



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2024년07월31일
(11) 등록번호 10-2691072
(24) 등록일자 2024년07월29일

- | | |
|--|--|
| <p>(51) 국제특허분류(Int. Cl.)
G06T 3/00 (2024.01) G06T 5/00 (2024.01)</p> <p>(52) CPC특허분류
G06T 3/14 (2024.01)
G06T 3/10 (2024.01)</p> <p>(21) 출원번호 10-2023-0047320</p> <p>(22) 출원일자 2023년04월11일
심사청구일자 2023년04월11일</p> <p>(56) 선행기술조사문헌
Pan Zhang et al., 'Cross-domain Correspondence Learning for Exemplar-based Image Translation', Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.*
Robin Rombach et al., 'High-Resolution Image Synthesis with Latent Diffusion Models', Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022.*
Seungryong Kim et al., 'Recurrent Transformer Networks for Semantic Correspondence', NeurIPS 2018, 2018.*
*는 심사관에 의하여 인용된 문헌</p> | <p>(73) 특허권자
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)</p> <p>(72) 발명자
김승룡
서울특별시 성북구 오패산로 46, 103동 1804호(하월곡동, 월곡두산위브아파트)
서준영
경기도 안양시 동안구 학의로 20, 118동 102호 (비산동, 관악아파트)
(뒷면에 계속)</p> <p>(74) 대리인
송인호, 윤형근, 최관락</p> |
|--|--|

전체 청구항 수 : 총 6 항

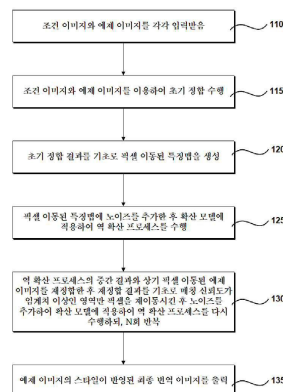
심사관 : 정성운

(54) 발명의 명칭 예제 이미지 기반 이미지 번역 방법 및 그 장치

(57) 요약

예제 이미지 기반 이미지 번역 방법 및 그 장치가 개시된다. 예제 기반 이미지 번역 방법은, 조건 이미지와 예제 이미지를 이용하여 초기 정합을 수행하는 단계; 상기 초기 정합 수행 결과를 기초로 픽셀 이동된 특징맵을 생성하는 단계; 상기 픽셀 이동된 특징맵에 노이즈를 추가한 후 확산 모델에 적용하여 역 확산 프로세스를 수행하는 단계; 상기 역 확산 프로세스의 중간 결과와 상기 픽셀 이동된 예제 이미지를 재정합한 후 재정합 결과를 기초로 매칭 신뢰도가 임계치 이상인 영역만 픽셀을 재이동시킨 후 노이즈를 추가하여 상기 확산 모델에 적용하여 상기 역 확산 프로세스를 다시 수행하는 단계를 반복 수행한 후 상기 예제 이미지의 스타일이 반영된 최종 번역 이미지를 생성할 수 있다.

대표도 - 도1



(52) CPC특허분류

G06T 5/70 (2024.01)

(72) 발명자

이규성

경기도 안양시 동안구 관악대로 121, 108동 2204호
(비산동, 비산 삼성 래미안)

조석주

서울특별시 성북구 성북로16가길 7-6(성북동)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711193663
과제번호	2020-0-01819-004
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성(R&D)
연구과제명	ICT명품인재양성(고려대학교)
기 여 율	1/1
과제수행기관명	고려대학교산학협력단
연구기간	2023.01.01 ~ 2023.12.31
공지예외적용	: 있음

명세서

청구범위

청구항 1

- (a) 조건 이미지와 예제 이미지를 입력받는 단계;
- (b) 상기 조건 이미지와 상기 예제 이미지를 이용하여 초기 정합을 수행하는 단계;
- (c) 상기 초기 정합 수행 결과를 기초로 픽셀 이동된 특징맵을 생성하는 단계;
- (d) 상기 픽셀 이동된 특징맵에 노이즈를 추가한 후 확산 모델에 적용하여 역 확산 프로세스를 수행하는 단계;
- (e) 상기 역 확산 프로세스의 중간 결과와 상기 픽셀 이동된 예제 이미지를 재정합한 후 재정합 결과를 기초로 매칭 신뢰도가 임계치 이상인 영역만 픽셀을 재이동시킨 후 노이즈를 추가하여 상기 확산 모델에 적용하여 상기 역 확산 프로세스를 다시 수행하는 단계; 및
- (f) 상기 (e) 단계를 복수회 반복 수행하여 상기 예제 이미지의 스타일이 반영된 최종 번역 이미지를 생성하는 단계를 포함하되,

상기 정합을 수행하는 단계는,

상기 조건 이미지를 제1a 인코더에 적용하여 임베딩 조건 특징맵을 생성하는 단계;

상기 예제 이미지를 제1b 인코더에 적용하여 임베딩 예제 특징맵을 생성하는 단계;

상기 임베딩 조건 특징맵을 제2a 인코더에 적용하여 영역 불변 조건 특징맵을 생성하는 단계;

상기 임베딩 조건 특징맵을 제2b 인코더에 적용하여 영역 불변 예제 특징맵을 생성하는 단계; 및

상기 영역 불변 조건 특징맵과 상기 영역 불변 예제 특징맵을 이용하여 상관 관계 맵을 생성하는 단계를 포함하는 것을 특징으로 하는 예제 이미지 기반 이미지 번역 방법.

청구항 2

삭제

청구항 3

제1 항에 있어서,

상기 임계치 미만인 영역은 픽셀을 재이동을 건너뛰는 것을 특징으로 하는 예제 이미지 기반 이미지 번역 방법.

청구항 4

제1 항에 있어서,

상기 (e) 단계는,

상기 역 확산 프로세스에 따른 중간 결과와 상기 픽셀 이동된 특징맵을 혼합하여 중간 불변 특징맵을 생성하는 단계; 및

상기 중간 불변 특징맵과 상기 영역 불변 예제 특징맵을 이용하여 중간 상관 관계 맵을 도출한 후 상기 중간 상관 관계 맵을 이용하여 상기 매칭 신뢰도가 임계치 이상인 영역만 픽셀을 재이동시키는 것을 특징으로 하는 예제 이미지 기반 이미지 번역 방법.

청구항 5

제1 항에 있어서,

상기 상관 관계 맵에 따라 상기 임베딩 예제 특징맵을 소프트맥스 함수에 적용하여 픽셀 이동시키는 것을 특징으로 하는 예제 이미지 기반 이미지 번역 방법.

청구항 6

제1 항, 제3 항 내지 제5 항 중 어느 하나의 항에 따른 방법을 수행하기 위한 프로그램 코드를 기록한 컴퓨터로 판독 가능한 기록매체.

