



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2024년06월27일
(11) 등록번호 10-2678473
(24) 등록일자 2024년06월21일

(51) 국제특허분류(Int. Cl.)
G06T 13/40 (2011.01) G06N 3/08 (2023.01)
G06T 11/40 (2006.01) G06T 3/40 (2024.01)
(52) CPC특허분류
G06T 13/40 (2013.01)
G06N 3/08 (2023.01)
(21) 출원번호 10-2022-0021764
(22) 출원일자 2022년02월18일
심사청구일자 2022년02월18일
(65) 공개번호 10-2023-0100533
(43) 공개일자 2023년07월05일
(30) 우선권주장
1020210189831 2021년12월28일 대한민국(KR)
(56) 선행기술조사문헌
KR1020210083276 A*
Zheng, Ziqiang, et al. "Representation-guided generative adversarial network for unpaired photo-to-caricature translation." Computers & Electrical Engineering 90 (2021): 106999. (2021.01.28.)*
KR1020200025062 A
KR102266165 B1
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
포항공과대학교 산학협력단
경상북도 포항시 남구 청암로 77 (지곡동)
(72) 발명자
이승용
서울특별시 마포구 서강대길 40, 104동 1902호
정유철
인천광역시 계양구 경명대로1114번길 6, 17동 207호
(74) 대리인
특허법인이상

전체 청구항 수 : 총 12 항

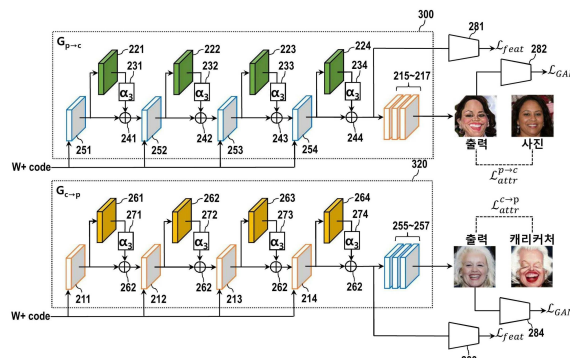
심사관 : 전세운

(54) 발명의 명칭 자동 캐리커처 생성 방법 및 장치

(57) 요약

세밀하고 사실적으로 얼굴 과장을 표현할 수 있고, 학습에 필요한 노동력과 비용을 크게 절감할 수 있는 캐리커처 생성 방법 및 장치를 제공한다. 예시적 실시예의 일 측면에 따른 캐리커처 생성 방법은 적대적 생성 신경망으로 생성된 형태 특징맵을 컨볼루션 신경망을 포함하는 얼굴 형태 과장 블록에 인가하여 캐리커처 형태 특징맵을 생성하는 단계; 상기 캐리커처 형태 특징맵을 상기 적대적 생성 네트워크에 인가하여, 상기 캐리커처 형태 특징맵에 따라 상기 형태 특징 맵을 변형시키는 단계; 및 변형된 형태 특징 맵에 따라 캐리커처 이미지를 생성하는 단계를 포함한다.

대표도



(52) CPC특허분류

G06T 11/40 (2013.01)
 G06T 3/4046 (2024.01)
 G06T 2207/20081 (2013.01)
 G06T 2207/20084 (2013.01)

장원중

경상북도 포항시 남구 효성로 11, 303호

(72) 발명자

주광진

전라남도 여수시 돌산읍 강남동로 46-7, 304동 201호

이 발명을 지원한 국가연구개발사업

과제고유번호 1711126500
 과제번호 2015-0-00174-007
 부처명 과학기술정보통신부
 과제관리(전문)기관명 정보통신기획평가원
 연구사업명 SW컴퓨팅산업원천기술개발
 연구과제명 (SW 스타랩) 빅 비주얼 데이터 기반의 고품질 사진 메이크업 SW 개발
 기여율 60/100
 과제수행기관명 포항공과대학교 산학협력단
 연구기간 2021.01.01 ~ 2021.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호 1711135289
 과제번호 2021-0-01696-001
 부처명 과학기술정보통신부
 과제관리(전문)기관명 정보통신기획평가원
 연구사업명 글로벌 핵심인재 양성 지원
 연구과제명 마이크로소프트연구소아시아와 공동연구협력 프로젝트
 기여율 30/100
 과제수행기관명 한국마이크로소프트(유)
 연구기간 2021.05.01 ~ 2022.04.30

이 발명을 지원한 국가연구개발사업

과제고유번호 1711125943
 과제번호 2019-0-01906-003
 부처명 과학기술정보통신부
 과제관리(전문)기관명 정보통신기획평가원
 연구사업명 정보통신방송혁신인재양성(R&D)
 연구과제명 인공지능대학원지원(포항공과대학교)
 기여율 10/100
 과제수행기관명 포항공과대학교 산학협력단
 연구기간 2021.01.01 ~ 2021.12.31

공지예외적용 : 있음

명세서

청구범위

청구항 1

적대적 생성 신경망으로 생성된 형태 특징맵을 컨볼루션 신경망을 포함하는 얼굴 형태 과장 블록에 인가하여 캐리커처 형태 특징맵을 생성하는 단계;

상기 캐리커처 형태 특징맵을 상기 적대적 생성 신경망에 인가하여, 상기 캐리커처 형태 특징맵에 따라 상기 형태 특징 맵을 변형시키는 단계; 및

변형된 형태 특징 맵에 따라 캐리커처 이미지를 생성하는 단계;

를 포함하고,

상기 적대적 생성 신경망의 일부인 복수의 제1 레이어들은 캐리커처로부터 이미지를 복원하는 기능이 사전 학습된 사진 복원 네트워크의 복수의 제1 레이어들로부터 형성되고,

상기 적대적 생성 신경망의 또 다른 일부인 복수의 제2 레이어들은 이미지로부터 캐리커처를 생성하는 기능이 사전 학습된 캐리커처 생성 네트워크의 복수의 제2 레이어들로부터 형성되고,

상기 캐리커처 형태 특징맵을 생성하는 단계 및 상기 형태 특징 맵을 변형시키는 단계는, 상기 얼굴 형태 과장 블록 및 상기 적대적 생성 신경망의 상기 복수의 제1 레이어들 간의 상호 작용에 의하여 수행되고,

상기 캐리커처 이미지를 생성하는 단계는, 상기 적대적 생성 신경망의 상기 복수의 제2 레이어들에 의하여 수행되며,

상기 캐리커처 형태 특징맵을 생성하는 단계는, 상기 사전 학습된 사진 복원 네트워크의 상기 복수의 제1 레이어들 각각에 대하여 하나씩 결합하는 얼굴 형태 과장 블록에 의하여 수행되는,

캐리커처 생성 방법.

청구항 2

삭제

청구항 3

삭제

청구항 4

청구항 1에 있어서, 상기 얼굴 형태 과장 블록의 입력과 출력의 크기는 서로 동일한, 캐리커처 생성 방법.

청구항 5

청구항 4에 있어서,

상기 캐리커처 형태 특징맵을 생성하기 이전에, 상기 적대적 생성 신경망에 대하여 기계학습을 시키는 단계;를 더 포함하며,

상기 적대적 생성 신경망에 대하여 기계학습을 시키는 단계가

이미지를 구별하는 제1 판별기에 의해 적대적 이미지 손실을 결정하고, 특징맵을 구별하는 제2 판별기에 의하여 적대적 특징맵 손실을 결정하는 단계;

를 더 포함하여, 상기 적대적 이미지 손실과 적대적 특징맵 손실을 토대로 상기 기계학습을 시키는 캐리커처 생성 방법.

청구항 6

삭제

청구항 7

청구항 5에 있어서, 상기 적대적 생성 신경망에 대하여 기계학습을 시키는 단계가

상기 형태 특징맵과, 상기 얼굴 형태 과장 블록으로 변환하고 얼굴 형태 복원 블록으로 복원한 형태 특징맵 사이의 차이를 최소화하도록 상기 기계학습을 수행하는 단계;

를 포함하는 캐리커처 생성 방법.

청구항 8

청구항 5에 있어서,

상기 형태 특징맵을 생성하는 기준이 된 입력 이미지의 속성들과 캐리커처의 속성들 간의 이진 교차 엔트로피 손실을 계산하는 단계;

를 더 포함하고, 상기 이진 교차 엔트로피 손실을 추가적으로 고려하여 상기 기계학습을 시키는 캐리커처 생성 방법.

청구항 9

청구항 1에 있어서, 상기 캐리커처 형태 특징맵을 생성하는 단계가

이용자로부터 상기 캐리커처의 변형 정도에 대한 선택 정보를 받아들이는 단계; 및

상기 선택 정보에 따라서 상기 얼굴 형태 과장 블록에서 상기 캐리커처 형태 특징맵의 가중치를 변경하는 단계;

를 포함하는 캐리커처 생성 방법.

청구항 10

신경망을 토대로 캐리커처를 생성하는 장치로서,

프로그램 명령들을 저장하는 메모리와; 상기 메모리에 접속되고 상기 메모리에 저장된 상기 프로그램 명령들을 실행하는 프로세서;를 구비하며,

상기 프로세서와 상기 메모리는 적대적 생성 신경망과 전자적으로 통신하고,

상기 프로그램 명령들은 상기 프로세서에 의해 실행될 때 상기 프로세서로 하여금:

상기 적대적 생성 신경망으로 생성된 형태 특징맵을 컨볼루션 신경망을 포함하는 얼굴 형태 과장 블록에 인가하여 캐리커처 형태 특징맵을 생성하고;

상기 캐리커처 형태 특징맵을 상기 적대적 생성 신경망에 인가하여, 상기 캐리커처 형태 특징맵에 따라 상기 형태 특징 맵을 변형시키고;

변형된 형태 특징 맵에 따라 캐리커처 이미지를 생성하되,

상기 적대적 생성 신경망의 일부인 복수의 제1 레이어들은 캐리커처로부터 이미지를 복원하는 기능이 사전 학습된 사진 복원 네트워크의 복수의 제1 레이어들로부터 형성되고,

상기 적대적 생성 신경망의 또 다른 일부인 복수의 제2 레이어들은 이미지로부터 캐리커처를 생성하는 기능이 사전 학습된 캐리커처 생성 네트워크의 복수의 제2 레이어들로부터 형성되고,

상기 얼굴 형태 과장 블록 및 상기 적대적 생성 신경망의 상기 복수의 제1 레이어들 간의 상호 작용에 의하여 상기 캐리커처 형태 특징맵이 생성되고 상기 형태 특징 맵이 변형되며,

상기 적대적 생성 신경망의 상기 복수의 제2 레이어들에 의하여 상기 캐리커처 이미지가 생성되고,

상기 사전 학습된 사진 복원 네트워크의 상기 복수의 제1 레이어들 각각에 대하여 하나씩 결합하는 얼굴 형태 과장 블록에 의하여 상기 캐리커처 형태 특징맵이 생성되는,

캐리커처 생성 장치.

청구항 11

삭제

청구항 12

삭제

청구항 13

청구항 10에 있어서, 상기 얼굴 형태 과장 블록의 입력과 출력의 크기는 서로 동일한, 캐리커처 생성 장치.

청구항 14

청구항 13에 있어서, 상기 프로그램 명령들은 상기 프로세서에 의해 실행될 때 상기 프로세서로 하여금:
 상기 캐리커처 형태 특징맵을 생성하기 이전에, 상기 적대적 생성 신경망에 대하여 기계학습을 수행하게 하며,
 상기 적대적 생성 신경망에 대하여 기계학습을 수행하게 하는 프로그램 명령들은

이미지를 구별하는 제1 판별기에 의해 적대적 이미지 손실을 결정하고, 특징맵을 구별하는 제2 판별기에 의하여 적대적 특징맵 손실을 결정하는 동작을 더 수행하게 하여, 상기 적대적 이미지 손실과 적대적 특징맵 손실을 토대로 상기 기계학습을 수행하는 캐리커처 생성 장치.

청구항 15

삭제

청구항 16

청구항 14 있어서, 상기 적대적 생성 신경망에 대하여 기계학습을 수행하게 하는 프로그램 명령들은
 상기 형태 특징맵과, 상기 얼굴 형태 과장 블록으로 변환하고 얼굴 형태 복원 블록으로 복원한 형태 특징맵 사이의 차이를 최소화하도록 상기 기계학습을 수행하게 하는 캐리커처 생성 장치.

청구항 17

청구항 14에 있어서, 상기 프로그램 명령들은 상기 프로세서에 의해 실행될 때 상기 프로세서로 하여금:
 상기 형태 특징맵을 생성하는 기준이 된 입력 이미지의 속성들과 캐리커처의 속성들 간의 이진 교차 엔트로피 손실을 계산하도록 하고;
 를 더 포함하고, 상기 이진 교차 엔트로피 손실을 추가적으로 고려하여 상기 기계학습을 수행하게 하는 캐리커처 생성 장치.

