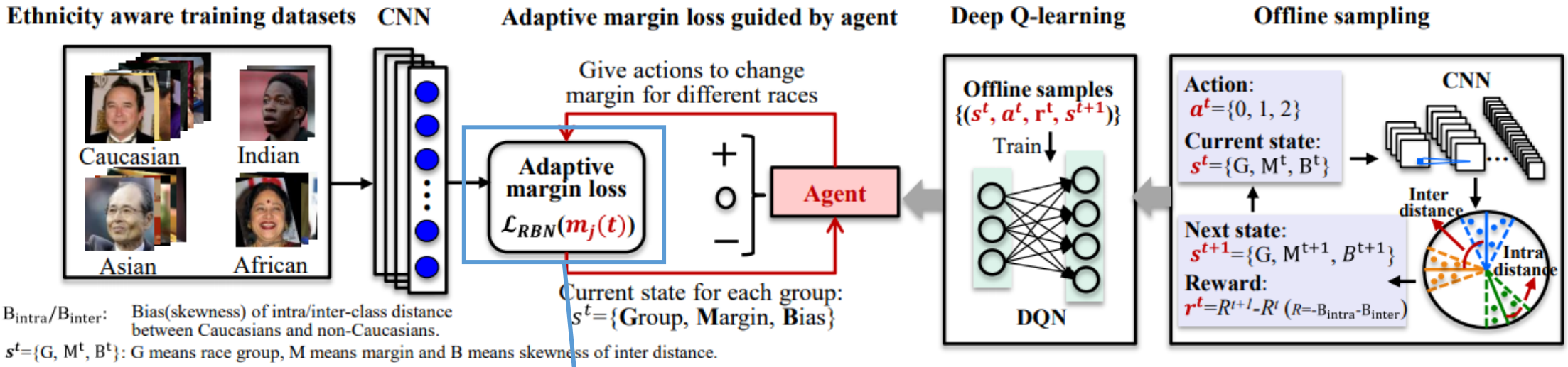
Fairness

: 얼굴 인식의 공정성에 대한 연구 조사

염지현

[18] Mitigating bias in face recognition using skewness-aware



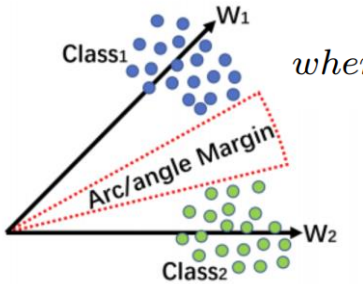
Positive pair: $y^{(j)}$ label(ID)과 $y^{(j)}$에 속하는 deep feature j 간 cosine 유사도

$$L_{RBN} = -\frac{1}{N}\sum_{j=1}^{N} log \frac{e^{s_c\left(cos\left(\theta_{y^{(j)}\,j}+\alpha_j(t)\right)\right)}}{e^{s_c\left(cos\left(\theta_{y^{(j)}\,j}+\alpha_j(t)\right)\right)} + \sum_{i=1,i\neq y^{(j)}}^{n} e^{s_c cos\theta_{ij}}}$$

Negative pair

$$where, \alpha_j(t) = \begin{cases} m, & if\ j \in Caucasian \\ m_j(t), & otherwise \end{cases}$$

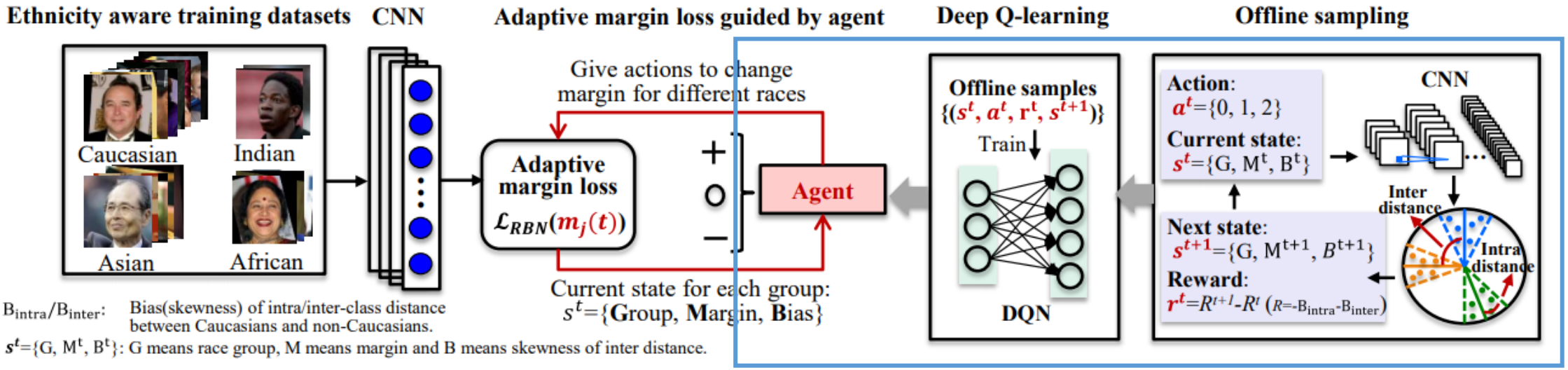
학습하고자 하는 것!



(a) ArcFace

Mei Wang and Weihong Deng. Mitigating bias in face recognition using skewness-aware reinforcement learning. In IEEE Conf. Comput. Vis. Pattern Recog., pages 9322–9331, 2020.

