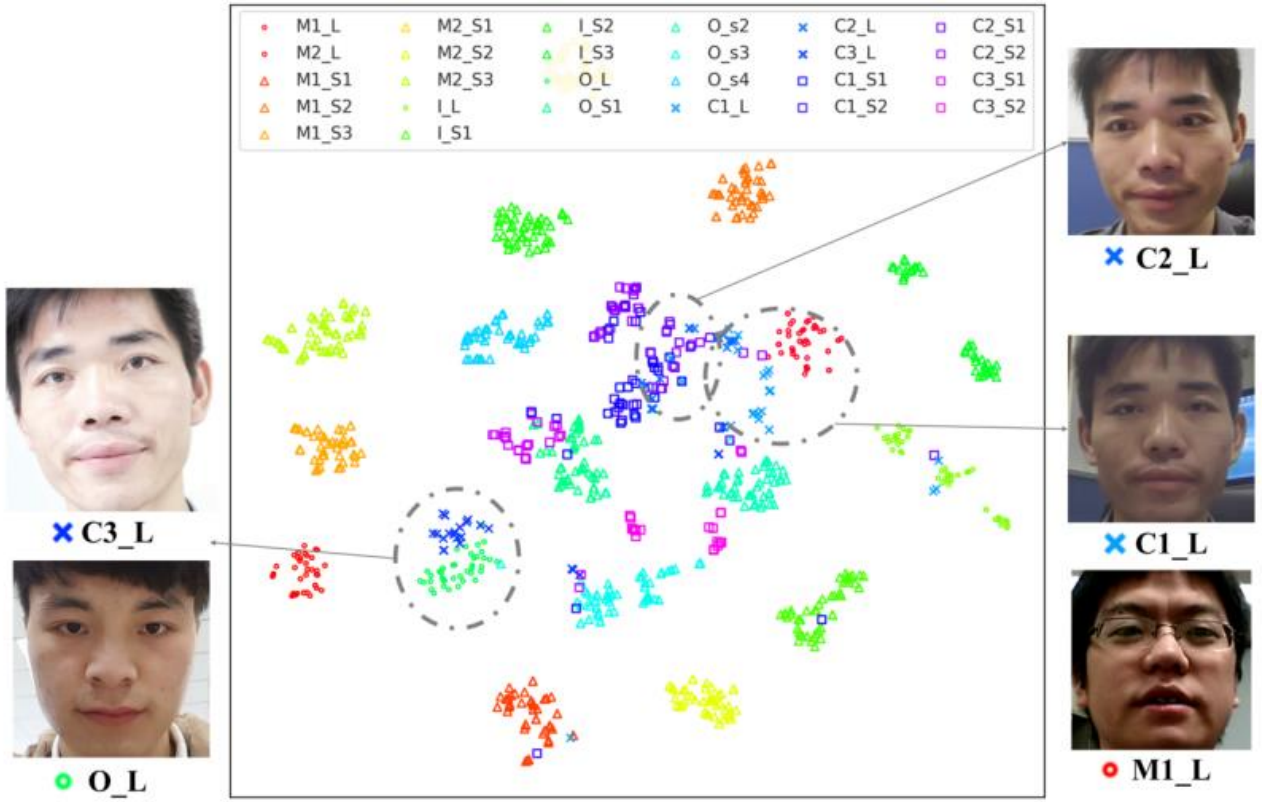
Table 7. Evaluations of different components of the proposed method on four cross-dataset protocols.

Method	O&C&I to M		O&M&I to C		O&C&M to I		I&C&M to O	
	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)	HTER(%)	AUC(%)
Auxiliary [18]	22.72	85.88	33.52	73.15	29.14	71.69	30.17	77.61
MADDG [20]	17.69	88.06	24.50	84.51	22.19	84.99	27.89	80.02
PAD-GAN [24]	17.02	90.10	19.68	87.43	20.87	86.72	25.02	81.47
RFM [21]	13.89	93.98	20.27	88.16	17.30	90.48	16.45	91.16
NAS-FAS [31]	16.85	90.42	15.21	92.64	**11.63**	**96.98**	13.16	94.18
SSDG-R [12]	7.38	97.17	10.44	95.94	11.71	96.59	15.61	91.54
ANRL [14]	16.03	91.04	**10.83**	**96.75**	17.85	89.26	15.67	91.90
DRDG [15]	15.56	91.79	12.43	95.81	19.05	88.79	15.63	91.75
PatchNet (Ours)	**7.10**	**98.46**	11.33	94.58	13.4	95.67	**11.82**	**95.07**

Table 8. Comparison results between the proposed PatchNet and state-of-the-art methods on four domain generalization protocols.

PatchNet: A Simple Face Anti-Spoofing Framework via Fine-Grained Patch Recognition

- Experiments



감사합니다