청구항 7

적어도 하나의 명령어를 저장하는 메모리; 및

상기 메모리에 저장된 명령어를 실행하는 프로세서를 포함하되,

상기 프로세서에 의해 실행된 명령어는 각각,

- (a) 조건 이미지와 예제 이미지를 입력받는 단계;
- (b) 상기 조건 이미지와 상기 예제 이미지를 이용하여 초기 정합을 수행하는 단계;
- (c) 상기 초기 정합 수행 결과를 기초로 픽셀 이동된 특징맵을 생성하는 단계;
- (d) 상기 픽셀 이동된 특징맵에 노이즈를 추가한 후 확산 모델에 적용하여 역 확산 프로세스를 수행하는 단계;
- (e) 상기 역 확산 프로세스의 중간 결과와 상기 픽셀 이동된 예제 이미지를 재정합한 후 재정합 결과를 기초로 매칭 신뢰도가 임계치 이상인 영역만 픽셀을 재이동시킨 후 노이즈를 추가하여 상기 확산 모델에 적용하여 상기 역 확산 프로세스를 다시 수행하는 단계; 및
- (f) 상기 (e) 단계를 복수회 반복 수행하여 상기 예제 이미지의 스타일이 반영된 최종 번역 이미지를 생성하는 단계를 수행하되,

상기 정합을 수행하는 단계는,

상기 조건 이미지를 제1a 인코더에 적용하여 임베딩 조건 특징맵을 생성하는 단계;

상기 예제 이미지를 제1b 인코더에 적용하여 임베딩 예제 특징맵을 생성하는 단계;

상기 임베딩 조건 특징맵을 제2a 인코더에 적용하여 영역 불변 조건 특징맵을 생성하는 단계;

상기 임베딩 조건 특징맵을 제2b 인코더에 적용하여 영역 불변 예제 특징맵을 생성하는 단계; 및

상기 영역 불변 조건 특징맵과 상기 영역 불변 예제 특징맵을 이용하여 상관 관계 맵을 생성하는 단계를 포함하는 것을 특징으로 하는 예제 기반 이미지 번역 장치.

발명의 설명

기술 분야

[0001] 본 발명은 확산 모델을 통한 예제 기반 이미지 번역 방법 및 그 장치에 관한 것이다.

배경 기술

[0003] 두개의 서로 다른 도메인 간의 매핑을 학습하는 것을 목표로 하는 이미지 대 이미지 번역은 최근 많이 발전되었다. 스케치나 사람의 자세 맵(pose map) 등이 묘사된 이미지를 조건 이미지, 스타일을 가져오고자 하는 이미지를 예제 이미지라 할 때, 예제 이미지의 스타일을 가져오면서 조건 이미지의 구체적 표현과 일치하는 새로운 사

실적인 이미지를 만드는 것을 예제 기반 이미지 번역 기법이라고 한다.

[0004] 그러나 종래의 예제 기반 이미지 번역 기법들은 예제 이미지의 구체적인 스타일을 잘 반영하지 못하는 단점이 있다.

[0005] 종래 기술들은 GAN(적대적 생성 모델, 이하 GAN이라 칭하기로 함)과 의미론적 이미지 정합(semantic Matching) 기술을 이용하여 예제 기반 이미지 번역을 진행한다. 이러한 종래 기술은 의미론적 이미지 정합 과정을 단 한번만 거친 후 생성 모델에 반영하여 완전한 정합이 이루어지지 못하는 단점이 있다.

발명의 내용

해결하려는 과제

[0007] 본 발명은 예제 이미지 기반 이미지 번역 방법 및 그 장치를 제공하기 위한 것이다.

[0008] 또한, 본 발명은 확산 모델을 이용하여 의미론적 이미지 정합 과정을 여러 번 반복하여 충분한 정합이 가능하도록 할 수 있는 예제 이미지 기반 이미지 번역 방법 및 그 장치를 제공하기 위한 것이다.

과제의 해결 수단

[0010] 본 발명의 일 측면에 따르면 예제 이미지 기반 이미지 번역 방법이 제공된다.

[0011] 본 발명의 일 실시예에 따르면, (a) 조건 이미지와 예제 이미지를 입력받는 단계; (b) 상기 조건 이미지와 상기 예제 이미지를 이용하여 초기 정합을 수행하는 단계; (c) 상기 초기 정합 수행 결과를 기초로 픽셀 이동된 특징맵을 생성하는 단계; (d) 상기 픽셀 이동된 특징맵에 노이즈를 추가한 후 확산 모델에 적용하여 역 확산 프로세스를 수행하는 단계; (e) 상기 역 확산 프로세스의 중간 결과와 상기 픽셀 이동된 예제 이미지를 재정합한 후 재정합 결과를 기초로 매칭 신뢰도가 임계치 이상인 영역만 픽셀을 재이동시킨 후 노이즈를 추가하여 상기 확산 모델에 적용하여 상기 역 확산 프로세스를 다시 수행하는 단계; 및 (f) 상기 (e) 단계를 복수회 반복 수행하여 상기 예제 이미지의 스타일이 반영된 최종 번역 이미지를 생성하는 단계를 포함하는 예제 이미지 기반 이미지 번역 방법이 제공될 수 있다.

[0012] 상기 정합을 수행하는 단계는, 상기 조건 이미지를 제1a 인코더에 적용하여 임베딩 조건 특징맵을 생성하는 단계; 상기 예제 이미지를 제1b 인코더에 적용하여 임베딩 예제 특징맵을 생성하는 단계; 상기 임베딩 조건 특징맵을 제2a 인코더에 적용하여 영역 불변 조건 특징맵을 생성하는 단계; 상기 임베딩 조건 특징맵을 제2b 인코더에 적용하여 영역 불변 예제 특징맵을 생성하는 단계; 및 상기 영역 불변 조건 특징맵과 상기 영역 불변 예제 특징맵을 이용하여 상관 관계 맵을 생성하는 단계를 포함될 수 있다.

[0013] 상기 임계치 미만인 영역은 픽셀을 재이동을 건너뛸 수 있다.

[0014] 상기 (e) 단계는, 상기 역 확산 프로세스에 따른 중간 결과와 상기 픽셀 이동된 특징맵을 혼합하여 중간 불변 특징맵을 생성하는 단계; 및 상기 중간 불변 특징맵과 상기 영역 불변 예제 특징맵을 이용하여 중간 상관 관계 맵을 도출한 후 상기 중간 상관 관계 맵을 이용하여 상기 매칭 신뢰도가 임계치 이상인 영역만 픽셀을 재이동시킬 수 있다.

[0015] 상기 상관 관계 맵에 따라 상기 예제 임베딩 특징맵을 소프트맥스 함수에 적용하여 픽셀 이동시킬 수 있다.

[0017] 본 발명의 다른 측면에 따르면, 적어도 하나의 명령어를 저장하는 메모리; 및 상기 메모리에 저장된 명령어를 실행하는 프로세서를 포함하되, 상기 프로세서에 의해 실행된 명령어는 각각, (a) 조건 이미지와 예제 이미지를 입력받는 단계; (b) 상기 조건 이미지와 상기 예제 이미지를 이용하여 초기 정합을 수행하는 단계; (c) 상기 초기 정합 수행 결과를 기초로 픽셀 이동된 특징맵을 생성하는 단계; (d) 상기 픽셀 이동된 특징맵에 노이즈를 추가한 후 확산 모델에 적용하여 역 확산 프로세스를 수행하는 단계; (e) 상기 역 확산 프로세스의 중간 결과와 상기 픽셀 이동된 예제 이미지를 재정합한 후 재정합 결과를 기초로 매칭 신뢰도가 임계치 이상인 영역만 픽셀을 재이동시킨 후 노이즈를 추가하여 상기 확산 모델에 적용하여 상기 역 확산 프로세스를 다시 수행하는 단계; 및 (f) 상기 (e) 단계를 복수회 반복 수행하여 상기 예제 이미지의 스타일이 반영된 최종 번역 이미지를 생성하는 단계를 수행하는 것을 특징으로 하는 예제 기반 이미지 번역 장치가 제공될 수 있다.