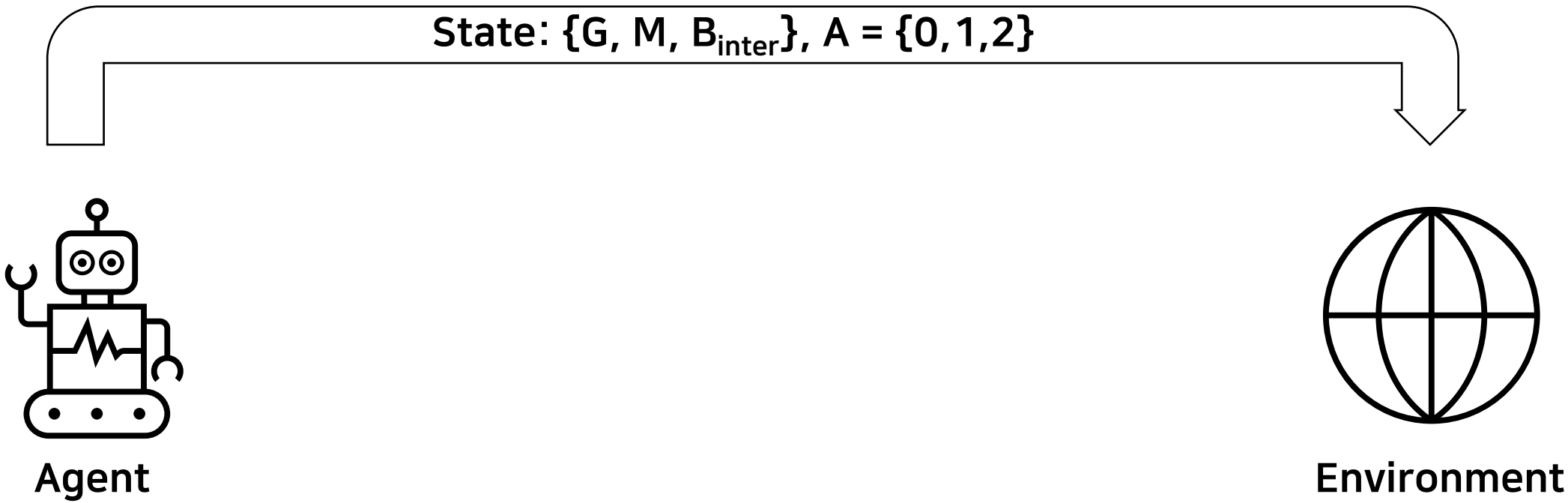
[18] Mitigating bias in face recognition using skewness-aware



Margin을 결정하는 reinforcement learning

Mei Wang and Weihong Deng. Mitigating bias in face recognition using skewness-aware reinforcement learning. In IEEE Conf. Comput. Vis. Pattern Recog., pages 9322–9331, 2020.

[18] Mitigating bias in face recognition using skewness-aware



Mei Wang and Weihong Deng. Mitigating bias in face recognition using skewness-aware reinforcement learning. In IEEE Conf. Comput. Vis. Pattern Recog., pages 9322–9331, 2020.

[18] Mitigating bias in face recognition using skewness-aware

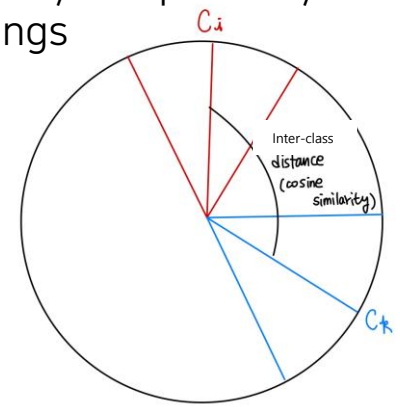
State: {G, M, B_{inter}}, A = {0,1,2}

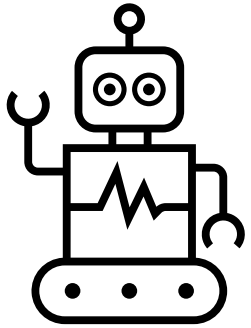
G: race
M: margin
B_{inter}: g-th group와 백인 사이 inter-class distance bias

$$B_{inter} = \left| d^{g}_{inter} - d^{Cau}_{inter} \right|$$

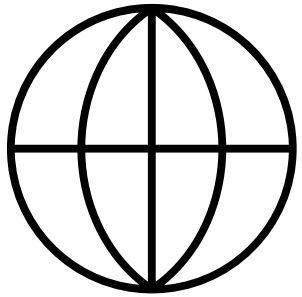
$$where,\ d^{g}_{inter} = \frac{1}{N_g} \sum_{i=1}^{N_g} \max_{k=1:N_g, k \neq i} cos(c_k, c_i)$$

