



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2020년08월28일

(11) 등록번호 10-2149184

(24) 등록일자 2020년08월24일

- (51) 국제특허분류(Int. Cl.)
G06T 3/40 (2006.01) **G06N 3/08** (2006.01)
G06T 1/20 (2018.01)
- (52) CPC특허분류
G06T 3/4076 (2013.01)
G06N 3/08 (2013.01)
- (21) 출원번호 10-2019-0049616
 (22) 출원일자 2019년04월29일
 심사청구일자 2019년04월29일
- (56) 선행기술조사문헌
 Xian et al. f-VAEGAN-D2 A Feature Generating Framework for Any Shot Learning. Arxiv, 2019년 3월, pp. 1-10. 1부.*
 Ye et al. SR GAN Semantic Rectifying Generative Adversarial Network for Zero shot Learning. Arxiv, 2019년 4월 15일, pp. 1-6. 1부.*
 KR1020190028235 A
 *는 심사관에 의하여 인용된 문헌

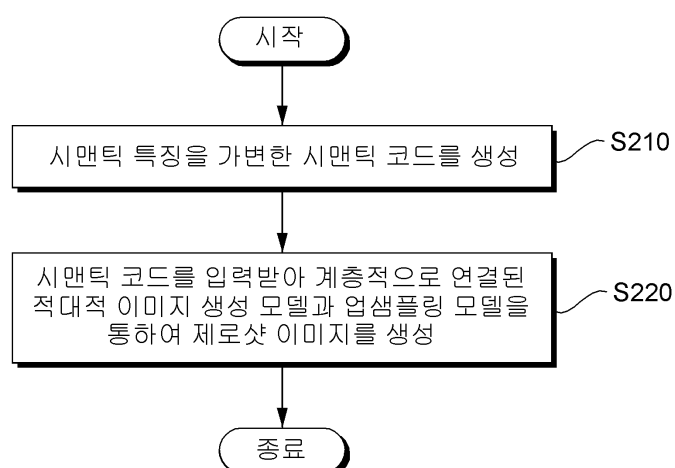
- (73) 특허권자
연세대학교 산학협력단
 서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
- (72) 발명자
변혜란
 서울특별시 서대문구 연세로 50, 제4공학관 810호(신촌동, 연세대학교)
김호성
 서울특별시 서대문구 연세로 50, 제4공학관 810호(신촌동, 연세대학교)
이제욱
 서울특별시 서대문구 연세로 50, 제4공학관 810호(신촌동, 연세대학교)
- (74) 대리인
특허법인우인

전체 청구항 수 : 총 13 항

심사관 : 김창원

(54) 발명의 명칭 **처음 보는 클래스의 설명 가능한 이미지를 생성하는 장치 및 방법****(57) 요약**

본 실시예들은 시맨틱 코드를 이용하여 시맨틱 특징을 조절하고, 계층적으로 연결된 적대적 이미지 생성 모델과 업샘플링 모델을 통하여 처음 보는(Unseen) 클래스의 이미지를 생성함으로써, 사용된(Seen) 클래스 및 처음 보는(Unseen) 클래스 간의 데이터 불균형 문제를 해결하고 제로샷 인식 성능을 향상시킬 수 있는 제로샷 이미지 생성 방법 및 장치를 제공한다.

대표도 - 도6

(52) CPC특허분류

G06T 1/20 (2013.01)

G06T 2207/20016 (2013.01)

G06T 2207/20081 (2013.01)

G06T 2207/20084 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호 2018-11-0646

부처명 과학기술정보통신부

연구관리전문기관 한국연구재단

연구사업명 원천기술개발사업

연구과제명 [Ezbaro] (2세부)딥러닝 기반 의미론적 상황 이해 원천기술 연구(1단계)(2/2)