청구항 18

청구항 10에 있어서, 상기 캐리커처 형태 특징맵을 생성하는 프로그램 명령들은 상기 프로세서로 하여금:
 이용자로부터 상기 캐리커처의 변형 정도에 대한 선택 정보를 받아들이고;
 상기 선택 정보에 따라서 상기 얼굴 형태 과장 블록에서 상기 캐리커처 형태 특징맵의 가중치를 변경하도록 하는 캐리커처 생성 장치.

발명의 설명

기술 분야

본 발명은 캐리커처 생성 방법 및 장치에 관한 것으로서, 보다 상세하게는 머신 러닝 기반으로 학습될 수 있고 전문화된 네트워크 아키텍처를 구비하는 신경망을 사용하여 자동으로 캐리커처를 생성하는 방법 장치에 관한 것이다.

배경 기술

[0001]

- [0002] 캐리커처는 예술가가 대상 인물의 가장 눈에 띄는 특성을 과장하고 다른 특성을 많이 단순화하는 초상화 유형이다. 캐리커처를 그리기 위해서는 만화 스타일 그림 그리기뿐만 아니라 가장 두드러진 얼굴 특징을 포착하고 과장하는 방법을 배워야 하므로 오랜 시간의 개발 훈련이 필요하다. 기술을 연마한 후에도, 예술가가 캐리커처 하나를 그리는 데에는 몇 시간 또는 며칠이 걸릴 수 있다. 컴퓨터를 사용하여 사진에서 캐리커처로의 자동 변환하는 방식은 교육과 제작에 대한 부담을 줄일 수 있고 일반 대중이 캐리커처를 쉽게 접할 수 있게 해준다.
- [0003] 캐리커처 생성에 대한 초기 연구 단계에서는, 얼굴의 기하학적 특징을 캡처하고 변형하는 방법이 제시된 바 있지만, 인상적인 결과를 가져오기 위해서는 사용자와의 상호작용에 의존해야 했다. 몇 가지 자동 캐리커처 생성 방법이 제안되었지만, 예술적 스타일은 미리 정의된 규칙에 한정되었다.
- [0004] 이미지-이미지 변환을 위한 최근의 딥러닝 알고리즘은 훈련 예제가 주어지면 입력 도메인에서 출력 도메인으로의 매핑을 찾을 수 있지만, 사진-캐리커처 자동 변환의 지도 학습에 사용할 수 있는 대응관계가 좋은 사진-캐리커처 쌍을 확보하기가 어렵다. 특히, 사람의 사진에 대응하는 캐리커처들 중 상당수는 포즈, 표정, 및 과장 스타일이 다를 수 있어서, 사진과 캐리커처의 대응 관계가 정형적이지 않다. 따라서 지도 학습을 위해서는 캐리커처 전문가가 수작업으로 만든 캐리커처를 필요로 할 수 있으며, 이러한 캐리커처를 다량 확보하기 위해서는 큰 비용지출이 요구된다.
- [0005] 최근의 자동 캐리커처 알고리즘 중 상당수는 기하학적 변형을 통해 사진을 캐리커처로 변환하는 전문화된 네트워크 아키텍처와 연관되어 있다. 이러한 방법들은 얼굴 사진을 캐리커처처럼 보이게 하기 위해 이미지 스타일을 전이(Image style transfer)한 후, 2D 이미지 와핑(warping)으로 얼굴을 과장하여 캐리커처를 생성한다. 하지만 이러한 방법들은 제어점의 개수의 한계로 세밀한 얼굴 과장을 표현하지 못한다. 부족한 제어점을 보충하기 위하여 제어점들을 보간하는 경우에는, 실제 캐리커처에 있는 세밀한 얼굴 형태가 변형되는 손실을 초래할 수 있다. 이를 해결하기 위해 밀집한 와핑 필드(dense warping field)를 이용해 얼굴을 과장하는 기법도 제안되었지만, 이 기법은 실제 캐리커처 아티스트가 만든 와핑 필드를 학습 데이터로 요구한다. 따라서 와핑 필드에 대한 학습 데이터를 많이 모으기 위해서는 상당한 노동력과 비용을 요구하기 때문에 실제 사용에 제약이 있다.

발명의 내용

해결하려는 과제

- [0006] 본 발명의 예시적 실시예는 이미지 공간에서 기하적인 와핑을 수행하는 것보다 더 세밀하고 사실적으로 얼굴 과장을 표현할 수 있고, 용이하게 캐리커처를 생성할 수 있으며, 학습에 필요한 노동력과 비용을 크게 절감할 수 있는 캐리커처 생성 방법을 제공한다.
- [0007] 본 발명의 예시적 실시예는 이미지 공간에서 기하적인 와핑을 수행하는 것보다 더 세밀하고 사실적으로 얼굴 과장을 표현할 수 있고, 용이하게 캐리커처를 생성할 수 있으며, 학습에 필요한 노동력과 비용을 크게 절감할 수 있는 캐리커처 생성 장치를 제공한다

과제의 해결 수단

- [0008] 예시적 실시예의 일 측면에 따른 캐리커처 생성 방법은 적대적 생성 신경망으로 생성된 형태 특징맵을 컨볼루션 신경망을 포함하는 얼굴 형태 과장 블록에 인가하여 캐리커처 형태 특징맵을 생성하는 단계; 상기 캐리커처 형태 특징맵을 상기 적대적 생성 네트워크에 인가하여, 상기 캐리커처 형태 특징맵에 따라 상기 형태 특징 맵을 변형시키는 단계; 및 변형된 형태 특징 맵에 따라 캐리커처 이미지를 생성하는 단계;를 포함한다.
- [0009] 상기 캐리커처 형태 특징맵을 생성하는 단계는 직렬 연결된 복수의 캐리커처 생성 학습 레이어들을 마련하고 학습용 캐리커처 이미지들로 사전학습시키고, 직렬 연결된 복수의 사진 복원 학습 레이어들을 마련하고 학습용 사진들로 사전학습시키는 단계; 및 상기 복수의 캐리커처 생성 학습 레이어들 중에서 하나 이상의 최저 해상도 캐리커처 생성 학습 레이어를, 상기 복수의 사진 복원 학습 레이어들 중에서 대응 위치에 있는 하나 이상의 사진 복원 학습 레이어로 교체하여 상기 적대적 생성 신경망을 수정하는 단계;를 포함할 수 있다.
- [0010] 상기 적대적 생성 신경망을 수정하는 단계는 상기 하나 이상의 사진 복원 학습 레이어들에 하나 이상의 얼굴 형태 과장 블록을 연결시키는 단계;를 포함할 수 있다.
- [0011] 상기 하나 이상의 사진 복원 학습 레이어들 각각에 입력과 출력의 크기가 동일한 얼굴 형태 과장 블록이 한 개씩 연결될 수 있다.

- [0012] 캐리커처 생성 방법은, 상기 캐리커처 형태 특징맵을 생성하기 이전에, 상기 적대적 생성 신경망에 대하여 기계 학습을 시키는 단계;를 더 포함할 수 있다. 상기 적대적 생성 신경망에 대하여 기계 학습을 시키는 단계는 이미지를 구별하는 제1 판별기에 의해 적대적 이미지 손실을 결정하고, 특징맵을 구별하는 제2 판별기에 의하여 적대적 특징맵 손실을 결정하는 단계;를 더 포함하여, 상기 적대적 이미지 손실과 적대적 특징맵 손실을 토대로 상기 기계 학습을 시킬 수 있다.
- [0013] 상기 캐리커처 형태 특징맵을 생성하는 단계는 상기 복수의 사진 복원 학습 레이어들 중에서 하나 이상의 최저 해상도 사진 복원 학습 레이어를, 상기 복수의 캐리커처 생성 학습 레이어들의 중에서 대응 위치에 있는 하나 이상의 대응 위치에 있는 하나 이상의 캐리커처 생성 학습 레이어로 교체하고, 얼굴 형태 복원 블록을 부가하여, 적대적 복원 신경망을 생성하는 단계;를 더 포함할 수 있다.
- [0014] 상기 적대적 생성 신경망에 대하여 기계 학습을 시키는 단계에서는, 상기 형태 특징맵과, 상기 얼굴 형태 과장 블록으로 변환하고 얼굴 형태 복원 블록으로 복원한 형태 특징맵 사이의 차이를 최소화하도록 상기 기계 학습을 수행할 수 있다.
- [0015] 캐리커처 생성 방법은 상기 형태 특징맵을 생성하는 기준이 된 입력 이미지의 속성들과 상기 캐리커처의 속성들 간의 이진 교차 엔트로피 손실을 계산하는 단계를 더 포함할 수 있다. 이 경우, 상기 이진 교차 엔트로피 손실을 추가적으로 고려하여 상기 기계 학습을 수행할 수 있다.
- [0016] 상기 캐리커처 형태 특징맵을 생성하는 단계는 이용자로부터 상기 캐리커처의 변형 정도에 대한 선택 정보를 받아들이는 단계; 및 상기 선택 정보에 따라서 상기 얼굴 형태 과장 블록에서 상기 캐리커처 형태 특징맵의 가중치를 변경하는 단계;를 포함할 수 있다.
- [0017] 예시적 실시예의 일 측면에 따르면, 신경망을 토대로 캐리커처를 생성하는 장치가 제공된다. 캐리커처 생성 장치는 프로그램 명령들을 저장하는 메모리와; 상기 메모리에 접속되고 상기 메모리에 저장된 상기 프로그램 명령들을 실행하는 프로세서;를 구비한다. 상기 프로그램 명령들은 상기 프로세서에 의해 실행될 때 상기 프로세서로 하여금: 적대적 생성 신경망으로 생성된 형태 특징맵을 컨볼루션 신경망을 포함하는 얼굴 형태 과장 블록에 인가하여 캐리커처 형태 특징맵을 생성하고; 상기 캐리커처 형태 특징맵을 상기 적대적 생성 네트워크에 인가하여, 상기 캐리커처 형태 특징맵에 따라 상기 형태 특징 맵을 변형시키고; 변형된 형태 특징 맵에 따라 캐리커처 이미지를 생성하도록 한다.

발명의 효과

- [0018] 본 발명의 일 실시예에 따르면, 얼굴 형태 과장 블록이 형태를 나타내는 특징 공간에서 얼굴 형태 특징 맵(feature map)을 학습 가능한 컨볼루션 신경망을 이용하여 캐리커처 형태 특징 맵으로 변환하는 방식으로 얼굴 과장을 수행한다. 이를 통해 이미지 공간에서 기하적인 왜곡을 수행하는 것보다 더 세밀하고 사실적으로 얼굴 과장을 표현할 수 있다. 특히, 다중 스케일 과장 조절을 가능하게 해주기 때문에, 이용자 취향에 맞게 그리고 용이하게 캐리커처를 생성할 수 있다.
- [0019] 본 발명에 의한 캐리커처 생성 장치는 왜핑 필드 데이터 없이 단순히 캐리커처 사진을 학습 데이터로 이용하여 학습할 수 있으며, 따라서 학습에 필요한 노동력과 비용이 현저히 절감될 수 있다.

도면의 간단한 설명

- [0020] 도 1은 본 발명의 일 실시예에 따른 캐리커처 생성 장치의 네트워크 환경을 보여주는 블록도이다.
- 도 2는 본 발명의 일 실시예에 따른 캐리커처 생성 장치의 개략적인 블록도이다.
- 도 3은 얼굴 형태 과장 블록의 학습과 추론에 의한 동작 과정에서의 컴포넌트 배치를 보여주는 블록도이다.
- 도 4는 얼굴 형태 과장 블록의 학습을 위한 목적 함수로 사용되는 여러 손실들을 보여준다.
- 도 5는 목적 함수 중 하나인 특징맵 순환 일관성 손실을 결정하기 위한 생성 학습 레이어들과 얼굴 형태 과장 블록들의 배치 예를 보여주는 도면이다.
- 도 6은 도 2에 도시된 캐리커처 생성 장치의 일 실시예의 물리적 구성을 보여주는 블록도이다.
- 도 7은 본 발명의 일 실시예에 따른 캐리커처 생성 방법을 보여주는 흐름도이다.

도 8은 레이어 스와핑의 경계에 따라 달라지는 출력 이미지들이 달라지는 예를 보여준다.

도 9는 이미지-이미지 변환 및 캐리커처 생성을 위한 다른 최신 기법들과의 비교 결과를 보여준다.

도 10은 사용자 연구 결과를 보여준다.

도 11은 얼굴 형태 과장에 대한 다중 스케일 조절 결과를 보여준다.

