Latent to Latent:

A Learned Mapper for Identity Preserving Editing of Multiple Face Attributes in StyleGAN-generated Images

WACV 2022

Siavash Khodadadeh, Shabnam Ghadar, Saeid Motiian, Wei-An Lin,
Ladislau Bölöni, Ratheesh Kalarot

발표자: 김상우

Face Editing?

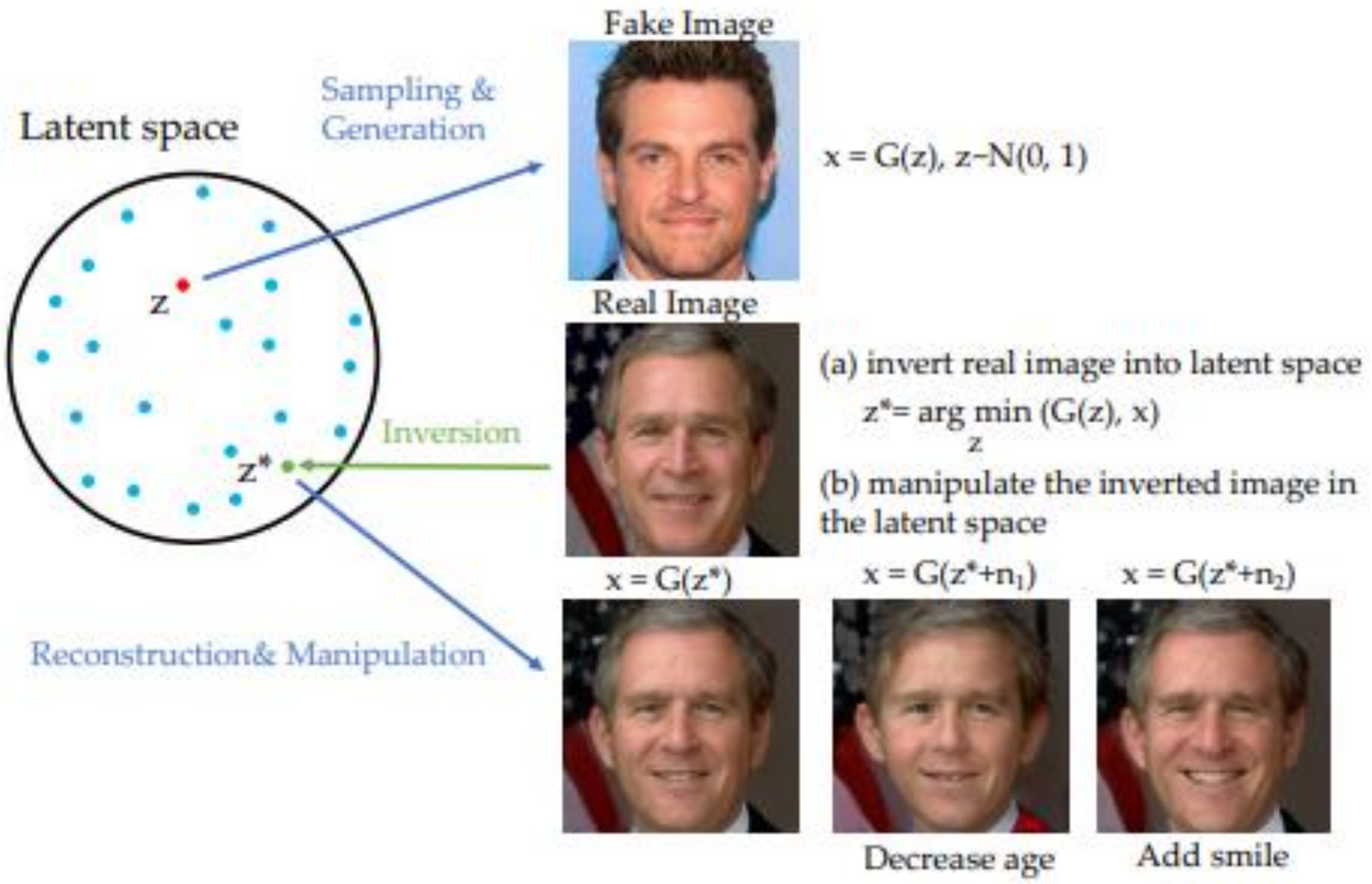
Figure 1: Examples of attribute edits with the proposed method. Notice that the background remains usually the same, and the edits are restricted to specified attribute.

1. Related Work – Face Editing

Image → Image′

GAN inversion:	$\mathbf{Image} \rightarrow \mathbf{w}$
Latent space manipulation:	$\mathbf{w} \rightarrow \mathbf{w}'$
StyleGAN:	$\mathbf{w}' \rightarrow \mathbf{Image}'$

1. Related Work – Gan Inversion



1. Related Work – Latent Space Manipulation

- Latent space manipulation?
  특정 attribute 변화와 연관된 latent vector의 변화를 찾는 것

Ex. InterFaceGAN: 특정 attribute 증가/감소 방향을 latent space에서 찾음

Attribute prediction model → 이미지 분석

Linear SVM classifier → decision boundary (hyperplane)

