

# **Consistent Instance False Positive Improves Fairness in Face Recognition**

Xu, X., Huang, Y., Shen, P., Li, S., Li, J., Huang, F., Li, Y., and Cui, Z.,  
Consistent Instance False Positive Improves Fairness in Face  
Recognition, Proc. of CVPR 2021, pp. 578-586.

	T	F
P	TP 동일한 사람인데 허용한 경우(정답)	FP 다른 사람인데 허용한 경우(오답)
N	TN 다른 사람인데 거부한 경우(정답)	FN 동일한 사람인데 거부한 경우(오답)

# 목차

---

- 00. Abstract
- 01. Introduction
- 02. Related Work
- 03. Proposed Approach
- 04. Experiments
- 05. Conclusion

# **00. Abstract**

# 00. Abstract

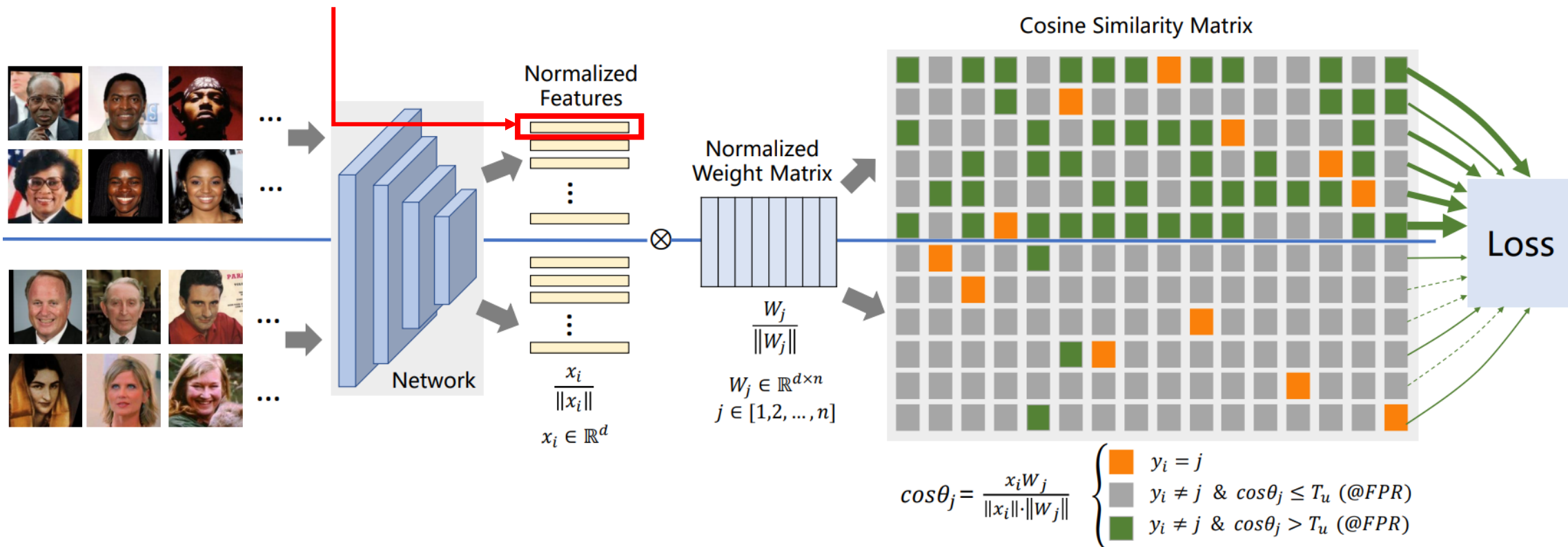
---

해결해야 하는 문제	얼굴 인식 시스템에서 인구통계학적 편향
기존 연구의 한계	인구통계학적 주석에 의존하기 때문에 실제 시나리오에서 사용 불가 ▶ 특정 인구 그룹을 위해 설계되었으므로 일반적이지 못함
해결 방법	<b>Instance False Positive Rate(FPR)</b> 의 일관성을 높여 얼굴 인식의 편향을 완화하는 <b>Penalty Loss</b> 제안 ▶ 인구 통계학적 그룹 간의 편향 완화
저자코드	<a href="https://github.com/Tencent/TFace">https://github.com/Tencent/TFace</a>

**Instance FPR0이란?**

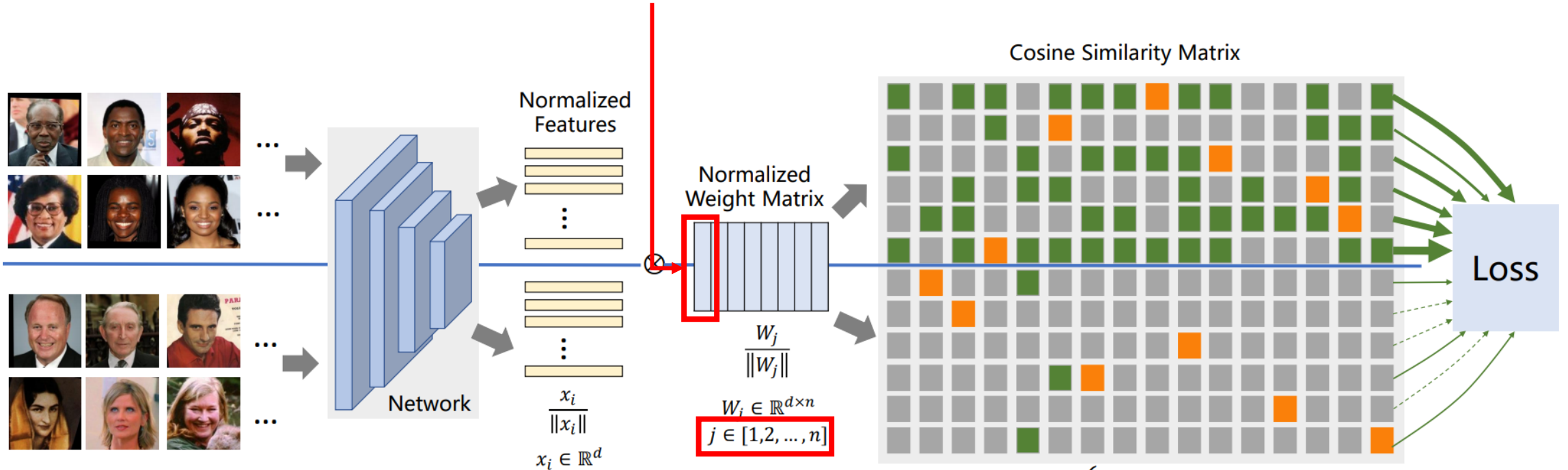
# 00. Abstract

Input image feature: 입력된 사람의 feature



# 00. Abstract

Given a weight matrix  $W$  that each column corresponds to one identity(본 논문 발제) 한 사람에 대한 IDENTITY이자 class

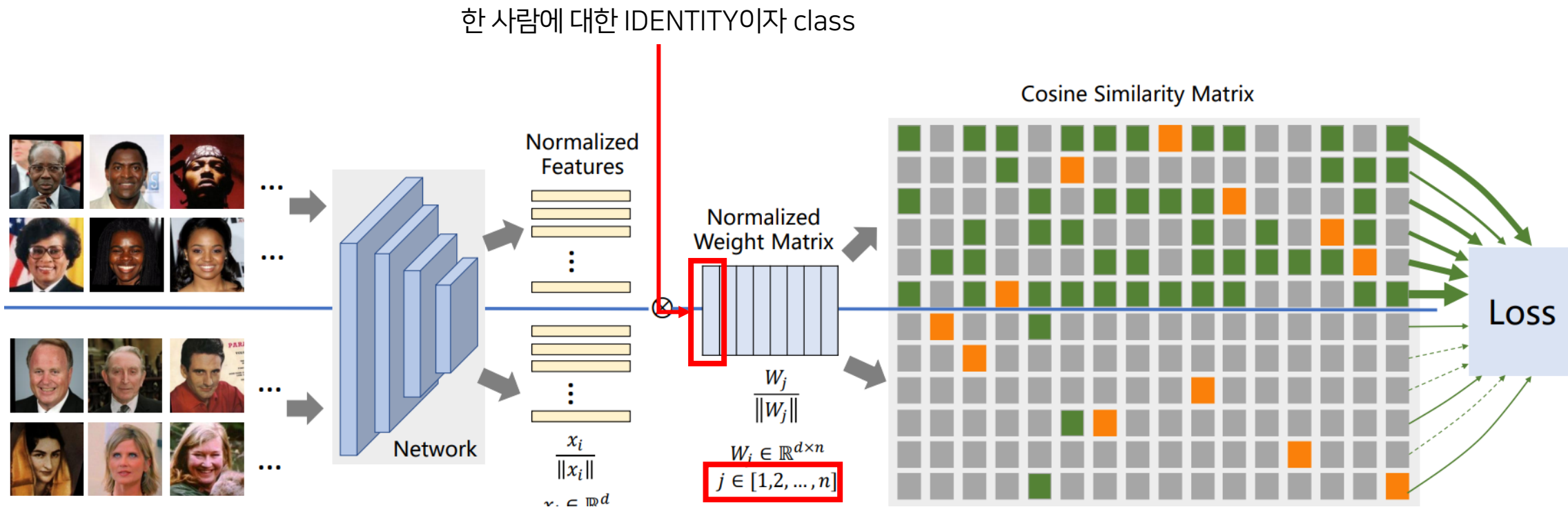


- 1~n 번에 해당하는 사람의 대한 class(identity)
- 따라서  $x_{y1}$ 번 사람의 이미지 feature \*  $w_{y1}$ 번 사람의 identity = POSITIVE PAIR

$$\cos\theta_j = \frac{x_i W_j}{\|x_i\| \cdot \|W_j\|} \begin{cases} \text{orange} & y_i = j \\ \text{grey} & y_i \neq j \ \& \ \cos\theta_j \leq T_u \ (\text{@FPR}) \\ \text{green} & y_i \neq j \ \& \ \cos\theta_j > T_u \ (\text{@FPR}) \end{cases}$$