발명을 실시하기 위한 구체적인 내용

- [0021] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세한 설명에 상세하게 설명하고자 한다. 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.
- [0022] 제1, 제2, 등의 서수가 다양한 구성요소들을 설명하는 데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다. "및/또는"이라는 용어는 복수의 관련된 기재된 항목들의 조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.
- [0023] 본 출원의 실시예들에서, "A 및 B 중에서 적어도 하나"는 "A 또는 B 중에서 적어도 하나" 또는 "A 및 B 중 하나 이상의 조합들 중에서 적어도 하나"를 의미할 수 있다. 또한, 본 출원의 실시예들에서, "A 및 B 중에서 하나 이상"은 "A 또는 B 중에서 하나 이상" 또는 "A 및 B 중 하나 이상의 조합들 중에서 하나 이상"을 의미할 수 있다.
- [0024] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는, 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다.
- [0025] 본 출원에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 출원에서, "포함하다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0026] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥 상 가지는 의미와 일치하는 의미를 가지는 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.
- [0027] 이하, 본 발명에 따른 바람직한 실시예를 첨부된 도면을 참조하여 상세하게 설명한다. 본 발명을 설명함에 있어 전체적인 이해를 용이하게 하기 위하여 도면상의 동일한 구성요소에 대해서는 동일한 참조부호를 사용하고 동일한 구성요소에 대해서 중복된 설명은 생략한다.
- [0028] 도 1은 본 발명의 일 실시예에 따른 캐리커처 생성 장치의 네트워크 환경을 보여주는 블록도이다. 예시적인 실시예에 있어서, 캐리커처 생성 장치는 인터넷을 통해 캐리커처 생성 서비스를 제공하는 서버 시스템을 기초로 구현될 수 있다. 상기 서버 시스템(10)은 인터넷을 통해서 접근하는 클라이언트(20)로부터 서비스를 요청할 수 있으며, 상기 요청에 응답하여 클라이언트(20)가 제공하는 입력 사진 이미지 또는 클라이언트(20)가 특정하는 인물에 대한 캐리커처를 생성하여 해당 클라이언트(20)에 제공할 수 있다. 클라이언트(20)는, 입력 사진 이미지를 제공할 때 또는 생성된 캐리커처를 받아본 후에, 캐리커처에서의 얼굴 형태 변형 정도를 지정할 수 있으며, 서버 시스템(10)은 클라이언트(20)가 지정한 얼굴 형태 변형 정도에 맞추어 캐리커처를 생성할 수 있다. 한편, 본 발명에 의한 캐리커처 생성 장치가 반드시 인터넷 서버를 토대로 구현되어야 하는 것은 아니며, 스탠드-얼론 컴퓨터 또는 소수의 폐쇄 사용자 그룹만이 접근가능한 서버에서 구현될 수도 있다.
- [0029] 도 2는 본 발명의 일 실시예에 따른 캐리커처 생성 장치의 블록도이다. 캐리커처 생성 장치는 인코더(100), 최적화부(120), 및 생성 네트워크(200)를 포함한다.

- [0030] 인코더(100)는 입력 이미지를 받아들이고 인코딩하여 입력 이미지를 나타내는 잠재 코드(latent code, W+ 코드)를 발생할 수 있다. 인코더(100)가 수행하는 인코딩은 다복수의 컨볼루션 연산을 통해 수행하는 특징맵 추출을 포함한 압축 작업일 수 있다. 다른 실시예에서는, 사진에서 잠재된 임베딩(photo-to-latent embedding)을 추출하여, 사진으로부터 잠재 코드(W+ 코드)를 획득할 수도 있다. 그렇지만, 본 발명이 이에 한정되는 것은 아니며, 다른 실시예에 있어서는, 인코딩 작업이 인코더(100)를 학습시키고, 주어진 입력 이미지를 그에 상응하는 잠재 코드(W+ 코드)로 맵핑하는 작업이 될 수 있다.
- [0031] 최적화부(120)는 가우시안 분포 즉, 정규분포에 따라 발생된 랜덤 넘버를 선택하고, 주어진 입력 이미지와의 차이가 최소화되도록 최적화하여 잠재 코드(W+ 코드)를 획득할 수 있다. 후술하는 바와 같이, 예시적인 실시예에 따르면, 생성 네트워크(200)의 캐리커처 생성 네트워크(210)와 사진 복원 네트워크(250)를 학습시킬 때에는 최적화부(120)가 사용될 수 있다. 한편, 예시적인 실시예에 있어서, 생성 네트워크(200)의 얼굴 형태 과장 블록들(221-224, 251-254)을 학습시킬 때 또는 캐리커처 생성 장치를 실제로 추론을 통해 동작시킬 때에는 인코더(100)가 사용될 수 있다. 이와 같이 인코더(100) 및/또는 최적화부(120)가 잠재 코드(W+ 코드)를 발생하여 생성 네트워크(200)에 공급하지만, 이들 각각이 잠재 코드(W+ 코드)를 공급하는 시점은 위와 다를 수 있다. 또한, 다른 실시예에 있어서는, 인코더(100)가 별도의 잠재 코드(W+ 코드)가 아닌 특징맵 또는 압축된 사진이나 캐리커처 이미지를 생성 네트워크(200)에 공급할 수도 있다.
- [0032] 생성 네트워크(200)은 잠재 코드(W+ 코드)를 받아들이고, 다중 스케일 특징맵과 궁극적으로 캐리커처 출력 이미지를 생성할 수 있다. 또한, 생성 네트워크(200)은 잠재 코드(W+ 코드)가 캐리커처에 관한 것인 경우, 잠재 코드(W+ 코드)로부터 상기 캐리커처에 상응하는 사진을 복원할 수 있다. 예시적인 실시예에 있어서, 생성 네트워크(200)은 캐리커처 생성 네트워크(210), 사진 복원 네트워크(250), 및 복수의 얼굴 형태 과장 블록들(221-224, 251-254)을 포함할 수 있다. 그밖에, 생성 네트워크(200)은 후술하는 바와 같이 손실 함수를 산출하는 복수의 판별기를 구비할 수 있다.
- [0033] 캐리커처 생성 네트워크(210)은 사진 이미지에 대한 잠재 코드(W+ 코드)를 받아들이고, 상기 잠재 코드(W+ 코드)에 상응한 캐리커처를 생성하도록 학습될 수 있다. 캐리커처 생성 네트워크(210)은 복수의 캐리커처 생성 학습 레이어들(211-217)을 포함할 수 있다. 복수의 캐리커처 생성 학습 레이어들(211-217) 각각은 두 개의 컨볼루션 레이어들, 두 개의 활성화 레이어들, 업샘플링 레이어, 및 어파인 변환 레이어를 포함하도록 구성될 수 있으며, 입력된 특징맵의 크기를 가로와 세로 각각 2배로 확장하여 출력할 수 있다.
- [0034] 사진 복원 네트워크(250)은 캐리커처 이미지에 대한 잠재 코드(W+ 코드)를 받아들이고, 상기 잠재 코드(W+ 코드)에 상응한 사진을 복원하도록 학습될 수 있다. 사진 복원 네트워크(250)은 캐리커처 생성 네트워크(210)과 유사하게 구성되며, 복수의 사진 복원 학습 레이어들(261-267)을 포함할 수 있다. 복수의 사진 복원 학습 레이어들(261-267) 각각은 두 개의 컨볼루션 레이어들, 두 개의 활성화 레이어들, 업샘플링 레이어, 및 어파인 변환 레이어를 포함하도록 구성될 수 있으며, 입력된 특징맵의 크기를 가로와 세로 각각 2배로 확장하여 출력할 수 있다.
- [0035] 얼굴 형태 과장 블록들(221-224, 251-254)은 입력 이미지에서 얼굴 형태를 과장시킬 수 있다. 얼굴 형태 과장 블록들(221-224, 251-254) 각각은 두 개의 컨볼루션 레이어들과 두 개의 활성화 레이어를 포함하도록 구성될 수 있으며, 입력된 특징맵의 크기를 동일한 크기의 특징맵을 출력할 수 있다. 후술하는 바와 같이, 얼굴 형태 과장 블록은 적대적 생성 네트워크가 생성하는 특징 공간 중에서 형태를 다루는 특징 공간의 특징 맵을 입력으로 받아 변환함으로써 얼굴 과장을 성취한다
- [0036] 도 3은 얼굴 형태 과장 블록의 학습과 추론에 의한 동작 과정에서의 컴포넌트 배치를 보여주는 블록도이다.
- [0037] 도 2에 도시된 생성 네트워크(200)은 2단계로 학습이 이루어질 수 있다. 먼저, 캐리커처 생성 네트워크(210)의 캐리커처 생성 학습 레이어들(211-217)과 사진 복원 네트워크(250)의 사진 복원 학습 레이어들(261-267)은 도 2에 도시된 순서로 배열된 상태로 학습될 수 있다. 캐리커처 생성 학습 레이어들(211-217)과 사진 복원 학습 레이어들(261-267)의 학습이 완료되면, 캐리커처 생성 학습 레이어들(211-217)과 사진 복원 학습 레이어들(261-267)의 일부가 서로 스와핑 즉 맞교환된다. 이와 같이 맞교환된 상태에서 얼굴 형태 과장 블록들(221-224, 251-254)의 학습이 이루어지고, 캐리커처 생성 장치가 입력 이미지에 대하여 추론을 하면서 동작하게 된다.
- [0038] 먼저, 레이어 스와핑과 이에 따른 네트워크 아키텍처 변화에 대하여 살펴본다. 도 2에 도시된 생성 네트워크(200)은 두 개의 생성 네트워크 즉, 캐리커처 생성 네트워크(210)와 사진 복원 네트워크(250)를 포함한다. 사전훈련을 통해서, 사진 복원 네트워크(250)은 사진을 생성하도록 훈련되고, 캐리커처 생성 네트워크(210)은 사진으로부터 캐리

커처 이미지를 생성하도록 사진 복원 네트워크(250)으로부터 미세하게 튜닝된다. 사전 훈련 후에 두 네트워크(200, 250)의 레이어들은 고정되고 프로세스 중에 업데이트되지 않는다. 캐리커처 이미지를 생성하도록 훈련된 캐리커처 생성 네트워크(210)을 '캐리커처 StyleGAN'으로 칭하고, 일반 얼굴 사진을 복원하도록 훈련된 사진 복원 네트워크(250)을 '사진 StyleGAN'으로 칭할 수도 있다.

[0039] 이와 같이 캐리커처 생성 네트워크(210)과 사진 복원 네트워크(250)의 학습이 완료되고 레이어들의 커널이 고정된 상태에서, 두 네트워크(210, 250)의 일부 레이어들을 맞교환하여, 잠재 코드로부터 캐리커처 이미지 또는 사진 이미지를 생성하는 레이어-혼합 모델들을 구성한다.

[0040] 도 2에 도시된 캐리커처 생성 네트워크(210)과 사진 복원 네트워크(250)은 StyleGAN이라는 적대적 생성 네트워크와 유사한 구조를 가진다. 일반적으로 적대적 생성 네트워크(Generative Adversarial Network: GAN)은 '생성적 대립 신경망'으로도 일컬어지는 기계학습 방식의 하나로써, 생성모델과 판별모델을 포함한다. 생성모델은 확률 분포를 학습하고 확률 분포를 토대로 가짜 예제를 만들어 판별모델을 최대한 속일 수 있도록 학습되고, 판별모델은 서로 다른 집합을 구분하며 특히 생성모델이 제시하는 가짜 예제와 실제 예제를 최대한 정확하게 구분할 수 있도록 학습된다. 이와 같이 판별모델을 속일 수 있도록 생성모델을 훈련하는 방식을 대립적 프로세스라고 한다. 생성 모델과 판별모델이 경쟁하는 대립적 프로세스를 통해 실제와 가까운 이미지나 음성 등을 만들어내게 된다.

[0041] 이와 같은 적대적 생성 네트워크는 낮은 해상도의 특징 맵은 주로 형태를, 높은 해상도의 특징 맵은 주로 텍스처를 생성하는 특징을 지닌다. 즉, 적대적 생성 네트워크는 특징 공간을 주로 형태를 다루는 공간과 텍스처를 다루는 공간으로 분리하는 속성을 가진다. 이 이 특징을 감안하여, 실제로 캐리커처를 생성하는 네트워크에서 낮은 해상도의 특징맵을 생성하는 레이어는 일반 얼굴사진에 대해 사전학습된 레이어를 이용하여 현실감을 높이고, 높은 해상도의 특징맵을 생성하는 레이어는 캐리커처에 대해 사전학습된 레이어로 구성한다. 이와 같이 구성된 네트워크도 3에서의 제1 레이어-혼합 네트워크(300)이다. 이 제1 레이어-혼합 네트워크(300)에서, 얼굴 형태 과장 블록들(221-224)은 형태를 다루는 특징 공간의 특징맵을 입력으로 받아 변환함으로써 얼굴 과장을 성취하게 된다.