발명의 효과

[0019] 본 발명은 예제 이미지 기반 이미지 번역 방법 및 그 장치를 제공함으로써, 확산 모델을 이용하여 의미론적 이

미지 정합 과정을 여러 번 반복하여 충분한 정합이 가능하도록 할 수 있다.

도면의 간단한 설명

- [0021] 도 1은 본 발명의 일 실시예에 따른 예제 이미지 기반 이미지 번역 방법을 나타낸 순서도.
 도 2는 종래와 본 발명의 일 실시예에 따른 예제 기반 이미지 번역 방법의 차이를 설명하기 위해 도시한 도면.
 도 3은 본 발명의 일 실시예에 따른 전체 플로우를 도시한 도면.
 도 4는 본 발명의 일 실시예에 따른 매칭 및 역 확산의 반복에 따른 예제 이미지 결과를 도시한 도면.
 도 5 내지 도 7은 본 발명의 일 실시예에 따른 조건 이미지에 따른 예제 번역 결과를 도시한 도면.
 도 8은 본 발명의 일 실시예에 따른 예제 기반 이미지 번역 방법에 대한 의사 코드를 도시한 도면.
 도 9는 본 발명의 일 실시예에 따른 이미지 번역 장치의 내부 구성을 개략적으로 도시한 블록도.

발명을 실시하기 위한 구체적인 내용

- [0022] 본 명세서에서 사용되는 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "구성된다" 또는 "포함한다" 등의 용어는 명세서상에 기재된 여러 구성 요소들, 또는 여러 단계들을 반드시 모두 포함하는 것으로 해석되지 않아야 하며, 그 중 일부 구성 요소들 또는 일부 단계들은 포함되지 않을 수도 있고, 또는 추가적인 구성 요소 또는 단계들을 더 포함할 수 있는 것으로 해석되어야 한다. 또한, 명세서에 기재된 "...부", "모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어 또는 소프트웨어로 구현되거나 하드웨어와 소프트웨어의 결합으로 구현될 수 있다.
- [0023] 이하, 첨부된 도면들을 참조하여 본 발명의 실시예를 상세히 설명한다.
- [0025] 도 1은 본 발명의 일 실시예에 따른 예제 이미지 기반 이미지 번역 방법을 나타낸 순서도이고, 도 2는 종래와 본 발명의 일 실시예에 따른 예제 기반 이미지 번역 방법의 차이를 설명하기 위해 도시한 도면이고, 도 3은 본 발명의 일 실시예에 따른 전체 플로우를 도시한 도면이고, 도 4는 본 발명의 일 실시예에 따른 매칭 및 역 확산의 반복에 따른 예제 이미지 결과를 도시한 도면이고, 도 5 내지 도 7은 본 발명의 일 실시예에 따른 조건 이미지에 따른 예제 번역 결과를 도시한 도면이며, 도 8은 본 발명의 일 실시예에 따른 예제 기반 이미지 번역 방법에 대한 의사 코드를 도시한 도면이다.
- [0026] 단계 110에서 이미지 번역 장치(100)는 예제 이미지와 조건 이미지를 입력받는다.
- [0027] 단계 115에서 이미지 번역 장치(100)는 입력된 예제 이미지와 조건 이미지를 초기 정합한다.
- [0028] 초기 정합 과정에 대해 간략하게 설명하기로 한다.

- [0029] 이하에서는 이해와 설명의 편의를 도모하기 위해 예제 이미지를 I_y 로 표기하며, 조건 이미지를 I_x 로 각각 표기하여 설명하기로 한다. 본 발명의 일 실시예에 따르면, 조건 이미지 I_x 의 콘텐츠와 예제 이미지 I_y 의 스타일을 따르는 이미지 $I_{x \leftarrow y}$ 를 생성하기 위한 것이다.
- [0030] 이를 위해, 이미지 번역 장치(100)는 조건 이미지 I_x 와 예제 이미지 I_y 에서 영역 불변 특징 S_x 와 S_y 를 각각 추출한 후 이를 일치시키는 정합 과정을 수행한다.

[0031]

우선, 이미지 번역 장치(100)은 조건 이미지 I_x 를 기학습된 제1a 인코더에 적용하여 임베딩 조건 특징맵을 추출하고, 예제 이미지 I_y 를 기학습된 제1b 인코더에 적용하여 임베딩 조건 특징맵을 추출한다. 이미지 번

역 장치(100)는 조건 이미지 I_x 와 예제 이미지 I_y 가 주어졌을 때, 사전 훈련된 인코더 모델에 적용하여 $D_x = \mathcal{E}(I_x)$ 및 $D_y = \mathcal{E}(I_y)$ 와 같이 임베딩 특징(즉, 임베딩 조건 특징맵, 임베딩 예제 특징맵)을 각각 추출한다.

[0032]

이어, 이미지 번역 장치(100)는 추가 인코더 모델을 통해 임베딩 조건 특징맵과 임베딩 예제 특징맵에서 공통 특징을 추출하여 도메인 불일치를 줄일 수 있다.

[0033]

예를 들어, 이미지 번역 장치(100)는 기학습된 제2a 인코더 모델에 임베딩 조건 특징맵을 적용하여 $S_x = \mathcal{F}_x(D_x)$ 와 같이 영역 불변 조건 특징맵을 추출할 수 있다. 또한, 이미지 번역 장치(100)는 기학습된 제2b 인코더 모델에 임베딩 예제 특징맵을 적용하여 $S_y = \mathcal{F}_y(D_y)$ 와 같이 영역 불변 예제 특징맵을 추출할 수 있다.

[0034]

이미지 번역 장치(100)는 추출된 영역 불변 특징맵 S_x 와 S_y 를 이용하여 상관 관계 맵 $C_{x \leftarrow y}$ 을 도출한다.

[0035]

예를 들어, 상관 관계 맵은 수학식 1과 같이 도출될 수 있다.

수학식 1

$$C_{x \leftarrow y}(u, v) = \frac{S_x(u)}{\|S_x(u)\|} \cdot \frac{S_y(v)}{\|S_y(v)\|}$$

[0036]

여기서, u와 v는 각각 영역 불변 조건 특징맵, 영역 불변 예제 특징맵 인덱스를 나타낸다.

[0037]

[0038]

단계 120에서 이미지 번역 장치(100)는 초기 정합 결과를 기초로 조건 이미지의 구조적 형태에 맞도록 임베딩 예제 특징맵의 픽셀을 이동하여 픽셀 이동된 특징맵을 생성할 수 있다.

[0039]

소프트맥스 작업을 통해, $C_{x \leftarrow y}$ 에 따라 임베딩 예제 특징맵(D_y)을 소프트하게 이동시킬 수 있다. 이를 수학식으로 나타내면, 수학식 2와 같이 나타낼 수 있다.

수학식 2

$$R_{\mathcal{X} \leftarrow \mathcal{Y}}(u) = \sum_v \text{softmax}(C_{\mathcal{X} \leftarrow \mathcal{Y}}(u, v) / \tau) D_{\mathcal{Y}}(v),$$

[0040]

[0041] 여기서, τ 는 소프트맥스 동작을 날카롭게 제어하는 온도 파라미터이다.

