(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2022년07월01일
(11) 등록번호 10-2416219
(24) 등록일자 2022년06월29일

(51) 국제특허분류(Int. Cl.)
G06T 3/00 (2019.01) *G06N 3/08* (2006.01)
(52) CPC특허분류
G06T 3/00 (2019.01)
G06N 3/08 (2013.01)
(21) 출원번호 10-2020-0093861
(22) 출원일자 2020년07월28일
심사청구일자 2020년07월28일
(65) 공개번호 10-2022-0014148
(43) 공개일자 2022년02월04일
(56) 선행기술조사문헌
Hwang & Byun. Unsupervised image-to-image translation via fair representation of gender bias, ICASSP, 2020년 5월, pp. 1953-1957. 1부.*
Sattigeri et al. FAIRNESS GAN: GENERATING DATASETS WITH FAIRNESS PROPERTIES USING A GENERATIVE ADVERSARIAL NETWORK. ICLR, 2019년 10월 15일. pp. 1-12. 1부.*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
변혜란
서울특별시 서대문구 연세로 50, 제4공학관 810호(신촌동, 연세대학교)
황선희
서울특별시 서대문구 연세로 50, 제4공학관 810호(신촌동, 연세대학교)
(뒷면에 계속)
(74) 대리인
특허법인우인

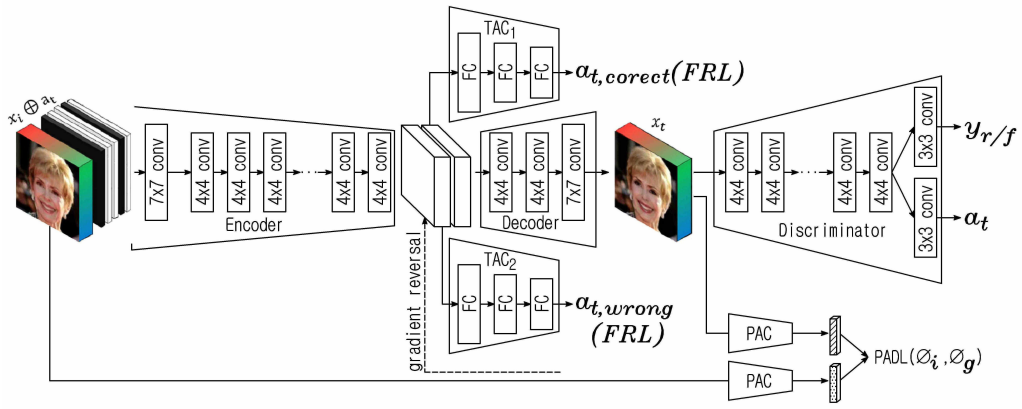
전체 청구항 수 : 총 10 항 심사관 : 남옥우

(54) 발명의 명칭 **보호 속성 표현 학습을 이용한 공정한 이미지 변환 장치 및 방법**

(57) 요 약

본 실시예들은 보호 속성 표현을 1차로 학습하고 학습된 보호 속성 표현을 이용하여 이미지 생성 모델을 2차로 학습하여 보호 속성에 따라 이미지를 공정하게 변환할 수 있는 공정한 이미지 변환 장치 및 방법을 제공한다.

대 표 도



(52) CPC특허분류
G06T 2207/20081 (2013.01)
(72) 발명자
박성호
서울특별시 서대문구 연세로 50, 제4공학관 810호 (신촌동, 연세대학교)
김도형
서울특별시 서대문구 연세로 50, 제4공학관 810호 (신촌동, 연세대학교)
도미래
서울특별시 서대문구 연세로 50, 제4공학관 810호 (신촌동, 연세대학교)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711102850
과제번호	2019-0-01396-002
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송연구개발사업
연구과제명	인공지능 모델과 학습데이터의 편향성 분석-탐지-완화 · 제거 지원 프레임워크 개발(2/4)
기 여 율	1/1
과제수행기관명	한국과학기술원
연구기간	2020.01.01 ~ 2020.12.31

명 세 서

청구범위

청구항 1

컴퓨팅 디바이스에 의한 공정한 이미지 변환 방법에 있어서,

원본 이미지와 변환 속성을 입력받는 단계; 및

상기 변환 속성을 기준으로 이미지 변환 모델을 통해 상기 원본 이미지를 변환 이미지로 변환하는 단계를 포함하며,

상기 이미지 변환 모델은 인코더를 통해 상기 원본 이미지로부터 특징을 추출하고 디코더를 통해 상기 특징으로부터 상기 변환 이미지로 변환하는 생성 모델을 포함하며, 상기 원본 이미지 및 상기 변환 이미지가 입력된 보호 속성 표현 모델의 결과를 반영하여 학습되며,

상기 원본 이미지를 상기 변환 이미지로 변환하는 단계는, 상기 이미지 변환 모델에 상기 변환 속성을 기준으로 상기 원본 이미지의 사이즈를 고려하여 깊이 차원을 확장한 후 확장된 이미지를 입력하는 것을 특징으로 하는 공정한 이미지 변환 방법.

청구항 2

삭제

청구항 3

제1항에 있어서,

상기 이미지 변환 모델이 학습되는 과정에서,

상기 생성 모델에 연결되며 상기 원본 이미지와 상기 생성된 이미지를 비교하여 상기 이미지 변환 모델의 파라미터를 설정하는 판별 모델;

상기 생성 모델에 연결되며 상기 특징이 상기 변환 속성과 관련된 것으로 분류하는 제1 잠재 공간; 및

상기 생성 모델에 연결되며 상기 특징이 상기 변환 속성과 관련되지 않은 것으로 분류하는 제2 잠재 공간의 결과를 반영하여 학습되는 것을 특징으로 하는 공정한 이미지 변환 방법.

청구항 4

제1항에 있어서,

상기 이미지 변환 모델이 학습되는 과정에서,

상기 보호 속성 표현 모델을 이용하여 상기 원본 이미지로부터 추출한 제1 보호 속성 표현 및 상기 보호 속성 표현 모델을 이용하여 상기 변환 이미지로부터 추출한 제2 보호 속성 표현을 비교하여 상기 이미지 변환 모델의 파라미터를 설정하는 것을 특징으로 하는 공정한 이미지 변환 방법.