[0042] 구체적으로, 제1 레이어-혼합 네트워크(300)에서, 거친(즉, 저해상도) 레이어들(251-254)은 사진 StyleGAN(250)에서 가져온 것이고, 세밀한(즉, 고해상도) 레이어들(211-214)만이 캐리커처 StyleGAN(210)에 있던 것이다. 제1 레이어-혼합 네트워크(300)에서 캐리커처의 형태를 만드는데 주로 기여하는 저해상도 레이어들(251-254)은 얼굴 사진으로 사전학습되어 있고, 캐리커처의 텍스처를 만드는데 주로 기여하는 고해상도 레이어들(211-214)은 캐리커처로 사전학습되어 있기 때문에, 제1 레이어-혼합 네트워크(300)은 사진 상의 얼굴 형태에 캐리커처 텍스처를 지니는 캐리커처 이미지를 생성해낼 수 있다. 실제로, 레이어의 스와핑은 해당하는 레이어들의 파라미터와 가중치 내지 커널을 서로 바꾸어 적용함으로써 이루어질 수 있다.

[0043] 얼굴 형태 과장 블록들(221-224)은 낮은 해상도의 특징맵을 생성하는 레이어들(251-254)의 각 계층에 삽입되어 접속된다. 얼굴 형태 과장 블록들(221-224)은 형태를 주로 생성하는 낮은 해상도의 특징맵을 입력으로 받아 캐리커처 형태를 지니는 특징맵으로 변환함으로써 얼굴 과장을 수행하도록 학습된다. 얼굴 형태 과장 블록들(221-224)이 학습되는 동안에, 앞에서 언급한 바와 같이 레이어들(251-254)의 커널 내지 가중치는 고정된다. 입력 얼굴 형태를 잘 유지하며 과장할 수 있도록, 얼굴 형태 과장 블록들(221-224)은 레지듀얼 블록(residual block) 형태로 구성될 수 있다.

[0044] 이와 같이 얼굴 형태 과장 블록들(221-224)은 사진에서 캐리커처로의 변환을 위해 특별히 고안된 지도방식으로 훈련되며, 사실감 있고 신원을 보존하는 캐리커처 과장을 통해 특징맵 변조를 도입하게 된다. 모델이 매우 다양한 캐리커처 지오메트리를 생성하도록 GAN 손실로 훈련되기 때문에, 얼굴 형태 과장 블록들(221-224)은 특징 공간의 평균 얼굴 지오메트리로부터의 편차를 암시적으로 증가시키며 상기 편차가 입력의 특징을 뚜렷하게 반영하도록 하게 된다.

[0045] 레이어 스와핑에서는 거친 레이어들과 세밀한 레이어들의 경계를 잘 선택하는 것이 중요할 수 있다. 도 8은 레이어 스와핑의 경계에 따라 달라지는 출력 이미지들이 달라지는 예를 보여준다. 도면에서, p는 제1 레이어-혼합 네트워크(300)에 사용된 거친 사진 StyleGAN 레이어들의 수를 나타내고, c는 세밀한 캐리커처 StyleGAN 레이어들의 수를 나타낸다. 예를 들어, (p=4, c=3)은 처음 네 개의 저해상도 레이어를 사진 StyleGAN(250)에서 가져오고 마지막 세 개의 세밀한 레이어들은 캐리커처 StyleGAN(210)에 있던 것을 유지함으로 의미한다. 경계 선택이 달라지면, 입력 구조 보존과 디테일에 대한 스타일링도 달라진다. 구체적으로, p가 감소함에 따라 입력의 얼굴 구조가 변형되고 정체성이 손실된다. 경계를 선택할 때의 목표는 레이어-혼합 모델이 얼굴 구조 및 신원

측면에서 입력 사진에 충실한 캐리커처 이미지를 생성할 수 있게 하는 것이다. 경험적으로 입력에 있는 얼굴 형태 정보를 유지하면서 세부사항을 충분히 스타일링할 수 있는 레이어-혼합 모델로서 ($p=4$, $c=3$)을 선택하였다.

[0046] 앞에서 언급한 바와 같이, 레이어들의 복사된 매개변수들은 레이어 맞교환 후 더 이상 업데이트되지 않는다. 즉, 모델의 거친 레이어들(251-254)과 미세한 레이어들(215-217)이 각각 얼굴 형태와 세부 스타일을 처리하는 원하는 속성을 이미 가지고 있기 때문에, 레이어-혼합 모델에 추가적인 엔드-투-엔드 미세 조정을 적용하지 않는다. 제1 레이어-혼합 네트워크(300)에서 훈련되는 유일한 구성 요소는 거친 레이어에 추가된 얼굴 형태 과장 블록들(221-224)이다.

[0047] 얼굴 형태 과장 블록들을 훈련시키기 위하여, 캐리커처에서 사진으로의 변환을 위한 제2 레이어-혼합 네트워크(320)을 정의하여, 순환 훈련을 가능하게 할 수 있다. 특히, 캐리커처 StyleGAN(210)에서 처음 4개의 거친 레이어들(211-214)를 가져오고 사진 StyleGAN(250)에서 나머지 세밀한 레이어들(255-257)을 가져와서, 이 새로운 제2 레이어-혼합 네트워크(320)을 구성할 수 있다. 제2 레이어-혼합 네트워크(320)은 $c2p$ 또는 $c-p$ 모델로 표시하고, 원래의 제1 레이어-혼합 네트워크(300)을 $p2c$ 또는 $p-c$ 모델로 표시할 수 있다. 본래의 StyleGAN 모델과 차별화되는 본 발명에 따른 생성 네트워크, 특히 제1 레이어-혼합 네트워크(300)을 StyleCariGAN으로 칭할 수도 있다.

[0048] 예시적 실시예에서 사진에서 캐리커처로의 변환을 위한 핵심적인 요소 중 하나는 얼굴 형태 과장 블록들(221-224)이다. 4개의 얼굴 형태 과장 블록들(221-224)이 $p2c$ 네트워크(300)에 구비되어 있다. 얼굴 형태 과장 블록들(221-224)은 StyleCariGAN의 거친 레이어들(251-254)에 대한 추가 특징 변조 맵을 생성하는 컨볼루션 레이어를 기반으로 구성된다. 각 얼굴 형태 과장 블록은 크기가 n^2 인 특징맵을 받아들이고, 잔차 학습과 유사하게 입력 특징 맵에 다시 추가되는 크기가 n^2 의 특징 변조 맵을 생성한다. 그리고, 변조된 특징맵은 더 세밀한 구조를 처리하기 위해 더 높은 해상도의 특징맵을 생성하는 각 가산기(241-242)를 통해서 다음 레이어에 공급된다. 가산기들(241-242)은 대응하는 요소들끼리 덧셈을 행하는 연산기이다.

[0049] 즉, 제1 얼굴 형태 과장 블록(221)은 레이어(251)에서 출력되는 크기가 4^2 인 특징맵을 받아들이고, 크기가 4^2 의 특징 변조 맵을 생성하며, 가산기(241)는 레이어(251)의 출력에 제1 얼굴 형태 과장 블록(221)으로부터의 특징 변조 맵을 더해서, 더한 결과를 레이어(212)에 공급한다. 제2 얼굴 형태 과장 블록(222)은 레이어(252)에서 출력되는 크기가 8^2 인 특징맵을 받아들이고, 크기가 8^2 의 특징 변조 맵을 생성하며, 가산기(242)는 레이어(252)의 출력에 제2 얼굴 형태 과장 블록(222)으로부터의 특징 변조 맵을 더해서, 더한 결과를 레이어(213)에 공급한다. 제3 얼굴 형태 과장 블록(223)은 레이어(253)에서 출력되는 크기가 16^2 인 특징맵을 받아들이고, 크기가 16^2 의 특징 변조 맵을 생성하며, 가산기(243)는 레이어(253)의 출력에 제3 얼굴 형태 과장 블록(223)으로부터의 특징 변조 맵을 더해서, 더한 결과를 레이어(214)에 공급한다. 제4 얼굴 형태 과장 블록(224)은 레이어(254)에서 출력되는 크기가 32^2 인 특징맵을 받아들이고, 크기가 32^2 의 특징 변조 맵을 생성하며, 가산기(244)는 레이어(254)의 출력에 제4 얼굴 형태 과장 블록(224)으로부터의 특징 변조 맵을 더해서, 더한 결과를 레이어(215)에 공급한다. 최종적으로 레이어(217)은 크기가 256^2 인 이미지를 출력한다. 순환 일관성을 견지하기 위하여 마련된 $c2p$ 네트워크(320)도 $p2c$ 네트워크(300)과 유사하게 구성되며, 4개의 얼굴 형태 과장 블록들(261-264)을 구비한다.

[0050] 도 4는 얼굴 형태 과장 블록의 학습을 위한 목적 함수로 사용되는 여러 손실들을 보여준다.

[0051] 다음 두 가지 목표를 달성하기 위해 얼굴 형태 과장 블록을 훈련시킨다. 첫 번째 목표는 실제 캐리커처를 닮은 얼굴 형태 변형을 도입하는 것으로서, 이는 두 가지 유형의 생성적 대립 손실들 즉, L_{feat} 과 L_{GAN} 에 의해 달성될 수 있다. 두 번째 목표는 입력 사진에서 중요한 시각적 특징을 보존하는 것이며, 이는 세 가지 손실들 즉, 특징맵 순환 일관성 손실 L_{fcyc} , 신원 순환 일관성 손실 L_{icyc} , 속성 정합 손실 L_{attr} 에 의해 달성될 수 있다. 즉, 사실적이면서 얼굴 속성을 보존하는 과장을 달성하기 위하여, 얼굴 형태 과장 블록은 적대적 손실들(L_{feat} , L_{GAN}), 순환 일관성 손실들(L_{fcyc} , L_{icyc}), 및 속성 정합 손실(L_{attr})로 지도된다.

[0052] 먼저, 얼굴 형태 과장 블록이 적용된 생성 네트워크의 최종 결과물은 캐리커처처럼 보여야 한다. 이를 성취하기 위해 얼굴 형태 과장 블록 학습 시 적대적 생성 네트워크의 학습 방법을 따른다. 즉, 판별기(282)는 출력되는 캐리커처가 진짜 예술가가 만든 캐리커처인지, 아니면 생성 네트워크에 의해 만들어진 캐리커처인지 구별하는 능력을 학습한다. 이와 적대적으로, 얼굴 형태 과장 블록을 포함한 제1 레이어-혼합 네트워크(300)는 최대한 실제 예술가가 그린 것과 같은 캐리커처를 만들고, 얼굴 형태 과장 블록을 포함한 생성 네트워크는 이 판별기(282)를

속여서 실제 캐리커처처럼 보이는 결과물을 생성해낸다. 적대적 학습이 더 효과적으로 진행될 수 있게 하기 위해, 이미지를 구별하는 판별기(282) 뿐만 아니라 형태 특징 맵을 구별하는 판별기(281)를 추가적으로 사용한다. 이 판별기(281)는 얼굴 형태 과장 블록에 의해 생성된 형태 특징 맵과 캐리커처에 대해 사전 학습된 레이어에 의해 생성된 캐리커처 형태 특징 맵을 구별하도록 학습한다. 얼굴 형태 과장 블록들은 이 판별기(281)를 속여 캐리커처 형태 특징 맵과 구별되지 않는 형태 특징 맵을 생성하도록 학습한다. 제2 레이어-혼합 네트워크(320)에 대해서도, 유사하게 이미지를 구별하는 판별기(284)와 형태 특징 맵을 구별하는 판별기(283)가 마련된다.

[0053] 두 개의 판별기를 도입하여, 변조된 특징맵들과 최종 이미지들 양자 모두에 적대적 학습을 적용한다. 양자 모두에 대하여 non-saturating logistic loss [Goodfellow et al. 2014] with L_1 regularization [Mescheder et al. 2018] (비포화 로지스틱 손실)을 사용하여, 수학식 1과 같이 두 손실의 합을 총 적대적 손실 \mathcal{L}_{adv} 로 계산한다.

수학식 1

[0054]
$$\mathcal{L}_{adv} = \mathcal{L}_{feat} + \lambda_{GAN} \mathcal{L}_{GAN}$$

[0055] 입력 얼굴 사진과 닮은 캐리커처를 만들도록 지도하기 위해 cycle-consistency 목적 함수를 사용한다. Cycle-consistency 목적 함수는 얼굴 형태 과장 블록이 입력 일반 얼굴 형태 특징 맵으로부터 캐리커처 형태 특징 맵을 만들고 다시 캐리커처 형태 특징 맵으로부터 일반 얼굴 형태 특징 맵을 만들었을 때 결과물이 입력 일반 얼굴 특징 맵과 동일해야 한다는 목적을 지닌다. 마찬가지로, 캐리커처 형태 특징 맵으로부터 일반 얼굴 형태 특징 맵을 만들고 다시 일반 얼굴 형태 특징 맵으로부터 캐리커처 형태 특징 맵을 만들었을 때 처음 캐리커처 형태 특징 맵과 결과 캐리커처 형태 특징 맵이 같아야 한다는 방식으로 적용할 수 있다.