[0042] 단계 125에서 이미지 번역 장치(100)는 픽셀 이동된 특징맵에 노이즈를 추가한 후 확산 모델에 적용하여 역 확산 프로세스를 수행한다.

[0043] 초기-워핑이 주어지면, 확산 모델의 정방향 프로세스에 따라 적절한 양의 노이즈를 추가하여 신뢰할 수 없는 대응에서 발생할 수 있는 원치 않는 인공물과 왜곡을 완화하는 동시에 이동된 특징(warped feature)의 구조적 정보를 보존할 수 있다.

[0044] 이어, 단계 130에서 이미지 번역 장치(100)는 역 확산 프로세스의 중간 결과와 픽셀 이동된 특징맵을 재정합한 후 재정합 결과를 기초로 매칭 신뢰도가 임계치 이상인 영역만 픽셀 이동한 후 노이즈를 추가하여 확산 모델에 적용하여 역 확산 프로세스를 다시 수행할 수 있다. 해당 단계 130은 복수회 반복될 수 있다.

[0045] 도 2의 (a)에 도시된 바와 같이, 종래의 적대적 생성 모델을 이용하는 경우, 한번의 정합 과정만을 거치므로 완전한 정합이 이루어지지 못하는 단점이 있었다. 그러나, 도 2의 (b)에 도시된 바와 같이, 본 발명의 일 실시예에 따른 경우, 확산 모델을 기반으로 복수의 의미론적 정합 과정을 반복하여 매칭 정확성을 높이고 더 좋은 결과를 도출할 수 있는 이점이 있다.

[0046] 단계 135에서 이미지 번역 장치(100)는 단계 130의 반복 수행 후 예제 이미지의 스타일이 반영된 최종 번역 이미지를 출력한다.

[0047] 이에 대해 보다 상세히 설명하기로 한다.

[0048] 이해와 설명의 편의를 도모하기 위해 확산 모델에 대해 간략하게 설명하기로 한다.

[0049] 확산 모델을 생성하면 점진적 노이즈 프로세스를 반전시켜 정규 분포에서 사실적인 이미지를 생성할 수 있다.

정방향 프로세스 $q(\cdot)$ 는 데이터 $x_0 \sim q(x_0)$ 에서 점진적으로 가우시안 분포로 변하는 마코프 체인이다. 정방향 프로세스의 한 단계는

$$q(x_t | x_{t-1}) := \mathcal{N}(x_t; \sqrt{1 - \beta_t} x_{t-1}, \beta_t \mathbf{I})$$

로 정의될 수 있다. 여기서, β_t 는 T 단계에서 미리 정의된 분산 스케줄이며, 정방향 프로세스는 닫힌 형식으로 임의의 타임

스텝 t에서 x_t 를 샘플링할 수 있다.

[0050] 이를 수학식으로 나타내면, 수학식 3과 같다.

수학식 3

$$x_t = \sqrt{\alpha_t}x_0 + \sqrt{1 - \alpha_t}\epsilon,$$

$$\alpha_t := \prod_{s=1}^t (1 - \beta_s), \quad \epsilon \sim \mathcal{N}(0, \mathbf{I}).$$

[0051]

[0052] 또한, 심층 신경망을 사용하여 매개변수화 할 수 있는 역 과정은

$$p_{\theta}(x_{t-1}|x_t) := \mathcal{N}(x_{t-1}; \mu_{\theta}(x_t, t), \sigma_{\theta}(x_t, t)\mathbf{I})$$

로 정의될 수 있다.

[0053]

DDPM 모델은 이전 노이즈를 데이터로 점차적으로 변환하는 데 $\mu_{\theta}(x_t, t)$ 를 사용하는 대신 노이즈 근사 모델 $\epsilon_{\theta}(x_t, t)$ 을 사용한다.

[0054]

따라서, 확산 모델의 샘플링은 수학식 4와 같다.

수학식 4

$$x_{t-1} = \frac{1}{\sqrt{1 - \beta_t}} \left(x_t - \frac{\beta_t}{\sqrt{1 - \alpha_t}} \epsilon_{\theta}(x_t, t) \right) + \sigma_t \epsilon.$$

[0055]

[0056] 최근 잠재 확산 모델(LDM:Latent Diffusion Models)은 잠재 공간에서 확산 모델을 학습하여 계산 비용을 줄이고

있다. 이미지를 잠재 공간에 삽입하기 위해 사전 훈련된 인코더 $\mathcal{E}$ 를 채택하고, 이미지를 재구성하기 위해 사

전 훈련된 디코더 $\mathcal{D}$ 를 채택한다. 이러한 잠재 확산 모델에서는 $\mathcal{X}$ 대신 $z = \mathcal{E}(x)$ 를 사용하여 확산 프로세스를 정의한다.

[0057]

디노이징 확산 암시적 모델(DDIM:Denoising Diffusion Implicit Models, 이하 DDIM이라 칭하기로 함)은 일부 신경 상미분방정식(ODE)의 오일러 이산화를 사용하여 빠르고 결정론적인 샘플링을 가능하게 하므로, 잠재 확산 모델도 DDIM 샘플링 프로세스를 채택했다.

[0058]

직관적으로 DDIM 샘플러는 z_t 에서 직접 z_0 을 예측한 다음 역 조건부 분포를 통해 z_{t-1} 을 생성한

다. 구체적으로 $f_{\theta}(z_t, t)$ 는 주어진 z_t 및 t 에 대한 z_0 의 예측이다.

[0059] 이를 수학식으로 나타내면 수학식 5와 같다.

수학식 5

$$f_{\theta}(z_t, t) := \frac{z_t - \sqrt{1 - \alpha_t} \epsilon_{\theta}(z_t, t)}{\sqrt{\alpha_t}}$$

[0060]

[0061] 잠재적 확산 모델에서 DDIM의 결정적 샘플링 프로세스는 수학식 6과 같이 나타낼 수 있다.

수학식 6

$$z_{t-1} = \sqrt{\alpha_{t-1}} f_{\theta}(z_t, t) + \sqrt{1 - \alpha_{t-1}} \epsilon_{\theta}(z_t, t)$$

[0062]

[0063] 확산 과정 후에 $x = \mathcal{D}(z)$ 가 되도록 이미지가 복원될 수 있다.

[0064] 이러한 확산 모델에 조건을 부여하기 위해, 조건부 생성이 제안되었다.

[0065] 조건부 기법은 예제 기반 이미지 번역에 바로 적용될 수 없다.

[0066] 따라서, 확산 과정에서 $R_{\mathcal{X} \leftarrow \mathcal{Y}}$ 를 확산 모델에 적용하여 확산 과정에서 어느 정도 전달하고 적절한 N 으로 노이즈 있는 잠재 변수 $r_{\mathcal{Y}}^N$ 를 획득할 수 있다. 이후, 이미지 번역 장치(100)는 가속 생성 프로세스에 따라 반복적으로 노이즈를 제거한다. 이를 수학식으로 나타내면, 수학식 7과 같다.