청구항 5

제1항에 있어서,

상기 이미지 변환 모델이 학습되는 과정에서,

상기 원본 이미지로부터 추출한 제1 보호 속성 표현 및 상기 보호 속성 표현 모델을 이용하여 상기 변환 이미지로부터 추출한 제2 보호 속성 표현 간의 보호 속성 측정 거리를 최소화하는 과정을 통해, 상기 원본 이미지 및 상기 변환 이미지 간의 보호 속성 표현이 동일해지도록 학습되는 것을 특징으로 하는 공정한 이미지 변환 방법.

청구항 6

제1항에 있어서,

상기 보호 속성 표현 모델은,

입력된 보호 속성을 분류하되 상기 보호 속성에 대한 레이블이 존재하는 제1 데이터 세트와 상기 보호 속성에 대한 레이블이 존재하지 않는 제2 데이터 세트의 도메인을 구분하지 못하도록 학습된 모델인 것을 특징으로 하는 공정한 이미지 변환 방법.

청구항 7

하나 이상의 프로세서 및 상기 하나 이상의 프로세서에 의해 실행되는 하나 이상의 프로그램을 저장하는 메모리를 포함하는 공정한 이미지 변환 장치에 있어서,

상기 프로세서는 원본 이미지와 변환 속성을 입력받고,

상기 프로세서는 상기 변환 속성을 기준으로 이미지 변환 모델을 통해 상기 원본 이미지를 변환 이미지로 변환하며,

상기 이미지 변환 모델은 인코더를 통해 상기 원본 이미지로부터 특징을 추출하고 디코더를 통해 상기 특징으로부터 상기 변환 이미지로 변환하는 생성 모델을 포함하며, 보호 속성 표현 모델의 결과를 반영하여 학습되며,

상기 프로세서는 상기 이미지 변환 모델에 상기 변환 속성을 기준으로 상기 원본 이미지의 사이즈를 고려하여 깊이 차원을 확장한 후 확장된 이미지를 입력하는 것을 특징으로 하는 공정한 이미지 변환 장치.

청구항 8

삭제

청구항 9

제7항에 있어서,

상기 이미지 변환 모델은,

상기 생성 모델에 연결되며 상기 원본 이미지와 상기 생성된 이미지를 비교하여 상기 이미지 변환 모델의 파라미터를 설정하는 판별 모델;

상기 생성 모델에 연결되며 상기 특징이 상기 변환 속성과 관련된 것으로 분류하는 제1 잠재 공간; 및

상기 생성 모델에 연결되며 상기 특징이 상기 변환 속성과 관련되지 않은 것으로 분류하는 제2 잠재 공간의 결과를 반영하여 학습되는 것을 특징으로 하는 공정한 이미지 변환 장치.

청구항 10

제7항에 있어서,

상기 프로세서는 상기 이미지 변환 모델이 학습되는 과정에서,

상기 보호 속성 표현 모델을 이용하여 상기 원본 이미지로부터 추출한 제1 보호 속성 표현 및 상기 보호 속성 표현 모델을 이용하여 상기 변환 이미지로부터 추출한 제2 보호 속성 표현을 비교하여 상기 이미지 변환 모델의 파라미터를 설정하는 것을 특징으로 하는 공정한 이미지 변환 장치.

청구항 11

제7항에 있어서,

상기 이미지 변환 모델은,

상기 원본 이미지로부터 추출한 제1 보호 속성 표현 및 상기 보호 속성 표현 모델을 이용하여 상기 변환 이미지로부터 추출한 제2 보호 속성 표현 간의 보호 속성 측정 거리를 최소화하는 과정을 통해, 상기 원본 이미지 및 상기 변환 이미지 간의 보호 속성 표현이 동일해지도록 학습되는 것을 특징으로 하는 공정한 이미지 변환 장치.

청구항 12

제7항에 있어서,

상기 보호 속성 표현 모델은,

상기 보호 속성을 분류하되 상기 보호 속성에 대한 레이블이 존재하는 제1 데이터 세트와 상기 보호 속성에 대한 레이블이 존재하지 않는 제2 데이터 세트의 도메인을 구분하지 못하도록 학습된 모델인 것을 특징으로 하는 공정한 이미지 변환 장치.

발명의 설명

기 술 분 야

[0001] 본 발명이 속하는 기술 분야는 보호 속성 표현 학습을 이용한 공정한 이미지 변환 장치 및 방법에 관한 것이다.

배 경 기 술

[0002] 이 부분에 기술된 내용은 단순히 본 실시예에 대한 배경 정보를 제공할 뿐 종래기술을 구성하는 것은 아니다.

[0003] 기존의 머신 러닝 기반의 이미지 변환 기술은 보호 속성을 고려하지 않고, 불공정한 스타일 변환을 수행한다. 예컨대, 얼굴 데이터 생성 모델이 성별, 인종, 나이에 대해서 차별적인 이미지를 생성하는 문제가 있다.

선행기술문헌

특허문헌

[0004] (특허문헌 0001) 한국공개특허공보 제10-2018-0134727호 (2018.12.19)

발명의 내용

해결하려는 과제

[0005] 본 발명의 실시예들은 보호 속성 표현을 1차로 학습하고 학습된 보호 속성 표현을 이용하여 이미지 생성 모델을 2차로 학습하여 이미지 생성 모델이 보호 속성에 따라 이미지를 공정하게 변환하는 데 주된 목적이 있다.

[0006] 본 발명의 명시되지 않은 또 다른 목적들은 하기의 상세한 설명 및 그 효과로부터 용이하게 추론할 수 있는 범위 내에서 추가적으로 고려될 수 있다.

과제의 해결 수단

[0007] 본 실시예의 일 측면에 의하면, 컴퓨팅 디바이스에 의한 공정한 이미지 변환 방법에 있어서, 원본 이미지와 변환 속성을 입력받는 단계, 및 상기 변환 속성을 기준으로 이미지 변환 모델을 통해 상기 원본 이미지를 변환 이미지로 변환하는 단계를 포함하며, 상기 이미지 변환 모델은 인코더를 통해 상기 입력 이미지로부터 특징을 추출하고 디코더를 통해 상기 특징으로부터 상기 변환 이미지로 변환하는 생성 모델을 포함하며, 상기 원본 이미지 및 상기 변환 이미지가 입력된 보호 속성 표현 모델의 결과를 반영하여 학습된 것을 특징으로 하는 공정한 이미지 변환 방법을 제공한다.