[0056] Cycle-consistency 목적 함수를 위해선 얼굴 사진으로부터 캐리커처를 만드는 네트워크와 대칭적인 캐리커처로부터 얼굴 사진을 만드는 네트워크가 필요하다. 캐리커처로부터 얼굴 사진을 만드는 네트워크는 형태를 생성하는 레이어로 캐리커처로 사전 학습된 레이어를 이용하고 텍스처를 생성하는 레이어는 일반 얼굴 사진으로 사전 학습된 레이어를 이용한다. 그리고 얼굴 형태 과장 블록과 반대의 기능을 수행하는 얼굴 형태 과장 복원 블록(shape de-exaggeration blocks)을 형태를 생성하는 레이어 각 계층에 삽입한다.

[0057] 변조된 특징맵에 대한 순환 일관성과 이미지에 대한 순환 일관성을 각각 특징맵 순환 일관성(\mathcal{L}_{fcyc})과 신원 순환 일관성(\mathcal{L}_{icyc})로 칭한다. 특징맵 일관성은 얼굴 형태 과장 블록의 효과가 각 특징맵 스케일에서 주기적으로 반전될 수 있게 한다. p2c-StyleCariGAN에 대한 얼굴 형태 과장 블록은 사진으로부터 캐리커처로의 특징 변조($\mathcal{S}^{p \rightarrow c}$)를 정의한다. 그에 상응하는 c2p-StyleCariGAN에서의 블록은 캐리커처로부터 사진으로의 특징 변조($\mathcal{S}^{c \rightarrow p}$)를 정의한다. 즉, c2p-StyleCariGAN의 블록들은 과장의 역동작을 수행하여 과장된 얼굴 형태를 일반 얼굴 형태로 되돌린다. 이 블록들을 얼굴 형태 과장복원 블록들(shape de-exaggeration blocks)로 칭하기로 한다. 상기 두 가지 특징 변조를 사용하여, 순환 일관성 손실(\mathcal{L}_{fcyc})은 다음과 같이 정의된다.

수학식 2

$$\mathcal{L}_{fcyc}^{p \rightarrow c} = \sum_{i=1}^4 \left(\mathbb{E}_{\mathcal{F}_p^i \sim G_p^i(w)} \left[\left\| \mathcal{S}_{c \rightarrow p}^i \left(\mathcal{S}_{p \rightarrow c}^i \left(\mathcal{F}_p^i \right) \right) - \mathcal{F}_p^i \right\|_2 \right] \right),$$

$$\mathcal{L}_{fcyc}^{c \rightarrow p} = \sum_{i=1}^4 \left(\mathbb{E}_{\mathcal{F}_c^i \sim G_c^i(w)} \left[\left\| \mathcal{S}_{p \rightarrow c}^i \left(\mathcal{S}_{c \rightarrow p}^i \left(\mathcal{F}_c^i \right) \right) - \mathcal{F}_c^i \right\|_2 \right] \right),$$

[0058]
$$\mathcal{L}_{fcyc} = \mathcal{L}_{fcyc}^{p \rightarrow c} + \mathcal{L}_{fcyc}^{c \rightarrow p}$$

[0059] 여기서, F_c^i 는 캐리커처 StyleGAN의 i -번째 블록(G_c^i)에 의해 생성된 캐리커처 특징맵이고, F_p^i 는 사진 StyleGAN의 i -번째 블록(G_p^i)에 의해 생성된 사진 특징맵이며, $S_{c \rightarrow p}^i$ 는 i -번째 사진-캐리커처 얼굴 형태 과장 블록이고, $S_{p \rightarrow c}^i$ 는 i -번째 캐리커처-사진 얼굴 형태 과장복원 블록이다. 그림 6은 이들 손실을 보여준다.

[0060] 신원 일관성 즉, 정체성에 대한 일관성은 특징맵으로부터 생성된 사진을 검사하여 얼굴 형태 과장 블록의 효과가 주기적으로 반전될 수 있게 한다. 사진들에 대해 훈련된 얼굴 임베딩 네트워크를 기반으로 신원 일관성 손실(L_{cyc})을 설계하여, 신원 일관성이 사진 특징맵에서 시작하는 주기에 대해서만 계산되도록 한다. 얼굴 임베딩 네트워크를 사용하여, 신원 일관성 손실(L_{cyc})을 2개의 임베딩 즉, 1) 시작 사진의 얼굴 임베딩과, 2) $S_{c \rightarrow p}^i$ 와 $S_{p \rightarrow c}^i$ 에 의해 순차적으로 변조된 거친 특징맵들을 사용하여 생성되는 cycle-reconstructed photo의 얼굴 임베딩, 사이의 L2 거리로 설계한다.

[0061] 도 5는 목적 함수 중 하나인 특징맵 순환 일관성 손실을 결정하기 위한 생성 학습 레이어들과 얼굴 형태 과장 블록들의 배치 예를 보여주는 도면이다. 본 발명에 따른 사진-캐리커처 맵핑과 캐리커처-사진 맵핑이 특징맵-특징맵 맵핑이기 때문에, 순환 일관성 손실은 캐리커처-특징맵 인코더가 필요하지 않다. 상기 맵핑들은 이미지를 입력으로 사용하지 않으며 특징맵을 직접 사용하여 순환 일관성 손실을 구현할 수 있다.

[0062] 마지막으로 주기 일관성 손실(L_{cyc})은 다음과 같이 계산된다.

수학식 3

[0063]
$$\mathcal{L}_{cyc} = \mathcal{L}_{fcyc} + \lambda_{icyc} \mathcal{L}_{icyc}$$

[0064] 순환 일관성이 있다고 하더라도, 결과적인 과장이 입력 사진의 중요한 특징을 보존한다는 보장이 없다. 따라서, 유효한 캐리커처 변형을 생성하도록 얼굴 형태 과장 블록을 제한하기 위하여, 사진들 및 캐리커처들에 대하여 얼굴 속성 분류기를 사용한다. 하나는 일반 얼굴 사진에 대해 학습된 얼굴 특징 분류기이며 다른 하나는 캐리커처에 대해 학습된 얼굴 특징 분류기이다. 이 사진 학습된 얼굴 특징 분류기를 이용하여 입력 얼굴 사진과 결과 캐리커처에 대해 얼굴 특징을 추출하고 이들이 서로 같아지도록 이진 크로스 엔트로피 손실(binary cross entropy loss) 형태로 목적 함수를 정의한다. 얼굴 형태 과장 복원 블록에 대해서는 이와 대칭적으로 입력 캐리커처와 결과 일반 얼굴에 대해 얼굴 특징을 추출하고 이들이 서로 같아지도록 한다. p2c-StyleCariGAN이 입력 사진과 동일한 속성을 가진 캐리커처를 생성하도록 제한한다. 유사한 속성 정합을 c2p-StyleCariGAN에 적용한다. 속성 정합 손실(L_{attr})은 사진 속성들과 캐리커처 속성들 간의 이진 교차 엔트로피 손실들을 사용하여 정의된다.

수학식 4

$$\begin{aligned} \mathcal{L}_{attr}^{p \rightarrow c} &= -\mathbb{E}_{w \sim W} [\phi_p(G_p(w)) \log \phi_c(G_{p \rightarrow c}(w)) \\ &\quad + (1 - \phi_p(G_p(w))) \log(1 - \phi_c(G_{p \rightarrow c}(w)))], \\ \mathcal{L}_{attr}^{c \rightarrow p} &= -\mathbb{E}_{w \sim W} [\phi_c(G_c(w)) \log \phi_p(G_{c \rightarrow p}(w)) \\ &\quad + (1 - \phi_c(G_c(w))) \log(1 - \phi_p(G_{c \rightarrow p}(w)))], \end{aligned}$$

[0065]
$$\mathcal{L}_{attr} = \mathcal{L}_{attr}^{p \rightarrow c} + \mathcal{L}_{attr}^{c \rightarrow p}$$

[0066] 여기서, ϕ_p 는 사진 속성 분류자, ϕ_c 는 캐리커처 속성 분류자, G_p 는 사진은 StyleGAN, G_c 는 캐리커처 StyleGAN, G_{p*c} 는 p2c-StyleCariGAN이며, G_{c*p} 는 c2p-StyleCariGAN이다.

[0067] 요약하자면, 훈련을 위한 총 목적 함수는 다음과 같다.

수학식 5

[0068]
$$\mathcal{L}_G = \lambda_{adv}\mathcal{L}_{adv} + \lambda_{cyc}\mathcal{L}_{cyc} + \lambda_{attr}\mathcal{L}_{attr}$$

[0069] 여기서, λ_{adv} , λ_{cyc} , λ_{attr} 은 손실 가중치들을 정의하는 상수들이다.

[0070] 얼굴 형태 과장 블록들은 생성된 이미지들에 대하여 다중-스케일 변형 조절을 가능하게 해준다. 이들 블록이 서로 다른 공간 스케일을 나타내는 4개의 특징맵들에 대해 변조를 수행하기 때문에, 각 블록의 출력에 스케일링 팩터(231-234)를 추가하기만 하면 변조 정도를 자유롭게 변경할 수 있으며, 이를 통해 과장 정도를 변경한다. 예를 들어, 전체적인 얼굴 형태의 변형을 줄이기 위하여, 첫 번째 얼굴 형태 과장 블록의 출력이 원래의 특징맵에 가산되기 이전에 첫 번째 얼굴 형태 과장 블록의 출력에 1보다 작은 가중치를 곱할 수 있다. 극단적인 변형으로 인한 주름을 제거하기 위하여, 세 번째 또는 네 번째 얼굴 형태 과장 블록의 출력에 0 또는 작은 스케일링 계수를 곱할 수 있다.

[0071] 도 6은 도 2에 도시된 캐리커처 생성 장치의 일 실시예의 물리적 구성을 보여주는 블록도이다. 캐리커처 생성 장치는 프로세서(400), 메모리(402), 저장 장치(404), 및 데이터 송수신부(406)를 포함할 수 있다. 또한, 캐리커처 생성 장치는 입력 인터페이스 장치(410) 및 출력 인터페이스 장치(412)를 더 포함할 수 있다. 캐리커처 생성 장치에 포함된 각각의 구성 요소들은 버스에 의해 연결되어 서로 통신할 수 있다.

[0072] 프로세서(400)는 메모리(402) 및/또는 저장 장치(404)에 저장된 프로그램 명령을 실행할 수 있다. 프로세서(400)는 적어도 하나의 중앙 처리 장치(central processing unit, CPU)나 그래픽 처리 장치(graphics processing unit, GPU)에 의해 구현될 수 있으며, 그밖에 본 발명에 따른 방법을 수행할 수 있는 여타의 프로세싱 디바이스일 수 있다. 프로세서(400)는 본 발명에 의한 디포커스 디블러링 방법을 구현하기 위한 프로그램 명령들을 실행할 수 있다.

[0073] 메모리(402)는 예컨대 RAM(Random Access Memory)와 같은 휘발성 메모리와, ROM(Read Only Memory)과 같은 비휘발성 메모리를 포함할 수 있다. 메모리(402)는 저장 장치(404)에 저장된 프로그램 명령을 로드하여, 프로세서(400)에 제공함으로써 프로세서(400)가 이를 실행할 수 있도록 할 수 있다. 특히, 본 발명에 따르면, 메모리(402)는 프로그램 명령 이외에, 입력 영상 데이터, 출력 영상 데이터, 커널 가중치들, 및 특징맵 데이터를 위시해서 디포커스 디블러링을 위한 프로그램 수행 과정에서 발생하는 데이터를 임시 저장할 수 있다.

[0074] 저장 장치(404)는 프로그램 명령과 데이터를 저장하기에 적합한 기록매체로서, 예컨대 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(Magnetic Media), CD-ROM(Compact Disk Read Only Memory), DVD(Digital Video Disk)와 같은 광 기록 매체(Optical Media), 플롭티컬 디스크(Floptical Disk)와 같은 자기-광 매체(Magneto-Optical Media), 플래시 메모리나 EPROM(Erasable Programmable ROM) 또는 이들을 기반으로 제작되는 SSD와 같은 반도체 메모리를 포함할 수 있다. 저장 장치(404)는 본 발명에 의한 캐리커처 생성 방법을 구현하기 위한 프로그램 명령을 저장할 수 있다. 또한, 저장 장치(404)는 커널 가중치들 및 특징맵 데이터를 포함하여 긴 시간동안 저장이 필요한 데이터를 저장할 수 있다.

[0075] 도 7은 본 발명의 일 실시예에 따른 캐리커처 생성 방법을 보여주는 흐름도이다.