기 여 율 1/1

주관기관 연세대학교 산학협력단

연구기간 2018.06.01 ~ 2019.05.31

명세서

청구범위

청구항 1

컴퓨팅 디바이스에 의한 제로샷 이미지 생성 방법에 있어서,

시맨틱 특징을 가변한 시맨틱 코드를 생성하는 단계; 및

상기 시맨틱 코드를 입력받아 계층적으로 연결된 적대적 이미지 생성 모델과 업샘플링 모델을 갖는 제로샷 이미지 생성 모델을 통하여 제로샷 이미지를 생성하는 단계를 포함하며,

상기 제로샷 이미지 생성 모델은 상기 시맨틱 코드와 랜덤 노이즈 분포를 입력받고 상기 제로샷 이미지를 출력하며, 제1 적대적 이미지 생성 모델에 제1 업샘플링 모델이 연결되고, 상기 제1 업샘플링 모델에 제2 적대적 이미지 생성 모델이 연결되고, 상기 제2 적대적 이미지 생성 모델에 제2 업샘플링 모델이 연결되고, 상기 제2 업샘플링 모델에 제3 적대적 이미지 생성 모델이 연결되며, 상기 제1 적대적 이미지 생성 모델, 상기 제2 적대적 이미지 생성 모델, 및 상기 제3 적대적 이미지 생성 모델을 순차적으로 거치면서 저해상도의 이미지를 고해상도의 이미지로 변환하는 것을 특징으로 하는 제로샷 이미지 생성 방법.

청구항 2

제1항에 있어서,

상기 시맨틱 코드를 생성하는 단계는,

특성 정보(Attribute)의 평균과 표준 편차를 산출하고, 정규 분포와 조절된 가우시안 분포 간의 엔트로피 변화를 산출하는 시맨틱 코드 조절 모델을 이용하여, 상기 특성 정보에 무작위성을 증가시킨 시맨틱 코드를 생성하는 것을 특징으로 하는 제로샷 이미지 생성 방법.

청구항 3

삭제

청구항 4

제1항에 있어서,

상기 적대적 이미지 생성 모델은 생성 모델과 판별 모델이 상호 작용하는 적대적 생성 신경망에서 상기 생성 모델의 생성 손실 함수와 상기 판별 모델의 판별 손실 함수를 최적화하여, 처음 보는 클래스의 이미지를 생성하는 것을 특징으로 하는 제로샷 이미지 생성 방법.

청구항 5

제4항에 있어서,

상기 생성 모델의 생성 손실 함수는 (i) 제1 조건부 생성 손실 함수, (ii) 제1 조건부 분류 손실 함수, (iii) 제1 무조건부 생성 손실 함수, 및 (iv) 제1 무조건부 분류 손실 함수가 결합되며,

상기 제1 조건부 생성 손실 함수와 상기 제1 조건부 분류 손실 함수는 특성 정보가 적용된 것을 특징으로 하는 제로샷 이미지 생성 방법.

청구항 6

제4항에 있어서,

상기 판별 모델의 판별 손실 함수는 (i) 제2 조건부 생성 손실 함수, (ii) 제2 조건부 분류 손실 함수, (iii) 제2 무조건부 생성 손실 함수, 및 (iv) 제2 무조건부 분류 손실 함수가 결합되며,

상기 제2 조건부 생성 손실 함수와 상기 제2 조건부 분류 손실 함수는 특성 정보가 적용된 것을 특징으로 하는

제로샷 이미지 생성 방법.

청구항 7

제6항에 있어서,

상기 판별 모델의 판별 손실 함수는 생성된 이미지가 실제 이미지에 수렴하도록 (i) 조건부 경사 페널티와 (ii) 무조건부 경사 페널티를 추가로 결합시켜, 상기 생성된 이미지를 정규화하는 것을 특징으로 하는 제로샷 이미지 생성 방법.

청구항 8

제1항에 있어서,

상기 업샘플링 모델은 이미지의 열과 행에 픽셀을 추가하여 상기 이미지의 크기를 증가시키는 것을 특징으로 하는 제로샷 이미지 생성 방법.