[0008] 본 실시예의 다른 측면에 의하면, 하나 이상의 프로세서 및 상기 하나 이상의 프로세서에 의해 실행되는 하나 이상의 프로그램을 저장하는 메모리를 포함하는 공정한 이미지 변환 장치에 있어서, 상기 프로세서는 원본 이미지와 변환 속성을 입력받고, 상기 프로세서는 상기 변환 속성을 기준으로 이미지 변환 모델을 통해 상기 원본 이미지를 변환 이미지로 변환하며, 상기 이미지 변환 모델은 인코더를 통해 상기 입력 이미지로부터 특징을 추출하고 디코더를 통해 상기 특징으로부터 상기 변환 이미지로 변환하는 생성 모델을 포함하며, 보호 속성 표현 모델의 결과를 반영하여 학습된 것을 특징으로 하는 공정한 이미지 변환 장치를 제공한다.

발명의 효과

[0009] 이상에서 설명한 바와 같이 본 발명의 실시예들에 의하면, 보호 속성 표현을 1차로 학습하고 학습된 보호 속성 표현을 이용하여 이미지 생성 모델을 2차로 학습하여 이미지 생성 모델이 보호 속성에 따라 이미지를 공정하게

변환할 수 있는 효과가 있다.

[0010] 여기에서 명시적으로 언급되지 않은 효과라 하더라도, 본 발명의 기술적 특징에 의해 기대되는 이하의 명세서에서 기재된 효과 및 그 잠정적인 효과는 본 발명의 명세서에 기재된 것과 같이 취급된다.

도면의 간단한 설명

[0011] 도 1은 본 발명의 일 실시예에 따른 공정한 이미지 변환 장치를 예시한 블록도이다.

도 2는 본 발명의 일 실시예에 따른 공정한 이미지 변환 장치의 보호 속성 표현 모델을 예시한 도면이다.

도 3은 본 발명의 일 실시예에 따른 공정한 이미지 변환 장치의 이미지 변환 모델을 예시한 도면이다.

도 4는 본 발명의 다른 실시예에 따른 공정한 이미지 변환 방법의 학습 동작을 예시한 흐름도이다.

도 5는 본 발명의 다른 실시예에 따른 공정한 이미지 변환 방법의 이미지 변환 동작을 예시한 흐름도이다.

도 6 및 도 7은 본 발명의 실시예들에 따른 시뮬레이션 결과를 예시한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0012] 이하, 본 발명을 설명함에 있어서 관련된 공지기능에 대하여 이 분야의 기술자에게 자명한 사항으로서 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략하고, 본 발명의 일부 실시예들을 예시적인 도면을 통해 상세하게 설명한다.

[0013] 도 1은 본 발명의 일 실시예에 따른 공정한 이미지 변환 장치를 예시한 블록도이다.

[0014] 공정한 이미지 변환 장치(110)는 적어도 하나의 프로세서(120), 컴퓨터 판독 가능한 저장매체(130) 및 통신 버스(170)를 포함한다.

[0015] 프로세서(120)는 공정한 이미지 변환 장치(110)로 동작하도록 제어할 수 있다. 예컨대, 프로세서(120)는 컴퓨터 판독 가능한 저장 매체(130)에 저장된 하나 이상의 프로그램들을 실행할 수 있다. 하나 이상의 프로그램들은 하나 이상의 컴퓨터 실행 가능 명령어를 포함할 수 있으며, 컴퓨터 실행 가능 명령어는 프로세서(120)에 의해 실행되는 경우 공정한 이미지 변환 장치(110)로 하여금 예시적인 실시예에 따른 동작들을 수행하도록 구성될 수 있다.

[0016] 컴퓨터 판독 가능한 저장 매체(130)는 컴퓨터 실행 가능 명령어 내지 프로그램 코드, 프로그램 데이터 및/또는 다른 적합한 형태의 정보를 저장하도록 구성된다. 컴퓨터 판독 가능한 저장 매체(130)에 저장된 프로그램(140)은 프로세서(120)에 의해 실행 가능한 명령어의 집합을 포함한다. 일 실시예에서, 컴퓨터 판독한 가능 저장 매체(130)는 메모리(랜덤 액세스 메모리와 같은 휘발성 메모리, 비휘발성 메모리, 또는 이들의 적절한 조합), 하나 이상의 자기 디스크 저장 디바이스들, 광학 디스크 저장 디바이스들, 플래시 메모리 디바이스들, 그 밖에 공정한 이미지 변환 장치(110)에 의해 액세스되고 원하는 정보를 저장할 수 있는 다른 형태의 저장 매체, 또는 이들의 적합한 조합일 수 있다.

[0017] 통신 버스(170)는 프로세서(120), 컴퓨터 판독 가능한 저장 매체(140)를 포함하여 공정한 이미지 변환 장치(110)의 다른 다양한 컴포넌트들을 상호 연결한다.

[0018] 공정한 이미지 변환 장치(110)는 또한 하나 이상의 입출력 장치(24)를 위한 인터페이스를 제공하는 하나 이상의 입출력 인터페이스(150) 및 하나 이상의 통신 인터페이스(160)를 포함할 수 있다. 입출력 인터페이스(150) 및 통신 인터페이스(160)는 통신 버스(170)에 연결된다. 입출력 장치(미도시)는 입출력 인터페이스(150)를 통해 공정한 이미지 변환 장치(110)의 다른 컴포넌트들에 연결될 수 있다.

[0019] 공정한 이미지 변환 장치(110)는 보호 속성 표현을 1차로 학습하고 학습된 보호 속성 표현을 이용하여 이미지 생성 모델을 2차로 학습하여 이미지 생성 모델이 보호 속성에 따라 이미지를 공정하게 변환한다.

[0020] 프로세서는 원본 이미지와 변환 속성을 입력받고, 프로세서는 변환 속성을 기준으로 이미지 변환 모델을 통해 원본 이미지를 변환 이미지로 변환한다. 보호 속성은 성별, 인종, 나이 등과 같이 집단에 대해서 차별적 요소로 취급되어 편향 가능한 속성을 의미한다.

[0021] 도 2는 본 발명의 일 실시예에 따른 공정한 이미지 변환 장치의 보호 속성 표현 모델을 예시한 도면이다.

[0022] 보호 속성 표현 모델은, 보호 속성을 분류하되 보호 속성에 대한 레이블이 존재하는 제1 데이터 세트와 보호 속

성에 대한 레이블이 존재하지 않는 제2 데이터 세트의 도메인을 구분하지 못하도록 학습된 모델이다.