[0076] 예시적 실시예에 따른 캐리커처 생성 장치에 의해 캐리커처를 생성함에 있어서는, 먼저 캐리커처 생성 네트워크(210)의 캐리커처 생성 학습 레이어들(211-217)과 사진 복원 네트워크(250)의 사진 복원 학습 레이어들(261-267)을 훈련시킨다(제500단계).

[0077] 캐리커처 생성 학습 레이어들(211-217)과 사진 복원 학습 레이어들(261-267)의 학습이 완료되면, 캐리커처 생성 학습 레이어들(211-217)과 사진 복원 학습 레이어들(261-267)의 일부를 스와핑하여, 제1 및 제2 레이어-혼합 네트워크(300, 320)를 구성한다(제510단계). 이때, 제1 레이어-혼합 네트워크(300)에 얼굴 형태 과장 블록들(221-224)이 포함되게 하고, 제2 레이어-혼합 네트워크(330)에 얼굴 형태 과장 블록들(261-264)이 포함되게 한다.

- [0078] 이어서, 얼굴 형태 과장 블록들(221-224, 261-264)을 훈련시킨다(제520단계).
- [0079] 얼굴 형태 과장 블록들(221-224, 261-264)의 훈련이 완료되면, 제1 레이어-혼합 네트워크(300)에 의해 형태 특징 맵을 생성하고, 얼굴 형태 과장 블록들(221-224)에 인가하여 캐리커처 형태 특징맵이 생성되게 한다(제530단계). 그리고, 캐리커처 형태 특징맵을 제1 레이어-혼합 네트워크(300)에 인가하여, 캐리커처 형태 특징맵에 따라서 상기 형태 특징 맵을 변형시킨다(제540단계). 이어서, 변형된 형태 특징 맵에 따라서 캐리커처 이미지를 생성한다(제550단계).
- [0080] 실험 및 평가
- [0081] 4개의 NVIDIA Quadro RTX 8000(48GB) GPU를 토대로 캐리커처 생성 장치를 구축하였다. 앞에서 언급한 바와 같이 각 얼굴 형태 과장 블록은 두 개의 컨볼루션 레이어로 구성되며, 각 레이어는 leaky ReLU를 활성화 함수로 갖는다. 각 컨볼루션 레이어에서 커널 크기=3, 스트라이드=1, 패딩=1이다. leaky ReLU 레이어는 음의 기울기 0.2를 가진다.
- [0082] PyTorch에서 $\beta_1=0$, $\beta_2=0.99$ 이고 학습속도가 0.002인 Adam 옵티마이저를 사용하였다. 배치(batch) 크기가 32인 각 미니 배치는 무작위로 생성된 사진들과 캐리커처들로 구성된다. 1,000번의 반복 후에 훈련을 중단하였다. 경험적으로 손실들에 대한 가중치를 $\lambda_{adv}=1$, $\lambda_{GAN}=10$, $\lambda_{cyc}=10$, $\lambda_{icyc}=1000$, 그리고 $\lambda_{attribute}=10$ 로 설정하였다. 4개의 NVIDIA Quadro RTX 8000(48GB) GPU를 사용한 훈련 시간은 약 8시간이었다.
- [0083] 캐리커처 StyleGAN과 사진 StyleGAN은 모두 StyleGAN2의 아키텍처와 학습 알고리즘을 사용하였다. 사진 StyleGAN은 256×256 해상도로 크기 조정된 FFHQ 데이터셋으로 훈련되었다. 캐리커처 StyleGAN은 FreezeD 및 ADA를 사용하여 WebCaricature의 캐리커처로 사진 모델로부터 미세 튜닝되었다. WebCaricature 데이터셋은 5개의 랜드마크(두 눈의 중심, 코의 끝, 입의 두 가장자리)를 사용하여 정렬되었고 크기도 256×256 으로 조정되었다. 두 모델의 훈련에는 PyTorch 구현을 사용하였다. 얼굴 임베딩 네트워크로는 사전 훈련된 FaceNet을 사용하였다.
- [0084] L_{attr} 을 구현하려면, 사진들에 대한 속성 분류기와 캐리커처들에 대한 속성 분류기가 필요하다. WebCaricature 데이터셋을 사용하여 속성 분류기를 훈련시켰는데, 이 데이터셋은 사진들과 캐리커처들 모두에 대한 레이블을 제공한다. 사진들과 캐리커처들의 레이블 분포가 데이터셋에서 상당히 유사하다는 것을 발견하였다. 속성 분류기의 백본 아키텍처는 ResNet-18이고 마지막의 완전연결 레이어의 출력 채널 크기만 변경하였다. 출력 채널 크기는 50이며, 이 숫자는 WebCaricature의 속성의 개수이다. PyTorch에 의해 제공된 사전 학습된 ResNet-18을 사전-튜닝하였다. WebCaricature 데이터셋의 테스트 셋에 대한 테스트 정확도는 사진의 경우 85%, 캐리커처의 경우 82%이었다.
- [0085] 테스트 시간으로서, 입력 사진에서 잠재 코드로의 GAN 반전은 3~4분이 소요된다. 반전 후에, 캐리커처 이미지를 생성하는 데에는 약 40ms가 소요되었다. 시간 측정은 NVIDIA Quadro RTX 8000과 Intel Xeon Gold 6226R이 설치된 서버에서 수행되었다. 입력 사진과 출력 캐리커처 모두 256×256 크기를 가진다.
- [0086] 두 종류의 방법들 즉, 일반적인 이미지-이미지 변환 기법과 딥러닝에 의한 캐리커처 생성 기법과 비교하여 본 발명에 의한 방법을 정성적으로 평가하였다. 도 9는 이미지-이미지 변환 및 캐리커처 생성을 위한 다른 최신 기법들과의 비교 결과를 보여준다. 예시적 실시예에 따른 방법은 입력 얼굴들을 변형하고 스타일화하여, 다른 기법과 비교할 때 더 사실적이고 디테일한 캐리커처를 생성할 수 있다. 입력 이미지는 모두 저작권이 있음(저작권자: Jordan Strauss/Invision/AP, AF archive/Alamy Stock Photo, Stuart C. Wilson/Getty Images, WENN Rights Ltd/Alamy Stock Photo, Danny Moloshok/Invision/AP, Slaven Vlastic/Getty Images).
- [0087] 다양한 이미지-이미지 변환 모델 중에서, U-GAT-IT [Kim et al. U-gat-it: unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation. In Proc. ICLR. 2020.]과 StarGAN v2[Choi et al. targan v2: Diverse image synthesis for multiple domains. In Proc. CVPR. 2020.]는 얼굴 형태 변화가 큰 두 도메인을 변환할 때 시각적으로 만족스러운 작업 결과를 보여주었다. 이 모델들에 대한 공식 구현 프로그램들을 사용하여 WebCaricature 데이터셋을 사용하여 U-GAT-IT 및 StarGAN v2를 훈련시켰다. U-GAT-IT는 입력 사진으로부터 캐리커처를 성공적으로 생성하지 못하고 흐릿한 이미지를 만들었다. 때로는 입력에 비해 출력의 변화가 미미하였다. StarGAN v2는 더 안정적인 결과를 생성하지만, 때로는 텍스처에서 아티팩트가 발견되었다. 그밖에, 과장 정도가 본 발명에 비하여 적었다. WarpGAN[Shi et al. "WarpGAN: Automatic caricature generation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.]과 AutoToon [Gong et al. . "AutoToon: Automatic

Geometric Warping for Face Cartoon Generation." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2020.]은 딥러닝 기반 캐리커처 생성을 위한 최신 방법이다. 저자가 발표한 사전 훈련된 WarpGAN 및 AutoToon 모델을 사용하였다. U-GAT-IT 및 StarGAN v2와 마찬가지로, WarpGAN을 WebCaricature 데이터셋으로 훈련시켰다. 이에 반하여, AutoToon은 지도 학습을 위해 사진 및 캐리커처 쌍을 포함하고 있는 자체 데이터셋을 사용하여 훈련시켰다. 그밖에 CariGAN [Cao et al. 2018] 및 CariGAN [Li et al. 2020]과 비교하였다.

[0088] 도 9에서 볼 수 있듯이, 이러한 딥러닝 기반 캐리커처 생성 방법들은 명시적인 2D 이미지 변형을 사용하여 큰 얼굴 형태 변화를 제공할 수 있다. 그러나 생성된 얼굴 형태 변형은 사실적인 캐리커처를 만들기에는 충분한 디테일을 제공하지 않는다. WarpGAN은 변형 처리에 위해 16개의 튜닝된 제어점만 사용할 뿐이고, AutoToon은 101개의 사진 캐리커처 쌍으로만 훈련된다. 제한된 제어점 밀도와 훈련 샘플의 적은 양으로 말미암아, 이들 기법을 사실감있는 캐리커처를 생성하는 수단으로서 일반화하기는 쉽지 않다. 이에 반하여, 육안으로 판별할 때 본 발명에 의한 결과는 더 만족스러울 뿐만 아니라 아티스트가 만든 콘텐츠에 더 가깝다.

[0089] 사용자 연구를 통해 손으로 그린 캐리커처 스타일의 충실도를 인지적으로 평가하였다. 사용자들에게 사진을 보고 사용자에게 최신 기법들과 본 발명에 의해 생성된 캐리커처들 중에서 손으로 그린 캐리커처처럼 보이는 최고의 캐리커처를 선택하도록 요청하였다. WarpGAN, AutoToon, U-GAT-IT, StarGANv2, Toonification에 대한 응답을 본 발명의 결과에 대한 응답과 비교하였다. 두 가지 다른 과장 팩터(0.5 및 1.0)를 사용하여 본 발명을 평가하였다. 우리는 우리 방법의 각 과장 팩터에 대하여 한번씩, 도합 두 번의 실험을 실행하였다. 학습 효과를 방지하기 위해 각 실험에 서로 다른 사용자들이 참여하였다. 비교할 수 있는 방법은 여러 가지가 있지만, 각 질문에 대해 모든 방법의 캐리커처 결과를 함께 제시하면 사용자가 일관되게 대답하기 어려울 수 있다. 따라서 특별히 설계된 캐리커처 생성기 기법(WarpGAN, AutoToon)과 일반적인 이미지-이미지 변환 기법(U-GAT-IT, StarGANv2, Toonification)의 두 그룹으로 기법들을 분할하여 사용자 연구를 실행하였다.

[0090] 각 그룹에 대하여, 각 사용자에게 71개의 질문 풀에서 무작위로 샘플링된 30개의 질문이 주어졌다. 질문 풀은 비교되는 방법들에 대하여 이전에 사용되었던 입력 사진과 그림 9의 입력 사진에 대한 캐리커처를 생성하여 구성하였다. 무작위 선택을 걸러내기 위하여 5개의 중복 질문을 추가하였다. 3개 이상의 중복 질문에서 일치하지 않는 답변을 보이는 사용자를 제외하여 각 그룹에 대해 60명의 사용자로부터 유효한 답변을 얻었다. 실험을 시작하기 전에, 손으로 그린 캐리커처를 사용하여 각 사용자에게 5가지 훈련용 질문들을 제시하여 사용자가 사실적인 캐리커처 스타일에 노출되도록 하였다. 각 훈련용 질문에서 사용자에게 입력 사진과 일치하는 손으로 그린 최고의 캐리커처를 선택해달라는 요청을 하였다. 사용자 연구는 Amazon Mechanical Turk를 사용하여 수행되었다.

[0091] 도 10은 사용자 연구 결과를 보여준다. 위쪽과 아래쪽 부분은 각각 과장 팩터가 0.5와 1.0일 때의 결과를 보여준다. (a, b)는 손으로 그린 캐리커처 스타일에 대한 충실도의 평가 결과를 보여준다. 본 발명의 방법은 캐리커처 스타일의 품질 측면에서 큰 차이로 다른 기법들을 능가한다. (c)는 신원 보존의 평가 결과를 보여준다. 본 발명의 방법은 다른 기법들과 비슷한 식별율을 보여준다. 과장 팩터에 관계없이 두 그룹 모두에서 본 발명의 방법이 이전 기법들을 훨씬 능가하였다.

[0092] 또한 다른 사용자 연구를 통해 본 발명에 의한 방법의 정체성 보존 정도를 인지적으로 평가하였다. 사용자에게 입력 캐리커처와 신원이 일치하는 사진을 선택하도록 요청하였다. 각 질문에서 입력 캐리커처는 Hand-drawn, WarpGAN, AutoToon, Toonification, U-GAT-IT, StarGANv2 및 본 발명에 의해 생성된 것들 중 하나였다. 두 가지 다른 과장 팩터(0.5 및 1.0)를 사용하여 방법을 평가하였다. 공정한 비교를 위하여, 방법당 노출 수의 균형을 맞출 수 있도록 두 결과로 실험을 한 번 실행하는 대신 다른 과장 팩터로 실험을 한번씩 도합 두 번 실행하였다. 질문당 5개의 사진 선택지를 제공하였는데, 하나는 캐리커처 입력과 동일인과 관련되어 있지만 다른 포즈를 가진 정답 선택지이고 나머지 선택지들은 정답과 유사한 얼굴을 포함한다. 이 사용자 연구도 Amazon Mechanical Turk를 사용하여 수행되었다.