청구항 9

제1항에 있어서,

상기 제로샷 이미지를 생성하면,

제로샷 학습 모델은 상기 생성된 제로샷 이미지와 실제의 이미지를 입력받고 학습하여 처음 보는 클래스를 인식하는 것을 특징으로 하는 제로샷 이미지 생성 방법.

청구항 10

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

상기 프로세서는 시맨틱 특징을 가변한 시맨틱 코드를 생성하고,

상기 프로세서는 상기 시맨틱 코드를 입력받아 계층적으로 연결된 적대적 이미지 생성 모델과 업샘플링 모델을 갖는 제로샷 이미지 생성 모델을 통하여 제로샷 이미지를 생성하며,

상기 제로샷 이미지 생성 모델은 상기 시맨틱 코드와 랜덤 노이즈 분포를 입력받고 상기 제로샷 이미지를 출력하며, 제1 적대적 이미지 생성 모델에 제1 업샘플링 모델이 연결되고, 상기 제1 업샘플링 모델에 제2 적대적 이미지 생성 모델이 연결되고, 상기 제2 적대적 이미지 생성 모델에 제2 업샘플링 모델이 연결되고, 상기 제2 업샘플링 모델에 제3 적대적 이미지 생성 모델이 연결되며, 상기 제1 적대적 이미지 생성 모델, 상기 제2 적대적 이미지 생성 모델, 및 상기 제3 적대적 이미지 생성 모델을 순차적으로 거치면서 저해상도의 이미지를 고해상도의 이미지로 변환하는 것을 특징으로 하는 제로샷 이미지 생성 장치.

청구항 11

제10항에 있어서,

상기 프로세서는 특성 정보(Attribute)의 평균과 표준 편차를 산출하고, 정규 분포와 조절된 가우시안 분포 간의 엔트로피 변화를 산출하는 시맨틱 코드 조절 모델을 이용하여, 상기 특성 정보에 무작위성을 증가시킨 시맨틱 코드를 생성하는 것을 특징으로 하는 제로샷 이미지 생성 장치.

청구항 12

삭제

청구항 13

제10항에 있어서,

상기 적대적 이미지 생성 모델은 생성 모델과 판별 모델이 상호 작용하는 적대적 생성 신경망에서 상기 생성 모델의 생성 손실 함수와 상기 판별 모델의 판별 손실 함수를 최적화하여, 처음 보는 클래스의 이미지를 생성하는 것을 특징으로 하는 제로샷 이미지 생성 장치.

청구항 14

제10항에 있어서,

상기 업샘플링 모델은 이미지의 열과 행에 픽셀을 추가하여 상기 이미지의 크기를 증가시키는 것을 특징으로 하는 제로샷 이미지 생성 장치.

청구항 15

제10항에 있어서,

상기 프로세서는 상기 제로샷 이미지를 생성하면,

제로샷 학습 모델은 상기 생성된 제로샷 이미지와 실제의 이미지를 입력받고 학습하여 처음 보는 클래스를 인식하는 것을 특징으로 하는 제로샷 이미지 생성 장치.

발명의 설명

기술 분야

[0001] 본 실시예가 속하는 기술 분야는 제로샷 이미지 생성 장치 및 방법에 관한 것이다.

배경 기술

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

[0003] 제로샷 학습(Zero Shot Learning)은 학습 데이터에 포함되지 않은 처음 보는(Unseen) 클래스를 인식할 수 있는 기술이다. 제로샷 학습 기술의 원리는 학습에 사용된(Seen) 클래스의 데이터로부터 해당 클래스와 관련된 특성 정보(Attribute)를 올바르게 추론할 수 있도록 모델을 학습하면, 처음 보는 클래스의 데이터에 대해서도 해당 클래스의 특성 정보를 추론하여 처음 보는 클래스를 인식할 수 있다.

[0004] 기존의 제로샷 학습은 학습 데이터에 포함되지 않은 처음 보는(Unseen) 클래스들이 학습에 사용된(Seen) 클래스로 편향(bias)되는 문제가 있다.

