

# [Xiao18-ECCV] ELEGANT: Exchanging Latent Encodings with GAN for Transferring Multiple Face Attributes

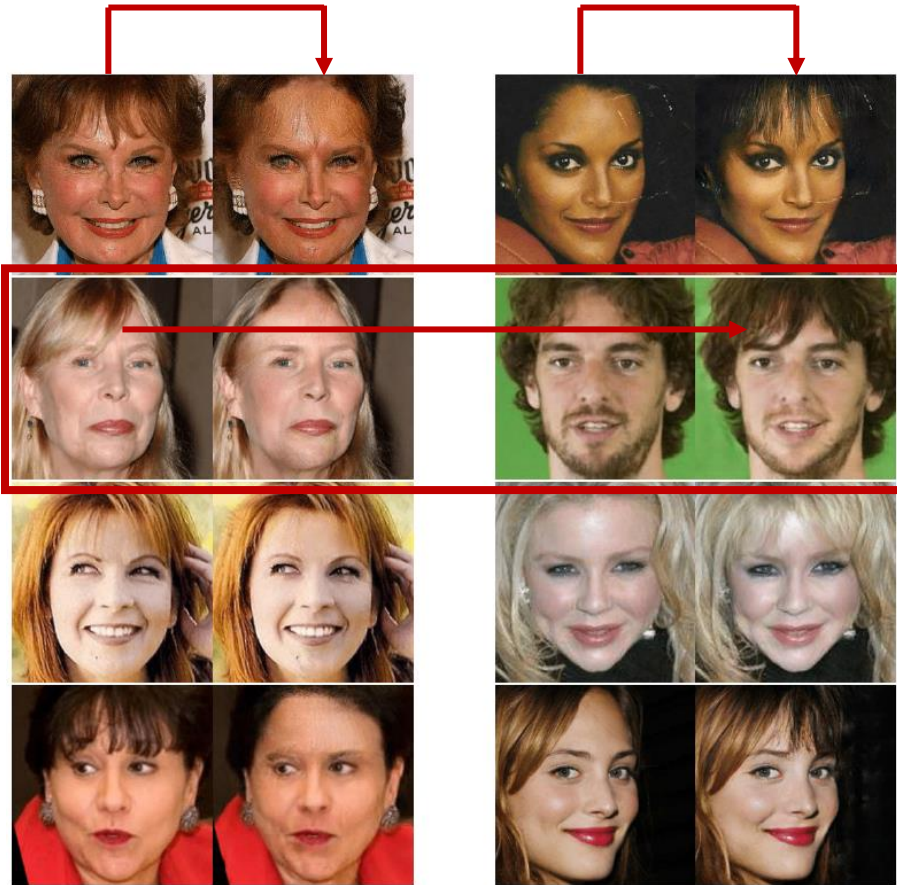
Presenter : Ji-In Kim

# Contents

- ✓ Introduction
- ✓ Purpose and Intuition
- ✓ Our Method
- ✓ Experiments
- ✓ Conclusion

# 1 Introduction

- Transferring face attributes 에서 source face image 는 targeted attribute 를 가져야 하고, person identity 는 보존되어야 한다.




(a) removing bangs

(b) adding bangs

- Person identity 는 바뀌지 않고 bangs attribute 가 조정된다.
- 각 이미지 쌍에서, 오른쪽 image 는 순전히 왼쪽 image 로부터 만들어진다. (Unsupervised learning)
- 하나의 행에 4개의 images 가 있다. 첫 번째 이미지의 bangs style 이 마지막 이미지로 전이되었다.

- Transferring face attributes 를 위해 많은 방법들이 제안되었지만, 여전히 많은 제약이 존재한다.

- Gardner et al. [3]
  - Deep Manifold Traversal 을 제안했다.
  - Maximum mean discrepancy (MMD)[6]를 사용해서 source domain 에서 target domain 까지 의 attribute vector 를 계산한다.
  - 그러나, 이 방법은 시간이 오래 걸리고 memory cost 가 비싸다.

- Upchurch et al. [24]
  - Linear Feature Space assumptions[1] 을 사용했다.
  - Example : no-bangs image B 에서 bangs image A 로 transferring 을 하자.
    - $A = f^{-1}(f(B) + v_{bangs})$
    - $f$  : Image space 로부터 feature space 까지의 mapping
    - $v_{bangs}$  : bangs images 와 no-bangs images 의 cluster centers of features 차이
  - Universal attribute vector ( $v_{bangs}$ ) 는 다양한 얼굴에 같은 스타일의 bangs 를 가지는 face images 를 생성한다. 
  - 그러나, 너무 다양한 스타일의 bangs 가 존재한다.



(a) removing bangs

(b) adding bangs

- Visual Analogy-Making [10]

- 다양성 문제를 해결하기 위해 attribute vector 를 특정하는 데에 한 쌍의 reference images 를 사용한다.
- 한 쌍의 reference images 는 동일 인물의 두 장의 사진으로 구성된다.
  - 한 장은 특정 attribute 를 가지고 있고 또다른 한 장은 attribute 를 가지고 있지 않다.
- 이 방법은 생성된 이미지들을 풍부하고 다양하게 만들 수 있지만, paired images 를 대량으로 구하는 것은 어렵다.

- 만약 face image 에서 attribute gender 를 transferring 한다고 하자.
- 그러면 동일 인물의 male 과 female images 를 함께 얻어야 하는데, 이것은 불가능하다.

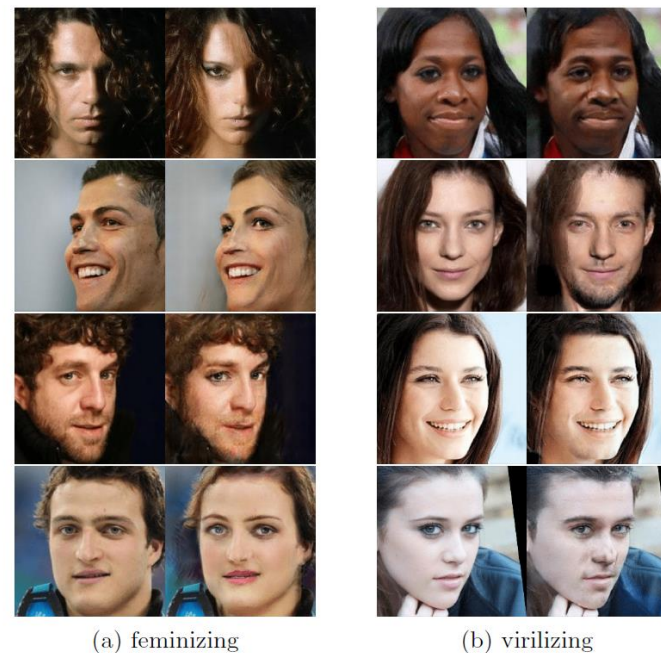


Fig. 2: Results of ELEGANT in transferring the gender attribute.



- 최근에 이런 어려움을 극복하기 위해 GANs[5] 에 기반한 많은 방법들이 제안되었다. [10, 18, 31]

- Dual learning approaches [7, 11, 21, 28, 32]

- Source image domain 과 target image domain 사이의 mapping 을 사용한다.
- Invariance of Domain Theorem 에 따르면, 두 이미지의 domains 의 고유한 차원은 같아야 한다.
- 그러나, 두 이미지 domains 의 고유한 차원은 항상 같지 않기 때문에 모순적이다.



(a) removing eyeglasses

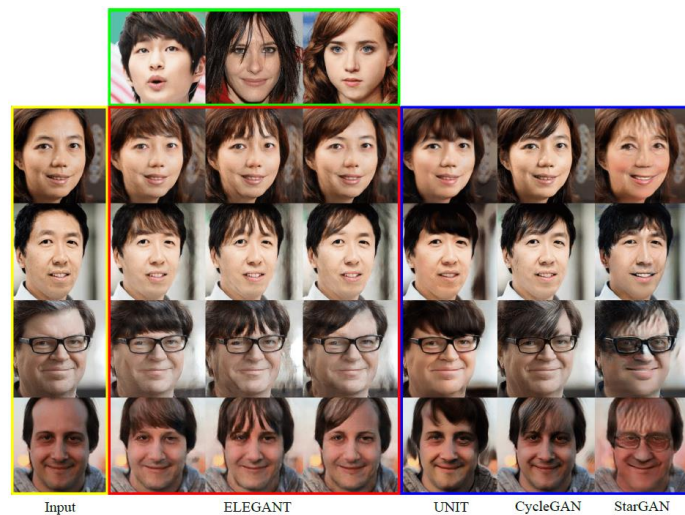


(b) adding eyeglasses

- Domain  $A$  : eyeglasses 를 착용한 face images
- Domain  $B$  : eyeglasses 를 착용하지 않은 face images
- eyeglasses 의 다양성 때문에  $A$  의 고유한 차원은  $B$  보다 크다.