c_i: feature centre of i-th identity computed by the mean vector of embeddings





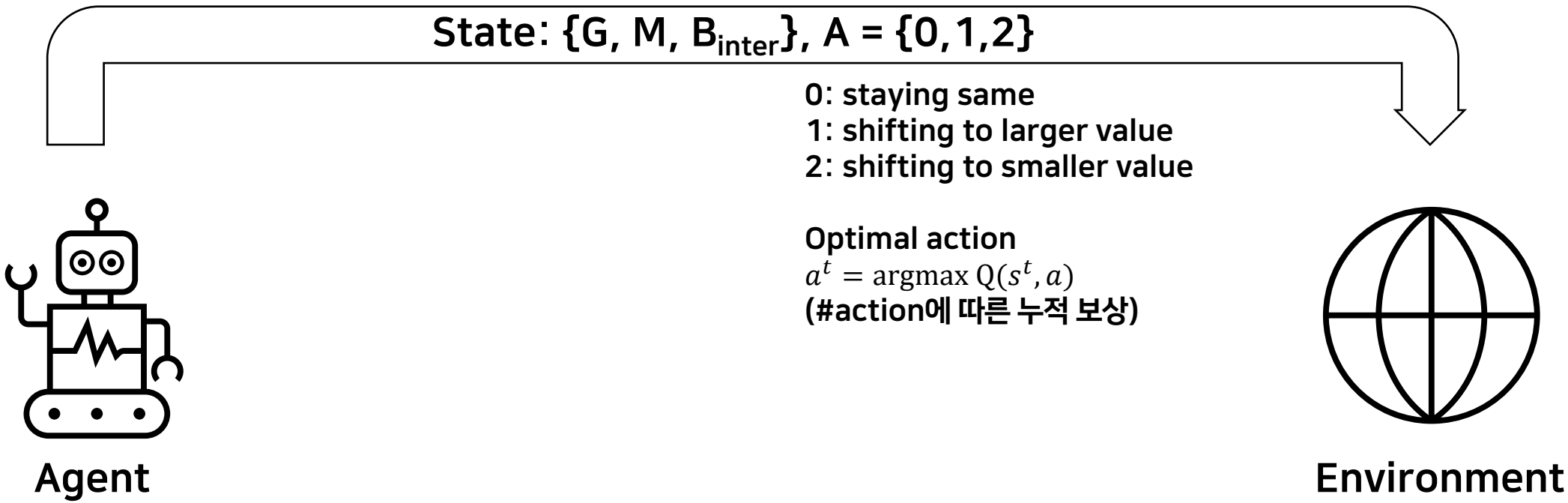
Agent



Environment

Mei Wang and Weihong Deng. Mitigating bias in face recognition using skewness-aware reinforcement learning. In IEEE Conf. Comput. Vis. Pattern Recog., pages 9322–9331, 2020.

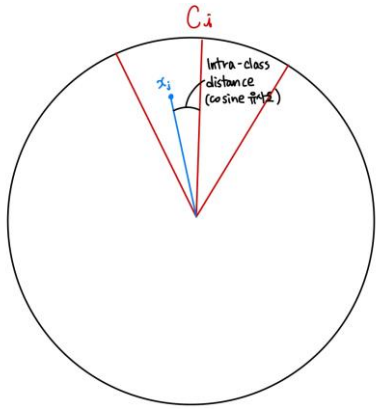
[18] Mitigating bias in face recognition using skewness-aware

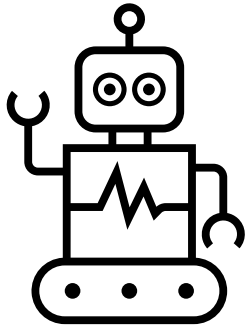


Mei Wang and Weihong Deng. Mitigating bias in face recognition using skewness-aware reinforcement learning. In IEEE Conf. Comput. Vis. Pattern Recog., pages 9322–9331, 2020.

[18] Mitigating bias in face recognition using skewness-aware

State: {G, M, B_{inter}}, A = {0,1,2}





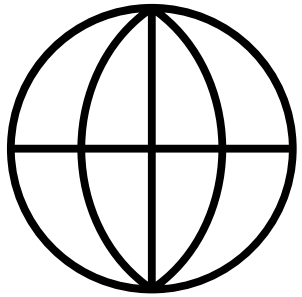
Agent

$$B_{intra} = \left| d^{g}_{intra} - d^{Cau}_{intra} \right|$$

$$where,\ d^{g}_{intra} = \frac{1}{N_g} \sum_{i=1}^{N_g} \frac{1}{|\mathcal{I}_i|} \sum_{x_j \in \mathcal{I}_i} cos(x_j, c_i)$$

$$r(s^t, a^t) = R^{t+1} - R^t$$

$$where,\ R = -\left(B_{inter} + B_{intra}\right)$$



Environment

B_{inter}, B_{intra} 값이 감소할수록 보상 증가

Reward: $r(s^t, a^t) = R^{t+1} - R^t$, Next state: s^{t+1}

Mei Wang and Weihong Deng. Mitigating bias in face recognition using skewness-aware reinforcement learning. In IEEE Conf. Comput. Vis. Pattern Recog., pages 9322–9331, 2020.

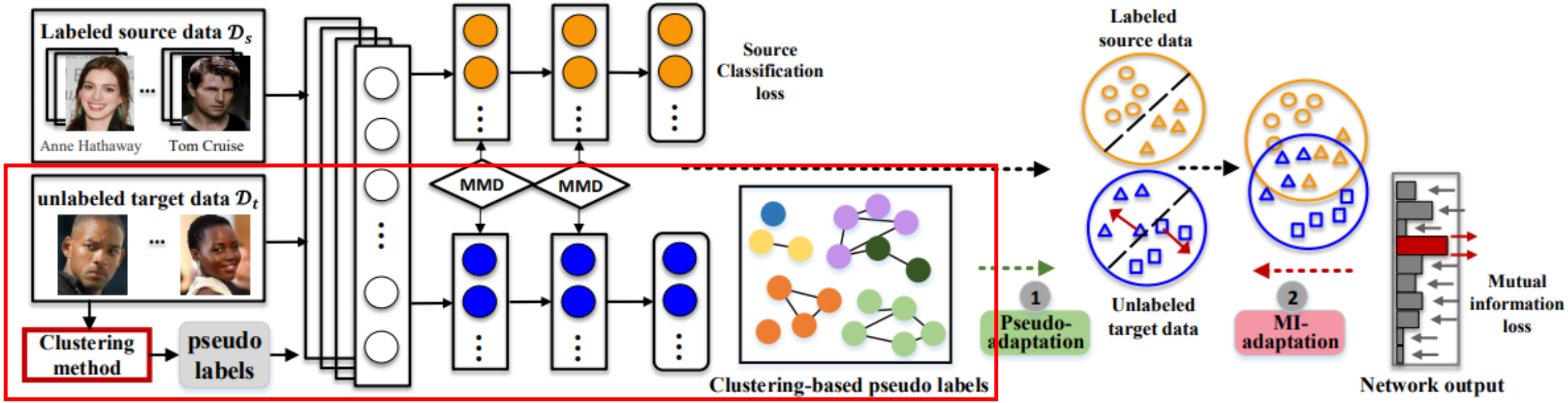
[18] Mitigating bias in face recognition using skewness-aware