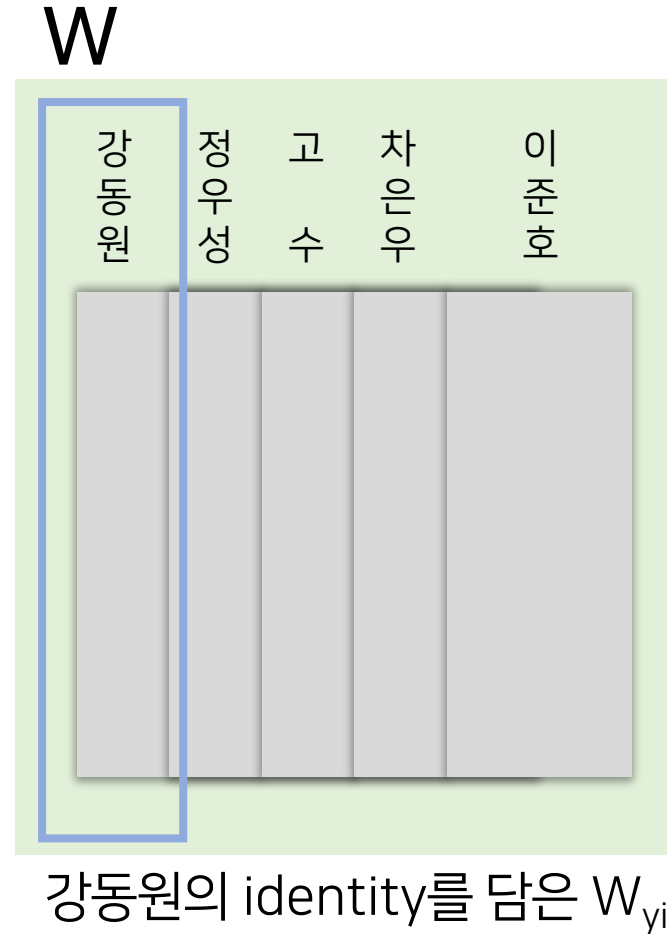
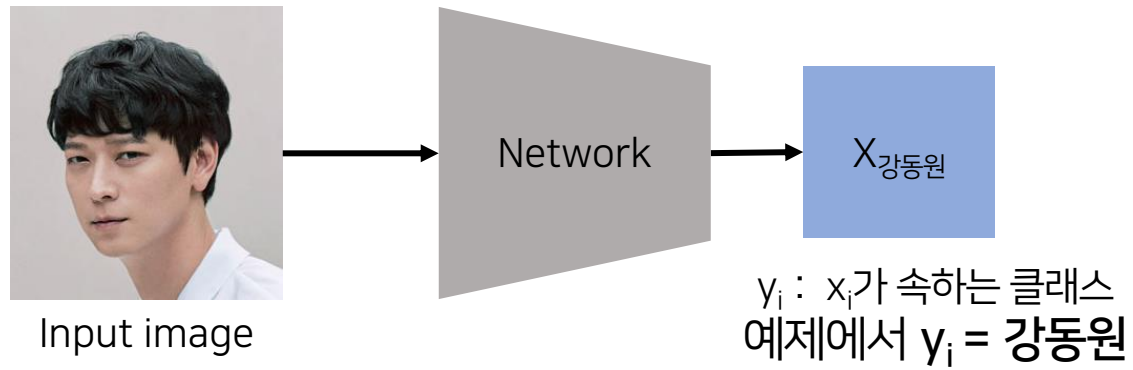
# 00. Abstract



- 1~n 번에 해당하는 사람의 대한 class(identity)
- 따라서  $x_{y1}$ 번 사람의 이미지 feature \*  $w_{y1}$ 번 사람의 identity = POSITIVE PAIR

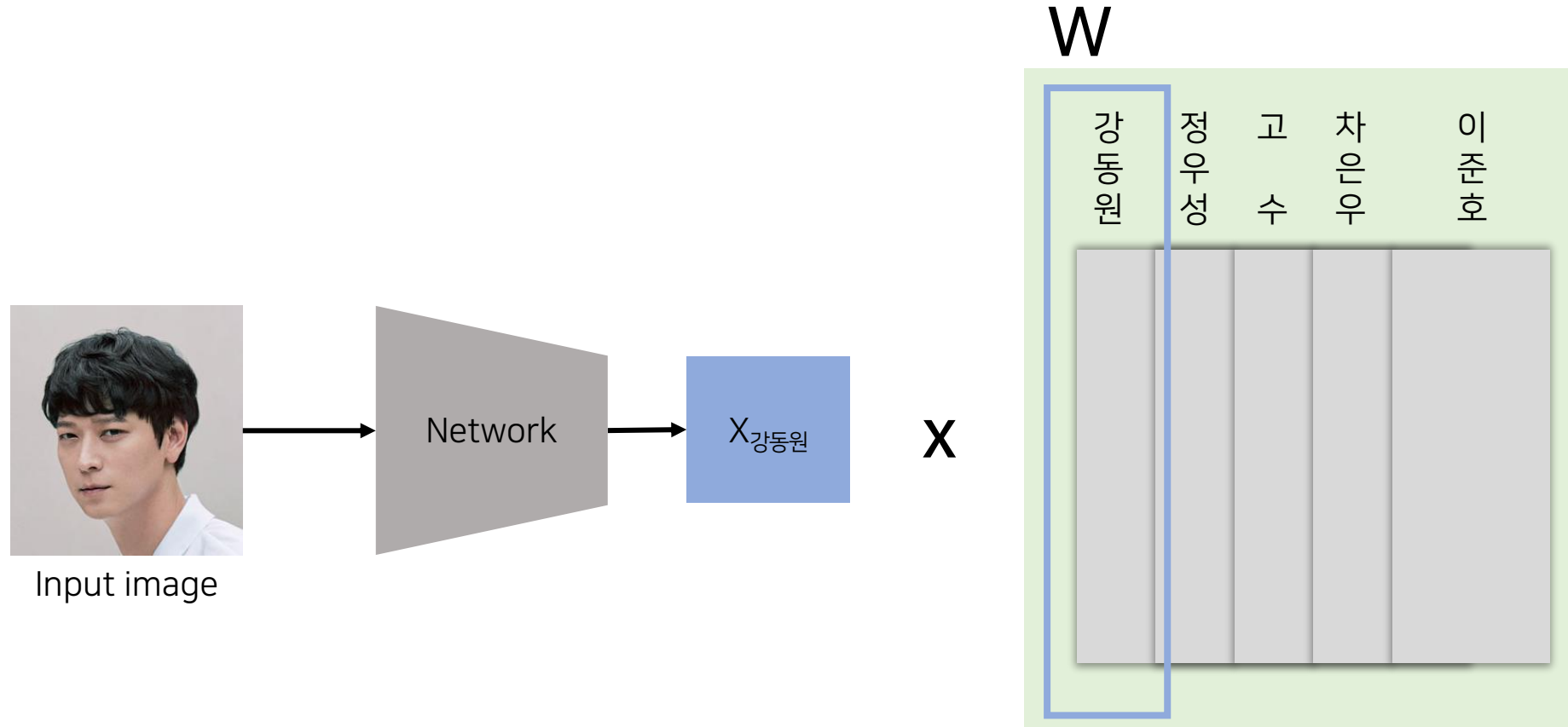
$$\cos\theta_j = \frac{x_i W_j}{\|x_i\| \cdot \|W_j\|} \begin{cases} \text{orange} & y_i = j \text{ \# Positive pair; 입력한 사람과 동일한 identity의 유사성} \\ \text{gray} & y_i \neq j \text{ \& } \cos\theta_j \leq T_u \text{ (@FPR) \# Negative pair; 입력한 사람과 다른 identity의 유사성이 임계치보다 작거나 같을 때} \\ \text{green} & y_i \neq j \text{ \& } \cos\theta_j > T_u \text{ (@FPR) \# Negative pair; 입력한 사람과 다른 identity의 유사성이 임계치보다 클 때} \\ & \text{\# 임계치 이상의 유사도를 갖는 non-target similarities} \end{cases}$$

# 00. Abstract



정리  
 $y_i$ :  $y_i$  class  
 $x_i$ : class  $y_i$ 에 속하는  $i$ 번째 sample의 deep feature

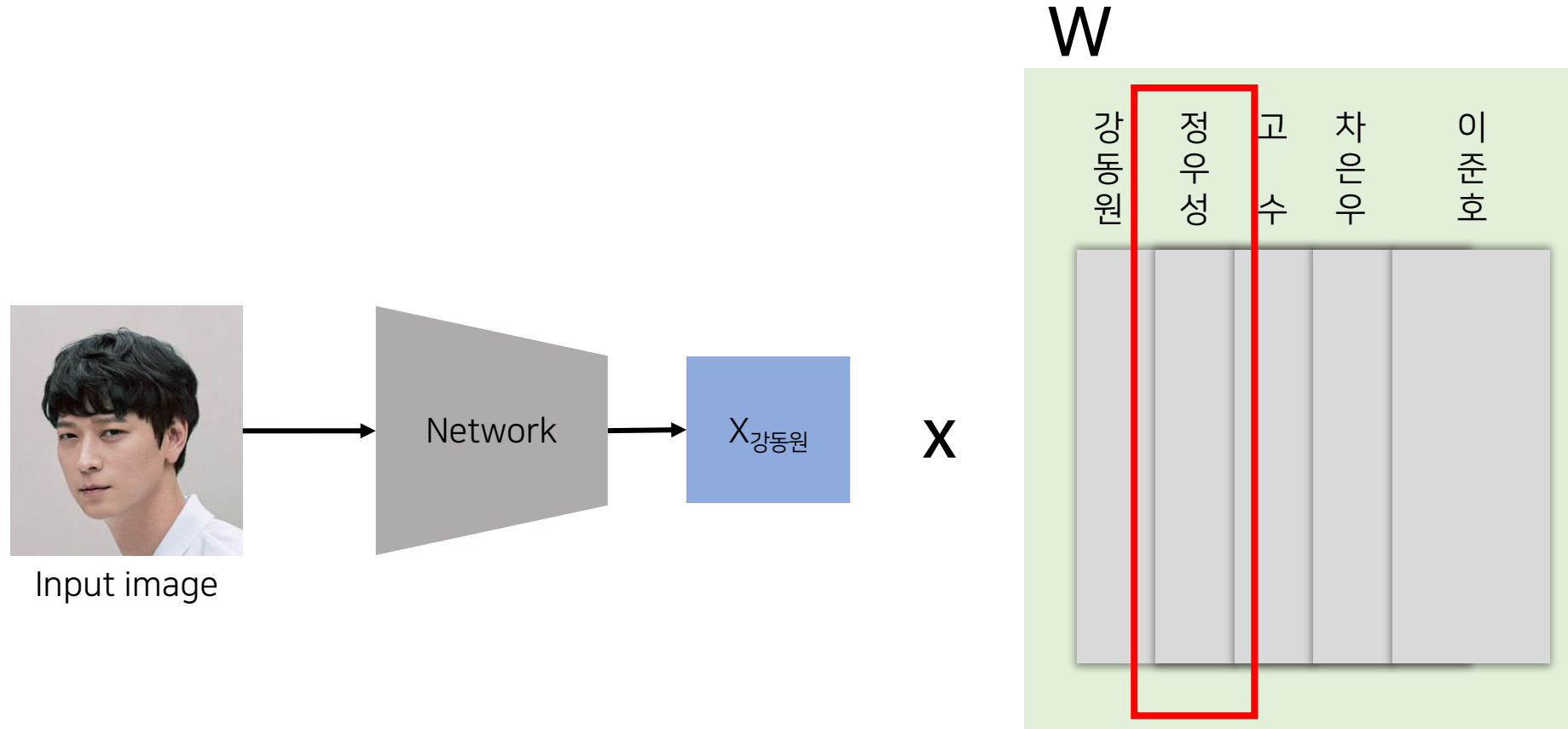
# 00. Abstract



➡ Positive pair!!