- 다른 방법들[15, 23, 30]
  - GAN과 VAE 조합의 변형이다.
  - Autoencoder 구조를 사용한다.
  - 동일하지 않은 고유한 차원들의 문제를 성공적으로 우회한다.
  - 그러나, 오직 하나의 face attribute 만 조정한다는 한계를 가진다.

- Conditional image generation methods [2, 13, 18, 29]
  - 여러 개의 attributes 를 동시에 제어하기 위해 image labels 를 conditions 로 받는다.
  - 그러나, exemplars 를 사용해서 이미지를 생성할 수 없다.
  - 결과적으로 생성된 image 에서 attributes 의 스타일은 비슷하게 되어 풍부함과 다양성이 부족해진다.



(a) bangs

Face image generation by exemplars

- BicycleGAN [33]
  - 다양성을 증가시키기 위해 noise term 을 도입하지만, 특정한 attributes 를 가진 images 를 생성하지 못한다.

- TD-GAN [25] & DNA-GAN [27]
  - 장점
    - Exemplars 를 사용해 images 를 생성할 수 있다.
  - 단점
    - TD-GAN
      - 명시적인 identity 정보를 label 로 사용해서 person identity 를 유지한다. Labeled identity information 이 없는 많은 dataset 에서 이 방법을 적용할 수 없다.
    - DNA-GAN
      - 고화질 images 에서 훈련하는 것이 어렵다.

- 또한 많은 다른 방법들 [14]이 존재하지만 그것들의 결과는 시각적으로 만족스럽지 않다.
- 생성된 이미지들은 저 화질이거나 artifacts 가 많다.

## 2 Purpose and Intuition

---



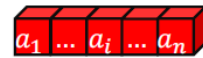
- Transferring face attributes 을 하는 많은 접근법 대부분은 한가지 이상의 제약을 가진다.
  1. Exemplars 로 image 를 생성할 수 없다.
  2. 여러 face attributes 를 동시에 transfer 할 수 없다.
  3. 생성된 이미지에서 low-resolution 또는 artifacts 와 같은 low quality 를 만든다.
- 이 세가지 제약들을 극복하기 위해, 우리는 multiple face attribute transfer 를 제안한다.

## 1. Exemplars 로 image 를 생성할 수 없다.

- Exemplars 로 images 를 생성하기 위해, model 은 conditional image generation 을 위한 reference 를 받아야 한다.
- Labels 를 reference 로 사용
  - 이전 방법들 [2, 13, 17, 18]의 대부분은 Conditional image generation 을 guiding 하기 위해 labels 를 직접적으로 사용한다.
  - 그러나, label 이 제공하는 정보는 매우 제한적이며 해당 label 의 다양한 이미지에 비례하지 않는다.
  - 즉, 다양한 종류의 smiling faces 는 smiling 으로 분류되지만, label smiling 으로부터 smiling faces 가 생성될 수 없다.
- Latent encodings 를 reference 로 사용
  - Encoder 가 image 의 unique identifier 로 간주될 수 있기 때문에, images 의 latent encodings 를 reference 로 설정한다.
  - 이런 방식으로 생성된 이미지는 reference images 의 attributes 와 정확히 같은 스타일의 attribute 를 갖는다.

## 2. 여러 face attributes 를 동시에 transfer 할 수 없다.

- 여러 개의 attributes 를 동시에 조정하기 위해서, image 의 latent encodings 는 여러 parts 로 나뉜다.



Latent Encodings

- 각 part ( $a_i$ ) 는 한 가지 attribute [27] 정보를 encode 한다.
- 이렇게 여러 attributes 는 disentangled manner 로 encode 된다.

### 3. 생성된 이미지에서 low-resolution 또는 artifacts 와 같은 low quality 를 만든다.

- 생성된 이미지의 quality 를 향상시키기 위해, Residual learning [8, 21]과 Multi-scale discriminators [26]를 사용한다.
- Residual Learning
  - Face attributes 의 local 속성은 face attributes transfer 에서 unique 하다.
  - Local 속성을 사용하면 image 의 local part 만 수정하여 face attributes 를 transfer 할 수 있으므로 훈련 난이도를 완화할 수 있다.
- Multi-scale discriminators
  - Different levels of information (local + global)을 포착할 수 있다.
  - 이것은 전체적인 부분과 지역적인 부분 모두를 transfer 하는 데에 유용하다.