[0023] 제1 도메인은 클래스 레이블이 매칭된 데이터 세트이고 제2 도메인은 클래스 레이블이 없는 데이터 세트로 설정될 수 있다.

[0024] 보호 속성 표현 모델은 손실 함수를 최소화하는 방향으로 네트워크 가중치를 갱신한다. 보호 속성 표현 모델의 손실 함수는 수학식 1 및 수학식 2와 같이 표현된다.

수학식 1

[0025] $$\mathcal{L}_{PAC} = \mathcal{L}_{PA} - \lambda \mathcal{L}_{ce}(y_d | fc_d(f(x)))$$

[0026] 손실 함수(L_{PAC})는 보호 속성을 올바르게 분류하도록 학습되는 보호 속성 손실 함수(L_{PA})와 정규화 상수를 곱하고 마이너스(-)를 취하여 도메인 분류시 도메인을 구분하지 못하도록 학습되는 도메인 손실 함수(L_{ce})로 정의된다. y_d는 도메인 속성을 의미한다.

수학식 2

[0027] $$\mathcal{L}_{PA} = \mathcal{L}_{ce}(y_g | f_g(h)) + \mathcal{L}_{ce}(y_a | f_a(h)) + \mathcal{L}_{ce}(y_r | f_r(h))$$

[0028] y_a는 나이 속성이고, y_g는 성별 속성이고, y_r은 인종 속성을 의미한다.

[0029] 보호 속성 표현 모델은 이미지 또는 비디오 등을 입력으로 하고, 시각적 특징 정보를 출력으로 한다. 예컨대, 보호 속성 표현 모델은 CNN(Convolutional Neural Network)으로 구현될 수 있다. 보호 속성 표현 모델은 다수의 레이어가 네트워크로 연결되며 히든 레이어를 포함한다. 레이어는 파라미터를 포함할 수 있고, 레이어의 파라미터는 학습가능한 필터 집합을 포함한다. 필터는 컨볼루션 필터를 적용할 수 있다. 파라미터는 노드 간의 가중치 및/또는 바이어스를 포함한다.

[0030] 보호 속성 표현 모델은 레지듀얼(Residual) 블록을 사용할 수 있다. 레지듀얼 블록은 네트워크 구조의 출력에 다시 입력을 더해서 다음 레이어로 넘기며, 레이어의 입력을 레이어의 출력에 바로 연결하는 스킵 구조를 가질 수 있다.

[0031] 도 3은 본 발명의 일 실시예에 따른 공정한 이미지 변환 장치의 이미지 변환 모델을 예시한 도면이다.

[0032] 이미지 변환 모델은 인코더를 통해 입력 이미지로부터 특징을 추출하고 디코더를 통해 특징으로부터 상기 변환 이미지로 변환하는 생성 모델을 포함하며, 보호 속성 표현 모델의 결과를 반영하여 학습된다.

[0033] 프로세서는 이미지 변환 모델에 변환 속성을 기준으로 입력 이미지의 사이즈를 고려하여 깊이 차원을 확장한 후 확장된 이미지를 입력한다.

[0034] 이미지 변환 모델은 생성 모델에 연결되며 원본 이미지와 생성된 이미지를 비교하여 이미지 변환 모델의 파라미터를 설정하는 판별 모델의 결과를 반영하여 학습된다.

[0035] 이미지 변환 모델은 생성 모델에 연결되며 특징이 변환 속성과 관련된 것으로 분류하는 제1 잠재 공간의 결과를 반영하여 학습된다.

[0036] 이미지 변환 모델은 생성 모델에 연결되며 특징이 변환 속성과 관련되지 않은 것으로 분류하는 제2 잠재 공간의 결과를 반영하여 학습된다.

[0037] 이미지 변환 모델은 원본 이미지로부터 추출한 제1 보호 속성 표현 및 보호 속성 표현 모델을 이용하여 변환 이미지로부터 추출한 제2 보호 속성 표현 간의 보호 속성 측정 거리를 최소화하는 과정을 통해, 원본 이미지 및 변환 이미지 간의 보호 속성 표현이 동일해지도록 학습된다.

[0038] 프로세서는 이미지 변환 모델이 학습되는 과정에서, 보호 속성 표현 모델을 이용하여 원본 이미지로부터 추출한

제1 보호 속성 표현 및 보호 속성 표현 모델을 이용하여 변환 이미지로부터 추출한 제2 보호 속성 표현을 비교하여 이미지 변환 모델의 파라미터를 설정한다.

[0039] 이미지 변환 모델은 적대적 생성 네트워크(Generative Adversarial Network, GAN)를 이용한다. 생성 모델의 목적은 손실을 최소화하고 판별 모델의 목적은 손실을 최대화한다. 이미지 변환 모델은 손실 함수를 최소화하는 방향으로 네트워크 가중치를 갱신한다.

[0040] 이미지 변환 모델의 손실 함수는 수학식 3 내지 수학식 6과 같이 표현된다. 이미지 변환 모델의 손실 함수는 생성 손실 함수, 복원 손실 함수, 대상 속성 관련도 손실 함수, 보호 속성 일치 손실 함수를 포함한다.

수학식 3

[0041]
$$\mathcal{L}_{acgan} = E_x[logD_{adv}(x_i)] + E_x[log(1 - D_{adv}(\tilde{x}_t)] + E_{x,a}[-logD_{cls}(a_i|x_i)] + E_{x,a}[-logD_{cls}(a_t|\tilde{x}_t)]$$

[0042] 생성 손실 함수는 이미지 생성의 정확도 향상 및 조건별 이미지 생성을 위한 손실 함수이다. x_i는 입력 이미지이고, x_t는 생성된 이미지이고, a_i는 입력 이미지의 대상 속성 벡터이고, a_t는 생성된 이미지의 대상 속성 벡터이고, Ex는 확률 분포이다.

수학식 4

[0043]
$$\mathcal{L}_{rec} = E_{x,a}[\|G(\tilde{x}_t, a_i) - x_i\|_1] + E_{x,a}[\|G(x_i, a_i) - x_i\|_1]$$

[0044] 복원 손실 함수는 동일 도메인 변환에서 불필요한 영역의 전달을 방지하여 입력 이미지의 형태를 보존하도록 학습한다.