1. Related Work – Latent Space Manipulation

Ex. InterFaceGAN

- Problem: entanglement

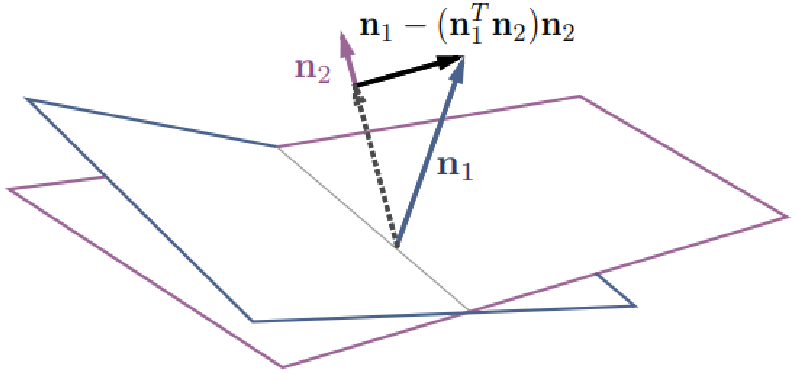
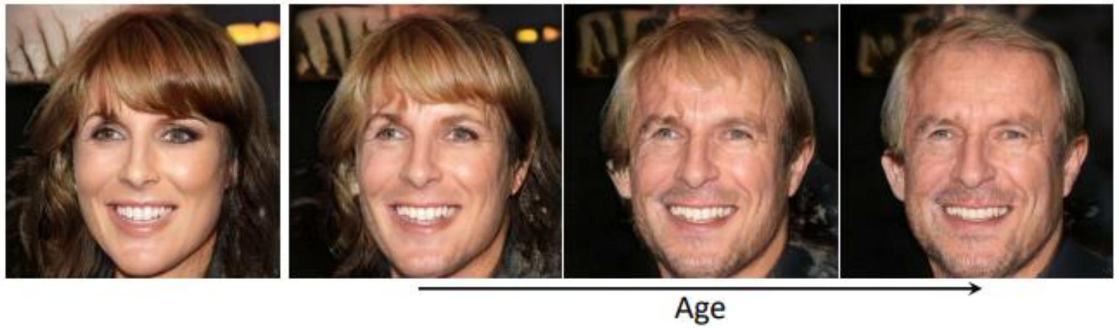


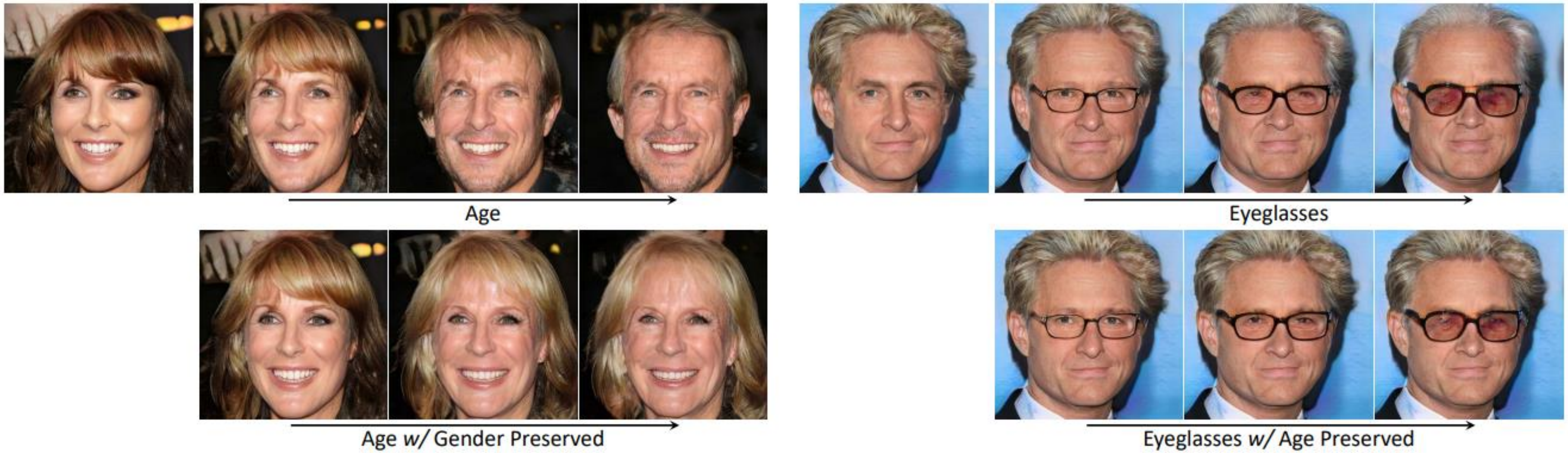
Fig. 2. Illustration of the **conditional manipulation via subspace projection**. The projection of $\mathbf{n}_1$ onto $\mathbf{n}_2$ is subtracted from $\mathbf{n}_1$, resulting in a new direction $\mathbf{n}_1 - (\mathbf{n}_1^T \mathbf{n}_2)\mathbf{n}_2$.



1. Related Work – Latent Space Manipulation

Ex. InterFaceGAN

- Solution: subspace projection을 통한 conditional manipulation



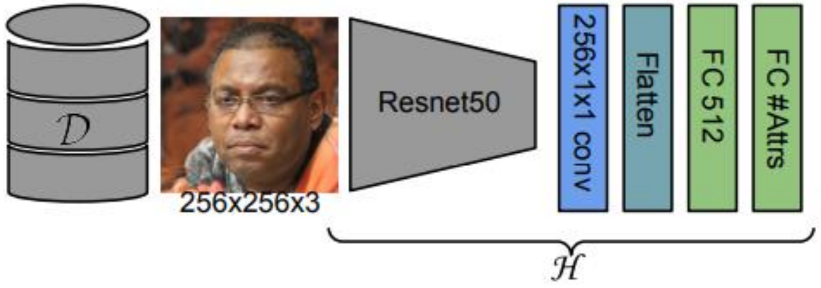
- But, entanglement is still a problem

2. Introduction

- 기존 방식: Latent space에서 specific attribute을 변화시키는 direction 탐색
- 문제점: Attributes are entangled and nonlinear in the latent space
  (ex. age 변화 → linear trajectory 위를 움직일수록 다른 attribute 생성)
- 최근 동향: 1. Non-linear mappings 찾으려고 함
  2. Changed latent vector을 generator의 특정 style layer에만 적용
- 본 논문: latent-to-latent transformation을 시행하는 NN 학습
  - 바뀐 attribute를 가진 image에 해당하는 latent encoding을 한 번에 찾음
  - Linear/nonlinear trajectory를 따르지 않고 one-shot