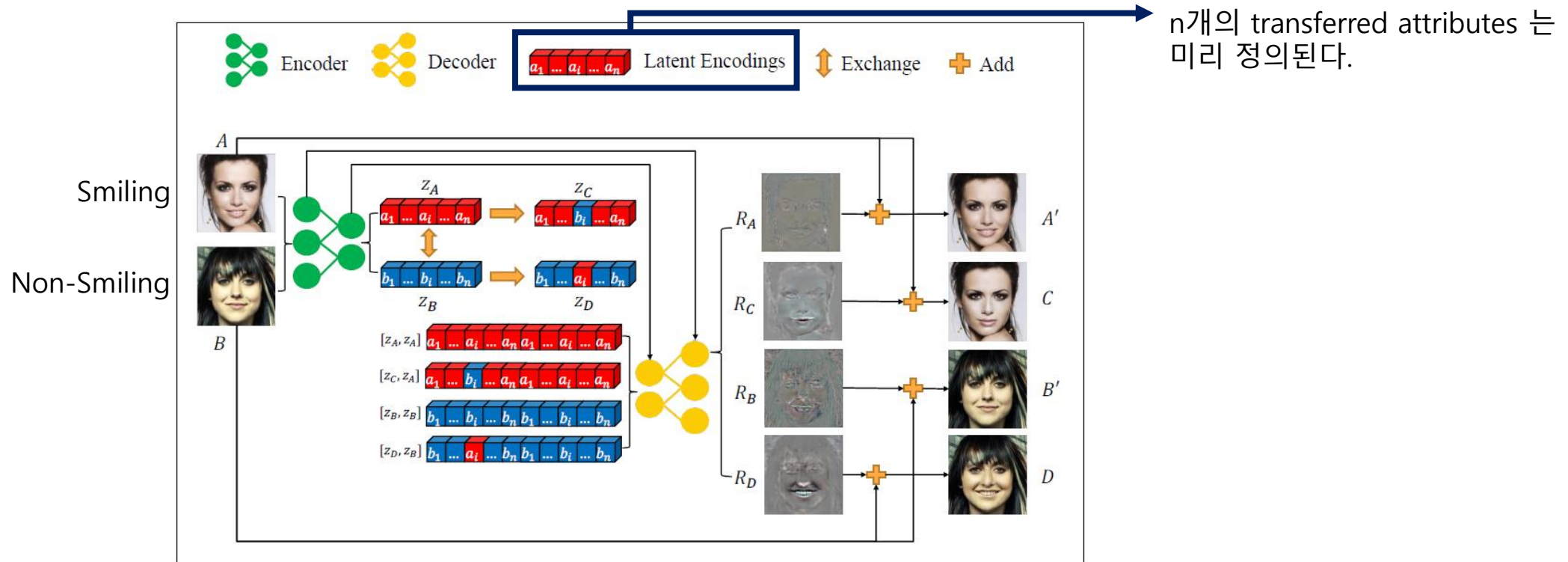
# 3 Our Method

---

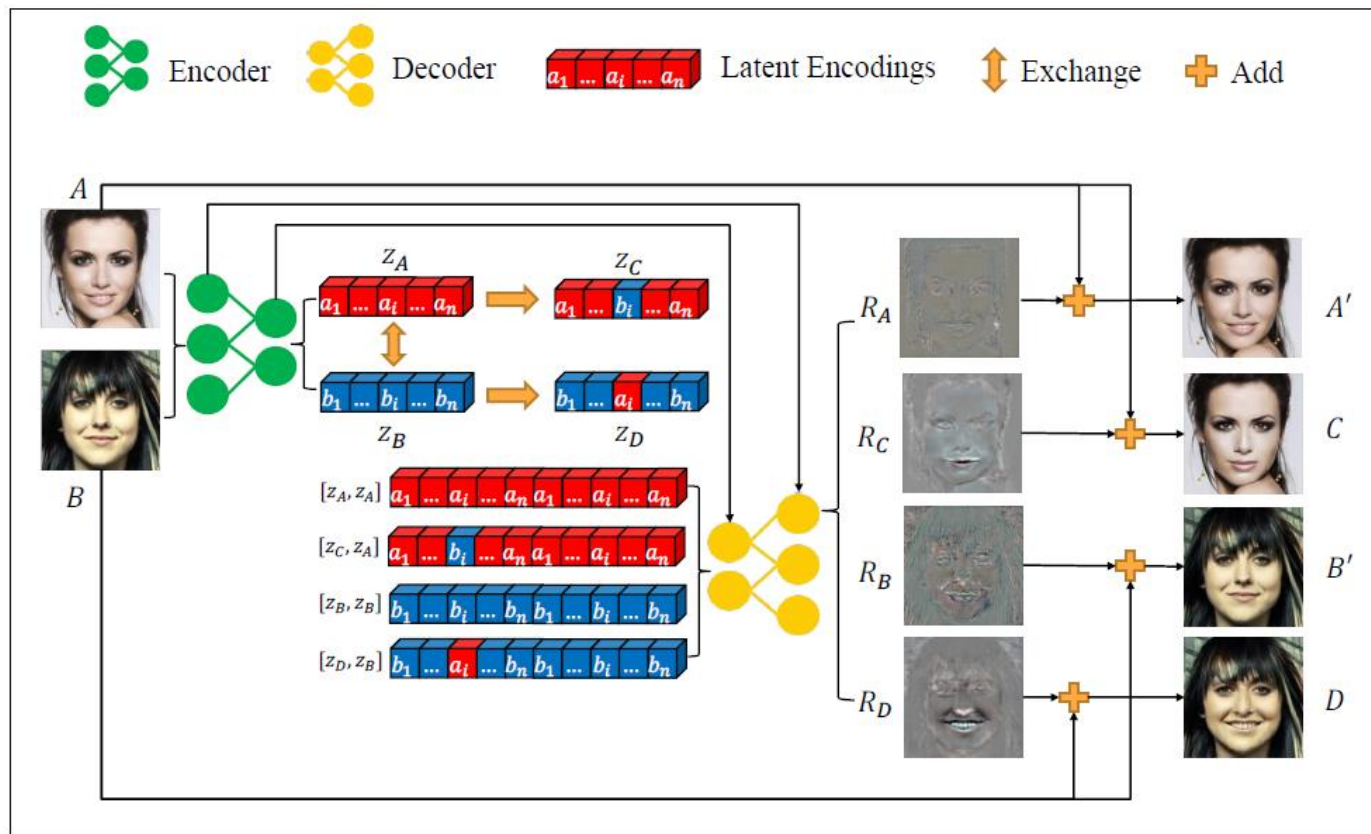
ELEGANT = Exchanging Latent Encodings with GAN for Transferring multiple face attributes

# 3.1 The ELEGANT Model

- ELEGANT model 의 inputs
  - Positive set : Attribute 를 가진다. ( $A$ )
  - Negative set : Attribute 를 가지지 않는다. ( $B$ )
  - Positive set 의 identity 가 Negative set 의 identity 와 같을 필요는 없다.



- Iterative training strategy 를 사용한다.
  - 반대의 attribute 를 가진 한 쌍의 이미지를 inputs 으로 한다. ( $A, B$ )
  - 매번 특정 attribute 에 대해 모델을 훈련시킨다. ( $a_i$ )
  - 모든 attribute 를 반복적으로 훈련한다. ( $a_1, \dots, a_n$ )



- 현재 iteration 에서 i-th attribute 인 smiling 에 대해서 ELEGANT 를 훈련시키는 경우
  - Inputs
    - Smiling images 의 집합
    - Non-smiling images 의 집합
  - A 와 B 의 attribute labels
    - $Y^A = (y_1^A, \dots, 1_i, \dots, y_n^A) : i\text{-th attribute 가 } 1 \rightarrow \text{Smiling}$
    - $Y^B = (y_1^B, \dots, 0_i, \dots, y_n^B) : i\text{-th attribute 가 } 0 \rightarrow \text{Non-smiling}$



## Encoder

- Encoder 로 Images  $A$  와  $B$  의 latent encodings( $z_A, z_B$ ) 를 얻는다.

$$z_A = \text{Enc}(A) = [a_1, \dots, a_i, \dots, a_n], \quad z_B = \text{Enc}(B) = [b_1, \dots, b_i, \dots, b_n] \quad (1)$$

- $a_i$ (or  $b_i$ ) : Image  $A$ (or  $B$ )의 smiling 정보를 encode 하는 feature tensor.

$$z_A = \text{Enc}(A) = [a_1, \dots, \boxed{a_i}, \dots, a_n], \quad z_B = \text{Enc}(B) = [b_1, \dots, \boxed{b_i}, \dots, b_n] \quad (1)$$

## Encoder

- $z_A$  와  $z_B$  의 latent encodings 에서 i-th part 를 교환해서  $z_C$  와  $z_D$  를 얻는다.

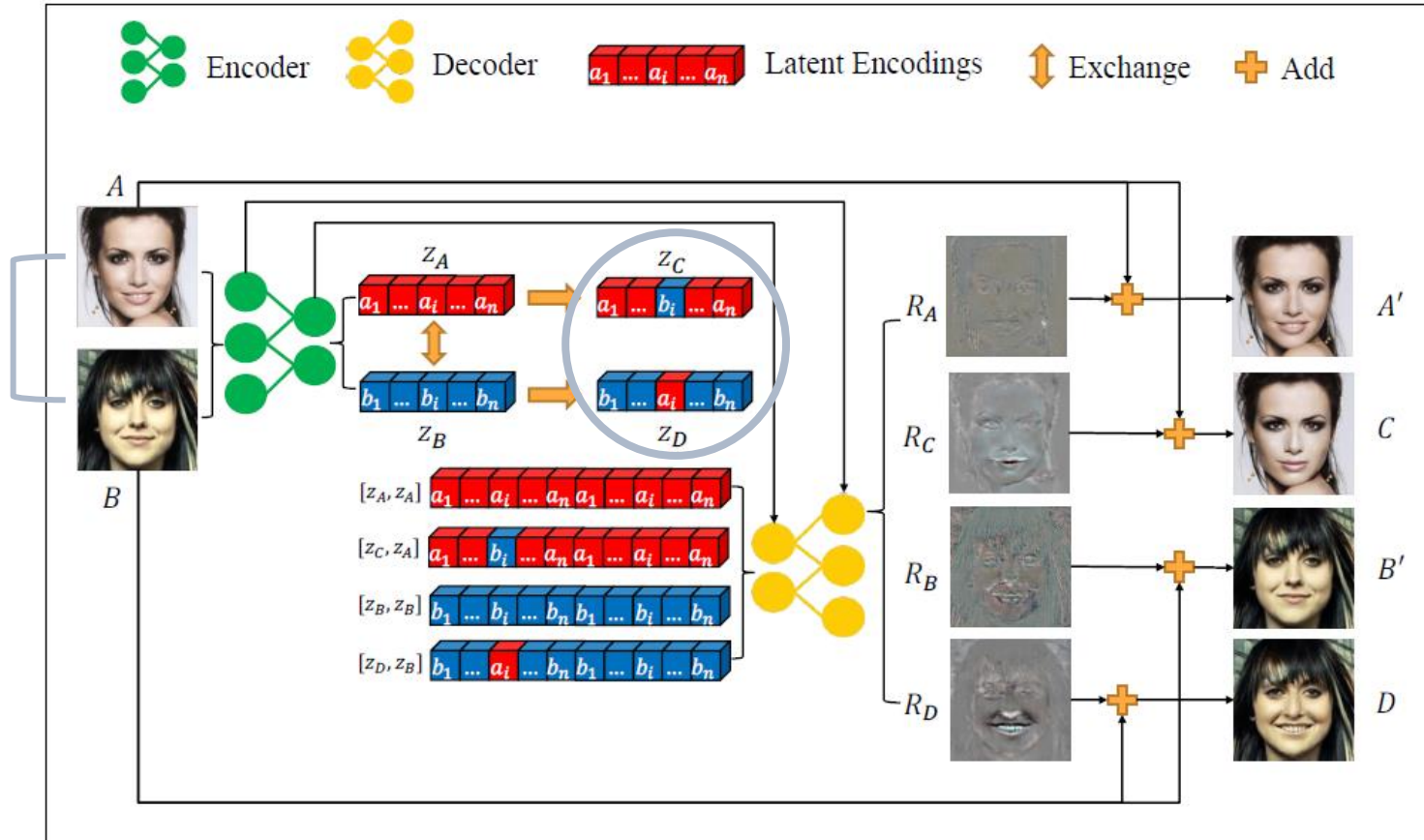


$$z_C = [a_1, \dots, \boxed{b_i}, \dots, a_n], \quad z_D = [b_1, \dots, \boxed{a_i}, \dots, b_n] \quad (2)$$

- $z_C$  가 image A 의 non-smiling version 의 encodings,  $z_D$  가 image B 의 smiling version 의 encodings 라고 기대된다.

## Encoder

$A$  와  $B$  는 서로의  
reference image 이다.



$C$  와  $D$  는 Latent encodings 를  
swap 해서 만든다.

Fig. 6: The ELEGANT model architecture.