수학식 7

$$r_{\mathcal{Y}}^n = \begin{cases} \tilde{r}_{\mathcal{Y}}^1 & (n = 0) \\ \sqrt{\alpha_{\tau_N}} R_{\mathcal{X} \leftarrow \mathcal{Y}} + \sqrt{1 - \alpha_{\tau_N}} \epsilon & (n = N) \\ \sqrt{\alpha_{\tau_n}} \tilde{r}_{\mathcal{Y}}^{n+1} + \sqrt{1 - \alpha_{\tau_n}} \epsilon_{\theta}(r_{\mathcal{Y}}^{n+1}, \tau_{n+1}) & (o.w.) \end{cases}$$

[0067]

[0068] 여기서, $\tilde{r}_{\mathcal{Y}}^n = f_{\theta}(r_{\mathcal{Y}}^n, \tau_n)$, $\epsilon \sim \mathcal{N}(0, \mathbf{I})$ 이고, $\{\tau_n\}$ 은 역 확산 프로세스의 시간 간격에 대한 서브시퀀스를 나타내고, 역 확산 프로세스의 전체 스텝의 개수는 $\{\tau_n\}$ 의 길이인 T로 줄어든다. $N \in (0, T)$ 는 역 확산 프로세스를 시작하기 위한 중간 단계이다.

[0069] 포워딩 확산 모델(확산 U-net) 및 매칭 모듈을 반복함으로써 정제된 잠재변수 r_y^0 를 획득할 수 있다.

[0070] 우선, r_y^N 를 얻기 위해 초기 픽셀 이동된 예제 $R_{x \leftarrow y}$ 를 노이즈 프로세스에 공급한다. 그런 다음

샘플링 프로세스의 한 단계에 추가하여 완전 노이즈가 제거된 예측 $\tilde{r}_y^N$ 를 획득한다. 수학적 5의 확산 모델 (DDIM)의 비-마코프 속성에 의해 완전히 노이즈가 제거된 $\tilde{r}_y^N$ 를 획득할 수 있다.

[0071] 인터리빙 대응 및 확산 과정에서 이를 가로채고 더 나은 픽셀 이동된 이미지를 생성하고, 수학적 6의 분포를 사용하여 노이즈 제거 궤적으로 돌아간다.(도 3 참조)

[0072] 이미지 번역 장치(100)는 각 단계에서 더 나은 대응을 달성하기 위해, S_y 과 $\tilde{r}_y^n$ 사이의 상관관계를 각각 계산한다. 이를 위해, 수학적 8과 같이 정의된 특징맵을 추출한다.

수학적 8

$$S_y^{\text{iter}} = \mathcal{F}_y^{\text{iter}}(\tilde{r}_y^n, D_x)$$

[0073]

[0074] 여기서, $\mathcal{F}_y^{\text{iter}}(\cdot)$ 는 반복을 위해 설계된 특징 추출기이며, 정제된 픽셀 이동된 예시 이미지 $\tilde{r}_y^n$ 와 조건 특징 D_x 을 입력받아 공간 적응 정규화를 사용하여 혼합한다.

[0075] $\tilde{r}_y^n$ 도 실제 분포에서 나온 것이므로, $\mathcal{F}_y^{\text{iter}}$ 대신 $\tilde{r}_y^n$ 를 $\mathcal{F}_y$ 에 공급할 수 있다. 그럼에도 불구하고, D_x 를 주입하는 경우 특징을 정렬하고 더 정확한 대응 관계를 구축하는데 도움이 될 수 있다.

[0076] 그런 다음, 이미지 번역 장치(100)는 S_y 와 S_y^{iter} 사이의 상관관계 맵 $C_{x \leftarrow y}^{\text{iter}}$ 를 계산하고,

$R_{\tilde{r}_y^n \leftarrow y}$ 를 추출할 수 있다. 수학적 6에 따라 $R_{\tilde{r}_y^n \leftarrow y}$ 를 노이즈 제거 궤적으로 되돌리면, r_y^{n-1} 를 얻을 수 있다. 상술한 과정을 반복하여 최종적으로 r_y^0 를 얻을 수 있다.

[0077] 다시 정리하면, $1 \leq n < N$ 에 대해 수학적 7에서 $\tilde{r}_y^{n+1}$ 를 수학적 9와 같이

$$R_{\tilde{r}_y^{n+1} \leftarrow y}$$

로 변경한다.

수학식 9

$$r_y^n = \sqrt{\alpha_{\tau_n}} R_{\tilde{r}_y^{n+1} \leftarrow y} + \sqrt{1 - \alpha_{\tau_n}} \epsilon_{\theta}(r_y^{n+1}, \tau_{n+1}).$$

[0078]

[0079] 예를 들어, 예제 이미지에 존재하지 않는 귀결이가 있는 조건 이미지의 경우, 예제의 세부 사항을 충실하게 가져오는 것과 조건 이미지에 일치하는 이미지를 생성하는 것 사이에는 트레이드-오프(trade-off) 관계가 있다. 이를 해결하기 위해, 본 발명의 일 실시예에서는 신뢰 기반 마스킹 기법을 이용한다.

[0080] 즉, 이미지 번역 장치(100)는 각 워핑 단계(warping step)에서 신뢰도 매칭으로 주기 일관성을 활용한다. 신뢰도 마스크를 수학식 10과 같이 정의한다.

수학식 10

$$M_{\tilde{r}_y^n \leftarrow y}(u) = \mathbb{1}(\|u - \psi_{y \leftarrow \tilde{r}_y^n}(\psi_{\tilde{r}_y^n \leftarrow y}(u))\|_2^2 < \gamma)$$

[0081]

[0082] 여기서, ψ 는 워핑 함수를 나타내고, γ 는 임계값이다. 신뢰도 마스크 $M_{\tilde{r}_y^n \leftarrow y}$ 를 사용하여 신뢰하는 영역만 리워핑하고 나머지 영역은 수학식 9의 리워핑 과정을 건너뛸 수 있다. 이를 수학식으로 나타내면 수학식 11과 같다.

수학식 11

$$r_y^n = \sqrt{\alpha_{\tau_n}} (M_{\tilde{r}_y^n \leftarrow y} \odot R_{\tilde{r}_y^{n+1} \leftarrow y} + (1 - M_{\tilde{r}_y^n \leftarrow y}) \odot \tilde{r}_y^{n+1}) + \sqrt{1 - \alpha_{\tau_n}} \epsilon_{\theta}(r_y^{n+1}, \tau_{n+1}),$$

[0083]

[0084] 즉, 매칭 신뢰도가 낮은 영역은 일반적인 확산 모델의 역 과정을 따르게 할 수 있다. 매칭 신뢰도에 따라 생성력을 선택적으로 제어할 수 있는 이점이 있다.

[0085] 최종적으로 이미지 번역 장치(100)는 $I_{x \leftarrow y} = \mathcal{D}(r_y^0)$ 와 같은 이미지 공간에 잠재 변수를 리턴하여 번역된 이미지를 생성한다(도 4 참조). 이에 대한 의사 코드는 도 8에 도시된 바와 같다.

[0086] 손실 함수에 대해 간략하게 설명하기로 한다.

[0087] 조건 이미지 I_x 의 실제 이미지를 I'_x 로 나타내기로 한다. 추출된 공통 특징 $\mathcal{S}_x$ 와 $\mathcal{S}'_x$ 이 동일한 도메인에 있는지를 확인해야 한다. 따라서, 도메인 손실은 수학식 12와 같이 계산될 수 있다.