수학식 5

[0045]
$$\mathcal{L}_{fp} = E_x[-logD(a_t|h_{tr})] + E_x[logD(a_t|h_{tu})]$$

[0046] 대상 속성 관련도 손실 함수는 대상 속성과 관련된 잠재 벡터(latent vector) h_{tr} 및 대상 속성과 관련되지 않은 잠재 벡터 h_{tu}를 통해 차별적 이미지 생성을 방지하도록 학습된다.

수학식 6

[0047]
$$\mathcal{L}_{pad} = E_x[\|\phi_i - \phi_g\|_1]$$

[0048] 보호 속성 일치 손실 함수는 원본 이미지로부터 추출한 제1 보호 속성 표현 Φ_i 및 변환 이미지로부터 추출한 제2 보호 속성 표현 Φ_g 간의 보호 속성 거리가 최소화되도록 학습된다.

[0049] 도 4는 본 발명의 다른 실시예에 따른 공정한 이미지 변환 방법의 학습 동작을 예시한 흐름도이다. 도 5는 본 발명의 다른 실시예에 따른 공정한 이미지 변환 방법의 이미지 변환 동작을 예시한 흐름도이다. 공정한 이미지 변환 방법은 컴퓨팅 디바이스에 의하여 수행될 수 있으며, 공정한 이미지 변환 장치에 의해 동작을 수행할 수 있다.

[0050] 단계 S11에서 프로세서는 1차적으로 보호 속성 표현 모델을 학습한다.

[0051] 단계 S12에서 프로세서는 보호 속성 표현 모델을 이용하여 2차적으로 이미지 변환 모델을 학습한다.

[0052] 단계 S21에서 프로세서는 원본 이미지와 변환 속성을 입력받는다.

[0053] 단계 S22에서 프로세서는 변환 속성을 기준으로 이미지 변환 모델을 통해 원본 이미지를 변환 이미지로 변환한다.

[0054] 이미지 변환 모델은 인코더를 통해 입력 이미지로부터 특징을 추출하고 디코더를 통해 특징으로부터 상기 변환 이미지로 변환하는 생성 모델을 포함하며, 원본 이미지 및 변환 이미지가 입력된 보호 속성 표현 모델의 결과를 반영하여 학습된다.

[0055] 원본 이미지를 변환 이미지로 변환하는 단계는, 이미지 변환 모델에 변환 속성을 기준으로 입력 이미지의 사이즈를 고려하여 깊이 차원을 확장한 후 확장된 이미지를 입력한다.

[0056] 이미지 변환 모델이 학습되는 과정에서, (i) 생성 모델에 연결되며 원본 이미지와 생성된 이미지를 비교하여 이미지 변환 모델의 파라미터를 설정하는 판별 모델, (ii) 생성 모델에 연결되며 특징이 변환 속성과 관련된 것으로 분류하는 제1 잠재 공간, 및 (iii) 생성 모델에 연결되며 특징이 변환 속성과 관련되지 않은 것으로 분류하는 제2 잠재 공간의 결과를 반영하여 학습된다.

[0057] 이미지 변환 모델이 학습되는 과정에서, 보호 속성 표현 모델을 이용하여 원본 이미지로부터 추출한 제1 보호 속성 표현 및 보호 속성 표현 모델을 이용하여 변환 이미지로부터 추출한 제2 보호 속성 표현을 비교하여 이미지 변환 모델의 파라미터를 설정한다.

[0058] 이미지 변환 모델이 학습되는 과정에서, 원본 이미지로부터 추출한 제1 보호 속성 표현 및 보호 속성 표현 모델을 이용하여 변환 이미지로부터 추출한 제2 보호 속성 표현 간의 보호 속성 측정 거리를 최소화하는 과정을 통해, 원본 이미지 및 변환 이미지 간의 보호 속성 표현이 동일해지도록 학습된다.

[0059] 보호 속성 표현 모델은, 입력된 보호 속성을 분류하되 보호 속성에 대한 레이블이 존재하는 제1 데이터 세트와 보호 속성에 대한 레이블이 존재하지 않는 제2 데이터 세트의 도메인을 구분하지 못하도록 학습된다.

[0060] 도 6 및 도 7은 본 발명의 실시예들에 따른 시뮬레이션 결과를 예시한 도면이다.

[0061] 도 6 및 도 7를 참조하면, 기존의 이미지 대 이미지 변환 방식이 성별, 인종, 나이에 대해서 차별적 결과를 나타내는 것과 달리, 본 실시예들은 사전 학습된 보호 속성을 활용하여 보존을 위한 레이블(예컨대, 성별, 인종, 나이 등)이 없는 데이터에 대해서 공정한 변환을 수행하여 공정한 이미지를 생성하는 것을 확인할 수 있다.

[0062] 공정한 이미지 변환 장치는 하드웨어, 펌웨어, 소프트웨어 또는 이들의 조합에 의해 로직회로 내에서 구현될 수 있고, 범용 또는 특정 목적 컴퓨터를 이용하여 구현될 수도 있다. 장치는 고정배선형(Hardwired) 기기, 필드 프로그램 가능한 게이트 어레이(Field Programmable Gate Array, FPGA), 주문형 반도체(Application Specific Integrated Circuit, ASIC) 등을 이용하여 구현될 수 있다. 또한, 장치는 하나 이상의 프로세서 및 컨트롤러를 포함한 시스템온칩(System on Chip, SoC)으로 구현될 수 있다.

[0063] 공정한 이미지 변환 장치는 하드웨어적 요소가 마련된 컴퓨팅 디바이스 또는 서버에 소프트웨어, 하드웨어, 또는 이들의 조합하는 형태로 탑재될 수 있다. 컴퓨팅 디바이스 또는 서버는 각종 기기 또는 유무선 통신망과 통신을 수행하기 위한 통신 모뎀 등의 통신장치, 프로그램을 실행하기 위한 데이터를 저장하는 메모리, 프로그램을 실행하여 연산 및 명령하기 위한 마이크로프로세서 등을 전부 또는 일부 포함한 다양한 장치를 의미할 수 있다.

[0064] 도 4 및 도 5에서는 각각의 과정을 순차적으로 실행하는 것으로 기재하고 있으나 이는 예시적으로 설명한 것에 불과하고, 이 분야의 기술자라면 본 발명의 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 도 4 및 도 5에 기재된 순서를 변경하여 실행하거나 또는 하나 이상의 과정을 병렬적으로 실행하거나 다른 과정을 추가하는 것으로 다양하게 수정 및 변형하여 적용 가능할 것이다.