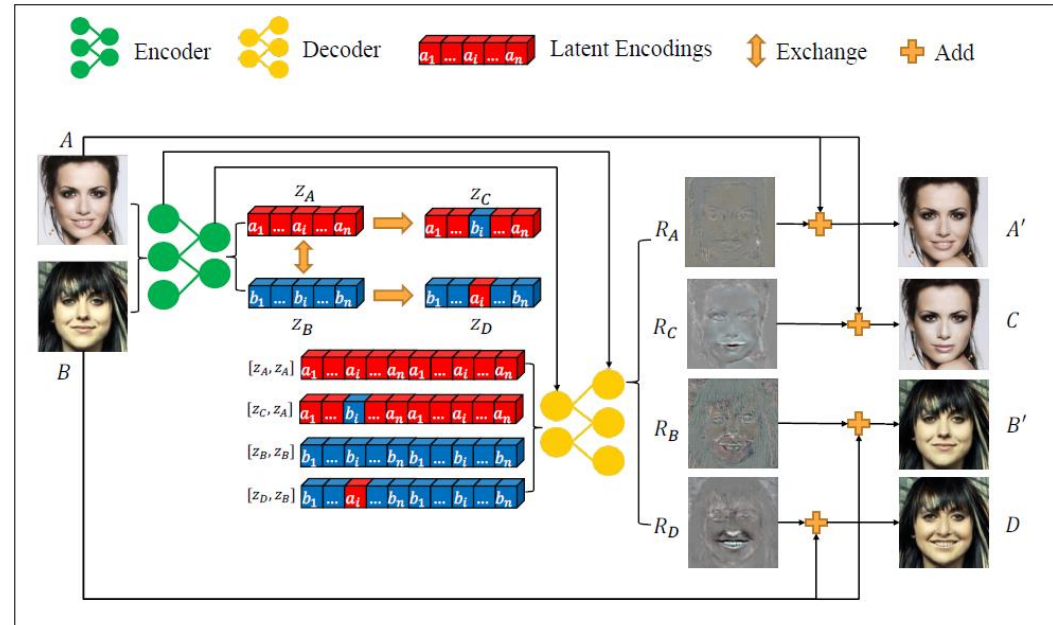
# Decoder

- Original image 를 학습하는 것보다 residual image 를 학습하는 것이 더 낫다. (Sec 2)

$$\text{Dec}([z_A, z_A]) = R_A, \quad A' = A + R_A \quad \text{Dec}([z_C, z_A]) = R_C, \quad C = A + R_C \quad (3)$$

$$\text{Dec}([z_B, z_B]) = R_B, \quad B' = B + R_B \quad \text{Dec}([z_D, z_B]) = R_D, \quad D = B + R_D \quad (4)$$

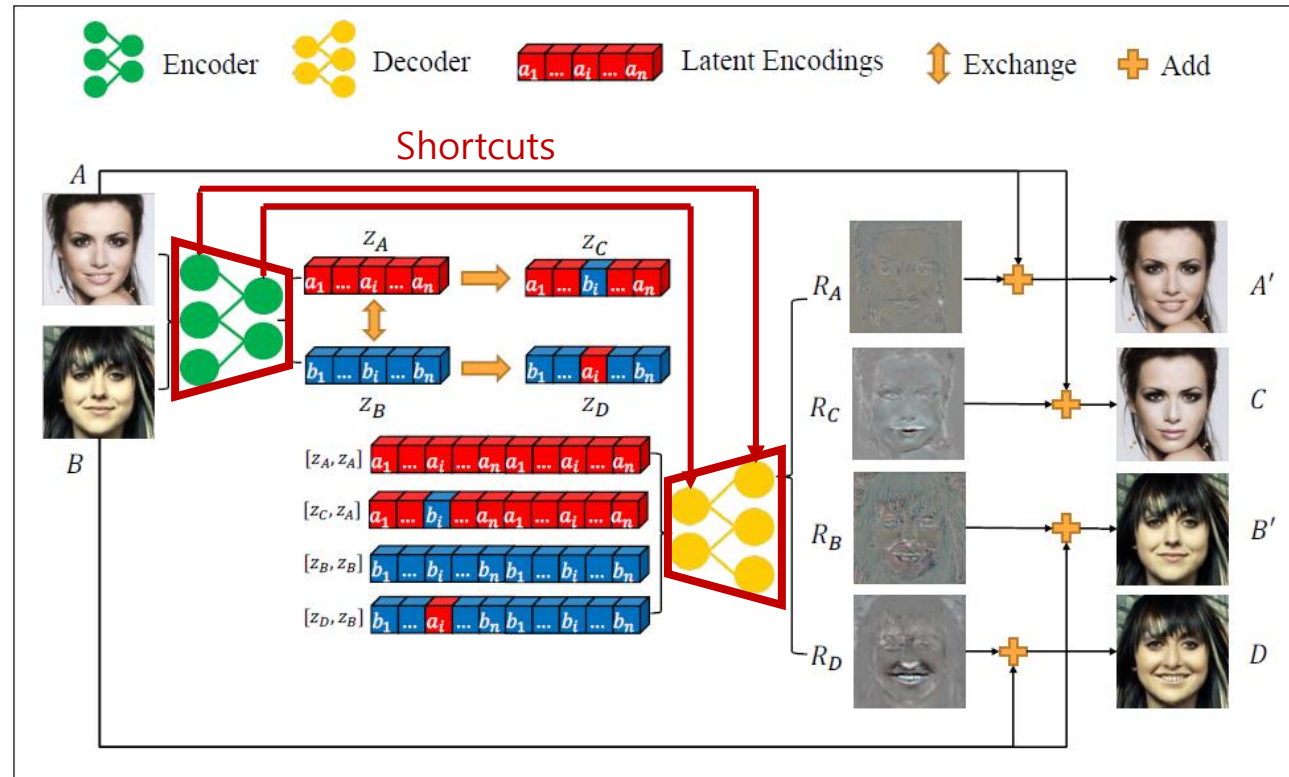
- $R_A, R_B, R_C, R_D$  : Residual images
- $A', B'$  : Reconstructed images
- $C, D$  : Images of novel attributes
- $[z_C, z_A]$  :  $z_C$  와  $z_A$  의 concatenation



- $[z_C, z_A]$  대신  $|z_C - z_A|$  를 사용할 수 있지만, concatenation 을 사용하는 이유는 subtraction 연산이 Dec 에 의해 학습될 수 있기 때문이다.

## Encoder + Decoder

- 더 나은 시각적 결과를 위해 U-NET[20] 구조를 사용한다.
- Enc 와 Dec 의 구조는 대칭적이고, 그들의 중간 layers 는 shortcuts 로 연결된다.
- Shortcuts 는 original images 를 content condition 으로 가져오고, 이것은 매끄러운 novel attributes 를 만든다.



# Discriminators

- Enc 와 Dec 는 모두 generator 역할을 하고, adversarial training 을 위해 discriminators 가 필요하다.
- 그러나, 단일 discriminator 의 receptive field 는 input 이미지의 크기가 커질 때 제한이 있다.

# Discriminators

- 이 문제를 해결하기 위해, multi-scale discriminators [26] 을 채택한다.
  - 두 개의 discriminators 는 동일한 network 구조를 가지지만, 서로 다른 크기의 image 에서 동작한다.
- Larger scale 에서 동작 :  $D_1$ 
  - $D_1$  은  $D_2$  보다 작은 receptive field 를 가진다.
  - $D_1$  은 세부적인 것을 만들기 위해 Enc 와 Dec 를 guiding 한다.
- Smaller scale 에서 동작 :  $D_2$ 
  - $D_2$  는 전체적인 image 를 다룬다.

# Discriminators

- Discriminators 역시 image labels 를 conditional inputs 으로 받아야 한다.
- 전체  $n$  개의 attributes 가 있다.
- 각 iteration 에서 discriminators 의 output 은 하나의 attribute 에 대해서 생성된 이미지들이 얼마나 실제와 같게 보이는지를 반영한다.
- 각 iteration 에서 어떤 attribute 를 다루는지 discriminators 가 알아야 한다.
- 수학적으로는  $D_1(A|Y^A)$ 로 나타낸다.
  - Label  $Y^A$  가 주어졌을 때 image  $A$  에 대한  $D_1$  output score



# Discriminators

- $C$ 와  $D$ 의 attribute labels 는 novel attributes 를 가지기 때문에 거기에 집중해야 한다.

$$Y^A = (y_1^A, \dots, 1_i, \dots, y_n^A) \quad Y^B = (y_1^B, \dots, 0_i, \dots, y_n^B) \quad (5)$$

$$Y^C = (y_1^A, \dots, 0_i, \dots, y_n^A) \quad Y^D = (y_1^B, \dots, 1_i, \dots, y_n^B) \quad (6)$$

- $Y^C$  는  $Y^A$  의  $i$ -th element 인 1을 0으로 바꾼 것이다.