Test→ Train↓	Method↓	Caucasian	Indian	Asian	African	Avg	Fairness STD	Fairness SER
$4:2:2:2$	N-Softmax [46]	89.67	87.97	84.68	84.17	86.62	**2.64**	**1.53**
	RL-RBN(soft)	91.35	90.77	89.87	90.13	90.53	**0.66**	**1.17**
$5:\frac{5}{3}:\frac{5}{3}:\frac{5}{3}$	N-Softmax [46]	89.88	88.52	85.13	83.42	86.74	**2.98**	**1.64**
	RL-RBN(soft)	90.33	90.23	88.97	89.37	89.73	**0.67**	**1.22**
$6:\frac{4}{3}:\frac{4}{3}:\frac{4}{3}$	N-Softmax [46]	90.43	88.32	84.75	83.32	86.70	**3.26**	**1.74**
	RL-RBN(soft)	90.17	90.02	87.67	88.27	89.03	**1.25**	**1.25**
$7:1:1:1$	N-Softmax [46]	90.67	87.77	84.37	82.97	86.44	**3.46**	**1.83**
	RL-RBN(soft)	90.63	90.73	87.72	87.53	89.15	**1.77**	**1.35**

Table 2. Verification accuracy (%) on RFW [49] trained with varying racial distribution. We boldface STD (lower is better) and skewed error ratio (SER) (1 is the best) since this is the important fairness criterion.

Mei Wang and Weihong Deng. Mitigating bias in face recognition using skewness-aware reinforcement learning. In IEEE Conf. Comput. Vis. Pattern Recog., pages 9322–9331, 2020.

[19] Racial faces in the wild: Reducing racial bias by information maximization adaptation network



Label이 없는 유색인종 이미지(target data)로부터 feature map 추출 후,
이미지별 cosine 유사도를 비교하여,

$$e(n_i, n_j) = \begin{cases} 1, & if\ s(i,j) > \lambda \\ 0, & otherwise \end{cases}$$

다음 식을 바탕으로 그래프 생성 ➔ 하나의 ID 생성 및 부여

Mei Wang, Weihong Deng, Jiani Hu, Xunqiang Tao, and Yaohai Huang. Racial faces in the wild: Reducing racial bias by information maximization adaptation network. In IEEE Conf. Comput. Vis. Pattern Recog., pages 692–702, 2019.

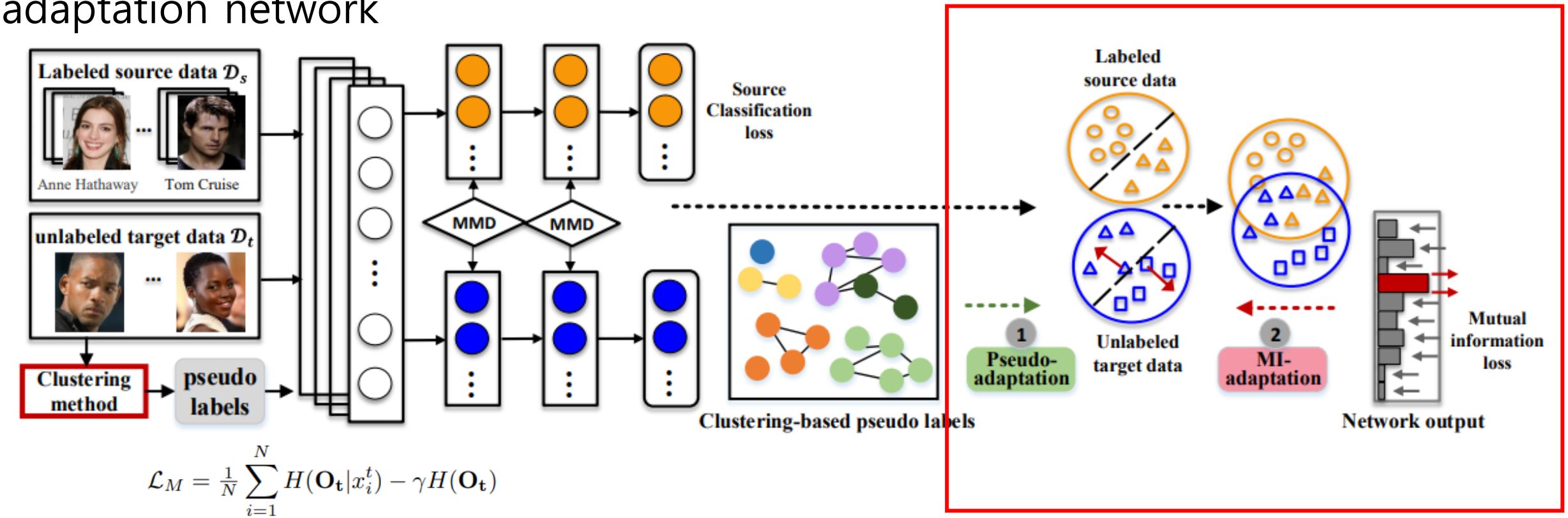
[19] Racial faces in the wild: Reducing racial bias by information maximization adaptation network

- UDA(Unsupervised Domain Adaptation)

- UDA 알고리즘 착안한 이유(추측)
  - ID가 없는 FACE도 인식할 수 있도록 하기 위해
  - 백인 얼굴 인식 시, Convolution layer마다 얼굴 인식에서 불변하는 특징을 잡아내는 것이 핵심
    → Source domain, target domain 가중치를 공유함으로써 백인 얼굴 인식에서 불변하는 특징을 잘 잡아내는 가중치를 다른 인종에서도 동일하게 사용하여 인종 편향 감소

Mei Wang, Weihong Deng, Jiani Hu, Xunqiang Tao, and Yaohai Huang. Racial faces in the wild: Reducing racial bias by information maximization adaptation network. In IEEE Conf. Comput. Vis. Pattern Recog., pages 692–702, 2019.