3. Latent-to-latent

- Image I has a collection of attributes $a = \{a_1, \dots, a_N\}\ \forall i\ a_i \in [0, 1]$
- Attribute regressor : $\mathcal{H}(I) \to \hat{\boldsymbol{a}}$



  - Network는 supervised multi-class training으로 학습, training dataset $D = \{(I_i, a_i)\}$
    - Datasets: CelebAMask-HQ [14], FFHQ [12] and a locally generated dataset

1. StyleGAN-v2의 Z space에서 400K vector을 sample
2. Microsoft Face API를 이용해 attribute들을 추출

3. Latent-to-latent

- Consider image I with attribute vector a
- Objective: User specifies $a' = a + \Delta a \rightarrow$ generate new image I'

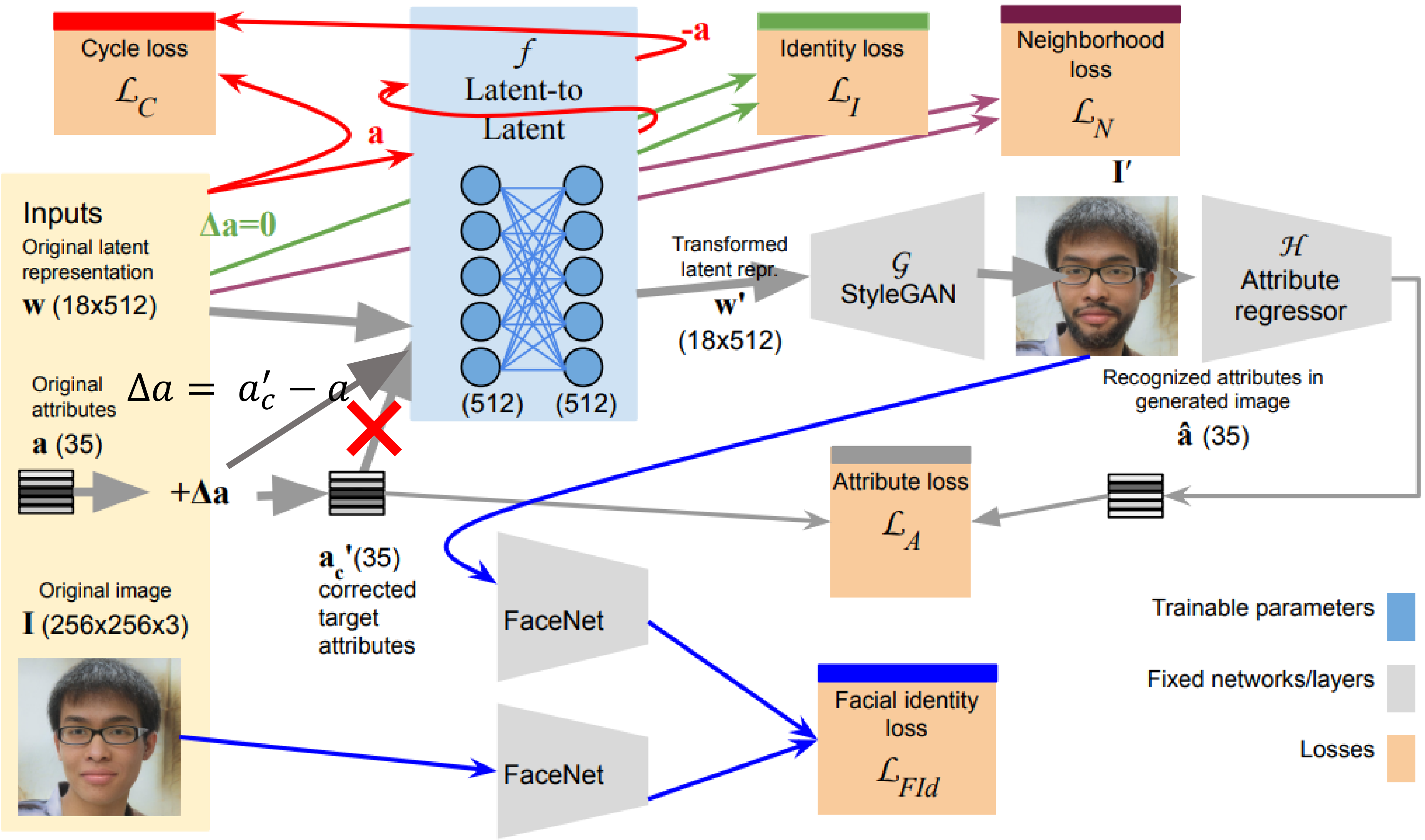
- Attribute edit performed by a latent-to-latent transformation using a network representing a parameterized function

$$f(\boldsymbol{w}, \Delta \boldsymbol{a}, \theta) \rightarrow \boldsymbol{w}', \text{ such that } \mathcal{H}(\mathcal{G}(\boldsymbol{w}')) = \boldsymbol{a}'$$

- Transformation f : fully connected, multilayer neural network
  - Hidden layers는 input, output과 같은 size (input image의 detail들을 모두 보존하기 위해)
  - Tanh nonlinearity 이용 (ReLU를 이용하면 정보의 일부가 suppress 됨)