## 3.2 Loss Functions

- Loss function for Discriminators

- Multi-scale discriminators  $D_1$  과  $D_2$  는 standard adversarial loss 를 사용한다.

$$L_{D_1} = -\mathbb{E}(\log(D_1(A|Y^A))) - \mathbb{E}(\log(1 - D_1(C|Y^C))) \\ - \mathbb{E}(\log(D_1(B|Y^B))) - \mathbb{E}(\log(1 - D_1(D|Y^D))) \quad (7)$$

$$L_{D_2} = -\mathbb{E}(\log(D_2(A|Y^A))) - \mathbb{E}(\log(1 - D_2(C|Y^C))) \\ - \mathbb{E}(\log(D_2(B|Y^B))) - \mathbb{E}(\log(1 - D_2(D|Y^D))) \quad (8)$$

$$L_D = L_{D_1} + L_{D_2} \quad (9)$$

- $L_D$  를 최소화할 때, 우리는 실제로 real images 에 대한 scores 를 최대화 시키는 동안 fake image 에 대한 scores 를 최소화 시킨다.

- Loss function for Generator (Enc + Dec)

- Reconstruction loss

- Encoding 과 Decoding 을 한 후 original image 가 얼마나 잘 reconstructed 되는지 측정한다.

$$L_{reconstruction} = ||A - A'|| + ||B - B'|| \quad (10)$$

- Standard adversarial loss

- 생성된 이미지가 얼마나 realistic 한지 측정한다.

$$\begin{aligned} L_{adv} = & -\mathbb{E}(\log(D_1(C|Y^C))) - \mathbb{E}(\log(D_1(D|Y^D))) \\ & - \mathbb{E}(\log(D_2(C|Y^C))) - \mathbb{E}(\log(D_2(D|Y^D))) \end{aligned} \quad (11)$$

- Total loss

$$L_G = L_{reconstruction} + L_{adv}. \quad (12)$$

# 4 Experiments

---

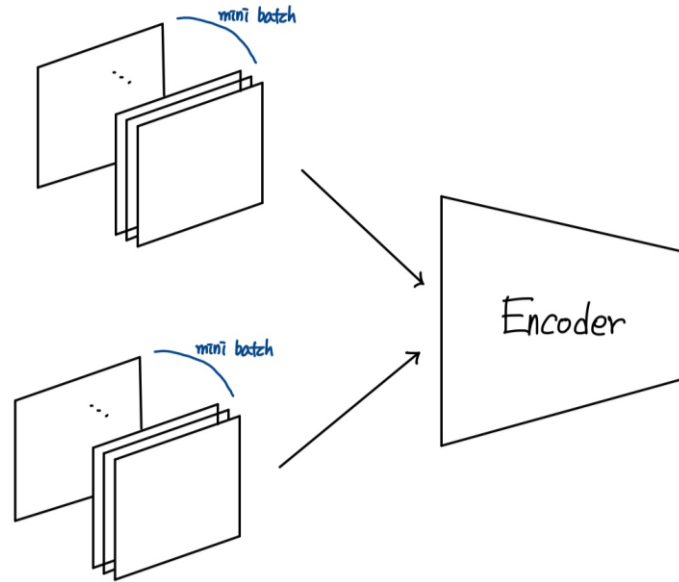
- Dataset
  - CelebaA[16]
    - large-scale for database including 202599 face images of 10177 identities
    - 각각은 40개의 attributes annotations 가 있고 5 landmark locations 가 있다.
    - 우리는 5-point landmarks 를 모든 face 를 align 하는데 쓰고 그것들을 256x256 크기로 crop 한다.

- Implementation

- Encoder : 5 layers of Conv-Norm-LeakyReLU block
- Decoder : 5 layers of Deconv-Norm-LeakyReLU block
- Multi-scale discriminators : 5 layers of Conv-Norm-LeakyReLU blocks + fully connected layer
- Optimizer : Adam[12]
- Learning rate :  $2e-4$
- $\beta_1 = 0.5, \beta_2 = 0.999$
- 모든 input images 는  $[-1, 1]$ 로 normalize 한다.
- Input image 와 output image 차이가 최대 2이기 때문에, Decoder 의 마지막 layers 는  $2 \cdot \tanh$  를 사용해서  $[-2, 2]$ 로 고정한다.
- Out-of-range error 를 피하기 위해 Input image 에 residual image 를 더한 후 output image value 를  $[-1, 1]$ 로 고정한다.

이 페이지는 완벽하기 이해하지 못했다.

- ELEGANT 는 inputs 으로 반대의 attribute 를 가진 두 개의 미니배치 크기의 이미지들을 받는다.



- 그러므로 각 레이어에서 두 미니배치 크기의 이미지들의 moving mean 과 moving variance 는 큰 차이를 만들어야 한다.
- 만약 Batch Normalization 을 사용하면, 각 layer 에서 이러한 running statistics 는 항상 진동할 것이다.

- 이 문제를 해결하기 위해, BN 대신  $\ell_2$  normalization 을 사용했다.
  - $\hat{x} = \frac{x}{\|x\|_2} \cdot \alpha + \beta$  ( $\alpha$  와  $\beta$  는 learnable parameters)
- Moving statistics 를 계산하지 않아도, ELEGANT 는 안정적으로 수렴되고 face attributes 를 효과적으로 swap 한다.