[19] Racial faces in the wild: Reducing racial bias by information maximization adaptation network



$$\mathcal{L}_M = \frac{1}{N}\sum_{i=1}^{N} H(\mathbf{O_t}|x_i^t) - \gamma H(\mathbf{O_t})$$

$$= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N_C} p(o_j^t|x_i^t) log p(o_j^t|x_i^t) - \gamma \sum_{j=1}^{N_C} p(o_j^t) log p(o_j^t)$$

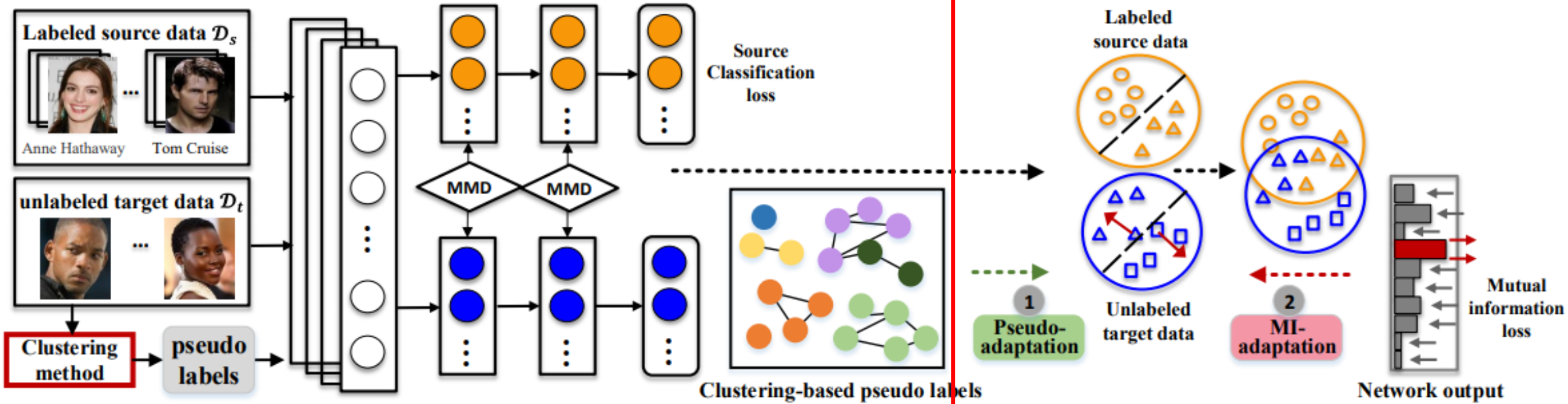
$$= \sum_{i=1}^{N}\sum_{j=1}^{N_C} p(x_i^t) p(o_j^t|x_i^t) log p(o_j^t|x_i^t) - \gamma \sum_{j=1}^{N_C} p(o_j^t) log p(o_j^t)$$

$$= H\left[\mathbf{O_t}|\mathbf{X_t}\right] - \gamma H\left[\mathbf{O_t}\right] \approx -I(\mathbf{X_t};\mathbf{O_t})$$

Target image 예측 시,
이상적인 조건부 분포: $p(O_t|x_i^t) = [0, 0, \ldots, 1, 0, 0]$와 같이 한 클래스의 출력을 확대하는 것이 목표
→ 모든 target image에 대해 비슷한 확률 분포가 나오길 바람
→ 엔트로피 관점에서 봤을 때 사건마다 확률 분포가 같을 때 가장 큰 엔트로피 출력

Mei Wang, Weihong Deng, Jiani Hu, Xunqiang Tao, and Yaohai Huang. Racial faces in the wild: Reducing racial bias by information maximization adaptation network. In IEEE Conf. Comput. Vis. Pattern Recog., pages 692–702, 2019.

[19] Racial faces in the wild: Reducing racial bias by information maximization adaptation network



$$\mathcal{L}_M = \frac{1}{N}\sum_{i=1}^{N} H(\mathbf{O_t}|x_i^t)$$ - 한 클래스 출력 확대, 나머지 클래스 출력 축소

$$= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N_C} p(o_j^t|x_i^t) log p(o_j^t|x_i^t) - \gamma \sum_{j=1}^{N_C} p(o_j^t) log p(o_j^t)$$

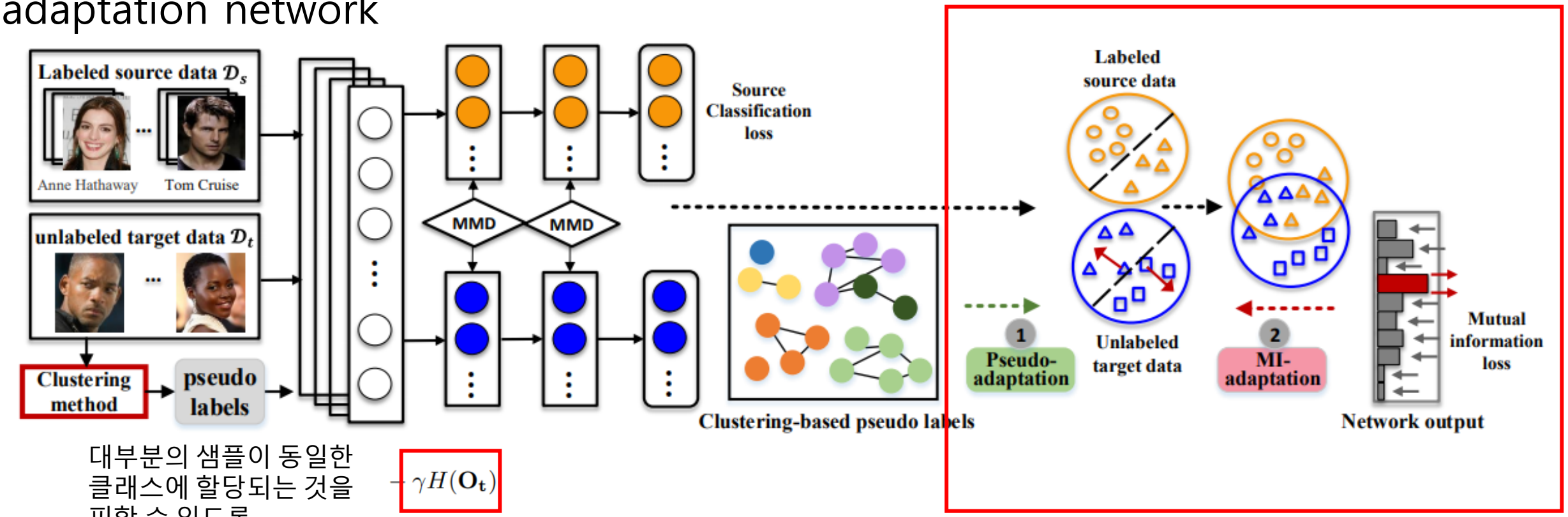
$$= \sum_{i=1}^{N}\sum_{j=1}^{N_C} p(x_i^t) p(o_j^t|x_i^t) log p(o_j^t|x_i^t) - \gamma \sum_{j=1}^{N_C} p(o_j^t) log p(o_j^t)$$

$$= H\left[\mathbf{O_t}|\mathbf{X_t}\right] - \gamma H\left[\mathbf{O_t}\right] \approx -I(\mathbf{X_t};\mathbf{O_t})$$