[0093] 실험 결과, 본 발명에 의한 방법 중 과장 팩터 0.5로 생성된 캐리커처가 손으로 그린 캐리커처를 포함하여 다른 방법으로 생성된 캐리커처들과 비슷한 정체성 보존율을 보여주었고, 과장 팩터 1.0으로 생성된 캐리커처는 낮은 정체성 보존율을 보여주었다(그림 10(c) 참조).

[0094] StyleCariGAN은 순환 일관성을 가지도록 훈련되기 때문에, 캐리커처에서 사진으로 변환하기 위한 역-맵핑 기능을 구비한다. 캐리커처를 사진으로 변환하려면 먼저 캐리커처 이미지를 캐리커처 StyleGAN의 W+ 공간으로 인코딩한다. 인코딩된 벡터가 입력으로 주어지면, 순환 일관성을 위해 훈련된 c2p-StyleCariGAN을 사용하는 것만으

로 주어진 캐리커처와 유사한 사진을 생성할 수 있다. 그림 12에서 볼 수 있듯이 본 발명의 방법은 캐리커처에서 사진으로의 설득력 있는 변환 결과를 생성할 수 있다.

[0095] 도 11은 얼굴 형태 과장에 대한 다중 스케일 조절 결과를 보여준다. StyleCariGAN은 과장 정도에 대해 4-스케일 조절을 제공한다. α_1 은 가장 거친 스케일의 과장을 조절하고, α_4 는 가장 미세한 스케일을 조절한다. 각 행은 편집의 다른 스케일을 보여준다. 각 행에서 레이블의 파라미터 집합은 변조되고 명시되지 않은 파라미터는 1로 고정된다. 첫 번째 행은 과장되지 않은 이미지에서 완전히 과장된 이미지에 이르기까지 얼굴 형태 과장 블록들의 효과를 보여준다. 얼굴 형태 과장 블록들을 제거하면, 사진과 같은 구조의 캐리커처가 생성된다(1행 1열). 모든 얼굴 형태 과장 블록들을 적용하면, 바람직하고 사실적인 변형이 있는 캐리커처가 생성된다(1행, 5열). 다른 행들은 과장의 다중 스케일 조절을 보여준다. α_1 에서 α_2 까지의 다른 스케일의 특징 변조를 추가하면, 스케일이 증가하는 구조 변형을 생성한다(2~4행, 1열). 가장 거친 특징 변조를 위한 파라미터 α_1 이 전체 얼굴 형태에 큰 영향을 미친다(4행, 1-4열). 맨 오른쪽 열의 캐리커처 결과는 파라미터 값들이 모두 1이므로 모두 동일하다.

[0096] 도 11의 첫 번째 행은 얼굴 형태 과장 블록의 기여도를 보여주는데, 여기서 맨 왼쪽과 맨 오른쪽 이미지들은 각각 얼굴 형태 과장 블록이 있을 때와 없을 때의 결과를 보여준다. 전체적인 색상 스타일을 유지하면서 캐리커처에서 요구되는 얼굴 형태 변경을 생성함에 있어 얼굴 형태 과장 블록이 중요한 역할을 한다는 것을 확인할 수 있다. 그리고 도 11은 얼굴 형태 과장 블록에 스케일링 팩터를 부가함으로써 캐리커처 결과에서 과장 정도를 조절하는 몇 가지 전형적인 예를 보여준다.

[0097] 위에서 언급한 바와 같이 본 발명의 실시예에 따른 동작은 컴퓨터로 읽을 수 있는 기록매체에 컴퓨터가 읽을 수 있는 프로그램 또는 코드로서 구현하는 것이 가능하다. 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의해 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 또한 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어 분산 방식으로 컴퓨터로 읽을 수 있는 프로그램 또는 코드가 저장되고 실행될 수 있다.

[0098] 상기 컴퓨터가 읽을 수 있는 기록매체는 롬(ROM), 램(RAM), 플래시 메모리(flash memory) 등과 같이 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치를 포함할 수 있다. 프로그램 명령은 컴파일러(compiler)에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터(interpreter) 등을 사용해서 컴퓨터에 의해 실행될 수 있는 고급 언어 코드를 포함할 수 있다.

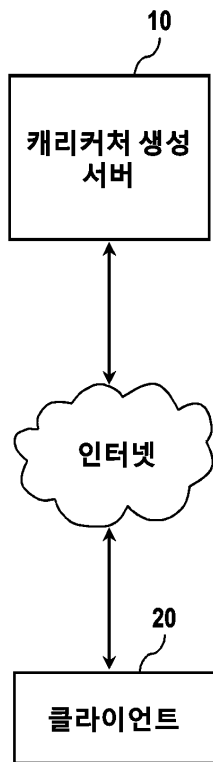
[0099] 본 발명의 일부 측면들은 장치의 문맥에서 설명되었으나, 그것은 상응하는 방법에 따른 설명 또한 나타낼 수 있고, 여기서 블록 또는 장치는 방법 단계 또는 방법 단계의 특징에 상응한다. 유사하게, 방법의 문맥에서 설명된 측면들은 또한 상응하는 블록 또는 아이템 또는 상응하는 장치의 특징으로 나타낼 수 있다. 방법 단계들의 몇몇 또는 전부는 예를 들어, 마이크로프로세서, 프로그램 가능한 컴퓨터 또는 전자 회로와 같은 하드웨어 장치에 의해(또는 이용하여) 수행될 수 있다. 몇몇의 실시예에서, 가장 중요한 방법 단계들의 하나 이상은 이와 같은 장치에 의해 수행될 수 있다.

[0100] 실시예들에서, 프로그램 가능한 로직 장치(예를 들어, 필드 프로그래머블 게이트 어레이)가 여기서 설명된 방법들의 기능의 일부 또는 전부를 수행하기 위해 사용될 수 있다. 실시예들에서, 필드 프로그래머블 게이트 어레이는 여기서 설명된 방법들 중 하나를 수행하기 위한 마이크로프로세서와 함께 작동할 수 있다. 일반적으로, 방법들은 어떤 하드웨어 장치에 의해 수행되는 것이 바람직하다.

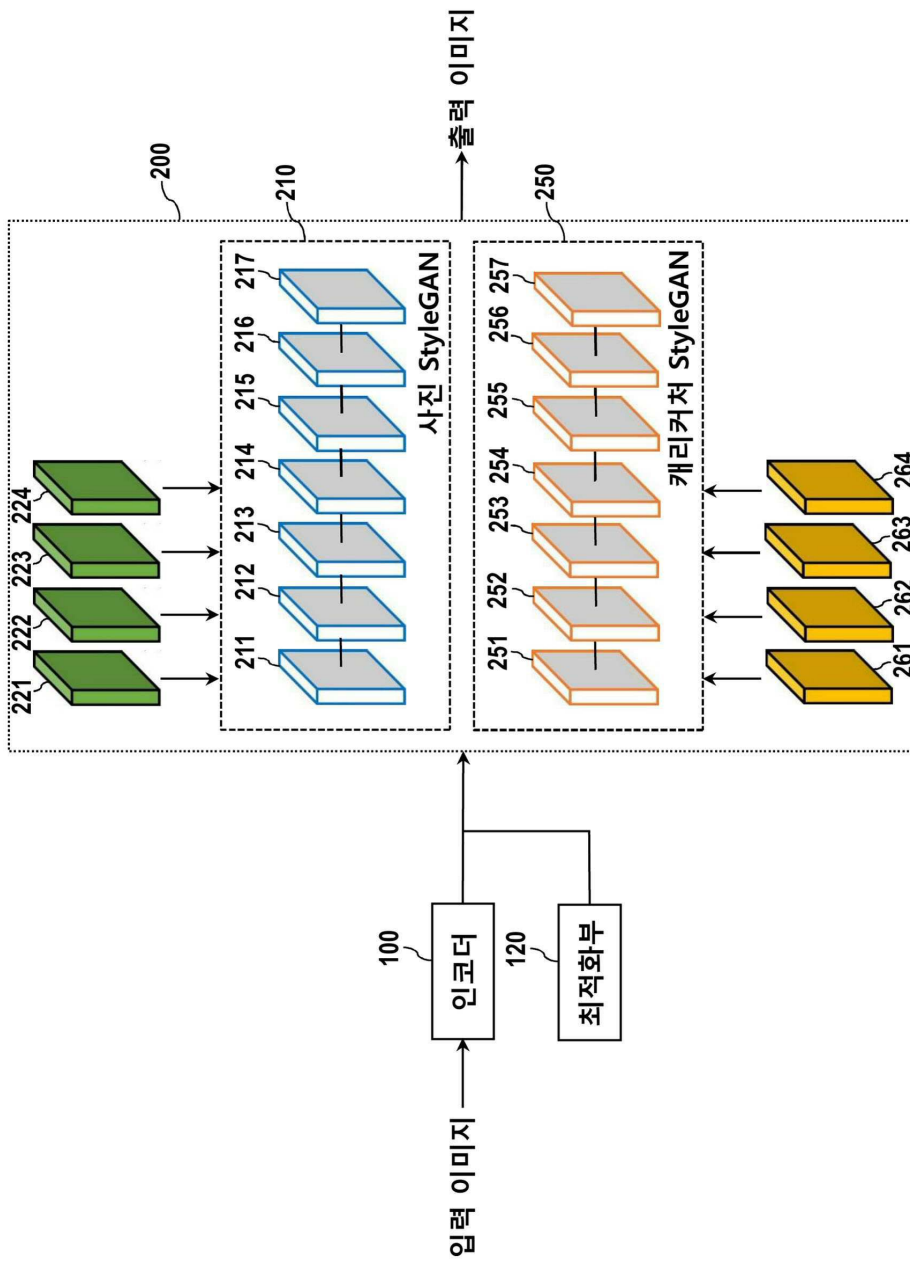
[0101] 위에서 본 발명의 바람직한 실시예를 참조하여 설명하였지만, 해당 기술 분야의 숙련된 당업자는 하기의 특허 청구의 범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 발명을 다양하게 수정 및 변경시킬 수 있음을 이해할 수 있을 것이다.