- Embedding framework for the training of the latent-to-latent network f

(I, w, a)
형식의 triplet

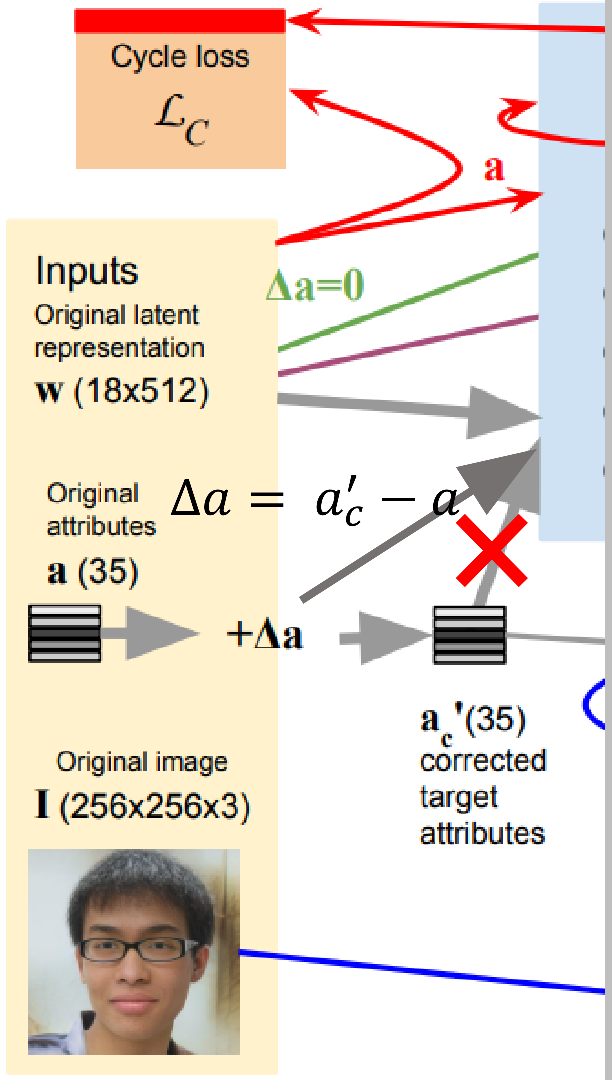


$\Delta a = a_c' - a$

- Information flows through a number of pre-trained components with frozen weights in addition to the trainable f network

- Embedding framework for the training of the latent-to-latent network f

(I, w, a)
형식의 triplet



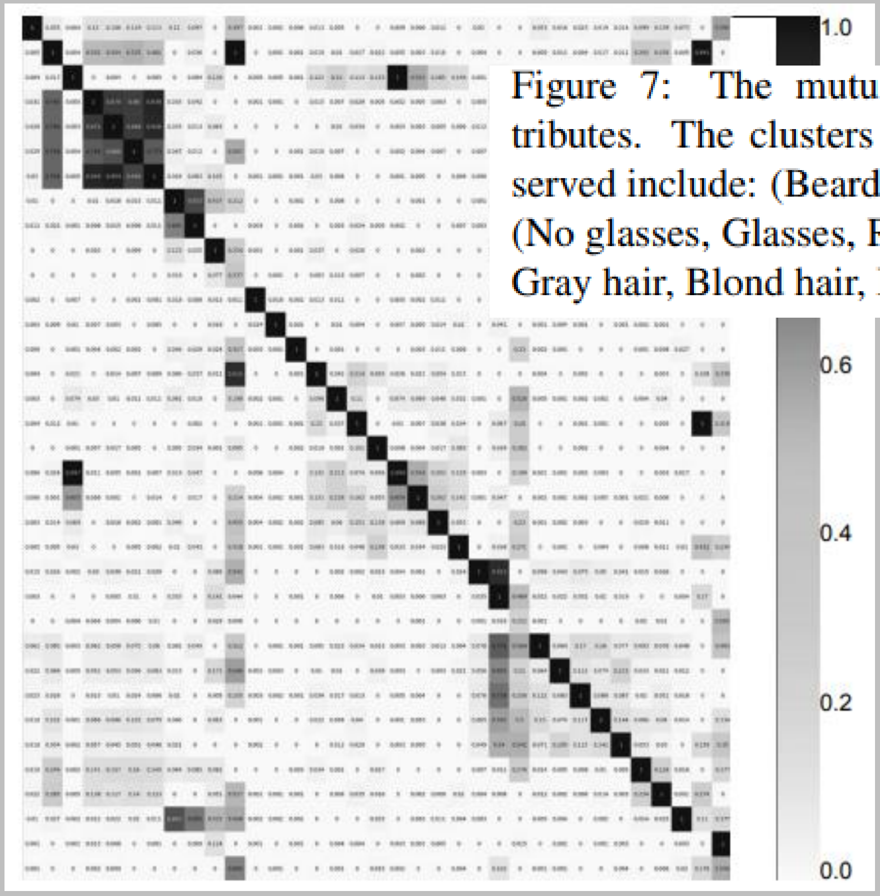
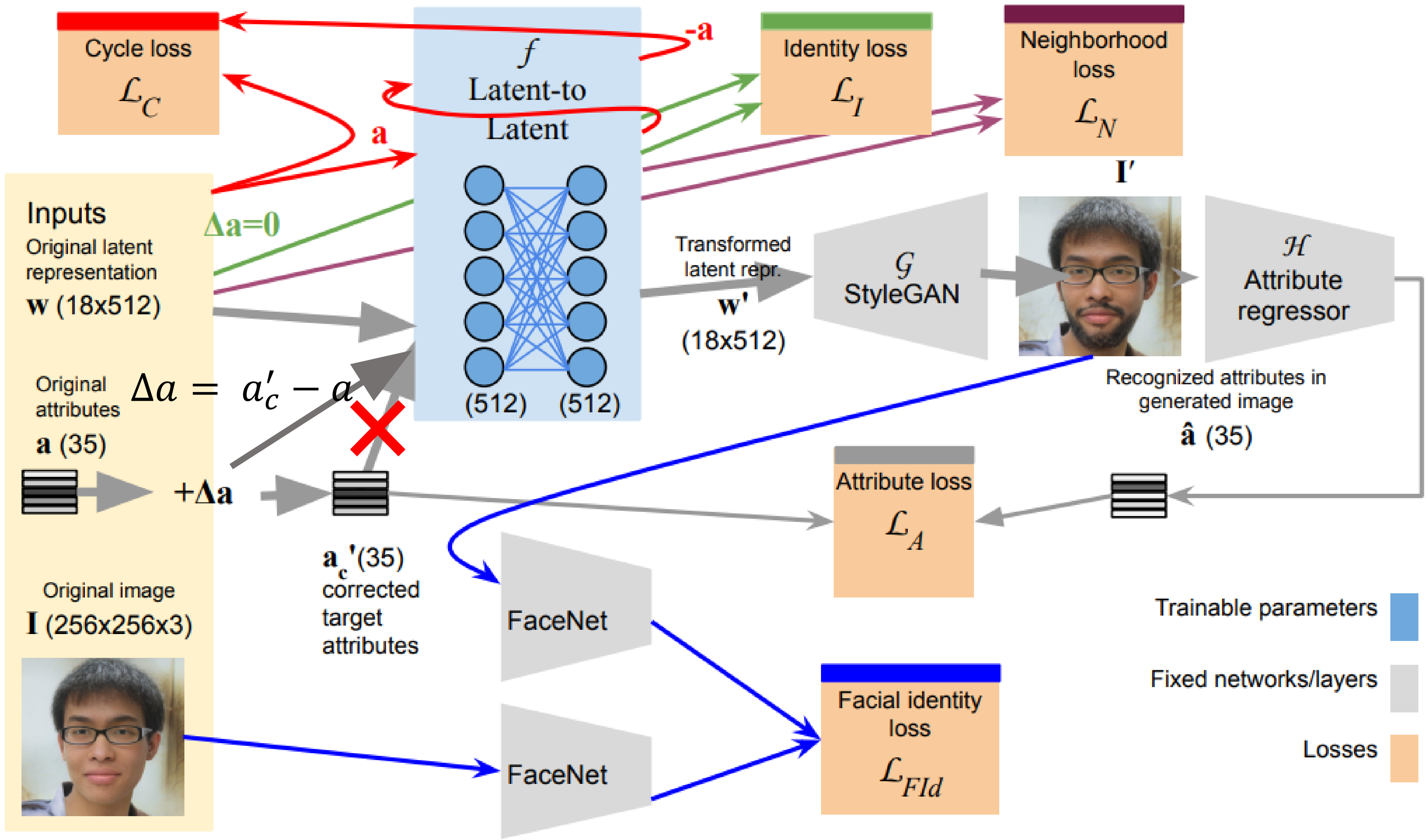