정리  
 $y_i$ :  $y_i$  class  
 $x_i$ : class  $y_i$ 에 속하는  $i$ 번째 sample의 deep feature

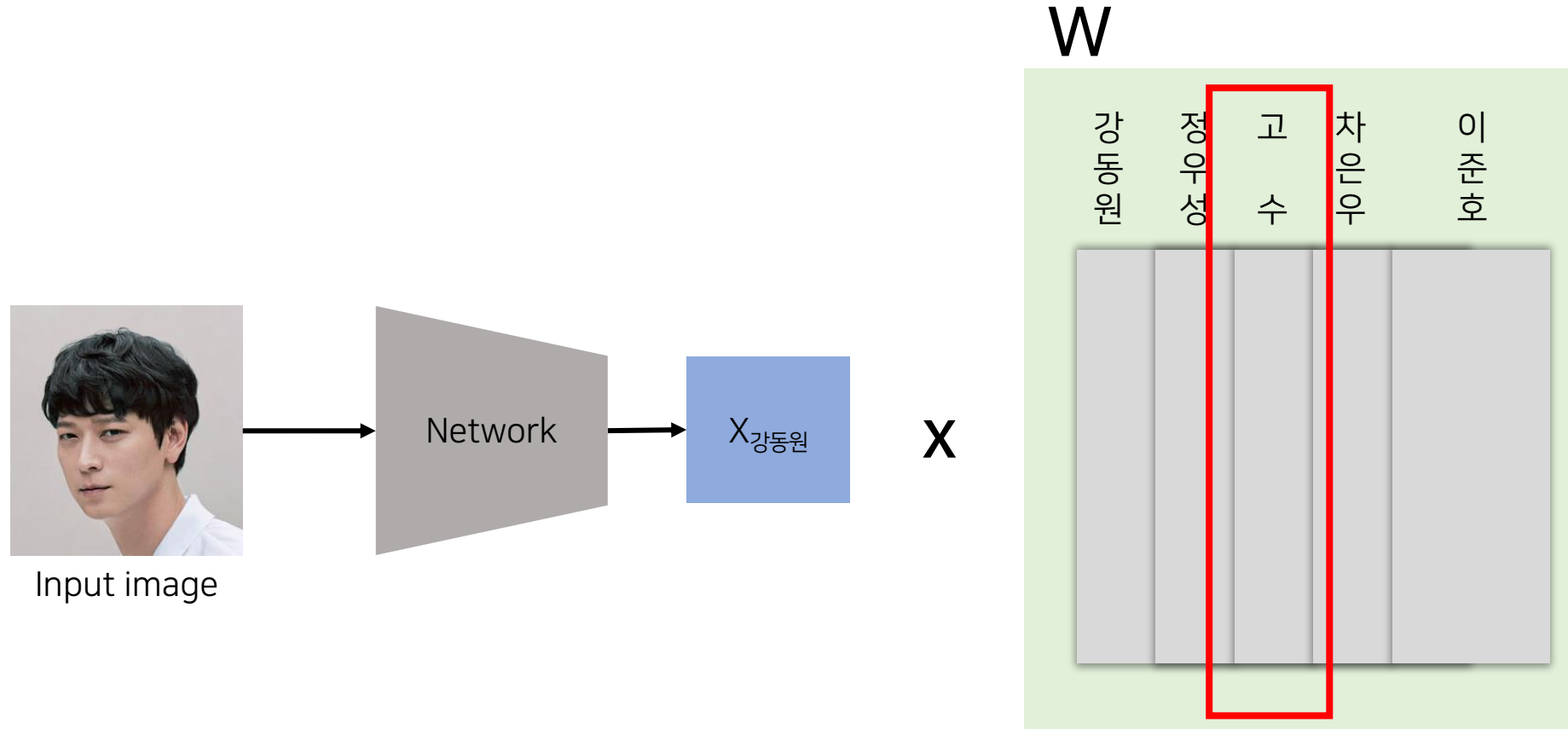
# 00. Abstract



**➡ Negative pair!!**

정리  
 $y_i$ :  $y_i$  class  
 $x_i$ : class  $y_i$ 에 속하는  $i$ 번째 sample의 deep feature

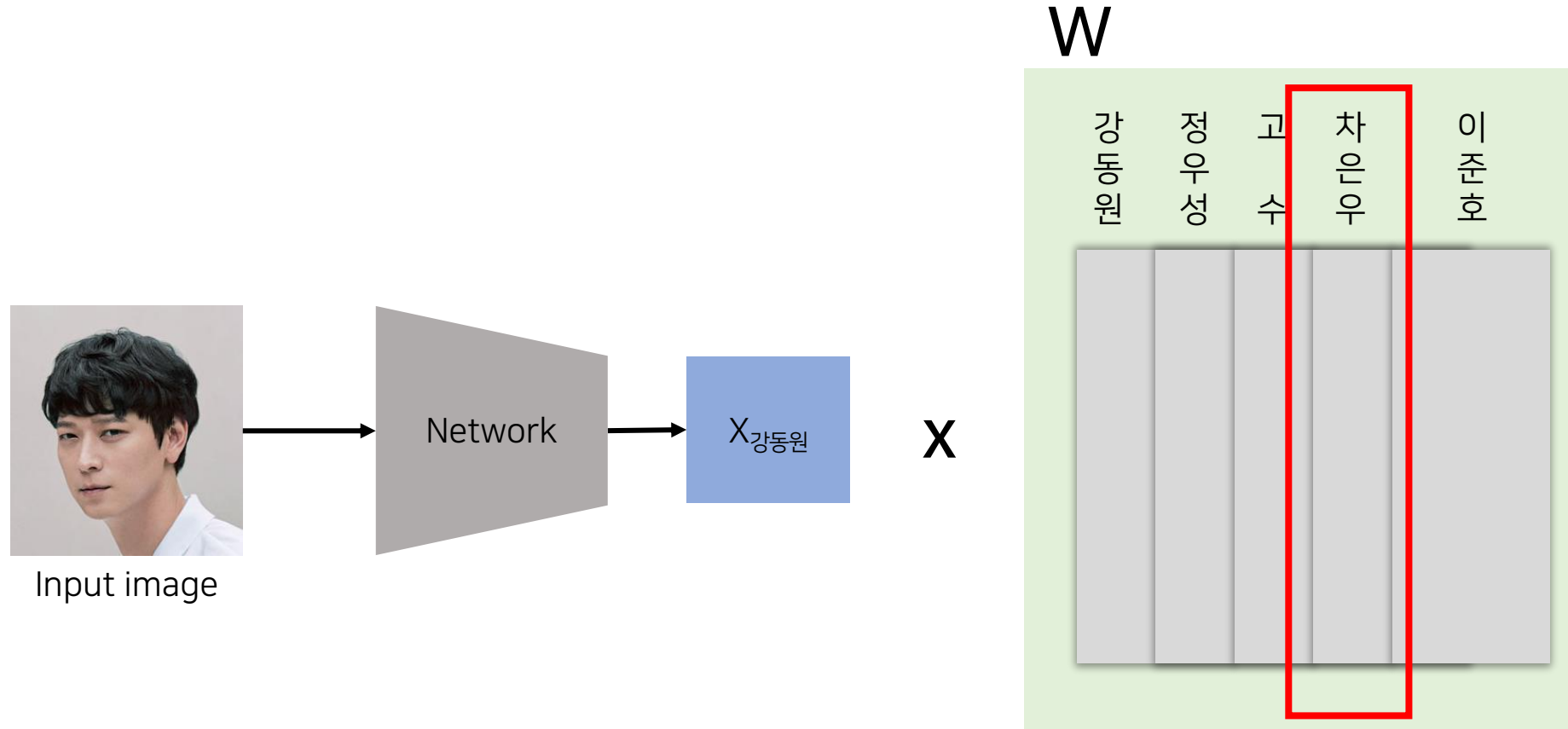
# 00. Abstract



 Negative pair!!