도면

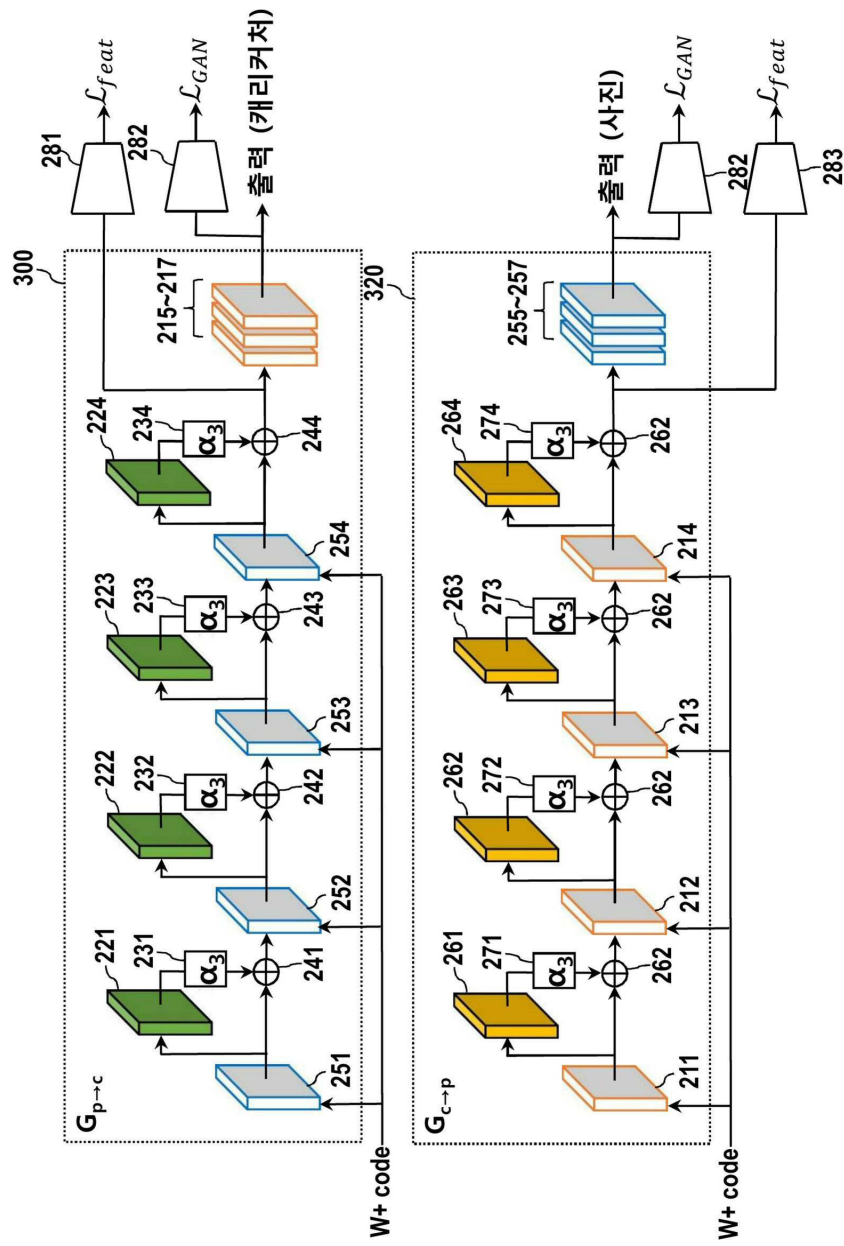
도면1



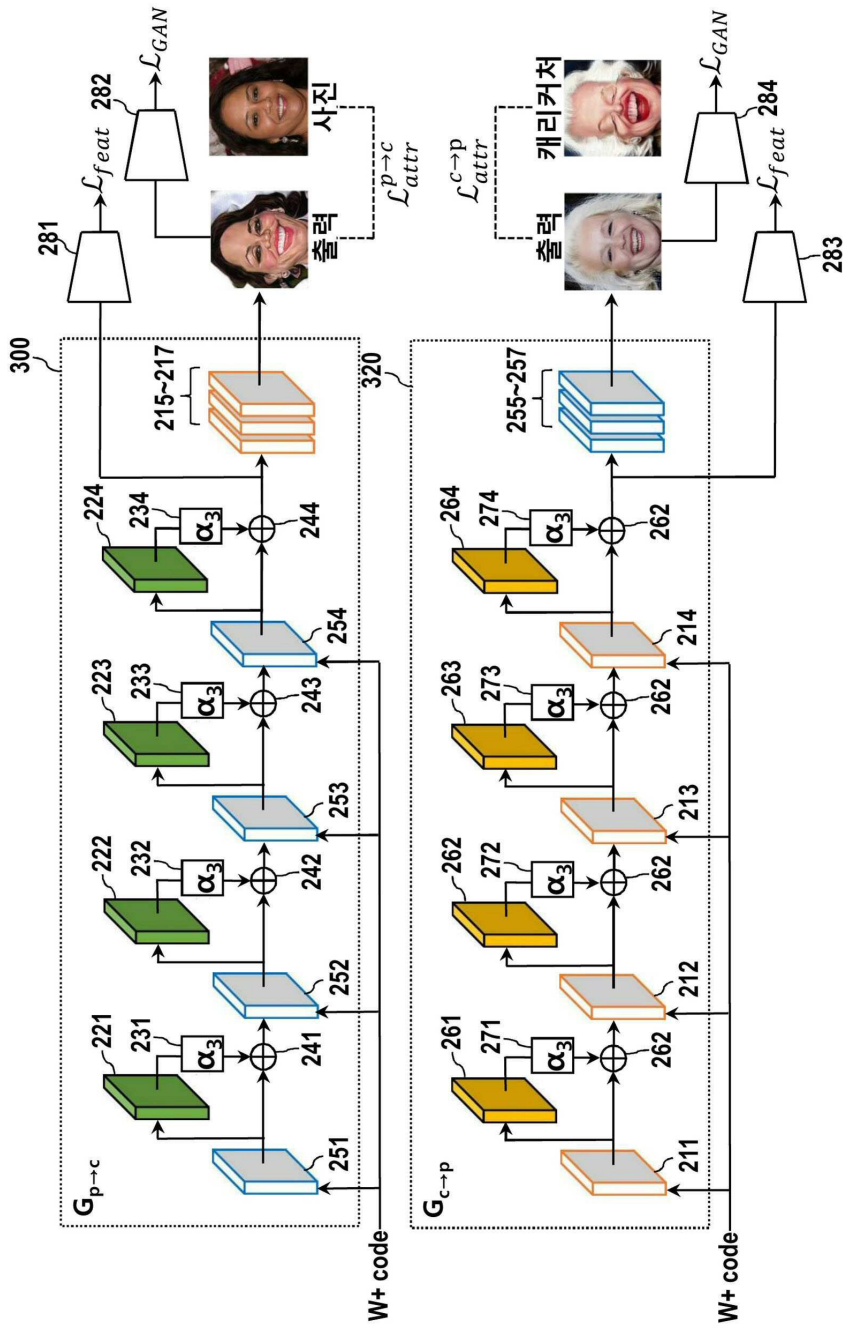
도면2



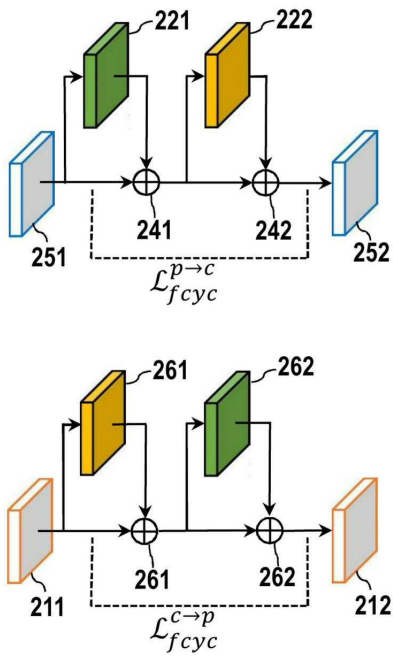
도면3



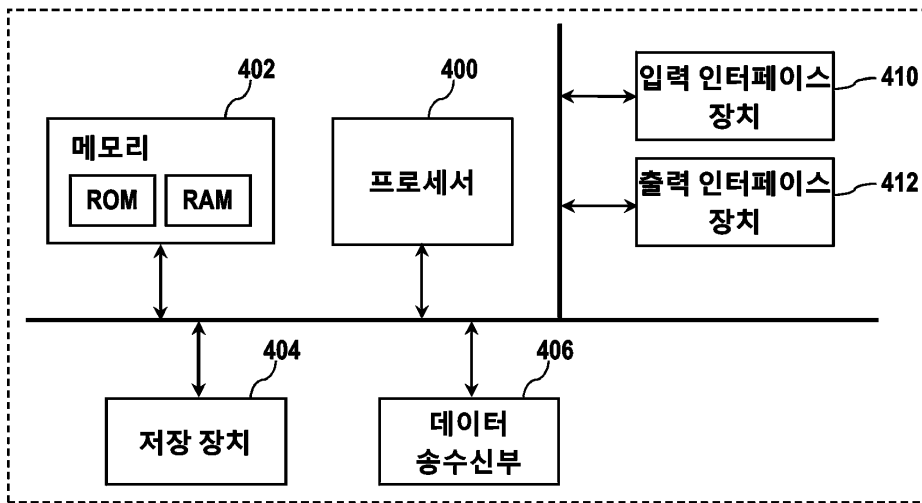
도면4



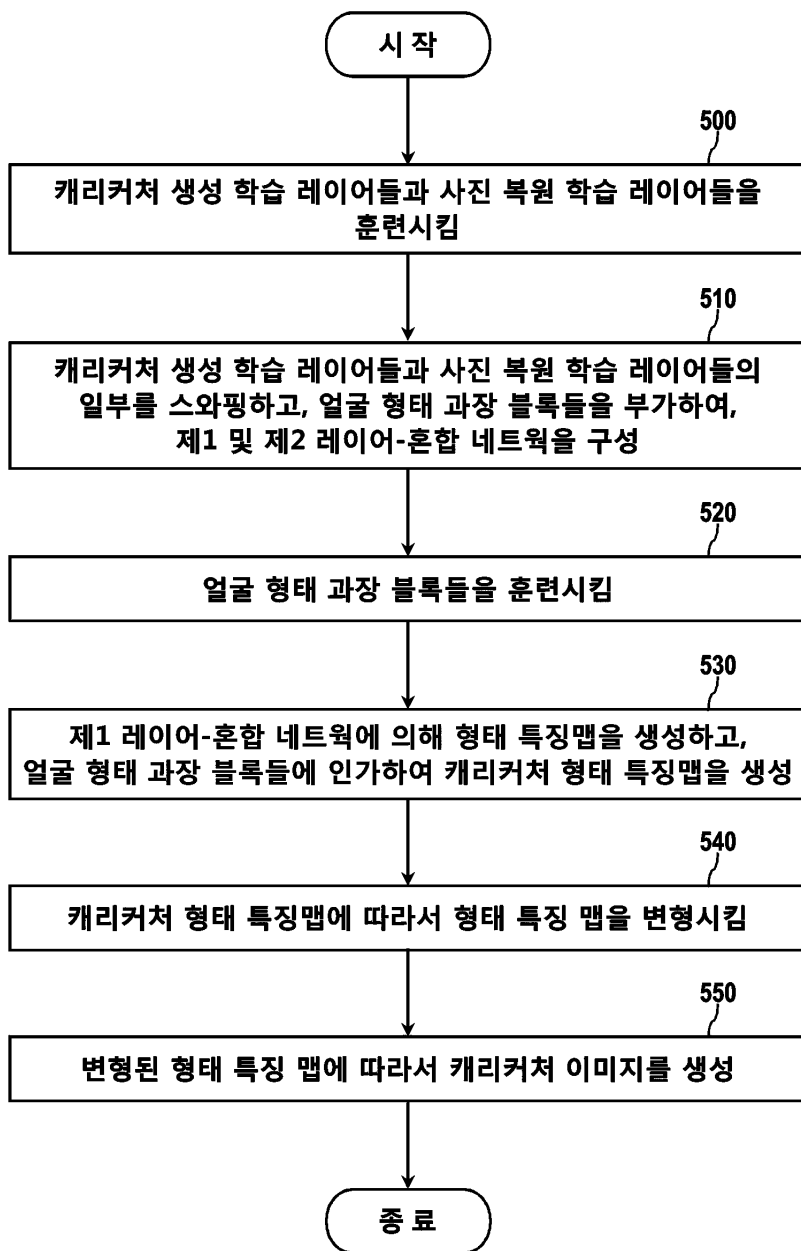
도면5



도면6



도면7



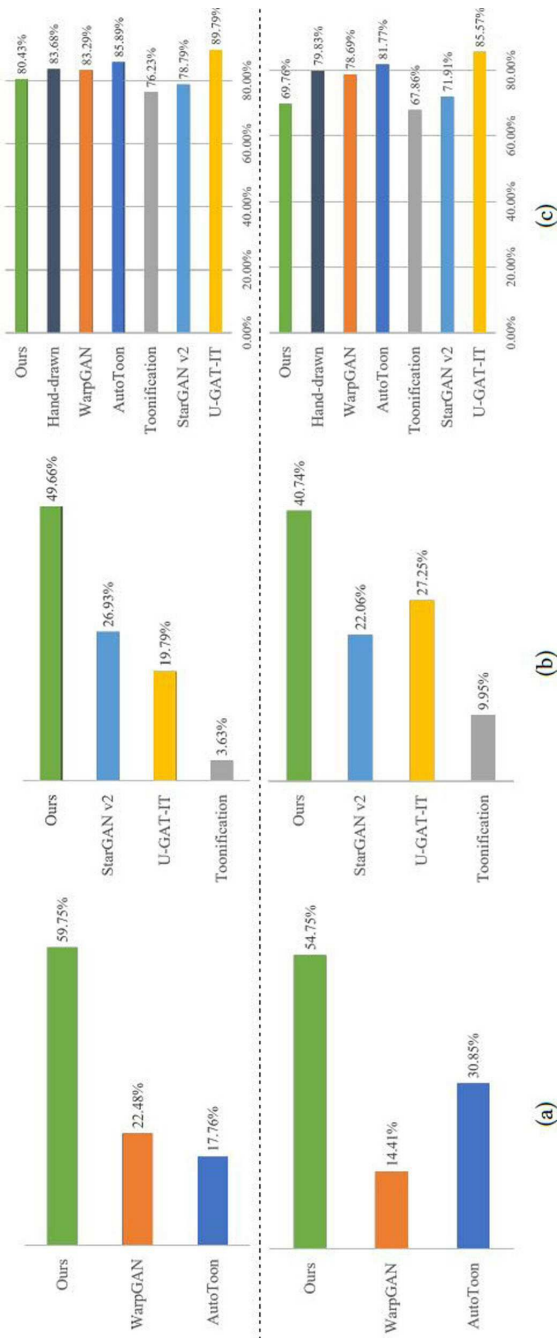
도면8



도면9



도면10



도면11