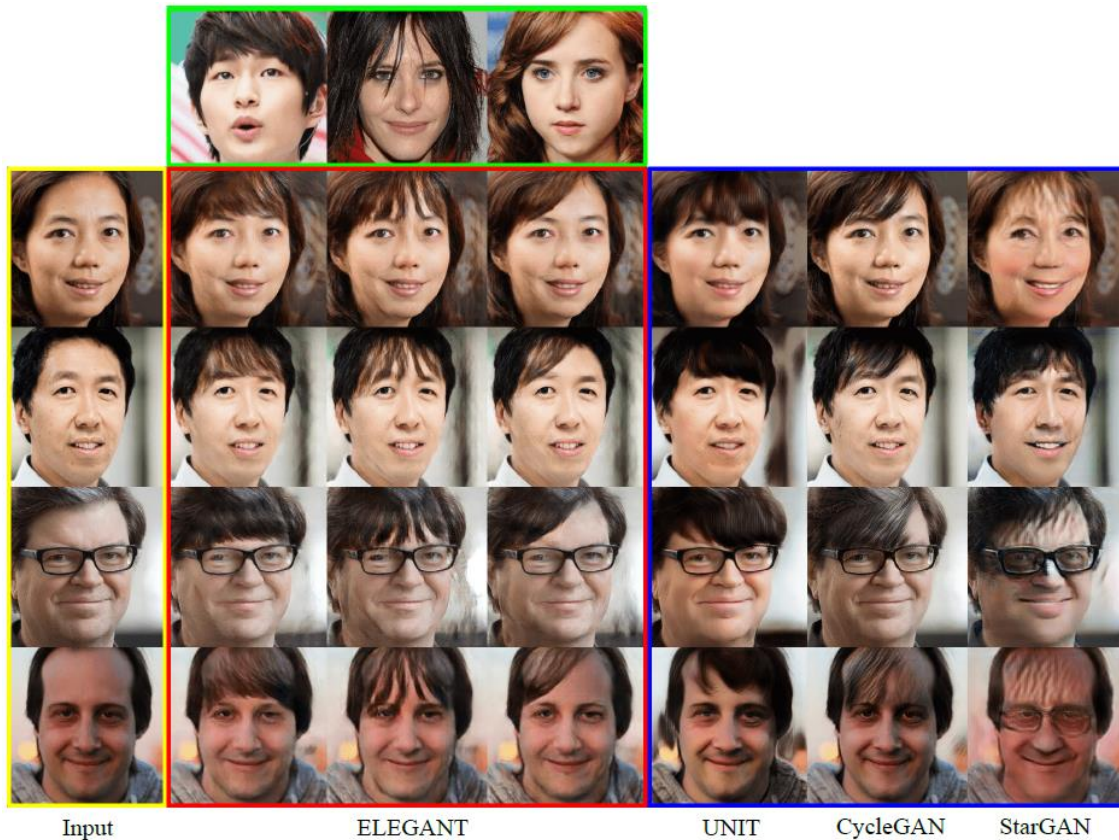


## 4.1 Face Image Generation by Exemplars

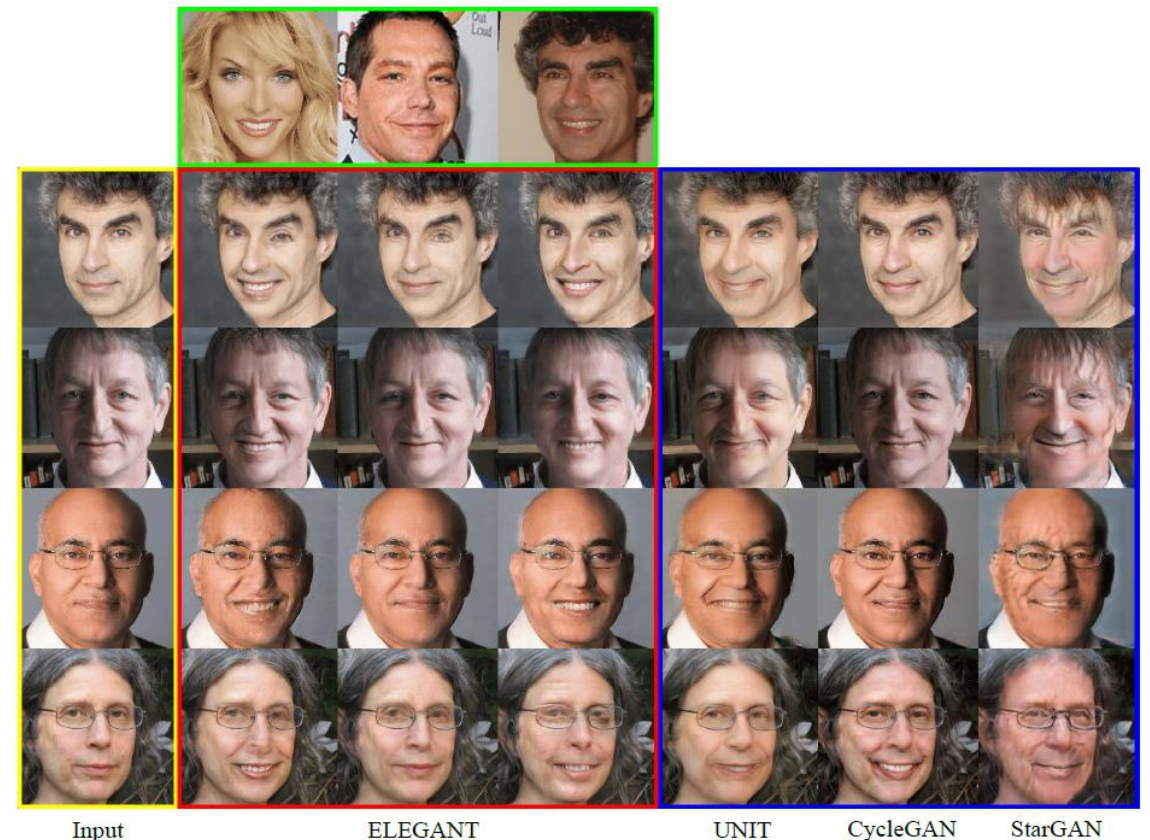
- Model 이 exemplars 에 의해 face images 를 생성할 수 있다는 것을 설명하기 위해, 대조군으로 UNIT[15], CycleGAN[32], StarGAN[2] 를 선택했다.

## Experiments

- Face image generation by exemplars.
  - Yellow box : Input images outside the training data
  - Green box : Reference images
  - Red box : Results of ELEGANT
  - Blue box : Results of other methods
- ELEGANT
  - Reference images 의 정확히 같은 스타일의 attribute 를 가진 다른 face images 를 만든다.
- 다른 방법들
  - 흔한 스타일의 attribute 을 가진 face images 를 만든다.



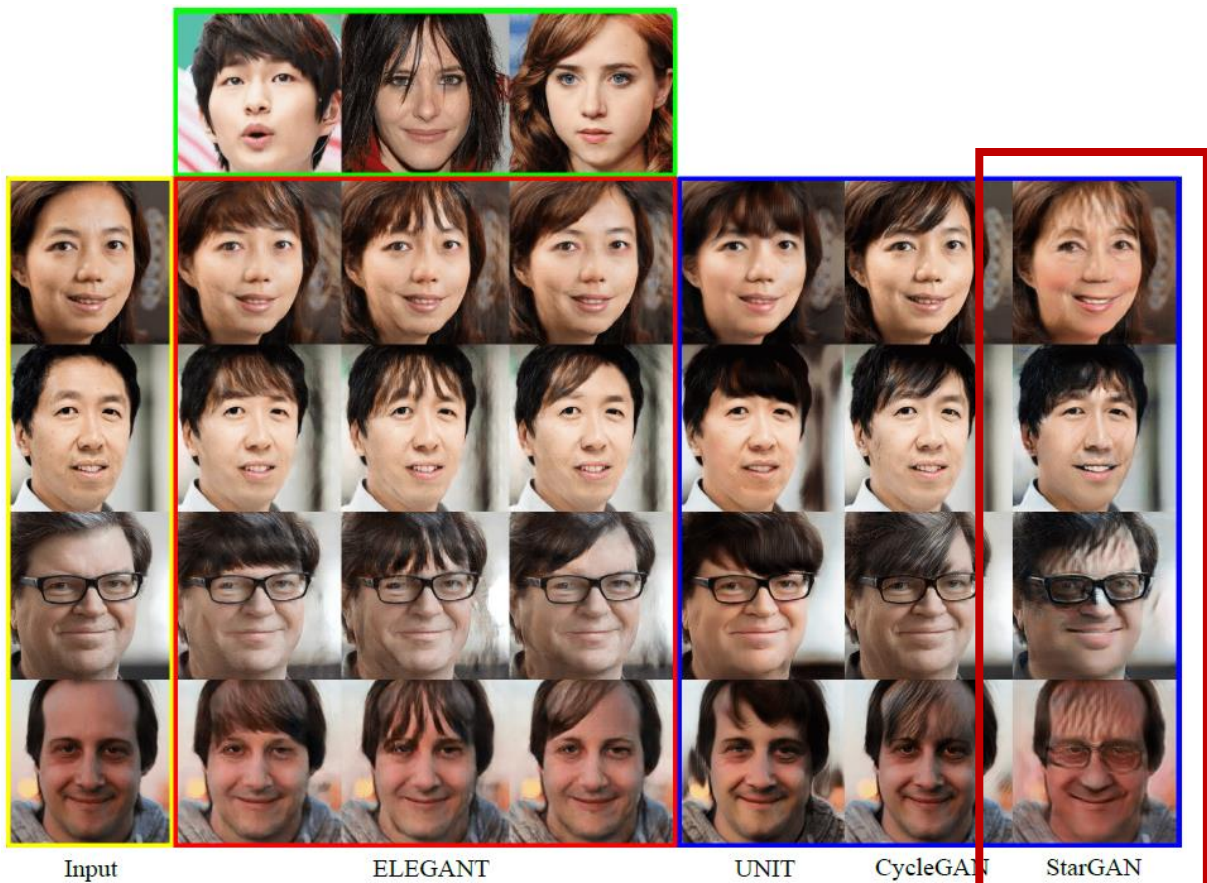
(a) bangs



(b) smiling

- 여기서 볼 수 있는 StarGAN 단점

- StarGAN 은 여러 개의 attributes 를 transfer 할 수 있다. 그러나, 한 가지 attribute 를 transferring 할 때 다른 attributes 도 바꾼다.



(a) bangs

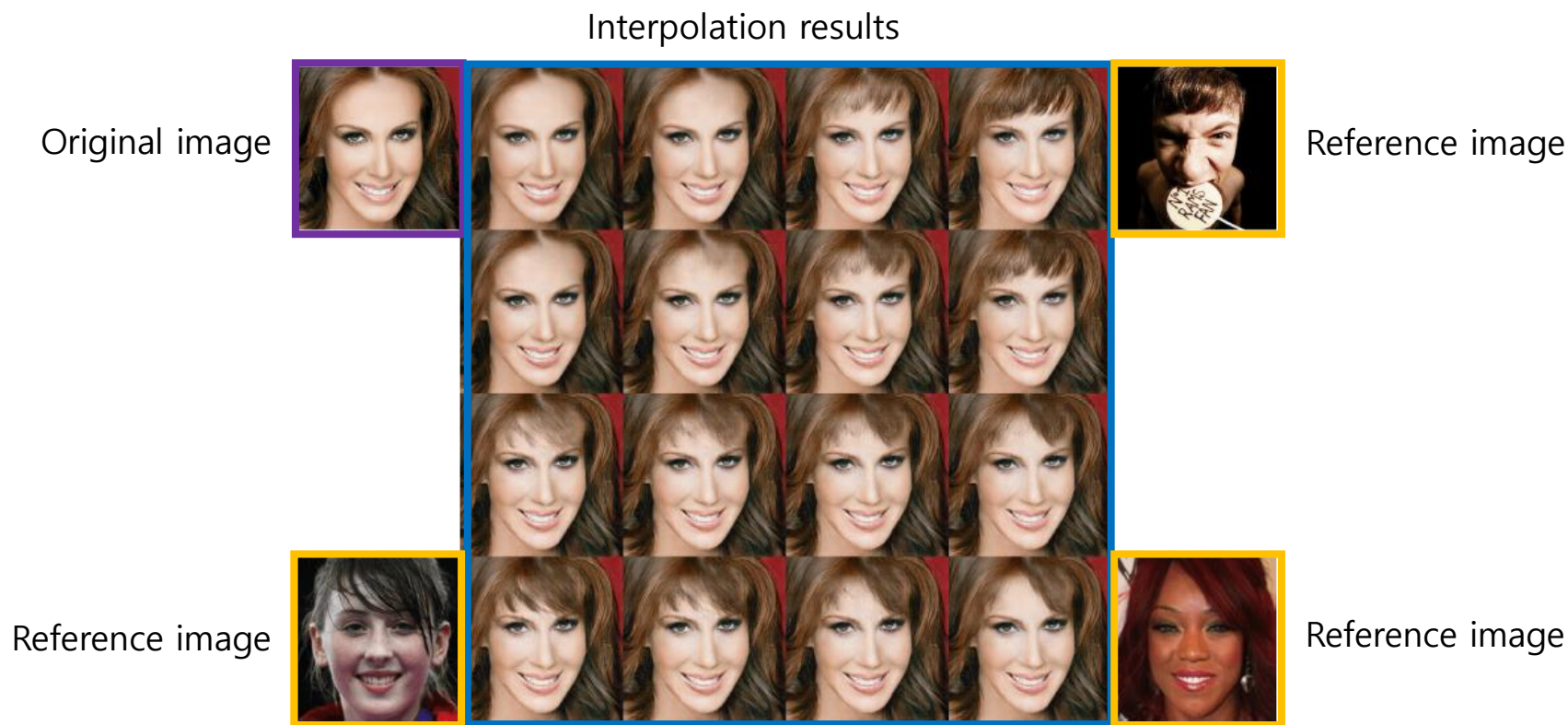
- Fei-Fei Li 와 Andrew Ng 에게 bangs 를 입혔을 때 age 도 함께 바뀐다. (더 어려졌다.)