정리  
 $y_i$ :  $y_i$  class  
 $x_i$ : class  $y_i$ 에 속하는  $i$ 번째 sample의 deep feature

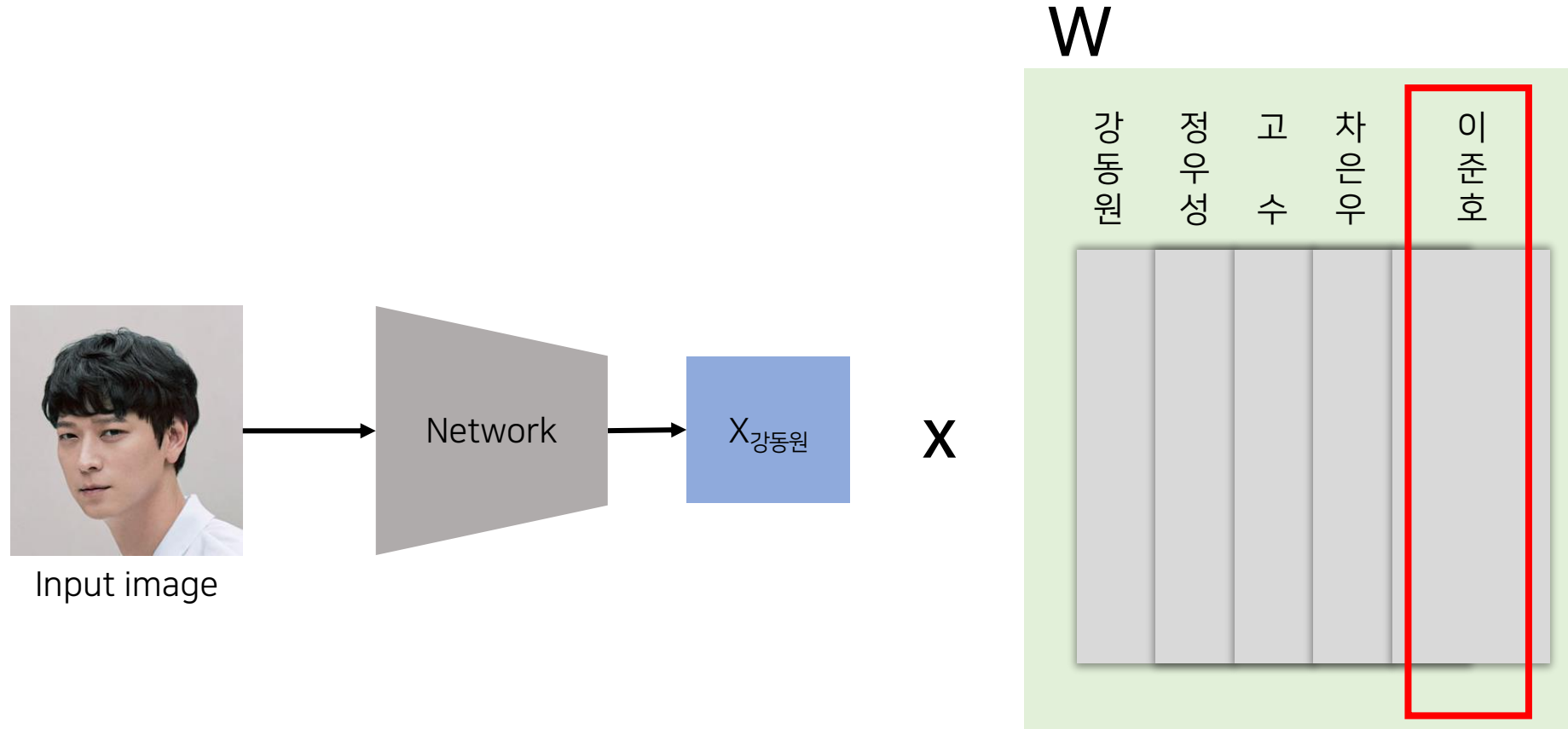
# 00. Abstract



**➡ Negative pair!!**

정리  
 $y_i$ :  $y_i$  class  
 $x_i$ : class  $y_i$ 에 속하는  $i$ 번째 sample의 deep feature

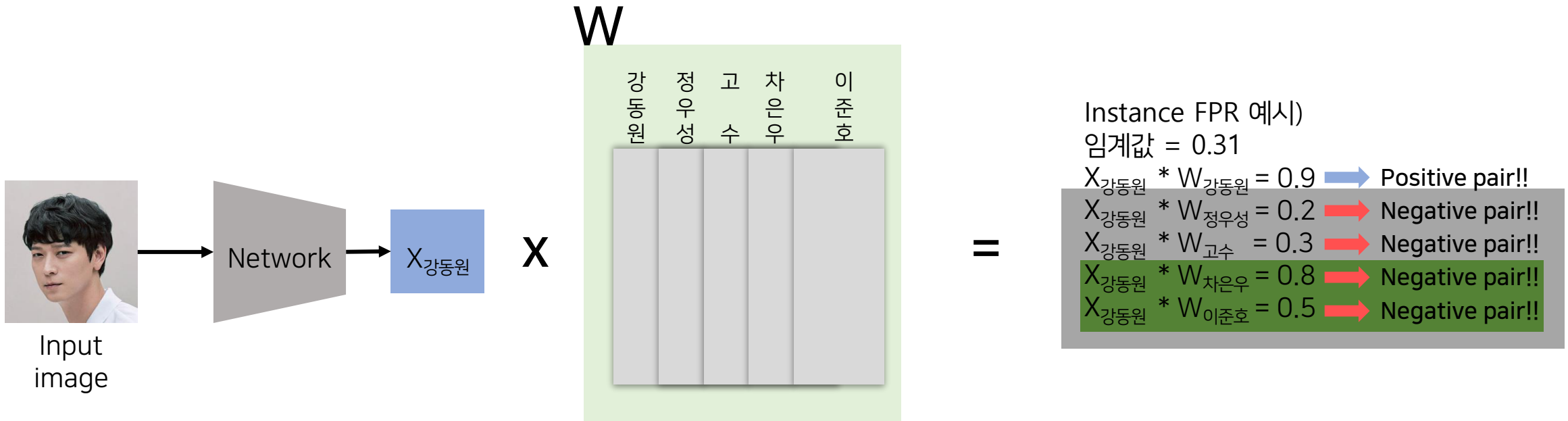
# 00. Abstract



 Negative pair!!

정리  
 $y_i$ :  $y_i$  class  
 $x_i$ : class  $y_i$ 에 속하는  $i$ 번째 sample의 deep feature

# 00. Abstract



$$\text{Instance FPR} = \frac{\text{임계치 이상의 유사도를 갖는 } non\text{-target similarities 수}}{non\text{-target similarities 수}} = \frac{2}{4}$$

정리

$y_i$ :  $y_i$  class

$x_i$ : class  $y_i$ 에 속하는  $i$ 번째 sample의 deep feature

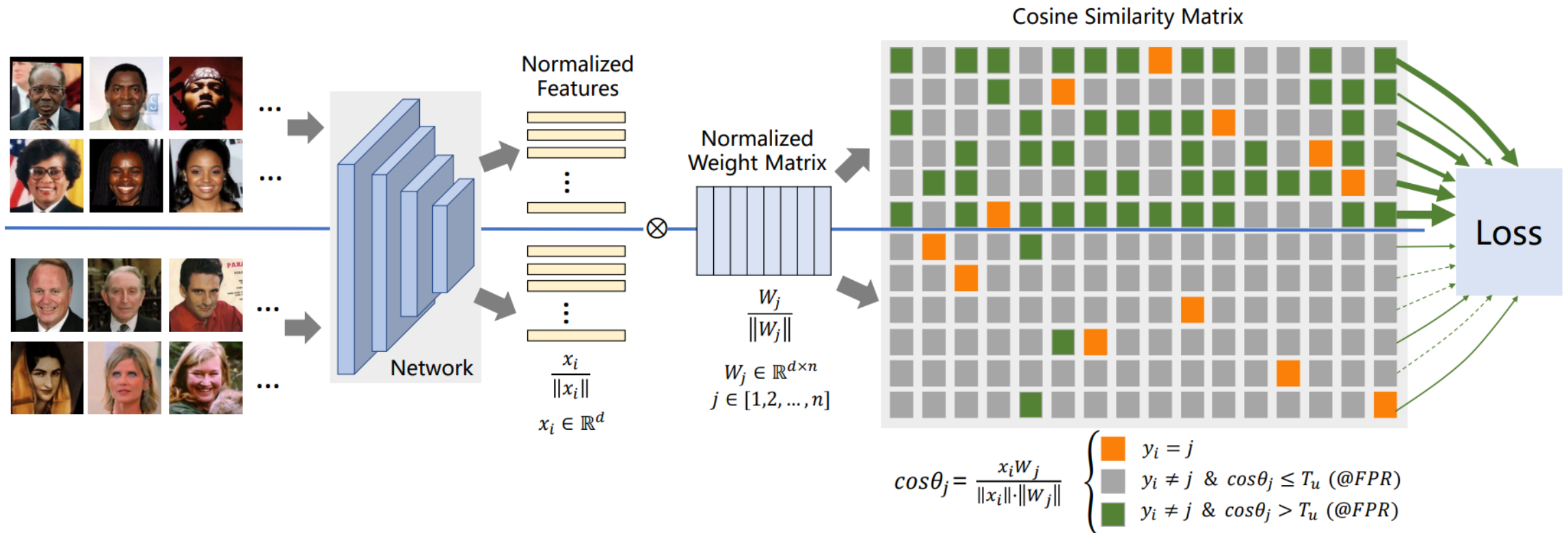
target similarities 수 = Positive pair

non-target similarities 수 = Negative pair

- 다르게 쓰는 이유: 전체 Positive/Negative pair를 보는 것이 아닌 미니 배치 사이즈만큼 보기 때문에 전체 positive/negative pair와 구별해줄 필요가 있어서 다르게 쓰는 것이 아닌지 추측



# 00. Abstract

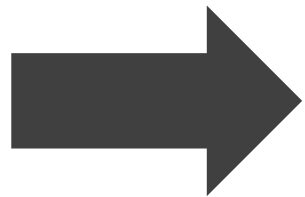
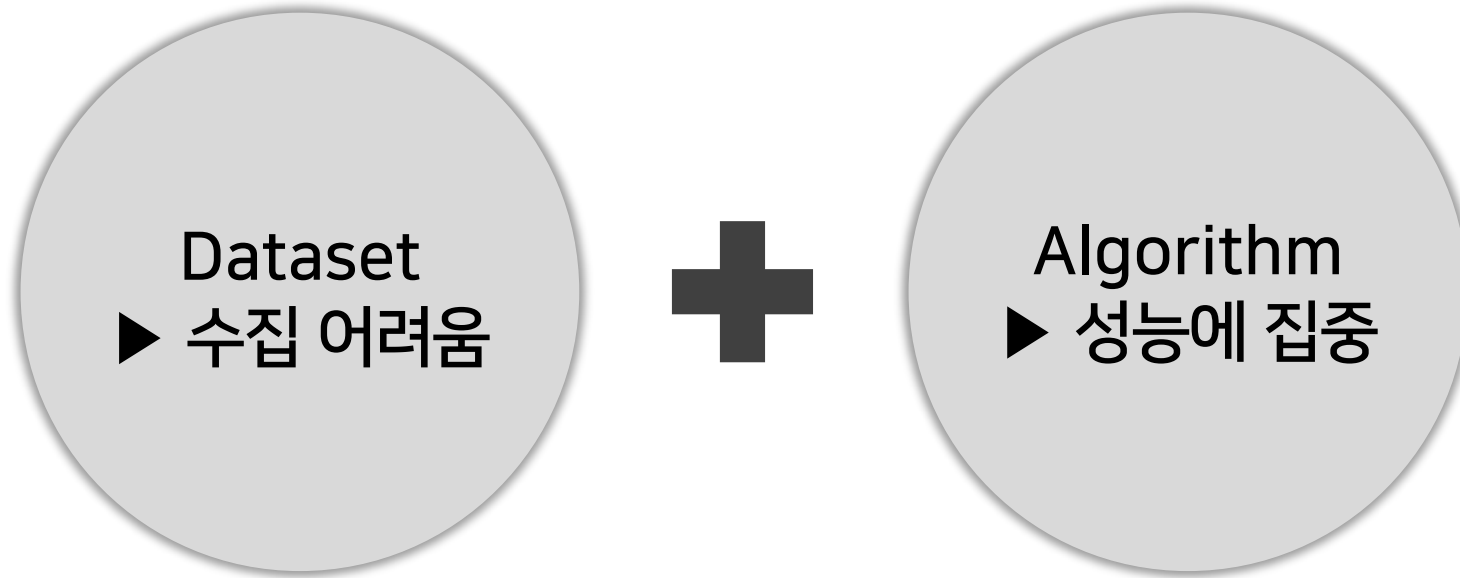


$$\text{Instance FPR} = \frac{\text{초록박스}(FP) \text{ 수}}{\text{회색박스 수} + \text{초록박스}(FP) \text{ 수}} = \frac{\text{임계치 이상의 유사도를 갖는 non-target similarities 수}}{\text{non-target similarities 수}}$$