선행기술문헌

비특허문헌

(비특허문헌 0001) Ye et al. SR GAN Semantic Rectifying Generative Adversarial Network for Zero shot Learning. Arxiv, 2019년 4월 15일, pp. 1-6.

(비특허문헌 0002) Xian et al. f-VAEGAN-D2 A Feature Generating Framework for Any Shot Learning. Arxiv, 2019년 3월, pp. 1-10.

발명의 내용

해결하려는 과제

[0005] 본 발명의 실시예들은 시맨틱 코드를 이용하여 시맨틱 특징을 조절하고, 계층적으로 연결된 적대적 이미지 생성 모델과 업샘플링 모델을 통하여 처음 보는(Unseen) 클래스의 이미지를 생성함으로써, 사용된(Seen) 클래스 및 처음 보는(Unseen) 클래스 간의 데이터 불균형 문제를 해결하고 제로샷 인식 성능을 향상하는 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 제로샷 이미지 생성 방법에 있어서, 시맨틱 특징을 가

변한 시맨틱 코드를 생성하는 단계, 및 상기 시맨틱 코드를 입력받아 계층적으로 연결된 적대적 이미지 생성 모델과 업샘플링 모델을 갖는 제로샷 이미지 생성 모델을 통하여 제로샷 이미지를 생성하는 단계를 포함하는 제로샷 이미지 생성 방법을 제공한다.

- [0008] 본 실시예의 다른 측면에 의하면, 하나 이상의 프로세서 및 상기 하나 이상의 프로세서에 의해 실행되는 하나 이상의 프로그램을 저장하는 메모리를 포함하는 제로샷 이미지 생성 장치에 있어서, 상기 프로세서는 시맨틱 특징을 가변한 시맨틱 코드를 생성하고, 상기 프로세서는 상기 시맨틱 코드를 입력받아 계층적으로 연결된 적대적 이미지 생성 모델과 업샘플링 모델을 갖는 제로샷 이미지 생성 모델을 통하여 제로샷 이미지를 생성하는 것을 특징으로 하는 제로샷 이미지 생성 장치를 제공한다.

발명의 효과

- [0009] 이상에서 설명한 바와 같이 본 발명의 실시예들에 의하면, 시맨틱 코드를 이용하여 시맨틱 특징을 조절하고, 계층적으로 연결된 적대적 이미지 생성 모델과 업샘플링 모델을 통하여 처음 보는(Unseen) 클래스의 이미지를 생성함으로써, 사용된(Seen) 클래스 및 처음 보는(Unseen) 클래스 간의 데이터 불균형 문제를 해결하고 제로샷 인식 성능을 향상시킬 수 있는 효과가 있다.
- [0010] 여기에서 명시적으로 언급되지 않은 효과라 하더라도, 본 발명의 기술적 특징에 의해 기대되는 이하의 명세서에 기재된 효과 및 그 잠정적인 효과는 본 발명의 명세서에 기재된 것과 같이 취급된다.