Figure 7: The mutual information matrix between attributes. The clusters of correlated attributes that we observed include: (Beard, Moustache, Facial hair, Sideburns), (No glasses, Glasses, ReadingGlasses), and (Hair invisible, Gray hair, Blond hair, Brown hair, Red hair, Black hair).

- Randomly changing a single attribute
  → might correspond to an unfeasible combination of attributes (ex. bearded man with no facial hair)
- Create corrected vector by multiplying corresponding row elements if above threshold

- Embedding framework for the training of the latent-to-latent network f

(I, w, a)
형식의 triplet



- Task?

3. Latent-to-latent

- Target attribute loss: 생성된 이미지의 $\hat{a}$가 a_c'와 최대한 비슷해야 함

$$\mathcal{L}_A = ||\hat{\boldsymbol{a}} - \boldsymbol{a}_c'||^2$$

  - 참고) 바뀌지 않아야 할 attribute가 바뀌는 경우도 penalize

- Cycle loss: 같은 사람에게 attribute change는 reversible해야 함
(ex. Age를 10년 증가시킨 후 10년 감소시키면 같은 image가 되어야 함)

$$\mathcal{L}_C = ||\boldsymbol{w} - f(f(\boldsymbol{w}, \Delta\boldsymbol{a}), -\Delta\boldsymbol{a})||$$

3. Latent-to-latent

- Identity loss: attribute change가 없을 경우, latent vector을 자기 자신에게 map 해야 함

$$\mathcal{L}_I = ||\boldsymbol{w} - f(\boldsymbol{w}, 0)||$$

- Neighborhood loss: attribute transformation을 최소 크기의 departure로 이루어 내도록 함
  - Latent space가 non-linear, entangled 하여 작은 attribute transformation도 latent space 상에서 먼 곳에 위치하게 될 수 있음.

$$\mathcal{L}_N = ||\boldsymbol{w} - f(\boldsymbol{w}, \Delta\boldsymbol{a})||^2$$

3. Latent-to-latent

- Face identity loss: FaceNet (facial recognition을 위해 학습된 네트워크) 을 이용한 loss

$$\mathcal{L}_{FId} = ||\mathcal{F}(\mathcal{G}(\boldsymbol{w})) - \mathcal{F}(\mathcal{G}(f(\boldsymbol{w}, \Delta\boldsymbol{a})))||^2$$

  - $F(I)$: image I로부터 FaceNet이 추출한 features

- 이 밖에도 user이 원하는 conservation properties가 나타나도록 loss를 추가할 수 있음
  - Ex) 배경을 똑같게 유지하기, 사람의 눈을 감은 채로 유지하기, 등등

3. Latent-to-latent

- Latent-to-latent를 이용한 face-editing

Figure 1: Examples of attribute edits with the proposed method. Notice that the background remains usually the same, and the edits are restricted to specified attribute.