# **01. Introduction**

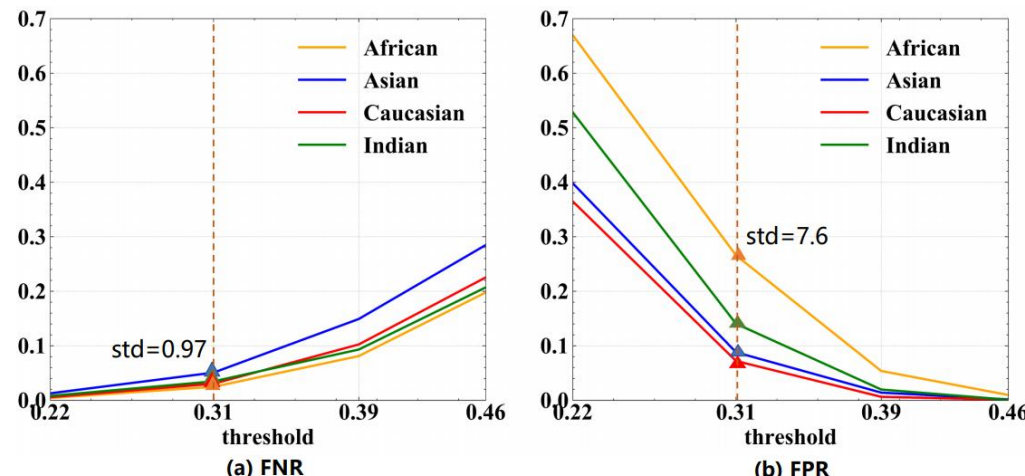
# 01. Introduction

---



균형 여부 관계 없이 편향을 완화할 수 있는 알고리즘 필요

# 01. Introduction

이전 연구 한계 극복 방법	Instance FPR의 일관성을 증가시켜 얼굴 인식의 편향을 완화하는 FPR 페널티 손실 제안
FNR과 FPR의 일관성 차이	 <p>Figure (a) FNR: The graph shows False Negative Rate (FNR) on the y-axis (0.0 to 0.7) versus threshold on the x-axis (0.22 to 0.46). Four lines represent different ethnic groups: African (orange), Asian (blue), Caucasian (red), and Indian (green). All lines show an upward trend as the threshold increases. A vertical dashed line is drawn at threshold 0.31, with a label 'std=0.97' indicating the standard deviation of the FNR values at this threshold.</p> <p>Figure (b) FPR: The graph shows False Positive Rate (FPR) on the y-axis (0.0 to 0.7) versus threshold on the x-axis (0.22 to 0.46). Four lines represent different ethnic groups: African (orange), Asian (blue), Caucasian (red), and Indian (green). All lines show a downward trend as the threshold increases. A vertical dashed line is drawn at threshold 0.31, with a label 'std=7.6' indicating the standard deviation of the FPR values at this threshold.</p>
Contributions	<ul style="list-style-type: none"><li>• 이미지의 인구통계학적 주석 불필요</li><li>• 얼굴 인식에서 일반적으로 사용되는 softmax 기반 손실 함수에 쉽게 적용 가능</li><li>• 인종/성별/연령과 같은 다양한 속성으로 구분된 모든 인구통계학적 그룹의 편향 완화</li><li>• SOTA 경쟁사보다 우수</li></ul>

## **02. Related Work**

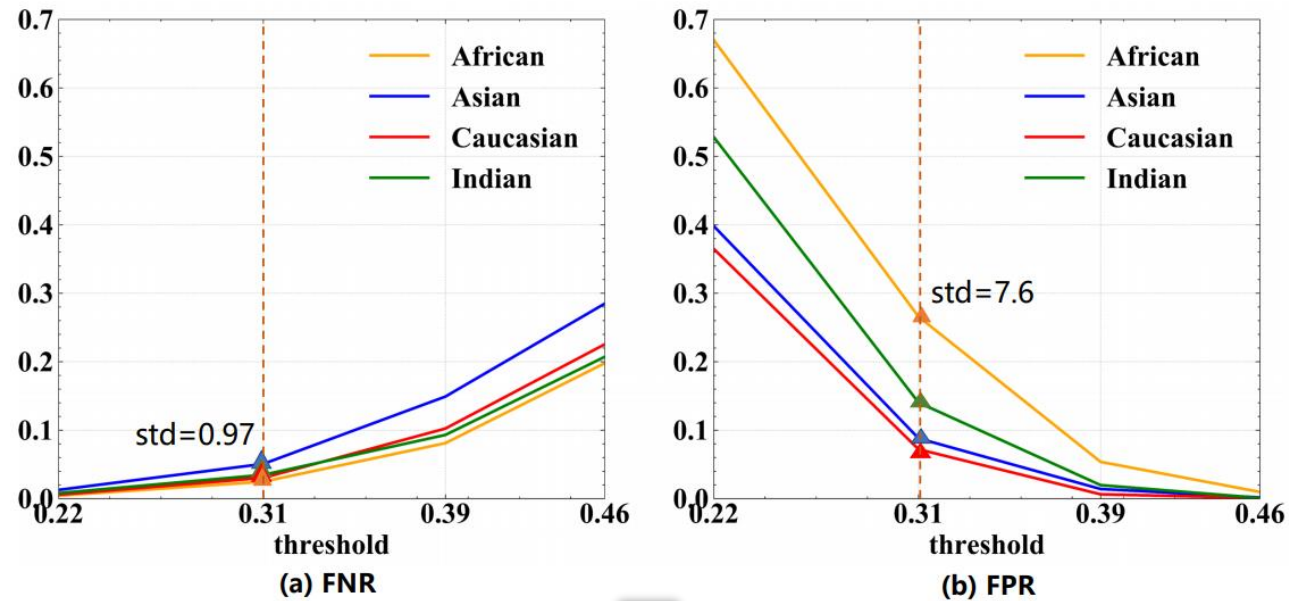
# 02. Related Work

Loss Functions	성능이 향상되지만 편향 고려 불가	
Dataset	DiF(Diversity in Faces)	100만 개의 얼굴 이미지에 대한 주석 제공
	RFW(Racial faces in-the-wild)	인종 편견 연구
	BUPT-balanced	인종에 대해 균형 잡힌 dataset 도입
	BUPT-Globalface	세계 인구 실제 분포 공개
	BFW	8개의 인구 통계학적 그룹 포함
Algorithm	<ul style="list-style-type: none"><li>• Deep Information Maximization Adaptation 네트워크</li><li>• 편향 제거 적대 네트워크: 4개의 특정 분류기 사용자</li></ul>	

## **03. Proposed Approach**

# 03-1. Demographic Bias

- FPR vs Bias
  - 인구통계학적 그룹마다 FNR에 비해 FPR의 일관성이 떨어짐



인구통계 전반에 걸쳐 FNR보다 FPR에서 높은 일관성 달성 필수



# 03-2. FPR Penalty Loss

- 기호

$$\text{FPR } \gamma^+ = \frac{\sum_{i=1}^{N^-} \mathbb{1}(S^-[i] > T_u)}{N^-},$$

# 임계치를 넘는 유사도를 갖는 Negative pair 수  
# 전체적인 Negative pair 수

$$\text{FNR } \gamma^- = \frac{\sum_{i=1}^{N^+} \mathbb{1}(S^+[i] < T_u)}{N^+},$$

# 임계치를 넘는 유사도를 갖는 Positive pair 수  
# 전체적인 Positive pair 수

그룹 내 FNR, FPR

$$\gamma_g^+ = \frac{\sum_{i=1}^{N_g^-} \mathbb{1}(S_g^-[i] > T_u)}{N_g^-},$$

$$\gamma_g^- = \frac{\sum_{i=1}^{N_g^+} \mathbb{1}(S_g^+[i] < T_u)}{N_g^+},$$

# 03-2. FPR Penalty Loss

---

- Softmax Loss Function

- 개별가중치는  $l_2$  norm에 의해  $\|W_j\| = 1$ ,  $b_j = 0$ 로 설정
- $X_i$  deep feature --정규화-->  $s$

$$\mathcal{L} = -\log \frac{e^{W_{y_i} x_i + b_{y_i}}}{\sum_{j=1}^n e^{W_j x_i + b_j}}, \quad \rightarrow$$

# 03-2. FPR Penalty Loss

---

- Softmax Loss Function

- 개별가중치는  $l_2$  norm에 의해  $\|W_j\| = 1$ ,  $b_j = 0$ 로 설정
- $X_i$  deep feature --정규화-->  $s$

$$\mathcal{L} = -\log \frac{e^{W_{y_i} x_i + b_{y_i}}}{\sum_{j=1}^n e^{W_j x_i + b_j}}, \quad \Rightarrow \quad \mathcal{L} = -\log \frac{e^{s(\cos \theta_{y_i})}}{e^{s(\cos \theta_{y_i})} + \sum_{j \neq y_i}^n e^{s(\cos \theta_j)}}.$$