수학식 12

$$\mathcal{L}_{\text{dom}} = \|S'_{\mathcal{X}} - S_{\mathcal{X}}\|_1$$

[0088]

[0089] 또한, 픽셀 이동된 특징은(warped feature) 주기 일관성을 가져야 한다. 이는 픽셀 이동된 특징에서 예제를 리턴할 수 있어야 함을 의미한다. 인터리브된 워핑 및 생성 프로세스로 인해, n번째 단계마다 순환-워핑된 특징을 획득할 수 있다.

[0090] 해당 손실함수는 수학식 13과 같이 계산될 수 있다.

수학식 13

$$\mathcal{L}_{\text{cycle}} = \sum_n \|R_{\mathcal{Y} \leftarrow \tilde{r}_y^{n+1} \leftarrow \mathcal{Y}} - D_{\mathcal{Y}}\|_1$$

[0091]

[0092] 여기서, $R_{\mathcal{Y} \leftarrow \tilde{r}_y^{n+1} \leftarrow \mathcal{Y}}$ 는 n 단계에서 순환 워핑된 참조 특징을 나타낸다.

[0093] 상관 관계 맵을 가지고 실제 특징 $\mathcal{D}'_{\mathcal{X}}$ 를 픽셀 이동시켰을 때, $R_{\mathcal{X} \leftarrow \mathcal{X}'}$ 를 획득할 수 있으며,

이는 실측 특징 $\mathcal{D}'_{\mathcal{X}}$ 와 시맨틱 측면에서 일치하여 수학식 14과 같이 소스-조건 손실 $\mathcal{L}_{\text{src}}$ 를 도출할 수 있다.

수학식 14

$$\mathcal{L}_{\text{src}} = \|\phi_l(I_{\mathcal{X} \leftarrow \mathcal{X}'} - I'_{\mathcal{X}})\|_1$$

[0094]

[0095] 여기서, ϕ_l 는 학습된 신경망 모델(VGG-19)의 1번째 활성화 계층을 나타낸다.

[0096] 생성된 이미지의 시맨틱이 조건 이미지 $I_{\mathcal{X}}$ 또는 실제 이미지 $I'_{\mathcal{X}}$ 와 일치해야 하므로 시맨틱 유사성을 최대화하기 위한 지각 손실이 계산된다. 이를 수학식으로 나타내면 수학식 15와 같다.

수학식 15

$$\mathcal{L}_{\text{perc}} = \|\phi_l(I_{\mathcal{X} \leftarrow \mathcal{Y}}) - \phi_l(I'_{\mathcal{X}})\|_1$$

[0097]

[0098]

또한, 생성된 이미지 $I_{\mathcal{X} \leftarrow \mathcal{Y}}$ 가 예제 이미지 $I'_{\mathcal{X}}$ 에 시맨틱적으로 대응하는 패치와 스타일 일관성을 가져야 한다. 따라서, 컨텍스트 손실을 스타일 손실로 도출할 수 있다. 이를 수학식으로 나타내면, 수학식 16과 같다.

수학식 16

$$\mathcal{L}_{\text{style}} = -\log \left(\sum_l \mu_l \text{CX}_{ij}(\phi_l(I_{\mathcal{X} \leftarrow \mathcal{Y}}), \phi_l(I_{\mathcal{Y}})) \right)$$

[0099]

[0100]

여기서, CX_{ij} 는 이미지간의 컨텍스트 유사성 함수를 나타낸다.

[0101]

이미지 도메인의 고품질 출력 이미지를 생성하기 위해 사전 훈련된 확산 모델을 미세 조정한다. 확산 모델의 목표 함수는 수학식 17과 같다.

수학식 17

$$\mathcal{L}_{\text{diff}} = \sum_n \|\epsilon_{\theta}(r_{\mathcal{Y}}^{n+1}, \tau_{n+1}) - \epsilon\|_2,$$

[0102]

[0103]

여기서, ϵ 는 확산 과정의 정방향 프로세스에서 사용되는 랜덤 노이즈를 나타낸다.

[0104]

도 5 내지 도 7은 본 발명의 일 실시예에 따른 예제 기반 이미지 번역 결과를 도시한 도면이다. 도 5는 스케치를 조건 이미지로, 얼굴 이미지를 예시 이미지로 번역한 결과를 도시한 도면이고, 도 6은 분류 맵을 조건 이미지로, 성당 이미지를 예시 이미지로 번역한 결과를 도시한 도면이며, 도 7은 자세(관절) 이미지를 조건 이미지로, 모델의 전신 이미지를 예시 이미지로 하여 번역한 결과를 도시한 도면이다.

[0105]

본 발명의 일 실시예와 같이 픽셀 이동된 특징맵에 반복적으로 노이즈를 추가하고 이를 제거하여 번역된 이미지를 생성함으로써 잠재 공간에서 도메인간 매칭 및 확산 단계를 인터리빙하여(즉, 재정합 과정과 역 확산 과정의 반복을 통해) 매칭 정확도가 높은 더 좋은 결과가 도출되는 것을 알 수 있다.

[0107]

도 9는 본 발명의 일 실시예에 따른 이미지 번역 장치의 내부 구성을 개략적으로 도시한 블록도이다.

[0108]

도 9를 참조하면, 본 발명의 일 실시예에 따른 이미지 번역 장치(100)는 메모리(910) 및 프로세서(920)를 포함하여 구성된다.

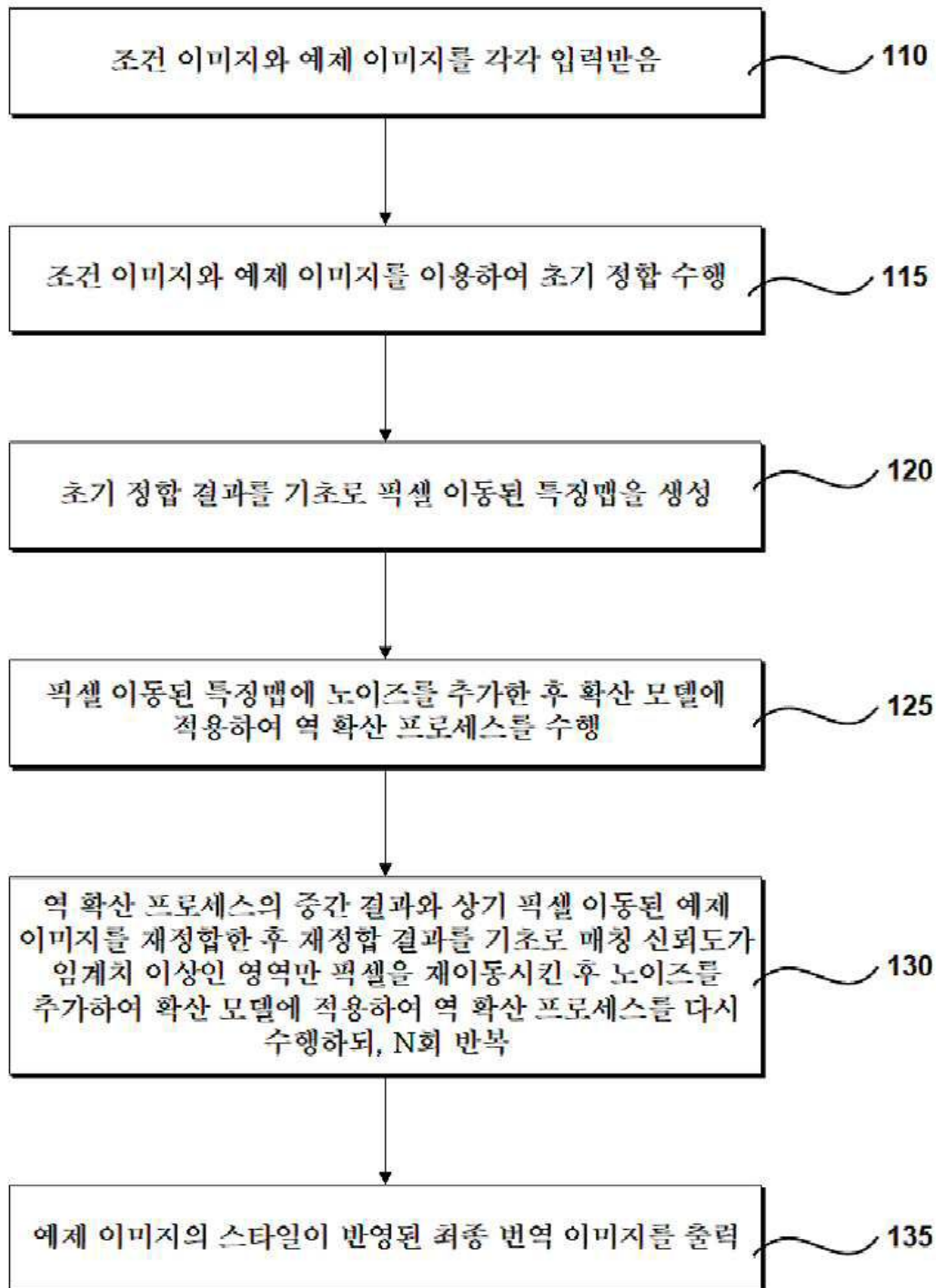
[0109]