Target image 예측 시,
이상적인 조건부 분포: $p(O_t|x_i^t) = [0, 0, \ldots, 1, 0, 0]$와 같이 한 클래스의 출력을 확대하는 것이 목표
→ 모든 target image에 대해 비슷한 확률 분포가 나오길 바람
→ 엔트로피 관점에서 봤을 때 사건마다 확률 분포가 같을 때 가장 큰 엔트로피 출력

Mei Wang, Weihong Deng, Jiani Hu, Xunqiang Tao, and Yaohai Huang. Racial faces in the wild: Reducing racial bias by information maximization adaptation network. In IEEE Conf. Comput. Vis. Pattern Recog., pages 692–702, 2019.

[19] Racial faces in the wild: Reducing racial bias by information maximization adaptation network



대부분의 샘플이 동일한 클래스에 할당되는 것을 피할 수 있도록 $-\gamma H(\mathbf{O_t})$

$$= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N_C} p(o_j^t|x_i^t) log p(o_j^t|x_i^t) - \gamma \sum_{j=1}^{N_C} p(o_j^t) log p(o_j^t)$$

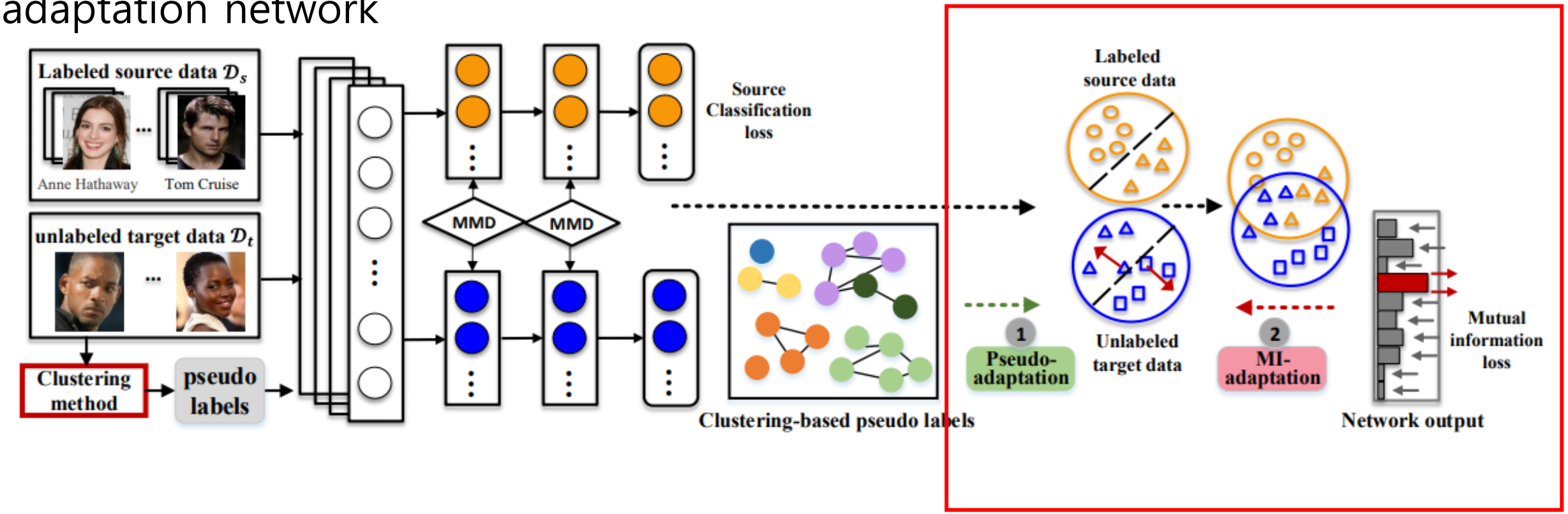
$$= \sum_{i=1}^{N}\sum_{j=1}^{N_C} p(x_i^t) p(o_j^t|x_i^t) log p(o_j^t|x_i^t) - \gamma \sum_{j=1}^{N_C} p(o_j^t) log p(o_j^t)$$

$$= H\left[\mathbf{O_t}|\mathbf{X_t}\right] - \gamma H\left[\mathbf{O_t}\right] \approx -I(\mathbf{X_t};\mathbf{O_t})$$

Target image 예측 시,
이상적인 조건부 분포: $p(O_t|x_i^t) = [0, 0, \ldots, 1, 0, 0]$와 같이 한 클래스의 출력을 확대하는 것이 목표
→ 모든 target image에 대해 비슷한 확률 분포가 나오길 바람
→ 엔트로피 관점에서 봤을 때 사건마다 확률 분포가 같을 때 가장 큰 엔트로피 출력

Mei Wang, Weihong Deng, Jiani Hu, Xunqiang Tao, and Yaohai Huang. Racial faces in the wild: Reducing racial bias by information maximization adaptation network. In IEEE Conf. Comput. Vis. Pattern Recog., pages 692–702, 2019.