# 03-2. FPR Penalty Loss

- Softmax Loss Function

- 개별가중치는  $l_2$  norm에 의해  $\|W_j\| = 1$ ,  $b_j = 0$ 로 설정
- $X_i$  deep feature --정규화-->  $s$

$$\mathcal{L} = -\log \frac{e^{W_{y_i} x_i + b_{y_i}}}{\sum_{j=1}^n e^{W_j x_i + b_j}}, \quad \Rightarrow \quad \mathcal{L} = -\log \frac{e^{s(\cos \theta_{y_i})}}{e^{s(\cos \theta_{y_i})} + \sum_{j \neq y_i}^n e^{s(\cos \theta_j)}}.$$

- Positive pair와 Negative pair 분리

# 03-2. FPR Penalty Loss

- Softmax Loss Function

- 개별가중치는  $l_2$  norm에 의해  $\|W_j\| = 1$ ,  $b_j = 0$ 로 설정
- $X_i$  deep feature --정규화-->  $s$

$$\mathcal{L} = -\log \frac{e^{W_{y_i} x_i + b_{y_i}}}{\sum_{j=1}^n e^{W_j x_i + b_j}}, \quad \Rightarrow \quad \mathcal{L} = -\log \frac{e^{s(\cos \theta_{y_i})}}{e^{s(\cos \theta_{y_i})} + \sum_{j \neq y_i}^n e^{s(\cos \theta_j)}}.$$

- Positive pair와 Negative pair 분리

- $G(\cos \theta_{y_i}) = \cos(\theta_{y_i} + m)$ : inter-class similarity 강조
- $H(\cos \theta_j)$ : mining-based loss functions, intra-class의 혼란을 감소

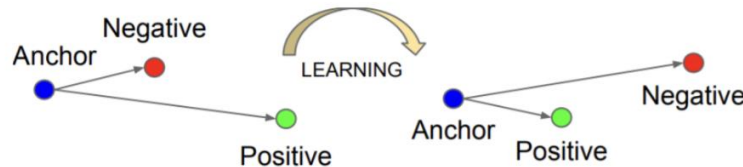


Figure 3. The **Triplet Loss** minimizes the distance between an *anchor* and a *positive*, both of which have the same identity, and maximizes the distance between the *anchor* and a *negative* of a different identity.

# 03-2. FPR Penalty Loss

- Softmax Loss Function

- 개별가중치는  $l_2$  norm에 의해  $\|W_j\| = 1$ ,  $b_j = 0$ 로 설정
- $X_i$  deep feature --정규화-->  $s$

$$\mathcal{L} = -\log \frac{e^{W_{y_i} x_i + b_{y_i}}}{\sum_{j=1}^n e^{W_j x_i + b_j}}, \quad \Rightarrow \quad \mathcal{L} = -\log \frac{e^{s(\cos \theta_{y_i})}}{e^{s(\cos \theta_{y_i})} + \sum_{j \neq y_i}^n e^{s(\cos \theta_j)}}.$$

- $G(\cos \theta_{y_i}) = \cos(\theta_{y_i} + m)$ : inter-class similarity 강조
- $H(\cos \theta_j)$ : mining-based loss functions, intra-class의 혼란을 감소

$$\Rightarrow \quad \mathcal{L} = -\log \frac{e^{s \cdot G(\cos \theta_{y_i})}}{e^{s \cdot G(\cos \theta_{y_i})} + \sum_{j \neq y_i}^n e^{s \cdot H(\cos \theta_j)}},$$

# 03-2. FPR Penalty Loss

- Extra Penalty on the FPR of Instance

Since the  $y_i$ -th column of the weight  $W$  usually could be regarded as a representative of the  $y_i$ -th class, for the  $i$ -th instance belonging to class  $y_i$ , the target logit  $\cos \theta_{y_i}$  could be considered as the similarity of a positive pair, while the non-target logits  $\cos \theta_j, j \neq y_i$  could be considered as the similarities of negative pairs.

Instance FPR

$$\gamma_i^+ = \frac{\sum_{j=1, j \neq y_i}^n \mathbb{1}(\cos \theta_j > T_u)}{n - 1}, \quad = \frac{\text{임계치 이상의 유사도를 갖는 } non\text{-target similarities 수}}{non\text{-target similarities 수}}$$

# 03-2. FPR Penalty Loss

Since the  $y_i$ -th column of the weight  $W$  usually could be regarded as a representative of the  $y_i$ -th class, for the  $i$ -th instance belonging to class  $y_i$ , the target logit  $\cos \theta_{y_i}$  could be considered as the similarity of a positive pair, while the non-target logits  $\cos \theta_j, j \neq y_i$  could be considered as the similarities of negative pairs.

- Extra Penalty on the FPR of Instance

Instance FPR

$$\gamma_i^+ = \frac{\sum_{j=1, j \neq y_i}^n \mathbb{1}(\cos \theta_j > T_u)}{n-1}, = \frac{\text{임계치 이상의 유사도를 갖는 } non\text{-target similarities 수}}{non\text{-target similarities 수}}$$

Loss function

$$\mathcal{L} = -\log \frac{e^{s \cdot G(\cos \theta_{y_i})}}{e^{s \cdot G(\cos \theta_{y_i})} + \sum_{j \neq y_i}^n e^{s \cdot \left( \cos \theta_j + \alpha \frac{\gamma_i^+}{\gamma_u^+} \right)}}.$$

Instance FPR 일관성 있게 하려면(= 모든  $\gamma_i^+$ 이  $\gamma_u^+$ 에 가깝게 만들기 위해)  $\alpha \frac{\gamma_i^+}{\gamma_u^+}$  추가

$e^{s \alpha \frac{\gamma_i^+}{\gamma_u^+}} > 1$  이므로  $\gamma_i^+$ 가 커질수록 Loss가 커지는 불평등



# 03-2. FPR Penalty Loss

Since the  $y_i$ -th column of the weight  $W$  usually could be regarded as a representative of the  $y_i$ -th class, for the  $i$ -th instance belonging to class  $y_i$ , the target logit  $\cos \theta_{y_i}$  could be considered as the similarity of a positive pair, while the non-target logits  $\cos \theta_j, j \neq y_i$  could be considered as the similarities of negative pairs.

- Extra Penalty on the FPR of Instance

Instance FPR

$$\gamma_i^+ = \frac{\sum_{j=1, j \neq y_i}^n \mathbb{1}(\cos \theta_j > T_u)}{n-1}, = \frac{\text{임계치 이상의 유사도를 갖는 } non\text{-target similarities 수}}{non\text{-target similarities 수}}$$

Loss function

$$\mathcal{L} = -\log \frac{e^{s \cdot G(\cos \theta_{y_i})}}{e^{s \cdot G(\cos \theta_{y_i})} + \sum_{j \neq y_i}^n e^{s \cdot \left( \cos \theta_j + \alpha \frac{\gamma_i^+}{\gamma_u^+} \right)}}.$$

Instance FPR 일관성 있게 하려면 (= 모든  $\gamma_i^+$ 이  $\gamma_u^+$ 에 가깝게 만들기 위해)  $\alpha \frac{\gamma_i^+}{\gamma_u^+}$  추가

$e^{s \alpha \frac{\gamma_i^+}{\gamma_u^+}} > 1$  이므로  $\gamma_i^+$ 가 커질수록 Loss가 커지는 불평등

Instance FPR

$$\bar{\gamma}_i^+ = \frac{\sum_{j=1, j \neq y_i}^n \mathbb{1}(\cos \theta_j > T_u) \cdot F(\cos \theta_j)}{n-1}.$$

$$F(z) = z^p$$

Negative pair 중 유사도 클수록  $(\cos \theta_j)^p$  항을 추가하여 높은 가중치 부여

# 03-2. FPR Penalty Loss

## Algorithm 1: FPR Penalty Loss

**Input:** The deep feature of  $i$ -th sample with its label  $y_i$ , cosine similarity  $\cos \theta_j$  of two vectors, last fully-connected layer parameters  $W$ , embedding network parameters  $\Theta$ , class number  $c$ , sample number  $n$ , learning rate  $\lambda$ , and overall false positive rate  $\gamma_u^+$

iteration number  $k \leftarrow 0$ , parameter  $t \leftarrow 0$ ,  $\gamma_u^+ \leftarrow 1e^{-4}$ ;

**while not converged do**

    Compute the  $\lceil \gamma_u^+ n(c-1) \rceil$ -th largest value of set  $\{\cos \theta_j \mid i \in [1, n], j \in [1, c], j \neq y_i\}$  as the temporary threshold  $T_u$ ;

**if**  $\cos(\theta_j) > T_u$  **then**

$I_j = 1$ ;

**else**

$I_j = 0$ ;

**end**

    Compute the weighted FPR  $\bar{\gamma}_i^+$  by Eq. 11;

    Compute the loss  $\mathcal{L}$  by Eq. 10 (replace  $\gamma_i^+$  by  $\bar{\gamma}_i^+$ );

    Compute the gradient of  $W_j$  and  $x_i$  by Eq. 12;

    Update the parameters  $W$  and  $\Theta$  by:

$$W^{(k+1)} = W^{(k)} - \lambda^{(k)} \frac{\partial \mathcal{L}_i}{\partial W},$$

$$\Theta^{(k+1)} = \Theta^{(k)} - \lambda^{(k)} \frac{\partial \mathcal{L}_i}{\partial x_i} \frac{\partial x_i}{\partial \Theta^{(k)}};$$

$k \leftarrow k + 1$ ;

**end**

**Output:**  $W, \Theta$ .

# Input:

- sample's deep feature & label  $y_i$  간 cosine 유사도
- last fully-connected layer parameters  $W$
- embedding network parameters  $\theta$
- class number  $c$
- sample number  $n$
- learning rate  $\lambda$
- overall false positive rate  $\gamma_u^+$

# 수렴되지 않을 동안

# sample 중에서 negative pair의 cosine 유사도 계산

# 만약 cosine 유사도가  $T_u$ 보다 클 때는 1; FP

# "  $T_u$ 보다 작을 때는 0; TN(정답)

# Instance FPR 계산  $\bar{\gamma}_i^+ = \frac{\sum_{j=1, j \neq y_i}^n \mathbb{1}(\cos \theta_j > T_u) \cdot F(\cos \theta_j)}{n-1}$ .