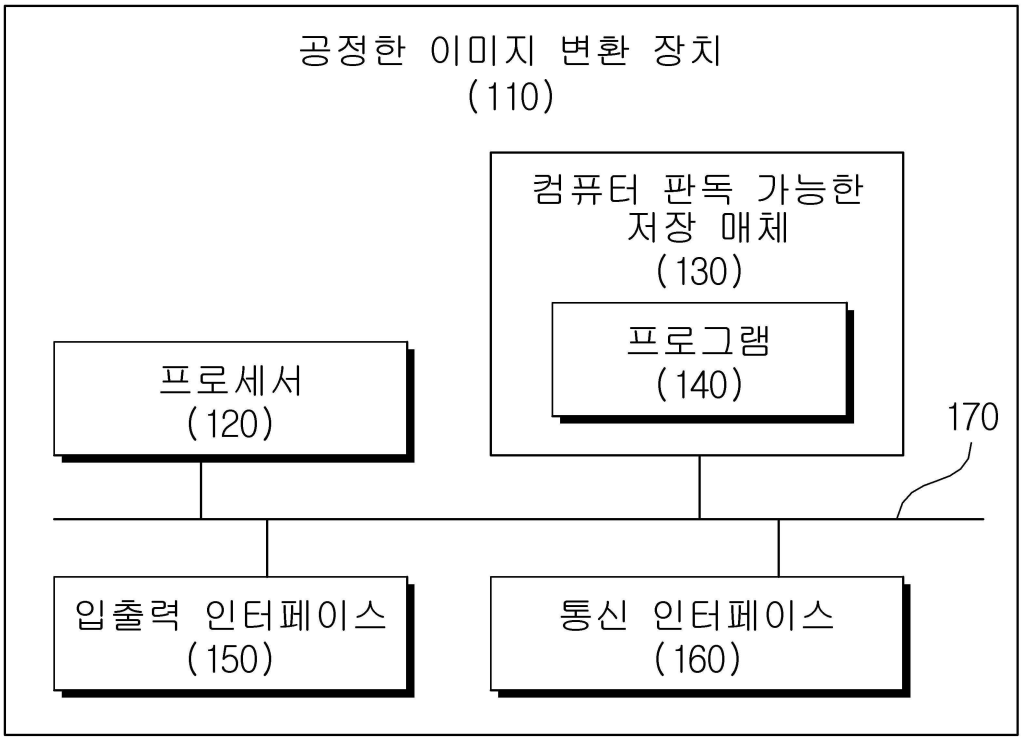
[0065] 본 실시예들에 따른 동작은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능한 매체에 기록될 수 있다. 컴퓨터 판독 가능한 매체는 실행을 위해 프로세서에 명령어를 제공하는 데 참여한 임의의 매체를 나타낸다. 컴퓨터 판독 가능한 매체는 프로그램 명령, 데이터 파일, 데이터 구조 또는 이들의 조합을 포함할 수 있다. 예를 들면, 자기 매체, 광기록 매체, 메모리 등이 있을 수 있다. 컴퓨터 프로그램은 네트워크로 연결된 컴퓨터 시스템 상에 분산되어 분산 방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수도 있다. 본 실시예를 구현하기 위한 기능적인(Functional) 프로그램, 코드, 및 코드 세그먼트들은 본

실시예가 속하는 기술분야의 프로그래머들에 의해 용이하게 추론될 수 있을 것이다.

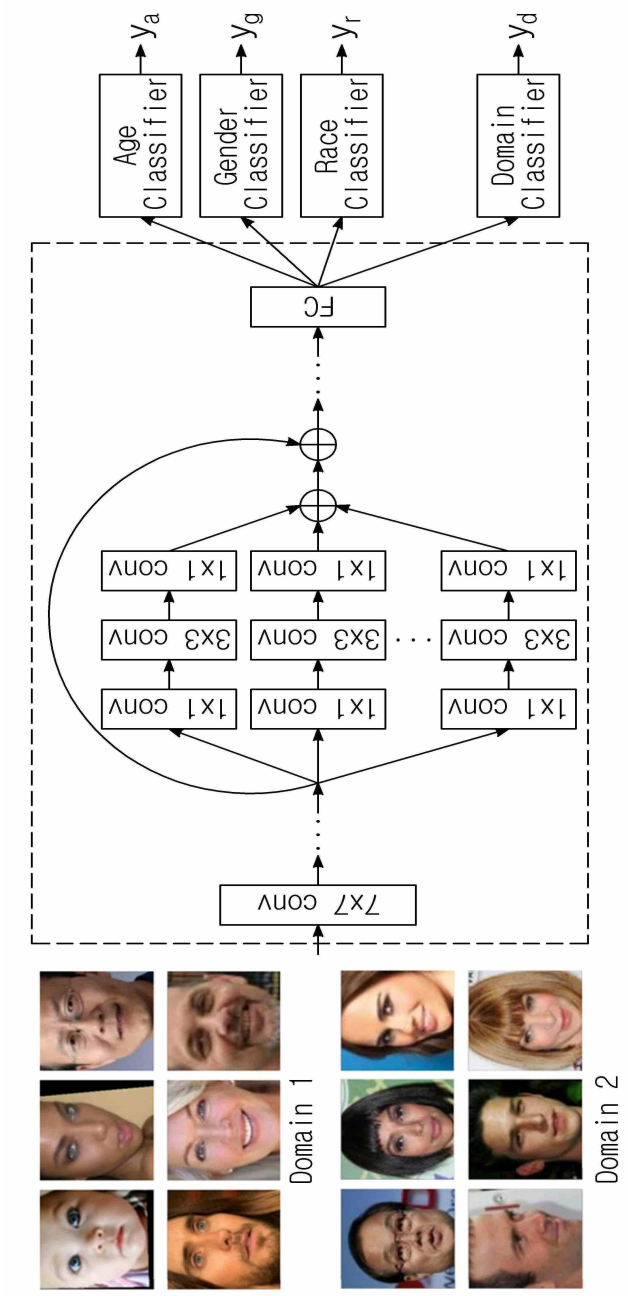
[0066] 본 실시예들은 본 실시예의 기술 사상을 설명하기 위한 것이고, 이러한 실시예에 의하여 본 실시예의 기술 사상의 범위가 한정되는 것은 아니다. 본 실시예의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 실시예의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