도면의 간단한 설명

- [0011] 도 1은 기존의 제로샷 학습 모델에 따른 시맨틱 특징 공간을 예시한 도면이다.
- 도 2는 본 발명의 일 실시예에 따른 제로샷 이미지 생성 장치를 예시한 블록도이다.
- 도 3은 본 발명의 일 실시예에 따른 제로샷 이미지 생성 장치의 제로샷 이미지 생성 모델을 예시한 블록도이다.
- 도 4는 본 발명의 일 실시예에 따른 제로샷 이미지 생성 장치의 시맨틱 코드 조절 모델을 예시한 블록도이다.
- 도 5는 본 발명의 일 실시예에 따른 제로샷 이미지 생성 장치의 적대적 이미지 생성 모델을 예시한 블록도이다.
- 도 6은 본 발명의 다른 실시예에 따른 제로샷 이미지 생성 방법을 예시한 흐름도이다.
- 도 7 및 도 8은 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] 기존의 딥러닝이 학습 데이터로 학습하지 않은 새로운 레이블을 인식할 수 없는 문제를 해결하기 위해 개발된 제로샷 학습은 다른 형식의 데이터를 이용하여 새로운 데이터를 인식할 수 있다.
- [0014] 기존의 제로샷 학습은 학습 데이터에 포함되지 않은 처음 보는(Unseen) 클래스들이 학습에 사용된(Seen) 클래스로 편향(Bias)되는 문제가 있다.
- [0015] 기존의 제로샷 학습 모델에 따른 시맨틱 특징 공간을 예시한 도 1을 참조하면, 테스트 과정에서 Groove-billed Ani라는 처음 보는(Unseen) 클래스의 새는 학습에 사용되지 않았기 때문에 사용된(Seen) 클래스인 Red-winged Blackbird로 오분류될 수 있다.
- [0016] 이를 해결하기 위하여 본 실시예들은 시맨틱 코드를 입력받아 계층적으로 연결된 적대적 이미지 생성 모델과 업샘플링 모델을 갖는 제로샷 이미지 생성 모델을 통하여 제로샷 이미지를 생성한다. 생성된 제로샷 이미지, 즉 생성된 처음 보는 클래스의 이미지를 실제의 이미지와 함께 입력받고 학습하여 제로샷 인식 성능을 향상시킬 수 있다.
- [0017] 본 실시예들은 특정 시맨틱 특징을 조절하여 설명 가능한 처음 보는 클래스의 이미지를 생성한다. 시맨틱 공간은 사용된 클래스와 처음 보는 클래스에 의해 공유될 수 있다. 시맨틱 코드는 처음 보는 클래스의 이미지를 생성하기 위해 조절된 시맨틱 특징이다.