# Loss 계산

$$\mathcal{L} = -\log \frac{e^{s \cdot G(\cos \theta_{y_i})}}{e^{s \cdot G(\cos \theta_{y_i})} + \sum_{j \neq y_i}^n e^{s \cdot \left( \cos \theta_j + \alpha \frac{\gamma_i^+}{\gamma_u^+} \right)}}.$$

# 역전파를 통한  $W_j$  계산

# parameters update

$$\frac{\partial \mathcal{L}_i}{\partial W_{y_i}} = \frac{\partial \mathcal{L}_i}{\partial G_i} \cdot x_i,$$

$$\frac{\partial \mathcal{L}_i}{\partial W_j} = \left( 1 + \frac{\alpha}{\gamma_u^+} \cdot I_j \frac{\partial F}{\partial \cos \theta_j} \right) \cdot \frac{\partial \mathcal{L}_i}{\partial H_j} \cdot x_i,$$

$$\frac{\partial \mathcal{L}_i}{\partial x_i} = \frac{\partial \mathcal{L}_i}{\partial G_i} \cdot W_{y_i} + \left( 1 + \frac{\alpha}{\gamma_u^+} \cdot \sum_{j \neq y_i} I_j \frac{\partial F}{\partial \cos \theta_j} \right) \cdot \frac{\partial \mathcal{L}_i}{\partial H_j} \cdot W_j,$$

# 03-2. FPR Penalty Loss

---

- Optimization

$$\frac{\partial \mathcal{L}_i}{\partial W_{y_i}} = \frac{\partial \mathcal{L}_i}{\partial G_i} \cdot x_i, \quad \# G_i = \cos \theta_{y_i} - m$$

$$\frac{\partial \mathcal{L}_i}{\partial W_j} = \left( 1 + \frac{\alpha}{\gamma_u^+} \cdot I_j \frac{\partial F}{\partial \cos \theta_j} \right) \cdot \frac{\partial \mathcal{L}_i}{\partial H_j} \cdot x_i, \quad \# H_j = \cos \theta_j + \alpha \frac{\bar{\gamma}_i^+}{\gamma_u^+}$$

$$\frac{\partial \mathcal{L}_i}{\partial x_i} = \frac{\partial \mathcal{L}_i}{\partial G_i} \cdot W_{y_i} + \left( 1 + \frac{\alpha}{\gamma_u^+} \cdot \sum_{j \neq y_i} \overset{\text{False positive case일 경우, } I_j = 1}{\boxed{I_j}} \frac{\partial F}{\partial \cos \theta_j} \right) \cdot \frac{\partial \mathcal{L}_i}{\partial H_j} \cdot W_j, \quad (12)$$

# 04. Experiments

# 04-1. Experimental setting

---

- Dataset

훈련	BUPT-Balancedface	<ul style="list-style-type: none"><li>28000명 유명인의 이미지(130만장)</li><li>인종 분포: 인종당 7000명</li></ul>
	BUPT-Globalface	<ul style="list-style-type: none"><li>38000명 유명인의 200장 이미지 포함</li><li>인종 분포: 세계 인구의 실제 분포와 거의 동일</li></ul>
테스트	RFW	<ul style="list-style-type: none"><li>4개 인종(아프리카, 아시아, 코카서스, 인도) 그룹</li><li>각 인종 3000명의 얼굴이 약 10000장 포함</li></ul>
	BFW	<ul style="list-style-type: none"><li>ID, 성별 및 인종을 포함한 더 많은 속성을 가진 균형 잡힌 얼굴 데이터 제공</li><li>2개의 성별 &amp; 4개 인종(흑인, 백인, 아시아인, 인도인) → 8개의 인구 통계학적 그룹</li></ul>

# 04-1. Experimental setting

---

- Training setting

Embedding network	ResNet34, ResNet50, ResNet100 채택	
Framework	Pytorch	
Batch size	512	
s	64	
m	0.35	
Train	BUPT-Balancedface	BUPT-Globalface
	<ul style="list-style-type: none"><li>Learning rate: 0.1 (20, 32, 36 epoch에서 10으로 나눔)</li><li>Epoch: 40</li></ul>	<ul style="list-style-type: none"><li>Learning rate: 0.1 (10, 18, 22 epoch에서 10으로 나눔)</li><li>Epoch: 24</li></ul>

# 04-1. Experimental setting

- Training setting

Embedding network	$\mathcal{L} = -\log \frac{e^{\boxed{s} \cdot G(\cos \theta_{y_i})}}{e^{\boxed{s} \cdot G(\cos \theta_{y_i})} + \sum_{j \neq y_i}^n e^{s \cdot H(\cos \theta_j)}},$	
Framework		
Batch size		
s		
m		
Train	BUPT-Balancedface	BUPT-Globalface
	<ul style="list-style-type: none"><li>Learning rate: 0.1 (20, 32, 36 epoch에서 10으로 나눔)</li><li>Epoch: 40</li></ul>	<ul style="list-style-type: none"><li>Learning rate: 0.1 (10, 18, 22 epoch에서 10으로 나눔)</li><li>Epoch: 24</li></ul>

# 04-1. Experimental setting

- Training setting

Embedding network	ResNet	$\frac{\partial \mathcal{L}_i}{\partial W_{y_i}} = \frac{\partial \mathcal{L}_i}{\partial G_i} \cdot x_i, \quad \# G_i = \cos \theta_{y_i} - m$ $\frac{\partial \mathcal{L}_i}{\partial W_j} = \left( 1 + \frac{\alpha}{\gamma_u^+} \cdot I_j \frac{\partial F}{\partial \cos \theta_j} \right) \cdot \frac{\partial \mathcal{L}_i}{\partial H_j} \cdot x_i, \quad \# H_j = \cos \theta_j + \alpha \frac{\overline{\gamma_i^+}}{\gamma_u^+}$ $\frac{\partial \mathcal{L}_i}{\partial x_i} = \frac{\partial \mathcal{L}_i}{\partial G_i} \cdot W_{y_i} + \left( 1 + \frac{\alpha}{\gamma_u^+} \cdot \sum_{j \neq y_i} I_j \frac{\partial F}{\partial \cos \theta_j} \right) \cdot \frac{\partial \mathcal{L}_i}{\partial H_j} \cdot W_j,$
Framework	Pytor	
Batch size	512	
s	64	
m	0.35	
Train	BUPT-Balancedface	BUPT-Globalface
	<ul style="list-style-type: none"> <li>Learning rate: 0.1 (20, 32, 36 epoch 에서 10으로 나눔)</li> <li>Epoch: 40</li> </ul>	<ul style="list-style-type: none"> <li>Learning rate: 0.1 (10, 18, 22 epoch에 서 10으로 나눔)</li> <li>Epoch: 24</li> </ul>

(12)



# 04-1. Experimental setting

---

- Training setting

Embedding network	ResNet34, ResNet50, ResNet100 채택	
Framework	Pytorch	
Batch size	512	
s	64	
m	0.35	
Train	BUPT-Balancedface	BUPT-Globalface
	<ul style="list-style-type: none"><li>Learning rate: 0.1 (20, 32, 36 epoch에서 10으로 나눔)</li><li>Epoch: 40</li></ul>	<ul style="list-style-type: none"><li>Learning rate: 0.1 (10, 18, 22 epoch에서 10으로 나눔)</li><li>Epoch: 24</li></ul>

# 04-2. Ablation Study

- Effect of the overall FPR  $\gamma_u^+$ 
  - $\gamma_u^+ = 10^{-4}$  에서 최상의 성능 달성  
( $\gamma_u^+ = 10^{-5}$  에서 백인의 높은 정확도를 제외)

Table 1. Verification performance (%) of different FPR parameter  $\gamma$ .

Methods (%)	African	Asian	Caucasian	Indian	Avg	Std
$\gamma_u^+ = 10^{-5}$	95.60	95.10	<b>97.18</b>	96.32	96.05	0.91
$\gamma_u^+ = 10^{-4}$	<b>95.95</b>	<b>95.17</b>	96.78	<b>96.38</b>	<b>96.07</b>	<b>0.69</b>
$\gamma_u^+ = 10^{-3}$	95.47	94.90	96.92	96.12	95.84	0.87
$\gamma_u^+ = 10^{-2}$	95.45	94.78	96.98	96.13	95.84	0.94
$\gamma_u^+ = 10^{-1}$	95.23	94.60	95.87	95.97	95.42	0.64

# 04-2. Ablation Study

- Effect of exponent  $p$  in  $F(z)$ 
  - 고정 FPR =  $10^{-4}$ 로 설정
  - $F(z) = z^p$ 에서  $p$  효과 조사
  - $p=2$ 가 적당

Table 2. Verification performance (%) of different exponent  $p$  in  $F(z)$ .

Methods (%)	African	Asian	Caucasian	Indian	Avg	Std
$p = 0.25$	95.35	95.10	96.97	96.07	95.87	0.84
$p = 0.5$	95.27	94.93	96.58	96.02	95.70	0.74
$p = 1.0$	95.18	94.92	96.90	95.83	95.71	0.88
$p = 1.5$	95.27	94.67	97.05	96.23	95.80	1.05
$p = 2.0$	<b>95.95</b>	95.17	96.78	<b>96.38</b>	<b>96.07</b>	<b>0.69</b>
$p = 2.5$	95.85	95.00	96.96	96.20	96.00	0.82
$p = 3.0$	95.60	<b>95.18</b>	<b>97.17</b>	95.98	95.98	0.85

# 04-3. Comparison with SOTA methods

- Accuracy on RFW
    - ResNet34 모델 사용
    - BUPT-Balancedface dataset 사용
    - SOTA에 비해 평균 정확도 약 0.77% 향상
    - SOTA에 비해 표준편차 0.69로 감소
    - ResNet50, 100에서도 비교적 좋은 성능
- ➔ 정확도 및 표준 편차를 통해 인종 균형/불균형 dataset에서 좋은 성능 달성

Table 3. Verification performance (%) of protocol on RFW with SOTA methods ([BUPT-Balancedface]).