4. Experimental validation - Qualitative

- 대체로 성공적
- Features not covered by the attributes (ex. 배경, 옷, illumination) 역시 보존

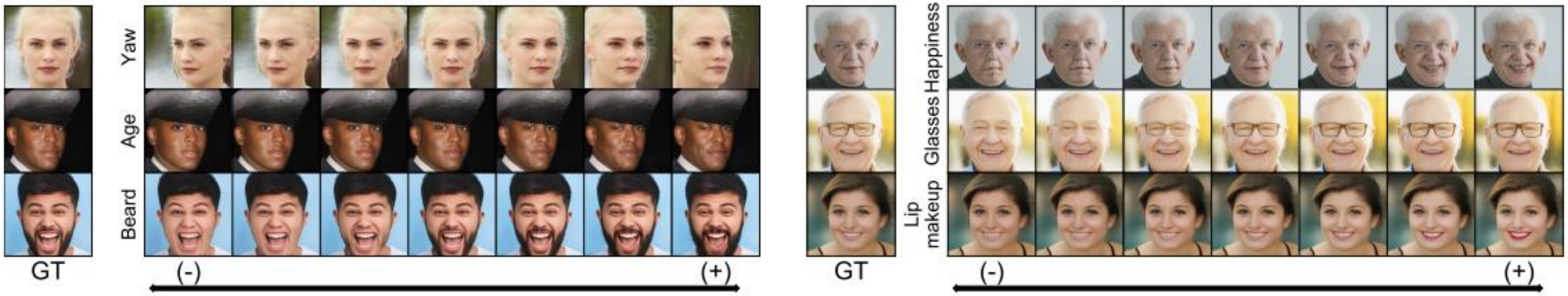


Figure 3: Changing six different attributes on six different faces.

4. Experimental validation - Qualitative

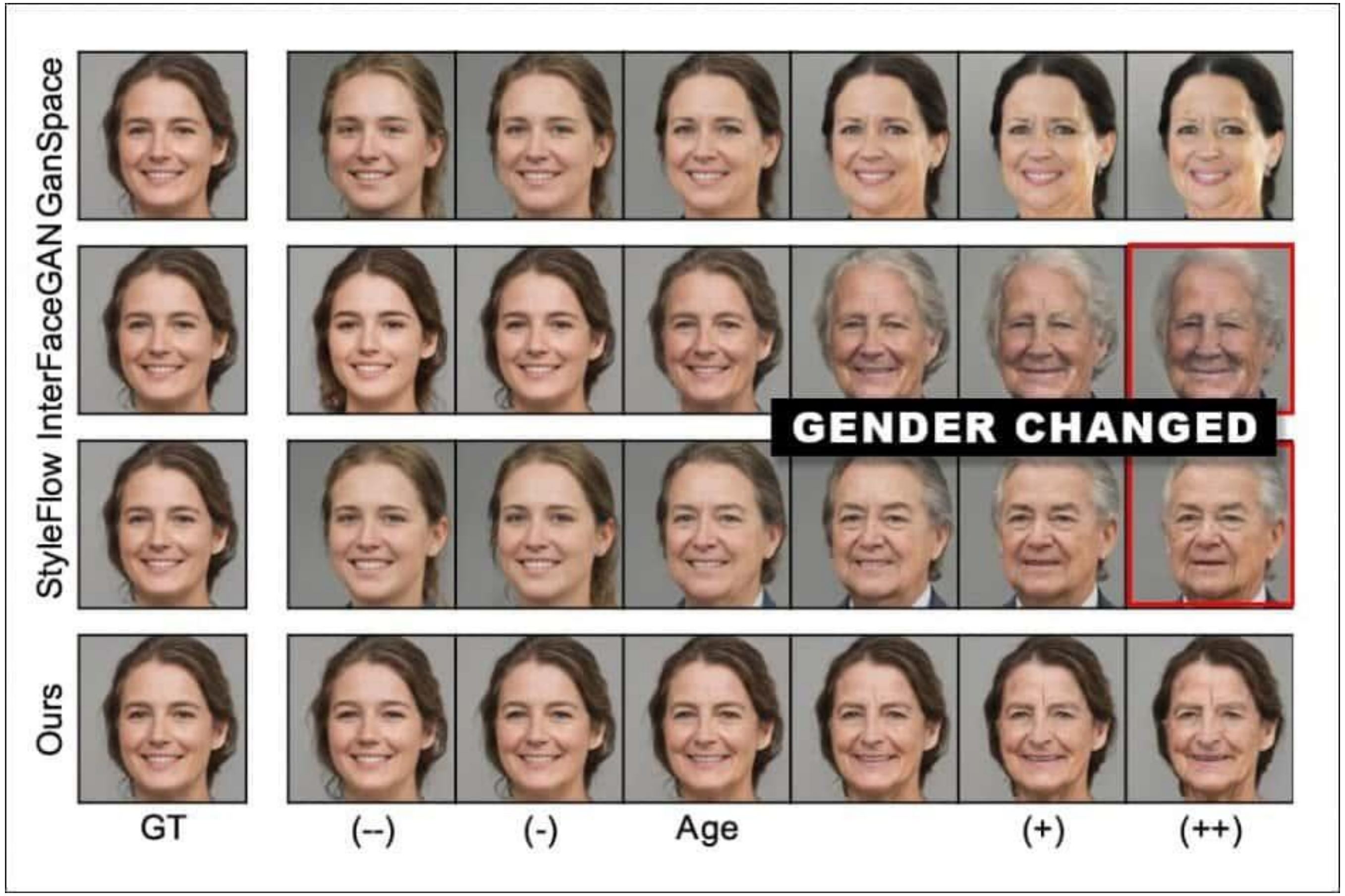
4. Experimental validation - Qualitative