- [0018] 시맨틱 특징은 특성 정보에 해당하며, 특성 정보(Attribute)는 클래스를 대표하는 주요 특성을 의미하다. 예를 들어, 새의 부리 길이, 날개 색상, 몸 크기 등이 있다. 또는 말의 다리 길이, 머리 모양, 꼬리 모양 등이 있다. 제로샷 학습 모델은 특성 정보를 핵심 단서로 사용하며 특성 정보는 클래스의 분류 기준으로 사용된다.
- [0019] 도 2는 본 발명의 일 실시예에 따른 제로샷 이미지 생성 장치를 예시한 블록도이다.
- [0020] 제로샷 이미지 생성 장치(110)는 적어도 하나의 프로세서(120), 컴퓨터 판독 가능한 저장매체(130) 및 통신 버스(170)를 포함한다.
- [0021] 프로세서(120)는 제로샷 이미지 생성 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 제로샷 이미지 생성 장치(110)로 하여금 예시적인 실시예에 따른 동작들을 수행하도록 구성될 수 있다.
- [0022] 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 판독 가능한 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 제로샷 이미지 생성 장치(110)에 의해 액세스되고 원하는 정보를 저장할 수 있는 다른 형태의 저장 매체, 또는 이들의 적합한 조합일 수 있다.
- [0023] 통신 버스(170)는 프로세서(120), 컴퓨터 판독 가능한 저장 매체(140)를 포함하여 제로샷 이미지 생성 장치(110)의 다른 다양한 컴포넌트들을 상호 연결한다.
- [0024] 제로샷 이미지 생성 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 제로샷 이미지 생성 장치(110)의 다른 컴포넌트들에 연결될 수 있다.
- [0025] 제로샷 이미지 생성 장치(110)는 시맨틱 코드 조절 모델과 제로샷 이미지 생성 모델을 통하여 처음 보는 클래스의 이미지인 제로샷 이미지를 생성한다.
- [0026] 도 3은 본 발명의 일 실시예에 따른 제로샷 이미지 생성 장치의 제로샷 이미지 생성 모델을 예시한 블록도이다.
- [0027] 제로샷 이미지 생성 장치는 시맨틱 특징을 가변한 시맨틱 코드를 생성하는 시맨틱 코드 조절 모델을 포함하고, 제로샷 이미지를 생성하는 제로샷 이미지 생성 모델을 포함한다. 제로샷 이미지 생성 모델은 계층적으로 연결된 적대적 이미지 생성 모델과 업샘플링 모델을 갖는다.
- [0028] 제로샷 이미지 생성 모델에서 사용되는 용어를 먼저 정의한다.
- [0029] $\mathcal{S} = \{(x, y, a^y) | x \in \mathcal{X}^s, y \in \mathcal{Y}^s, a^y \in \mathcal{A}\}$ 는 사용된 클래스의 학습 데이터 세트이다. x 는 실제 이미지이다. y 는 $\mathcal{Y}^s = \{y_1, \dots, y_{C_s}\}$ 내의 클래스 레이블이다. C_s 는 사용된 클래스의 개수이다. $a^y \in \mathbb{R}^{d_a}$ 는 클래스 y 에 대응하는 특성 정보(Attribute)와 같은 시맨틱 특징이다.
- [0030] $\mathcal{U} = \{(\tilde{x}, \tilde{y}, a^{\tilde{y}}) | \tilde{x} \in \mathcal{X}^u, \tilde{y} \in \mathcal{Y}^u, a^{\tilde{y}} \in \mathcal{A}\}$ 는 처음 보는 클래스의 테스트 데이터 세트이다. \tilde{x} 는 \mathcal{X}^u 내의 실제 이미지이다. \tilde{y} 는 $\mathcal{Y}^u = \{y_{C_s+1}, \dots, y_{C_s+C_u}\}$ 내의 클래스 레이블이다. C_u 는 처음 보는 클래스의 개수이다. $a^{\tilde{y}} \in \mathbb{R}^{d_a}$ 는 클래스 \tilde{y} 와 관련된 시맨틱 특징이다. \mathcal{A} 는 정의된 특성 정보 공간이다. GZSL(Generalized Zero-Shot Learning)은 $\mathcal{Y}^t = \mathcal{Y}^s \cup \mathcal{Y}^u$ 및 $\mathcal{Y}^s \cap \mathcal{Y}^u = \emptyset$ 조건에서 분류 모델 $f_{gzsl} : \mathcal{X} \rightarrow \mathcal{Y}^t$ 을 학습한다. 학습 데이터와 평가 데이터의 클래스 간의 교집합이 공집합으로

서로 독립 집합인 경우, 학습되지 않은 데이터라도 클래스에 맞게 해당하는 처리를 수행한다.

[0031] 적대적 이미지 생성 모델은 생성 모델과 판별 모델이 상호 작용하는 적대적 생성 신경망이다. 적대적 생성 신경망의 목적 함수는 수학식 1과 같이 표현된다.

수학식 1

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}} [\log D(x, a^y)] + \mathbb{E}_{z \sim p_z} [\log(1 - D(\hat{x}, a^y))]$$

[0032]

[0033] x 는 실제 데이터의 분산 p_{data} 로부터 추출되는 사용된 클래스 y 의 진짜 이미지이다. $\hat{x} = G(z, a^y)$ 는 사용된 클래스 y 의 가짜 이미지이다. z 는 가우시안 분산 p_z 로부터 추출된 노이즈 벡터이다. a^y 는 추가적인 조건 변수인 특성 정보 벡터이다.

[0034] 판별 모델 D 는 다수의 레이어를 갖고 마지막 레이어는 시그모이드 함수를 갖는 퍼셉트론(Perceptron)으로 구현될 수 있다. 생성 모델 G 의 목적은 손실을 최소화하고 판별 모델 D 의 목적은 손실을 최대화한다.

[0035] 제로샷 이미지 생성 장치는 시맨틱 코드를 생성하여 시맨틱 특징을 조절한다. 제로샷 이미지 생성 장치는 시맨틱 특징을 조절하여 시맨틱 특징의 무작위성을 증가시킨다.