- 이것은 StarGAN 이 input image 에 확실한 label 을 요구하기 때문이다.
- 이 두개의 images (Fei-Fei Li 와 Andrew Ng)는 attribute young 에서 1로 확실하게 labeled 된다.
- 그러나, 그것들 둘 다 middle-aged 이고 young 이나 old 둘 중 하나로 간단히 labeled 될 수 없다.

## Experiments

- ELEGANT model 에서 latent encodings 를 교환하는 구조는 효과적으로 StarGAN 의 문제를 해결한다.
- ELEGANT 는 현재 다루고 있는 attribute 에 초점을 맞추고, testing 단계에서 input images 를 위한 labels 를 요구하지 않는다.



- 더욱이, ELEGANT 는 reference images 의 다른 bangs style 사이에서 미묘한 차이를 학습할 수 있다.

## 4.2 Dealing with Multiple Attributes Simultaneously

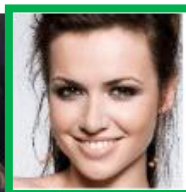
- ELEGANT 와 DNA-GAN[27] 을 비교한다.
  - 공통점
    - 다수의 face attribute 를 조정할 수 있다.
    - Exemplars 에 의해 images 를 생성할 수 있다.
- 세 개의 attributes(bangs, smiling, mustache)에 대해서 실험한다.
- 두개의 모델에 동일한 face images 와 reference images 를 사용해서 실험한다.

## Experiments

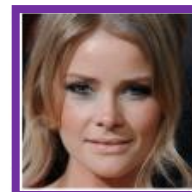
ELEGANT

DNA-GAN

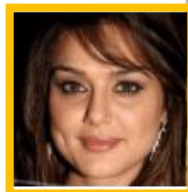
Original image



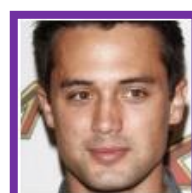
Reference images  
of the second  
attributes



Reference images  
of the first  
attributes

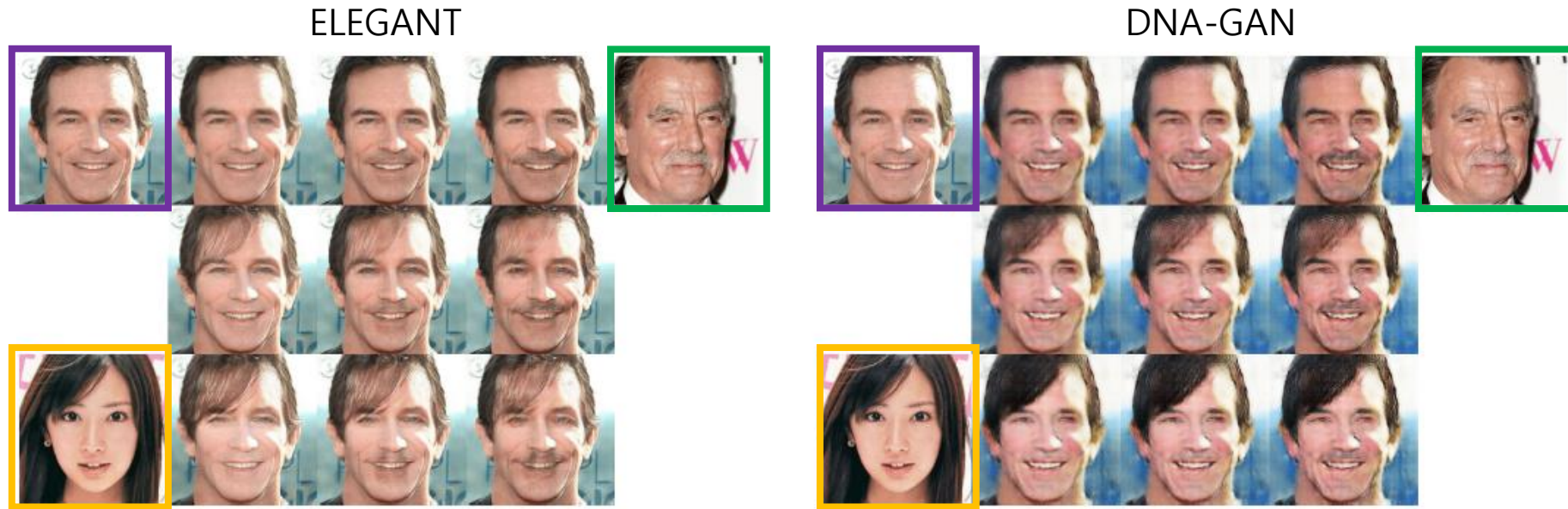


(a) Bangs and Smiling



(b) Smiling and Mustache





(c) Bangs and Mustache

- ELEGANT 는 DNA-GAN 의 결과보다 훨씬 낫다.
- ELEGANT 는 특히 미세한 details 를 잘 만든다.
- DNA-GAN 과 비교해서 개선점은 주로 residual learning 의 결과와 multi-scale discriminators 에 있다.



- Residual learning 은 훈련의 어려움을 감소시킨다.
  - DNA-GAN
    - DNA-GAN 은 훈련이 불안정하다. (특히 고화질 images 에서)
    - 그것은 generator 와 discriminator 사이가 불균형하기 때문이다.
    - DNA-GAN 의 훈련의 초기 단계에서, generator 는 터무니 없는 것을 출력한다.
    - 그래서 discriminator 는 실제 이미지로부터 생성된 이미지를 구별 하는 방법을 쉽게 배우게 되고, 이것은 균형을 깨뜨린다.
  - ELEGANT
    - ELEGANT 는 residual learning 의 idea 를 채택했기 때문에, generator 의 outputs 은 초기 단계에서 거의 original image 와 동일하다.
    - 이러한 방식으로, discriminator 는 빠르게 훈련될 수 없게 되고, 그것은 training process 를 안정화 시킨다.
- 반면, 이미지의 크기가 더 커지면서 generator 의 부담이 discriminator 보다 커지게 된다.
- Generator 의 출력 크기는 더 커지지만, discriminator 는 여느 때와 같이 오직 숫자만을 출력하기 때문이다.
- 그러나, ELEGANT 는 적은 수의 pixels 를 수정해야 하는 residual images 를 학습함으로써 generator 의 출력 크기의 차원을 효과적으로 줄인다.

- Multi-scale discriminators 는 생성된 이미지의 quality 를 향상시킨다.
  - 작은 크기의 이미지에서 동작하는 discriminator 는 전체적인 image content generation 을 guide 한다.
  - 큰 크기의 이미지에서 동작하는 discriminator 는 generator 가 finer details 를 만드는 것을 돕는다.

이 페이지는 완벽하기 이해하지 못했다.

- 더욱이,
  - DNA-GAN
    - Additional part 를 사용해서 face id 와 background information 을 encode 한다.
    - 이 방법은 문제점을 가진다. : Loss constraints 를 충족하기 위해 두 개의 입력 이미지들이 직접적으로 swap 될 수 있다.
    - Xiao et al. [27] 은 이 문제를 해결하기 위해 annihilating operation 을 제안했다. 그러나 이 연산은 parameter spaces 에서 왜곡을 만들어 훈련을 어렵게 한다.
  - ELEGANT
    - 변화를 책임지는 residual images 를 학습해서 face id 와 background information 을 자동으로 보존한다.
    - ELEGANT 는 latent encodings 에서 annihilating operation 과 additional part 를 제거함으로써 framework 전체를 더 멋들어지고 이해하기 쉽게 만든다.