- InterFaceGAN, GANSpace, StyleFlow (SOTA baselines) 와 비교
  - 8 attributes: Age, Baldness, Beard, Expression, Gender, Glasses, Pitch, and Yaw
  - Used 50 images sampled from the StyleFlow test set

- 각각의 technique에 대해 range of valid values를 calibrate
  - ex. Method에 따라 age에 0.2를 더하거나 0.3을 빼는 것이 다른 의미를 가짐
  - MTCNN face detector가 valid face를 찾을 수 있었던 attribute value의 최솟값, 최댓값 구함
  - 최솟값, 최댓값 range를 40 interval로 나누고 interval boundaries에 41개의 image 생성
  - Used 8 $\times$ 50 $\times$ 41 = 16400 generated images

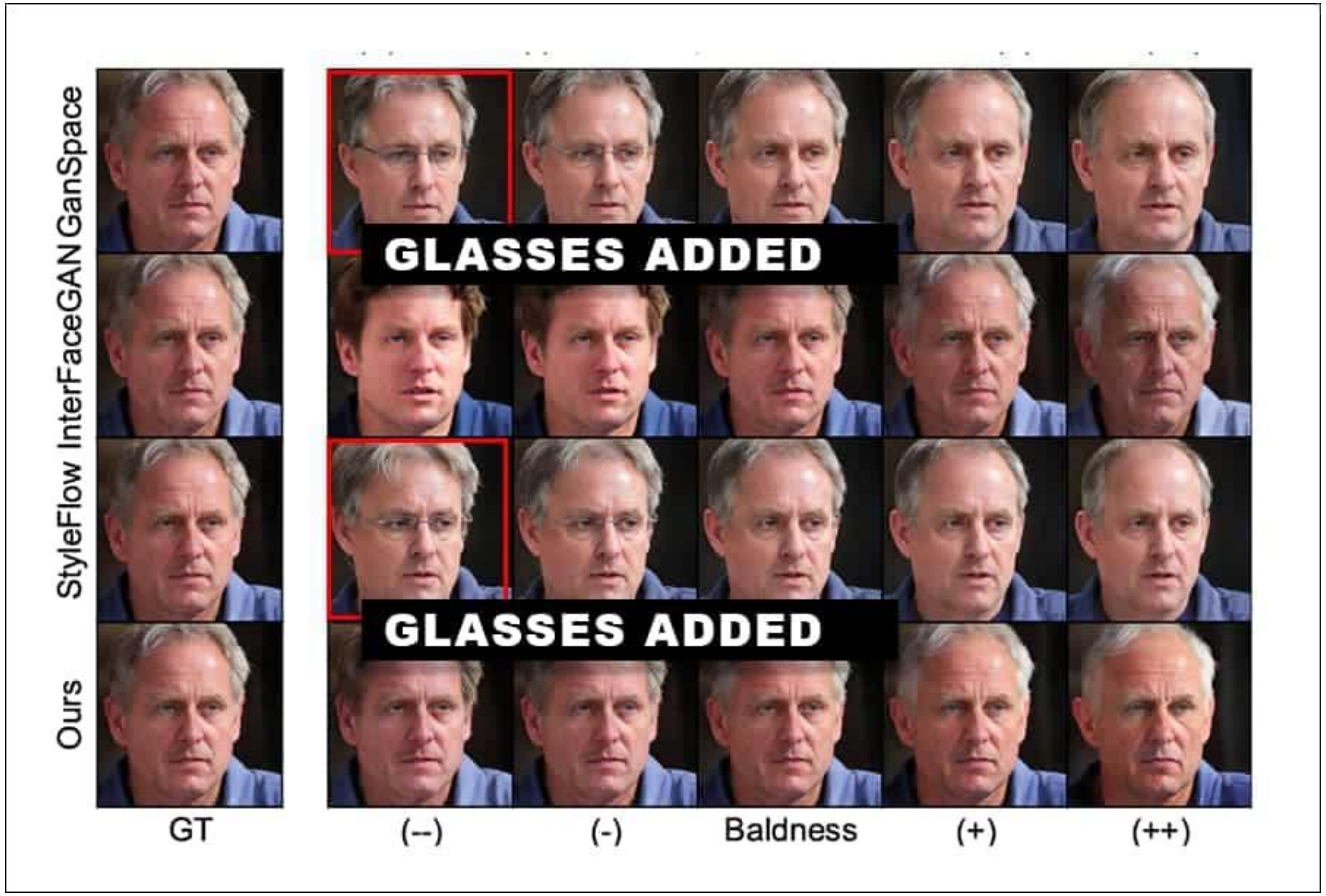
4. Experimental validation - Qualitative