도면

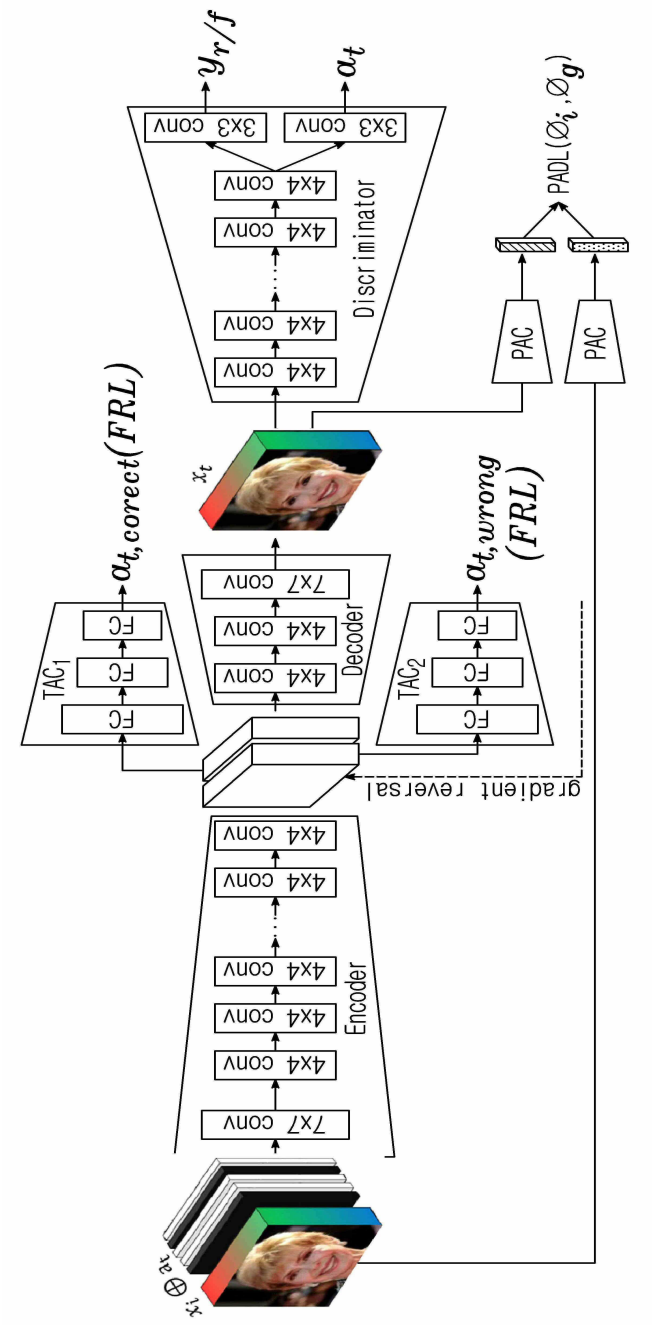
도면1



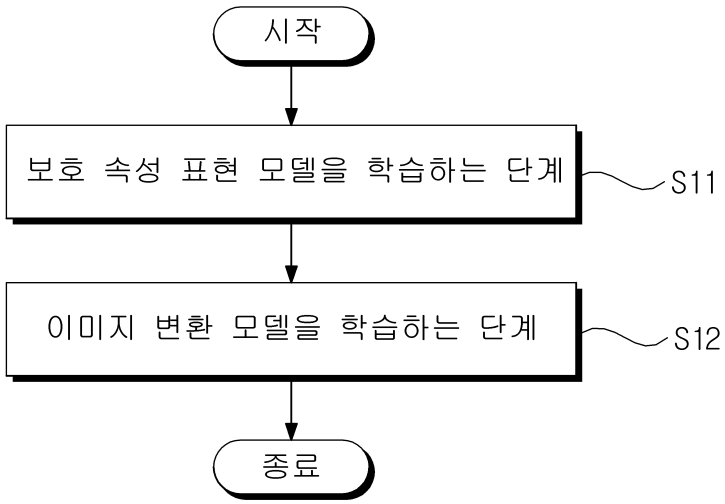
도면2


y_a
y_g
y_r
y_d
Age Classifier
Gender Classifier
Race Classifier
Domain Classifier
FC
1x1 conv
3x3 conv
7x7 conv
Domain 1
Domain 2

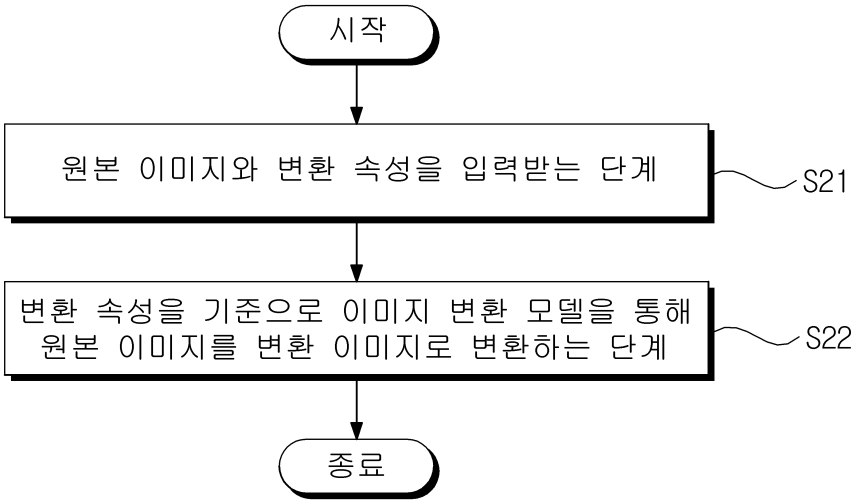
도면3


$y_{r/f}$
a_t
3x3 conv
3x3 conv
4x4 conv
4x4 conv
4x4 conv
4x4 conv
Discriminator
PADL($\emptyset_i$,$\emptyset_g$)
PAC
PAC
x_t
$a_{t,correct}(FRL)$
$a_{t,wrong}$ (FRL)
7x7 conv
4x4 conv
4x4 conv
Decoder
FC
FC
FC
TAC_1
FC
FC
FC
TAC_2
gradient reversal
4x4 conv
4x4 conv
4x4 conv
4x4 conv
4x4 conv
Encoder
7x7 conv
$x_t \oplus a_t$