Methods (%)	African	Asian	Caucasian	Indian	Avg	Std
ArcFace-R34 [18]	93.98	93.72	96.18	94.67	94.64	1.11
CosFace-R34 [18]	92.93	92.98	95.12	93.93	93.74	1.03
DebFace-R34 (ECCV'20)	93.67	94.33	95.95	94.78	94.68	0.83
PFE-R34 [5]	95.17	94.27	96.38	94.60	95.11	0.93
GAC-R34 [5]	94.65	94.93	96.23	95.12	95.23	0.60
RL-RBN-R34(cos) (CVPR'20)	95.27	94.52	95.47	95.15	95.10	<b>0.41</b>
RL-RBN-R34(arc) (CVPR'20)	95.00	94.82	96.27	94.68	95.19	0.93
<b>Ours-R34</b>	<b>95.95</b>	<b>95.17</b>	<b>96.78</b>	<b>96.38</b>	<b>96.07</b>	0.69
ArcFace-R50	95.55	94.95	96.68	95.47	95.66	0.73
<b>Ours-R50</b>	<b>96.47</b>	<b>95.75</b>	<b>97.08</b>	<b>96.77</b>	<b>96.52</b>	<b>0.57</b>
ArcFace-R100	96.43	94.98	97.37	96.17	96.24	0.98
<b>Ours-R100</b>	<b>97.03</b>	<b>95.65</b>	<b>97.6</b>	<b>96.82</b>	<b>96.78</b>	<b>0.82</b>

# 04-3. Comparison with SOTA methods

- Accuracy on RFW
  - ResNet34 모델 사용
  - BUPT-Globalface dataset 사용
  - ResNet50, 100에서도 비교적 좋은 성능
- ➔ 정확도 및 표준 편차를 통해 인종 균형/불균형 dataset에서 좋은 성능 달성

Table 4. Verification accuracy (%) of protocol on RFW with SOTA methods ([BUPT-Globalface]).

Methods (%)	African	Asian	Caucasian	Indian	Avg	Std
ArcFace-R34 [18]	93.87	94.55	97.37	95.86	95.37	1.53
CosFace-R34 [18]	92.17	93.50	96.63	94.68	94.25	1.90
RL-RBN-R34(cos) (CVPR'20)	94.27	94.58	96.03	95.15	95.01	0.77
RL-RBN-R34(arc) (CVPR'20)	94.87	95.57	97.08	95.63	95.79	0.93
<b>Ours-R34</b>	<b>95.77</b>	<b>95.85</b>	<b>97.92</b>	<b>96.70</b>	<b>96.56</b>	<b>0.75</b>
ArcFace-R50	96.23	96.43	97.98	96.92	96.89	0.78
<b>Ours-R50</b>	<b>96.85</b>	<b>96.75</b>	<b>98.30</b>	<b>96.95</b>	<b>97.21</b>	<b>0.73</b>
ArcFace-R100	96.68	96.10	98.17	97.32	97.07	0.89
<b>Ours-R100</b>	<b>97.37</b>	<b>96.48</b>	<b>98.57</b>	<b>97.4</b>	<b>97.45</b>	<b>0.85</b>

# 04-3. Comparison with SOTA methods

- SOTA methods와 Bias degree 비교
  - 전체 FPR에 따라 편향 비교
  - 본 논문에 제시한 방법의 편향이 모든 FPR에 있어서 훨씬 낮음

Table 5. Bias degree of protocol on RFW with SOTA methods.

overall FPR	$10^{-5}$	$10^{-4}$	$10^{-3}$	$10^{-2}$
RL-RBN-R34(arc)	351.98	208.44	92.18	16.70
<b>Ours-R34</b>	<b>257.53</b>	<b>185.91</b>	<b>59.25</b>	<b>10.33</b>

Table 6. Bias degree of protocol on BFW with SOTA methods.

overall FPR	$10^{-7}$	$10^{-6}$	$10^{-5}$	$10^{-4}$	$10^{-3}$
RL-RBN-R34(arc)	2.44	2.01	2.49	2.91	2.43
<b>Ours-R34</b>	<b>1.18</b>	<b>1.08</b>	<b>1.18</b>	<b>1.67</b>	<b>1.80</b>

# 04-3. Comparison with SOTA methods

- FPR on RFW
  - (a)에서 아프리카 ROC curve를 통해 높은 성능 확인
  - RFW에 대해 제안한 loss가 RL-RBN(arc)보다 더 좋은 성능임을 증명

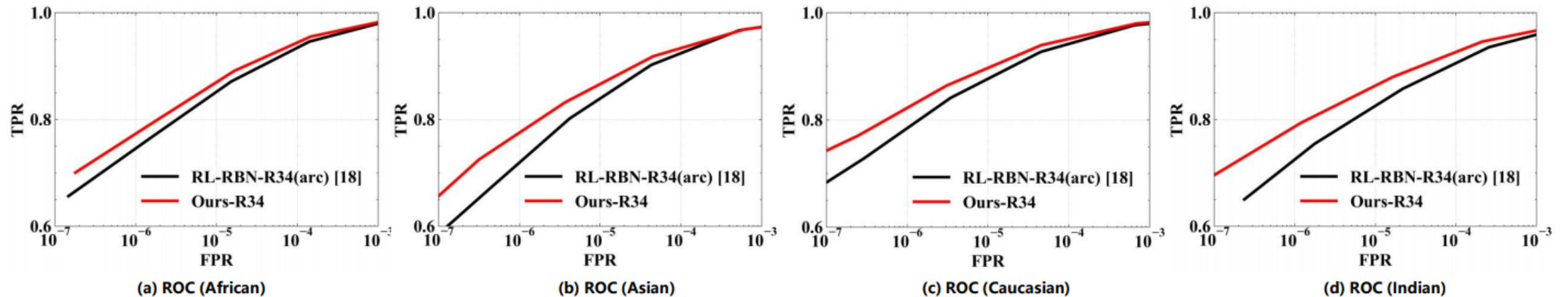
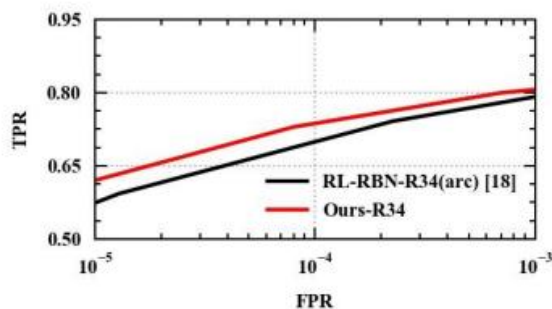


Figure 5. ROC for RFW.

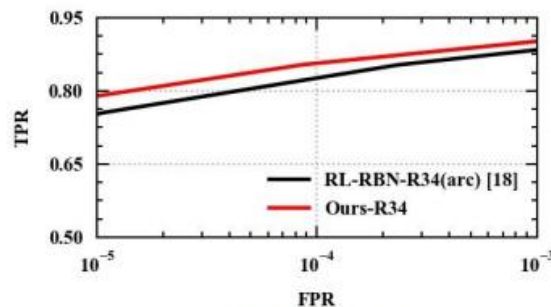


# 04-3. Comparison with SOTA methods

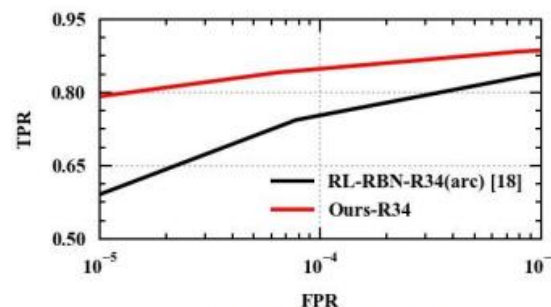
- FPR on BFW
  - BFW의 8개의 모든 인구 통계학적 그룹에 대한 ROC curve
  - 모든 인종에 걸쳐 여성 그룹과 남성 그룹 모두 더 나은 성능 달성



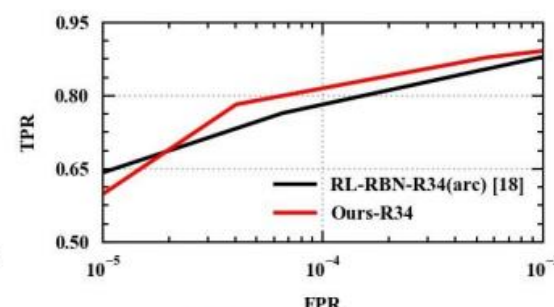
(a) ROC (Asian female)



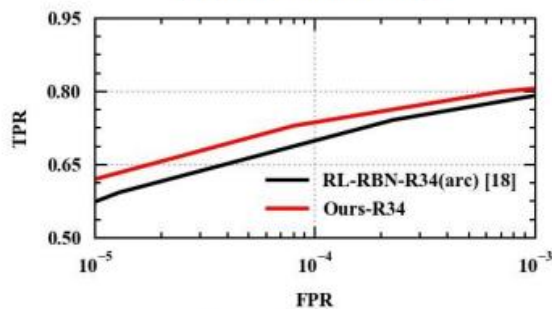
(c) ROC (Black female)



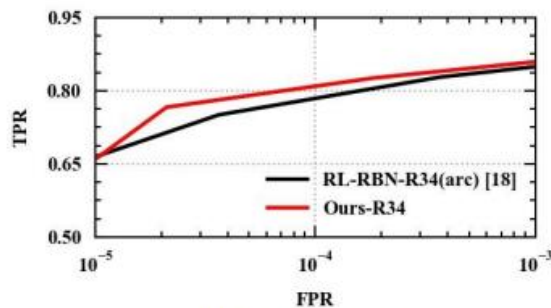
(e) ROC (Indian female)



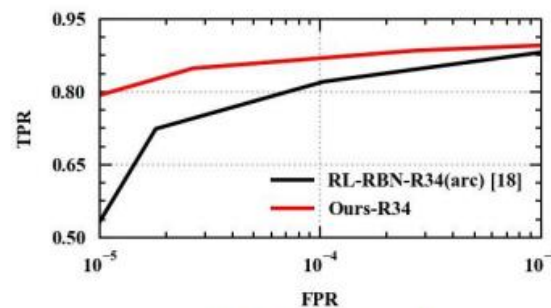
(g) ROC (White female)



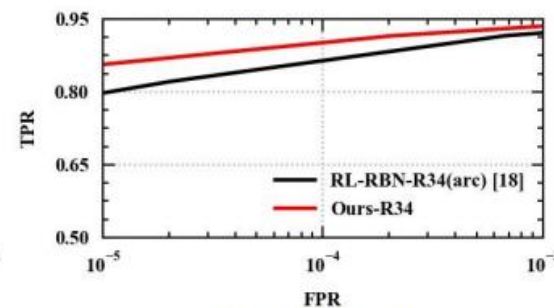
(b) ROC (Asian male)



(d) ROC (Black male)



(f) ROC (Indian male)



(h) ROC (White male)



# **05. Conclusion**

# 05. Conclusion

---

1

얼굴 인식에서 편향 완화 및 공정성 향상

2

SOTA 경쟁사와 비교하여 제안한 방법의 효과 입증

3

향후 FP 사례로 인한 노이즈 샘플 영향 조사 등 다양한 측면으로 확장 가능