- Age가 변할 때 Gender / Hair color 보존



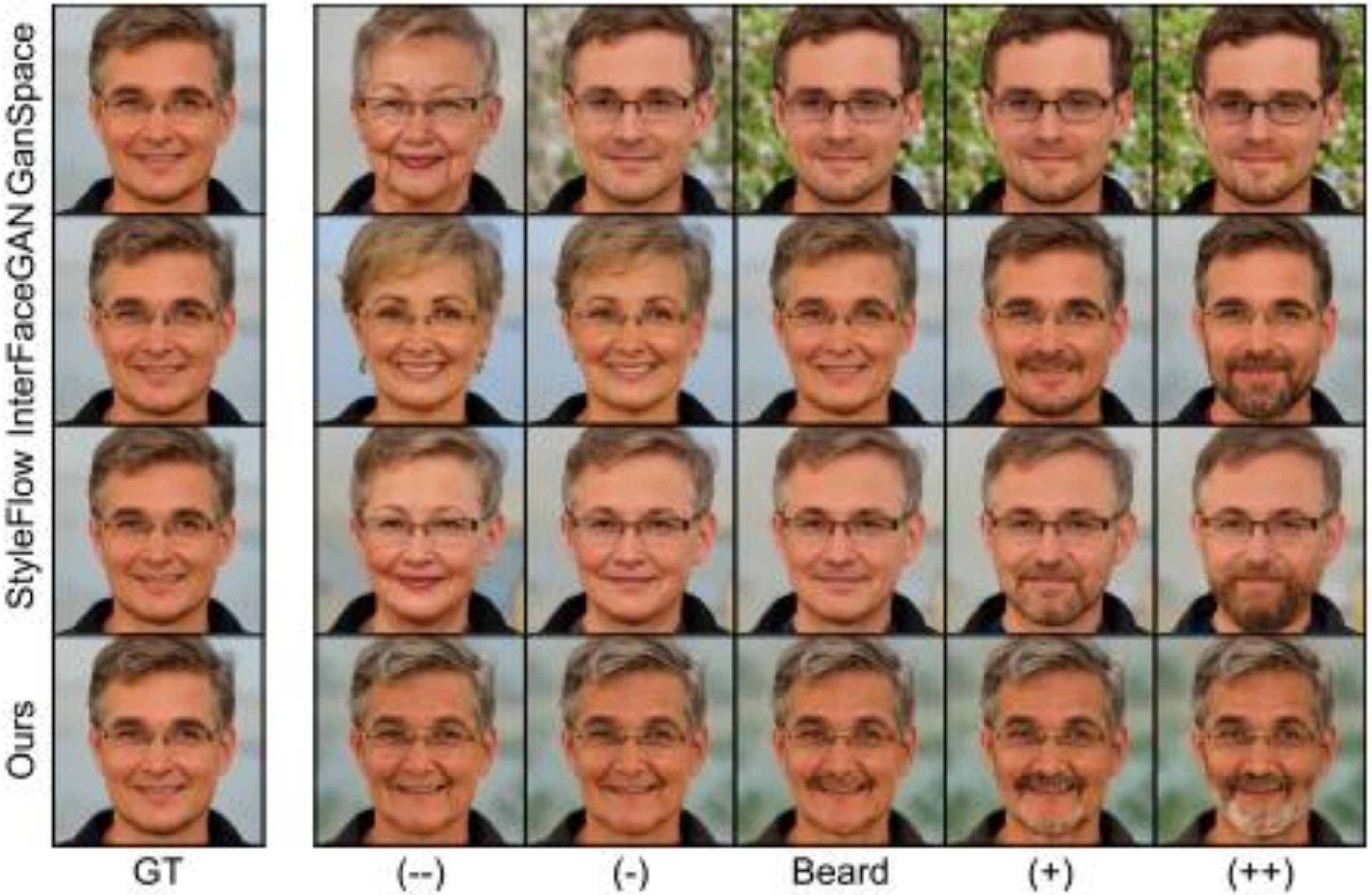
4. Experimental validation - Qualitative

- Baldness가 변할 때 Glasses / Hair color 보존



4. Experimental validation - Qualitative

- Beard가 변할 때 Gender 보존



4. Experimental validation - Quantitative

- How well is facial identity preserved?
  - External facial recognition system이 image들을 인식하는 정도를 measure
  - MTCNN이 추출한 feature 이용 (training 때 이용한 FaceNet features과 다름)
  - Original vs. edited images의 MTCNN feature vectors 간의 average cosine distance 측정 (값이 낮을수록 좋음)

ATTRIBUTE	IG	SF	GS	L2L
AGE	0.77	0.65	**0.46**	0.57
BALDNESS	0.53	0.64	0.46	**0.21**
BEARD	0.57	0.55	0.53	**0.28**
EXPRESSION	0.60	0.21	0.23	**0.14**
GENDER	0.54	0.52	0.58	**0.28**
GLASSES	0.59	0.46	**0.24**	0.25
PITCH	0.54	**0.51**	**0.51**	**0.51**
YAW	**0.39**	0.46	0.41	0.46

4. Experimental validation - Quantitative

- How good is the quality of the edited images?
  - Generated images vs. FFHQ StyleGAN-v2 images
  - Frechet Inception Distance(FID), Inception Score(ISD), Kernel Inception Distance(KID) 비교

METRIC	IG	SF	GS	L2L
FID	43.16	41.64	45.44	39.83
KID	0.0118	0.0086	0.0122	0.0062
ISD	0.0025	0.0014	0.0044	0.0085

4. Experimental validation - Quantitative

- How fast can we create the edited images?
  - Interactive applications, mass generation에 중요한 지표
  - Real-time interactive editing에 사용 가능

(a) Inference speed (seconds)

METHOD	IG	SF	GS	L2L
TIME	0.1080	0.6783	0.1095	0.1654

5. Contributions

- Latent space transformation을 통해 변형된 latent encoding을 찾는 NN 도입

- End-to-end training을 통해 latent-to-latent NN 학습
  - Manual image labeling이 불필요 (어떠한 attribute recognizer을 사용해도 가능)
    본 논문은 35개의 facial attributes를 다룰 수 있도록 NN 학습
  - 사용자가 원하는 conservation properties를 정할 수 있음
    (ex. Facial identity 보존을 위해 FaceNet을 이용한 identity loss 도입)

- 다른 SOTA 방법들에 비해 entanglement, facial identity loss가 적음

6. Online Demo

Thank You