[19] Racial faces in the wild: Reducing racial bias by information maximization adaptation network

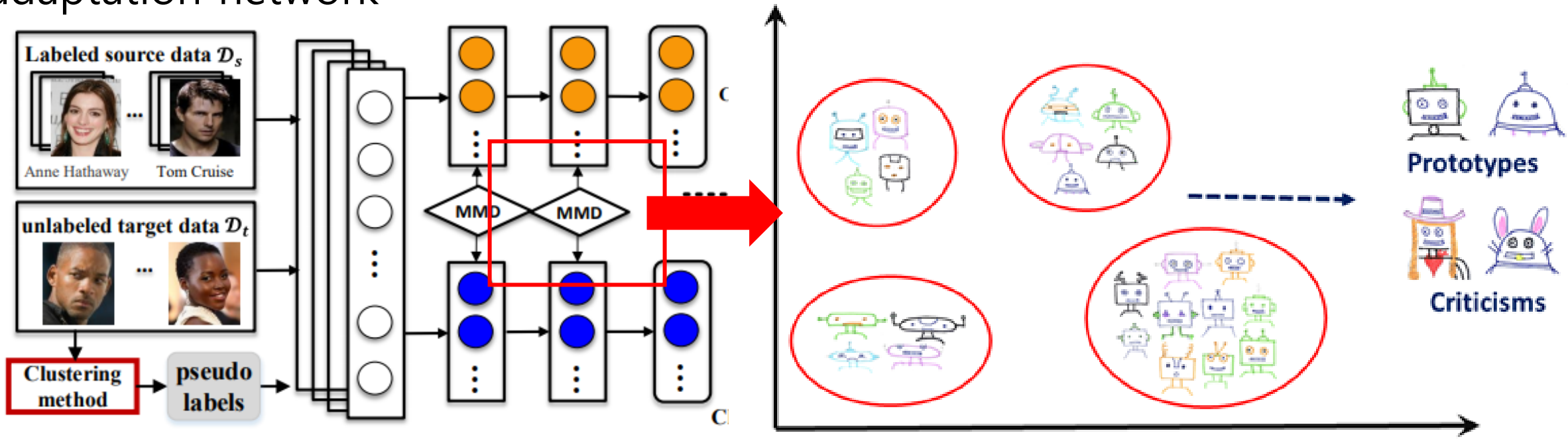


Target domain label이 존재하지 않기 때문에
이 단계에서 pseudo-label을 사용하여 예측 값을 비교!

Target image 예측 시,
이상적인 조건부 분포: $p(O_t|x_i^t) = [0, 0, \dots, 1, 0, 0]$와 같이 한 클래스의 출력을 확대하는 것이 목표
→ 모든 target image에 대해 비슷한 확률 분포가 나오길 바람
→ 엔트로피 관점에서 봤을 때 사건마다 확률 분포가 같을 때 가장 큰 엔트로피 출력

Mei Wang, Weihong Deng, Jiani Hu, Xunqiang Tao, and Yaohai Huang. Racial faces in the wild: Reducing racial bias by information maximization adaptation network. In IEEE Conf. Comput. Vis. Pattern Recog., pages 692–702, 2019.

[19] Racial faces in the wild: Reducing racial bias by information maximization adaptation network

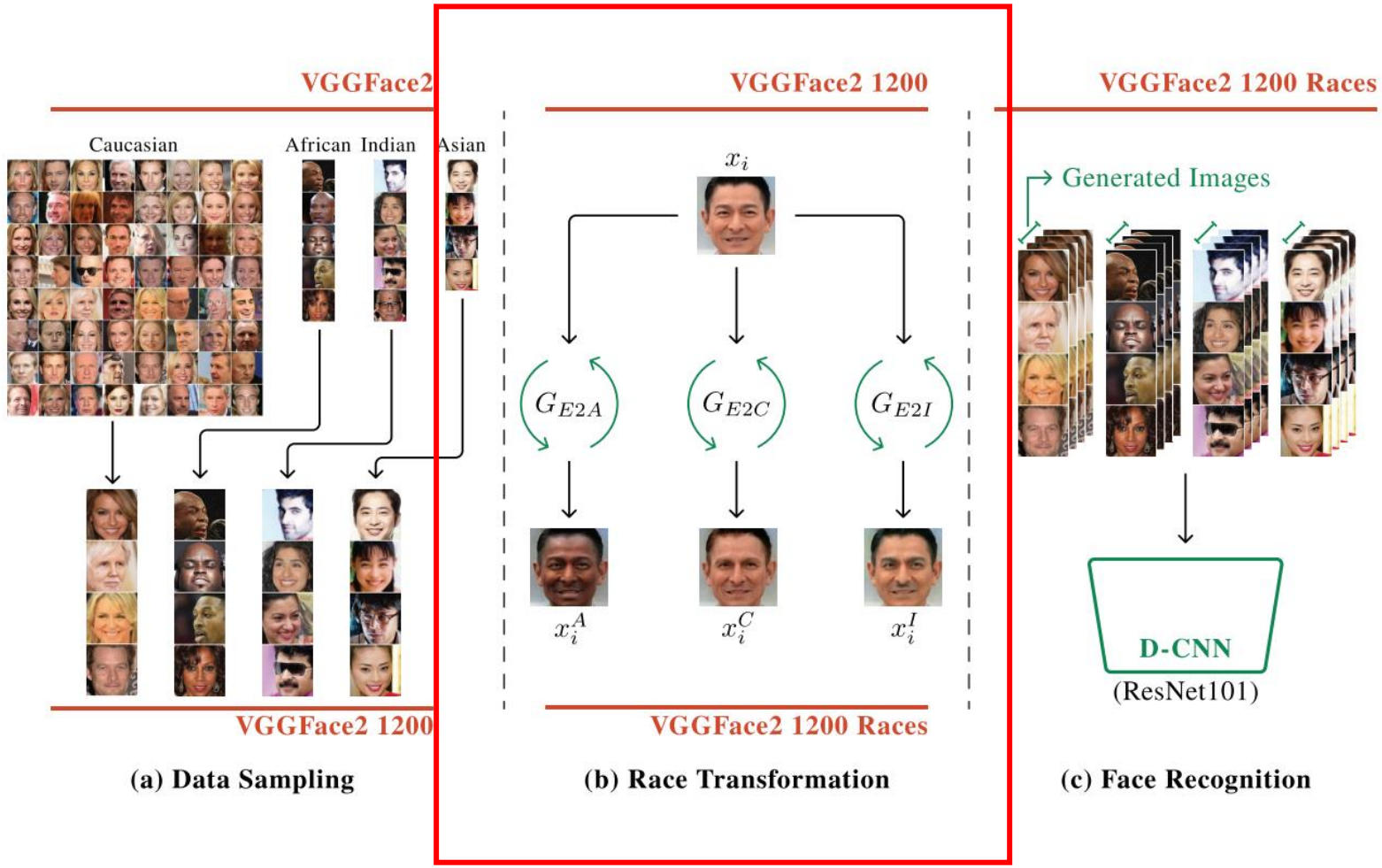


MMD(Maximum Mean Discrepancy)