## 4.3 High-quality Generated Images

- 더 면밀하게 살펴보기 위해, Large 크기의 이미지에서 다른 여러가지 attributes 에 대한 ELEGANT 의 결과를 보여준다.
- 생성된 이미지의 quality 를 측정하기 위해 Fréchet Inception Distance [9] (FID) 를 사용한다.
- FID 는 두 분포 사이의 거리를 다음과 같이 측정한다.

$$d^2 = ||\mu_1 - \mu_2||^2 + \text{Tr}(C_1 + C_2 - 2(C_1 C_2)^{1/2}). \quad (13)$$

- $(\mu_1, C_1)$  과  $(\mu_2, C_2)$  는 두 distributions 의 means 와 variance metrics 이다.

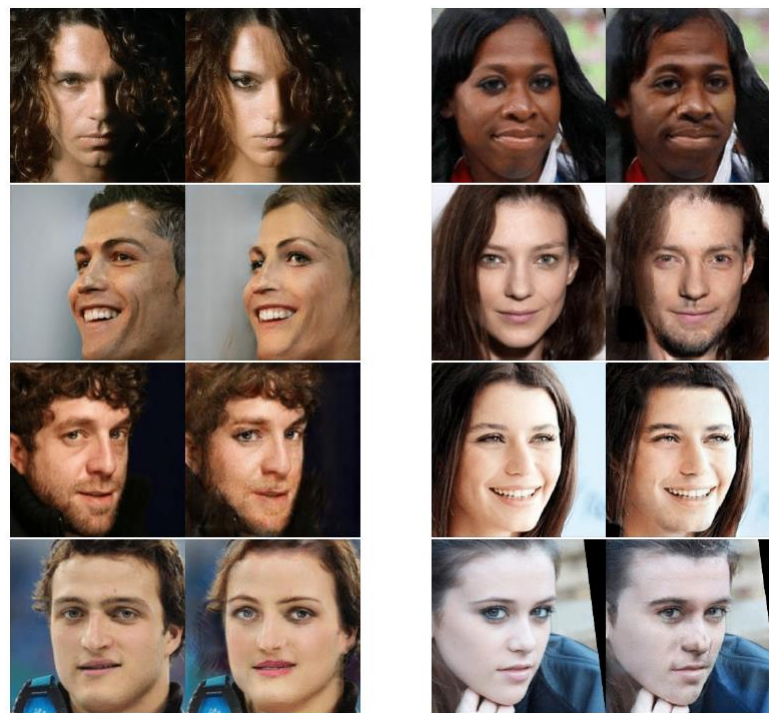
Fig. 1



(a) removing bangs

(b) adding bangs

Fig. 2



(a) feminizing

(b) virilizing

Fig. 3

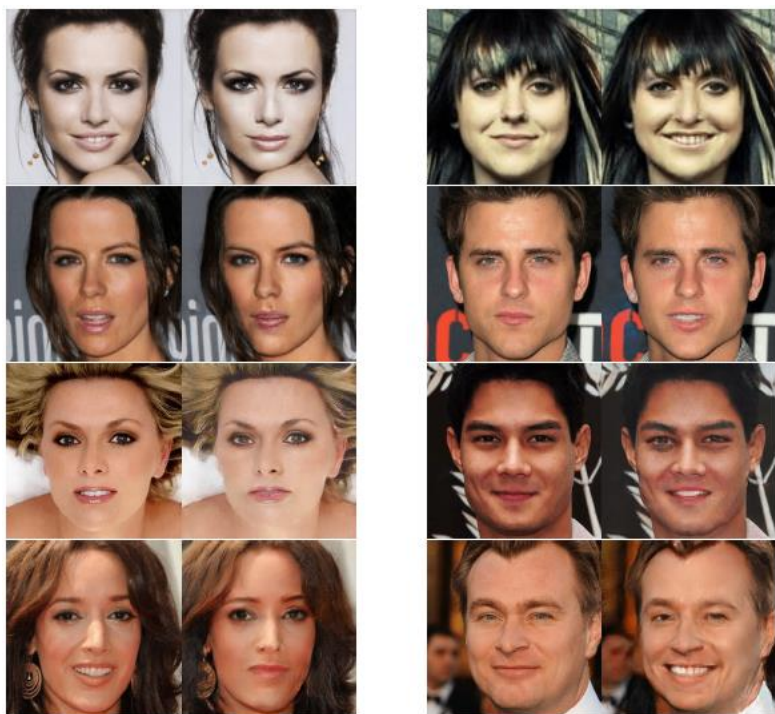


(a) removing eyeglasses

(b) adding eyeglasses



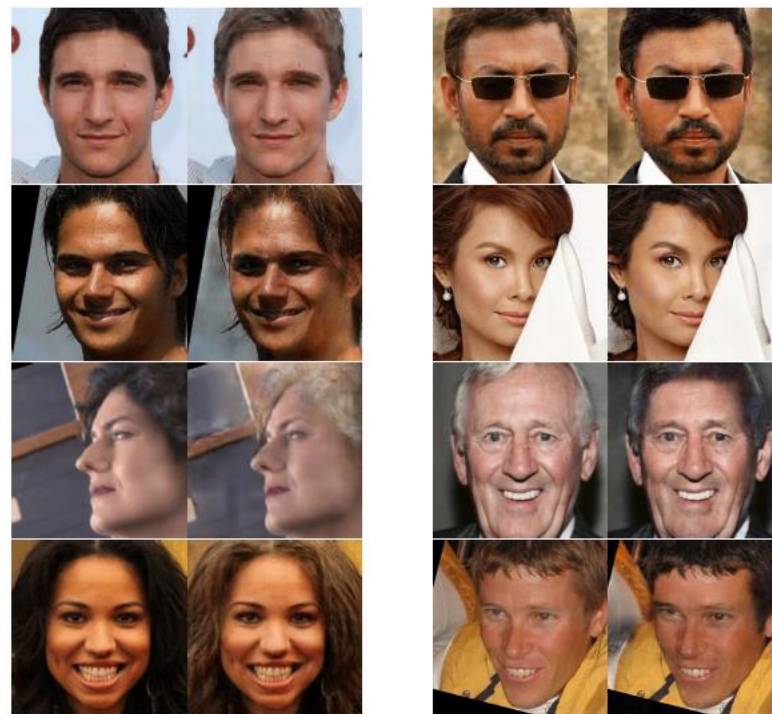
Fig. 4



(a) removing smile

(b) adding smile

Fig. 5



(a) black hair to non-black

(b) non-black hair to black

## Experiments

Table 1: FID of Different Methods with respect to five attributes. The + (−) represents the generated images by adding (removing) the attribute.

FID	bangs		smiling		mustache		eyeglasses		male	
	+	−	+	−	+	−	+	−	+	−
UNIT	135.41	137.94	120.25	125.04	119.32	131.33	111.49	139.43	152.16	154.59
CycleGAN	<b>27.81</b>	33.22	<b>23.23</b>	<b>22.74</b>	43.58	55.49	<b>36.87</b>	<b>48.82</b>	60.25	<b>46.25</b>
StarGAN	59.68	71.07	51.36	78.87	99.03	176.18	70.40	142.35	70.14	206.21
DNA-GAN	79.27	76.89	77.04	72.35	126.33	127.66	75.02	75.96	121.04	118.67
ELEGANT	30.71	<b>31.12</b>	25.71	24.88	<b>37.51</b>	<b>49.13</b>	47.35	60.71	<b>59.37</b>	56.80

- 다른 여러가지 attributes 에 대해서 실제 이미지의 분포와 생성된 이미지의 분포 사이에서 FID 를 계산한다.
- ELEGANT 는 다른 방법들과 비교해서 좋은 결과를 낸다. (낮을 수록 quality 가 좋다.)
- FID score 는 다음 두 가지 이유 때문에 참고만 하는 용도로 쓰인다.
  - 이유 1: ELEGANT 와 DNA-GAN 은 exemplars 에 의해 images 를 생성할 수 있고, 이것은 다른 종류의 image translation methods 보다 더 일반적이고 어렵다. 그래서 어떤 종류의 qualitative measures 를 사용하는 것이 여전히 불공평하다.
  - 이유 2: GAN 을 위한 합리적인 qualitative measure 는 아직 정해지지 않았다.

## 5 Conclusion



1. Transferring multiple face attributes 를 위해 ELEGANT 를 만들었다.
2. ELEGANT 는 다른 attributes 를 구분된 부분으로 encode 하고, latent encodings 의 특정 부분을 교환해서 novel attributes 와 함께 이미지들을 생성한다.
3. 오직 이미지의 local 부분만 바뀌어야 하기 때문에, 고화질 이미지에서 훈련을 쉽게 하기 위해 residual learning 을 채택한다.
4. U-Net 구조와 multi-scale discriminators 는 이미지의 품질을 향상시킨다.
5. CelebA face database 에서의 실험 결과는 ELEGANT 가 세가지의 일반적인 제약을 성공적으로 극복했다는 것을 보여준다.