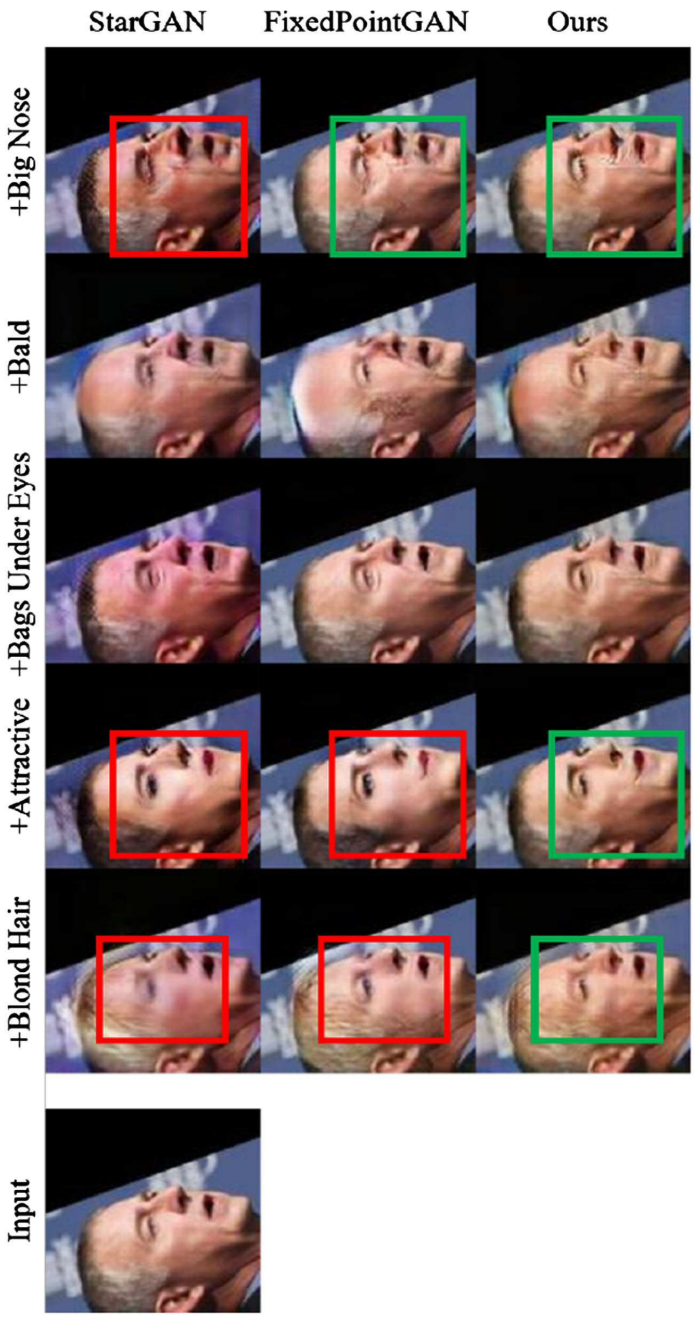
도면4


시작
보호 속성 표현 모델을 학습하는 단계
S11
이미지 변환 모델을 학습하는 단계
S12
종료

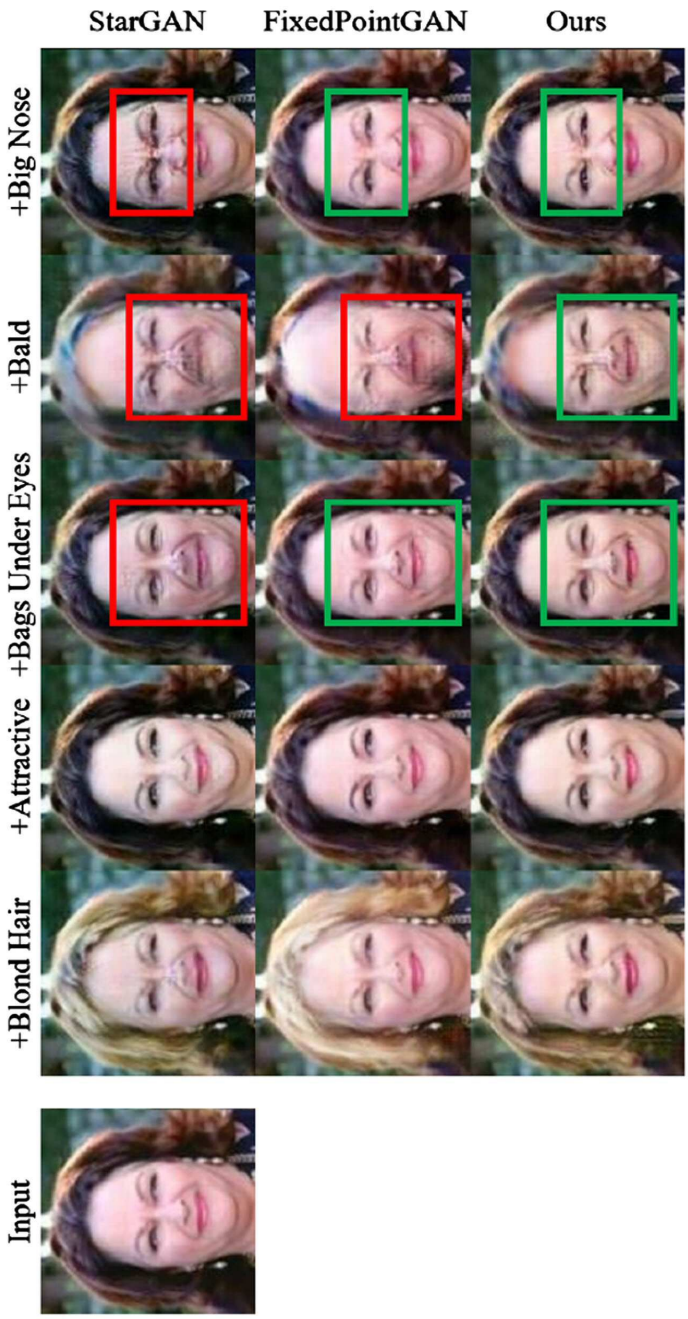
도면5


시작
원본 이미지와 변환 속성을 입력받는 단계
S21
변환 속성을 기준으로 이미지 변환 모델을 통해
원본 이미지를 변환 이미지로 변환하는 단계
S22
종료

도면6


StarGAN
FixedPointGAN
Ours
+Big Nose
+Bald
+Bags Under Eyes
+Attractive
+Blond Hair
Input

도면7


StarGAN
FixedPointGAN
Ours
+Big Nose
+Bald
+Bags Under Eyes
+Attractive
+Blond Hair
Input