메모리(910)는 본 발명의 일 실시예에 따른 예제 기반 이미지 번역 방법을 수행하기 위한 다양한 명령어를 저장한다.

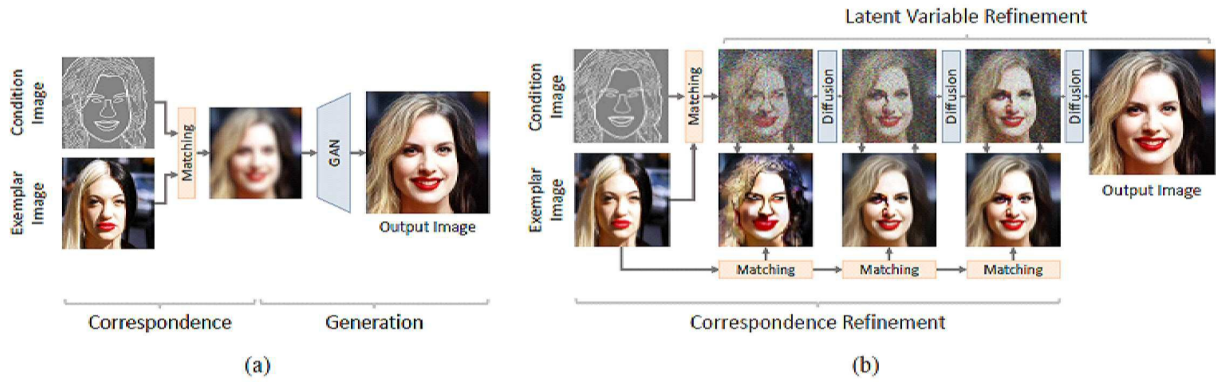
- [0110] 프로세서(920)는 본 발명의 일 실시예에 따른 이미지 번역 장치(100)의 내부 구성 요소들(910)을 제어하기 위한 수단이다. 또한, 프로세서(920)는 메모리(910)에 저장된 명령어를 실행하되, 프로세서(920)에 의해 실행된 명령어는 도 1에서 설명된 각각의 단계를 수행할 수 있다. 이는 중복되므로 중복되는 설명은 생략하기로 한다.
- [0112] 본 발명의 실시 예에 따른 장치 및 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 컴퓨터 판독 가능 매체에 기록되는 프로그램 명령은 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 분야 통상의 기술자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media) 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.
- [0113] 상술한 하드웨어 장치는 본 발명의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.
- [0114] 이제까지 본 발명에 대하여 그 실시 예들을 중심으로 살펴보았다. 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명이 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 구현될 수 있음을 이해할 수 있을 것이다. 그러므로 개시된 실시 예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 한다. 본 발명의 범위는 전술한 설명이 아니라 특허청구범위에 나타나 있으며, 그와 동등한 범위 내에 있는 모든 차이점은 본 발명에 포함된 것으로 해석되어야 할 것이다.