- Class를 대표하는 prototypes과 class에서 예외적인 criticisms를 찾아 데이터를 직관적으로 확인 가능
- Source domain과 target domain 간의 MMD를 통해 두 분포 간의 불일치를 측정하는 도구(두 분포 간의 차이를 측정)

Mei Wang, Weihong Deng, Jiani Hu, Xunqiang Tao, and Yaohai Huang. Racial faces in the wild: Reducing racial bias by information maximization adaptation network. In IEEE Conf. Comput. Vis. Pattern Recog., pages 692–702, 2019.

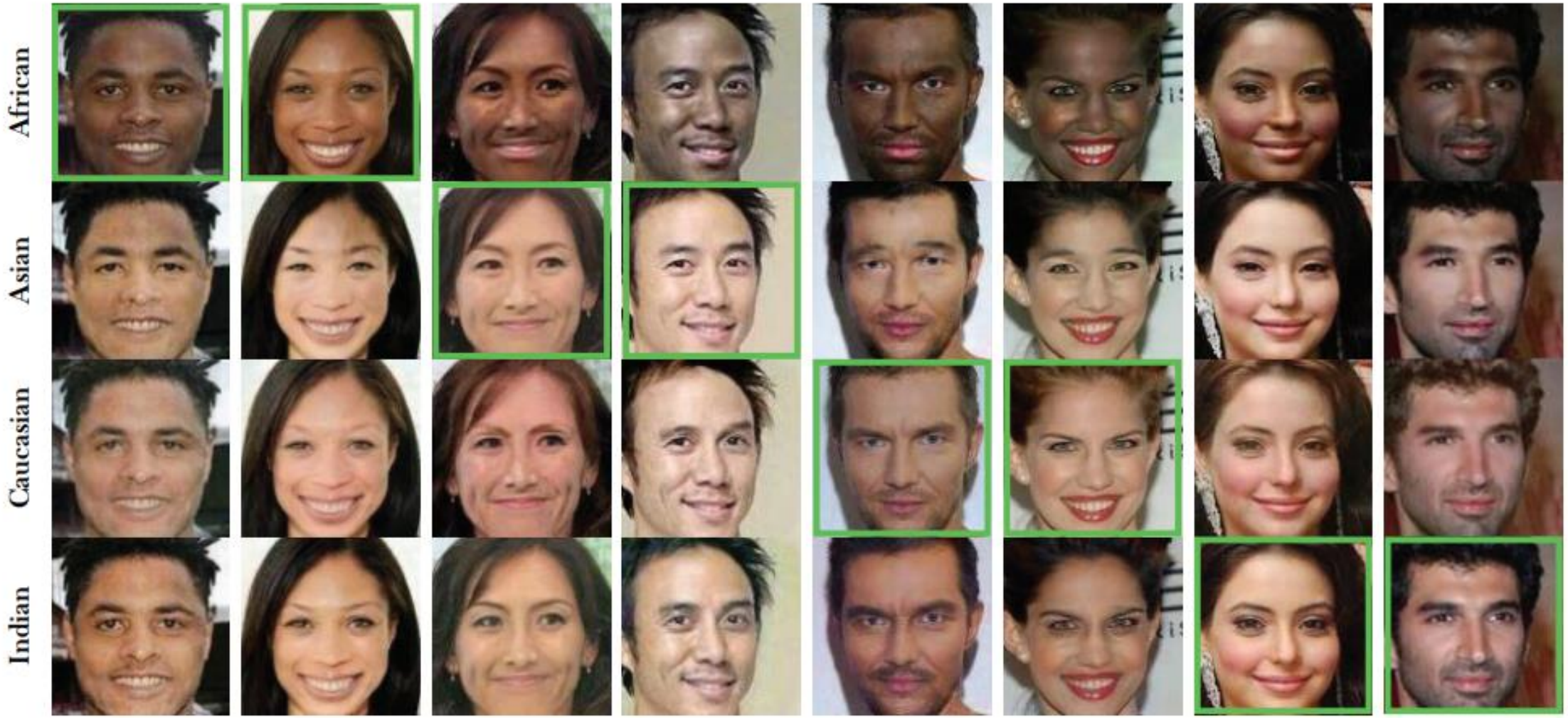
[22] Exploring racial bias within face recognition via per-subject adversarially-enabled data augmentation



Data augmentation 을 통한 dataset 생성 하는 것이 주목적

Seyma Yucer, Samet Akc¸ay, Noura Al-Moubayed, and Toby P Breckon. Exploring racial bias within face recognition via per-subject adversarially-enabled data augmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pages 18–19, 2020.

[22] Exploring racial bias within face recognition via per-subject adversarially-enabled data augmentation



Seyma Yucer, Samet Akc¸ay, Noura Al-Moubayed, and Toby P Breckon. Exploring racial bias within face recognition via per-subject adversarially-enabled data augmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pages 18–19, 2020.

[22] Exploring racial bias within face recognition via per-subject adversarially-enabled data augmentation

Seyma Yucer, Samet Akc¸ay, Noura Al-Moubayed, and Toby P Breckon. Exploring racial bias within face recognition via per-subject adversarially-enabled data augmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pages 18–19, 2020.