도면

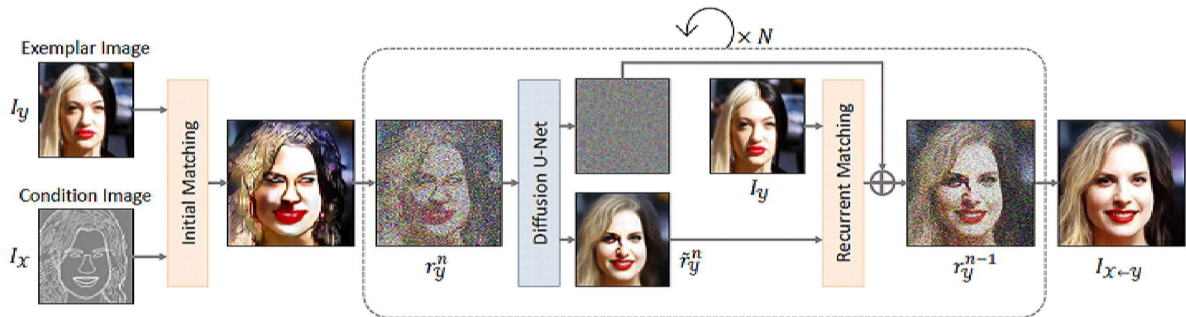
도면1



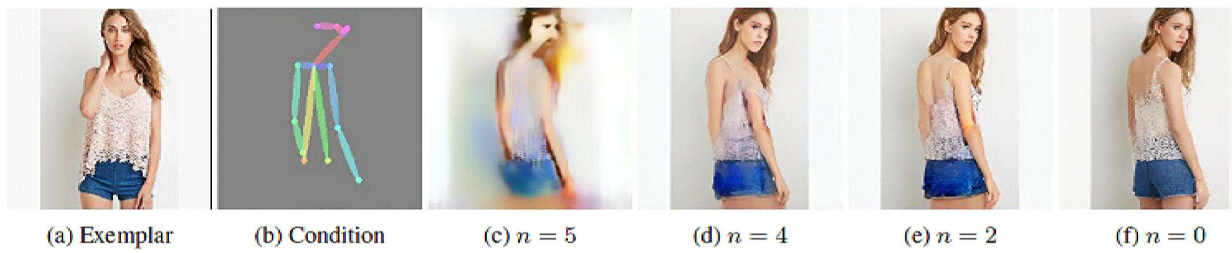
도면2



도면3



도면4



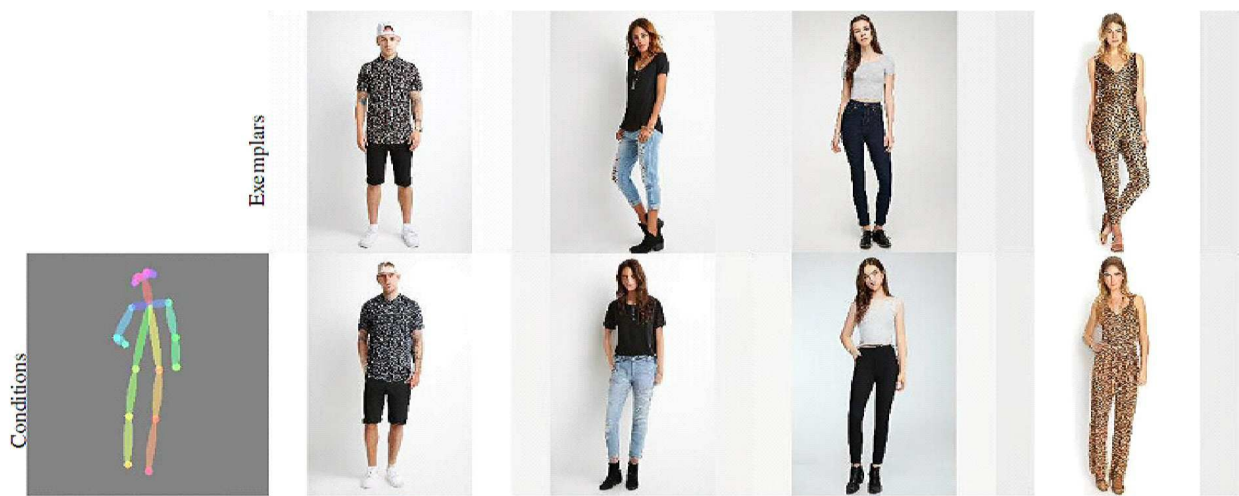
도면5



도면6



도면7



도면8

Algorithm 1: MIDMs' sampling, given a diffusion model $\epsilon_\theta(x_t)$, pretrained encoder $\mathcal{E}$, pretrained decoder $\mathcal{D}$, feature extractor $\mathcal{F}_X$ for condition, $\mathcal{F}_Y$ for exemplar and $\mathcal{F}_Y^{\text{iter}}$ for recurrent matching.

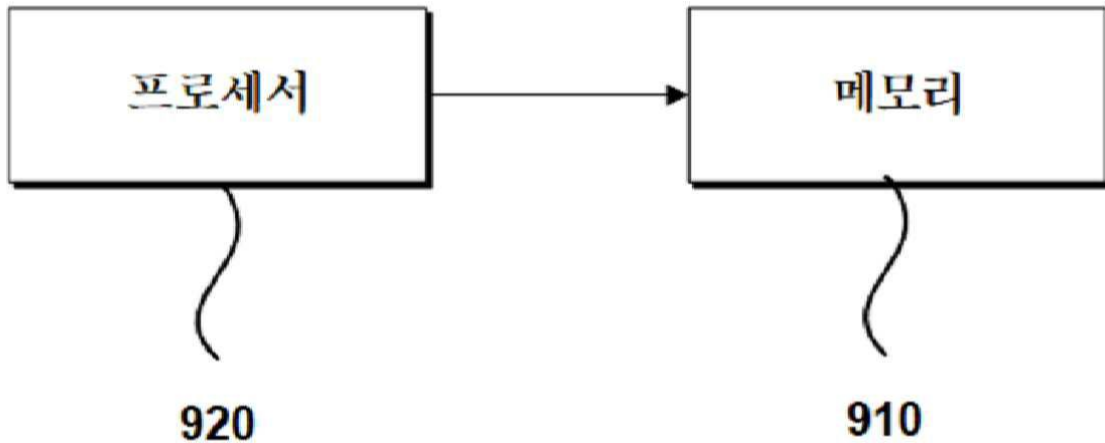
```

1: Input: conditional image  $I_X$  and exemplar image  $I_Y$ 
2: Output: generated image  $I_{X \leftarrow Y}$ 
3:  $\text{warp}(\cdot)$  : soft-warping function stated in Eq.6
4:  $\text{confidence\_mask}(\cdot)$  : confidence mask stated in Eq.10
5:  $D_X, D_Y \leftarrow \mathcal{E}(I_X), \mathcal{E}(I_Y)$ 
6:  $S_X, S_Y \leftarrow \mathcal{F}_X(D_X), \mathcal{F}_Y(D_Y)$ 
7:  $C_{X \leftarrow Y} \leftarrow \text{normalize}(S_X) \cdot \text{normalize}(S_Y)^T$ 
8:  $R_{X \leftarrow Y} \leftarrow \text{warp}(C_{X \leftarrow Y}, D_Y)$ 
9:  $\epsilon \leftarrow \text{sample from } \mathcal{N}(0, \mathbf{I})$ 
10:  $r_Y^N \leftarrow \sqrt{\alpha_{\tau_N}} R_{X \leftarrow Y} + \sqrt{1 - \alpha_{\tau_N}} \epsilon$ 
11: for all  $n$  from  $N$  to 2 do
12:    $\tilde{r}_Y^n \leftarrow f_\theta(r_Y^n, \tau_n)$ 
13:    $S_Y^{\text{iter}} \leftarrow \mathcal{F}_Y^{\text{iter}}(\tilde{r}_Y^n, D_X)$ 
14:    $C_{X \leftarrow Y}^{\text{iter}} \leftarrow \text{normalize}(S_Y^{\text{iter}}) \cdot \text{normalize}(S_Y)^T$ 
15:    $R_{\tilde{r}_Y^n \leftarrow Y} \leftarrow \text{warp}(C_{X \leftarrow Y}^{\text{iter}}, D_Y)$ 
16:    $M_{\tilde{r}_Y^n \leftarrow Y} \leftarrow \text{confidence\_mask}(C_{X \leftarrow Y}^{\text{iter}})$ 
17:    $r_Y^{n-1} \leftarrow \sqrt{\alpha_{\tau_{n-1}}} (M_{\tilde{r}_Y^n \leftarrow Y} \odot R_{\tilde{r}_Y^n \leftarrow Y} + (1 - M_{\tilde{r}_Y^n \leftarrow Y}) \odot \tilde{r}_Y^n) + \sqrt{1 - \alpha_{\tau_{n-1}}} \epsilon_\theta(r_Y^n, \tau_n)$ 
18: end for
19:  $r_Y^0 \leftarrow f_\theta(r_Y^1)$ 
20:  $I_{X \leftarrow Y} \leftarrow \mathcal{D}(r_Y^0)$ 
21: return  $I_{X \leftarrow Y}$ 

```

도면9

100



【심사관 직권보정사항】

【직권보정 1】

【보정항목】 청구범위

【보정세부항목】 청구항 5

【변경전】

제1 항에 있어서,

상기 상관 관계 맵에 따라 상기 예제 임베딩 특징맵을 소프트맥스 함수에 적용하여 픽셀 이동시키는 것을 특징으로 하는 예제 이미지 기반 이미지 번역 방법.

【변경후】

제1 항에 있어서,

상기 상관 관계 맵에 따라 상기 임베딩 예제 특징맵을 소프트맥스 함수에 적용하여 픽셀 이동시키는 것을 특징으로 하는 예제 이미지 기반 이미지 번역 방법.