[0036] 도 4는 본 발명의 일 실시예에 따른 제로샷 이미지 생성 장치의 시맨틱 코드 조절 모델을 예시한 블록도이다.

[0037] 제로샷 이미지 생성 장치는 시맨틱 코드 조절 모델을 이용하여 시맨틱 코드를 생성한다. 시맨틱 코드 조절 모델은 특성 정보(Attribute)의 평균과 표준 편차를 산출하고, 정규 분포와 조절된 가우시안 분포 간의 엔트로피 변화를 산출하는 방식으로 특성 정보에 무작위성을 증가시킨 시맨틱 코드를 생성한다.

[0038] 시맨틱 코드 조절 모델은 쿨백-라이블러 발산(Kullback-Leibler Divergence, KLD)을 적용하여 정규 분포와 가우시안 분포 간의 엔트로피 변화를 산출한다.

[0039] 시맨틱 코드 조절 모델은 수학식 2와 같이 표현된다.

수학식 2

$$D_{KL}(\mathcal{N}(\mu(a^y), \Sigma(a^y)) || \mathcal{N}(0, \mathcal{I}))$$

[0040]

[0041] 다시 도 3을 참조하면, 제로샷 이미지 생성 장치는 시맨틱 코드를 입력받아 계층적으로 연결된 적대적 이미지 생성 모델과 업샘플링 모델을 갖는 제로샷 이미지 생성 모델을 통하여 처음 보는 클래스의 이미지인 제로샷 이미지를 생성한다.

[0042] 제로샷 이미지 생성 모델은 시맨틱 코드와 랜덤 노이즈 분포를 입력받고 제로샷 이미지를 출력한다. 제로샷 이미지 생성 모델은 제1 적대적 이미지 생성 모델, 제2 적대적 이미지 생성 모델, 및 제3 적대적 이미지 생성 모델을 포함한다.

[0043] 제1 적대적 이미지 생성 모델에 제1 업샘플링 모델이 연결되고, 제1 업샘플링 모델에 제2 적대적 이미지 생성 모델이 연결되고, 제2 적대적 이미지 생성 모델에 제2 업샘플링 모델이 연결되고, 제2 업샘플링 모델에 제3 적대적 이미지 생성 모델이 연결된다.

[0044] 제1 적대적 이미지 생성 모델, 제2 적대적 이미지 생성 모델, 및 제3 적대적 이미지 생성 모델을 순차적으로 거치면서 저해상도의 이미지는 고해상도의 이미지로 변환된다.

[0045] 업샘플링 모델은 이미지의 열과 행에 픽셀을 추가하여 이미지의 크기를 증가시킨다. 예컨대, 제1 적대적 이미지

생성 모델은 64x64 픽셀의 이미지를 처리하고, 제1 업샘플링 모델은 64x64 픽셀의 이미지를 128x128 픽셀의 이미지로 변환하고, 제2 적대적 이미지 생성 모델은 128x128 픽셀의 이미지를 처리하고, 제2 업샘플링 모델은 128x128 픽셀의 이미지를 256x256 픽셀의 이미지로 변환하고, 제3 적대적 이미지 생성 모델은 256x256 픽셀의 이미지를 처리할 수 있다. 즉, 제로샷 이미지 생성 모델을 통해 생성된 이미지가 크기가 커지면서 동시에 구체적으로 표현된다.

[0046] 도 5는 본 발명의 일 실시예에 따른 제로샷 이미지 생성 장치의 적대적 이미지 생성 모델을 예시한 블록도이다.

[0047] 적대적 이미지 생성 모델은 생성 모델과 판별 모델이 상호 작용하는 적대적 생성 신경망에서 생성 모델의 생성 손실 함수와 판별 모델의 판별 손실 함수를 최적화하여, 처음 보는 클래스의 이미지를 생성한다.

[0048] 생성 모델의 생성 손실 함수는 수학식 3과 같이 표현된다.

수학식 3

$$L_G = \alpha D_{KL} + \sum_{i=1}^k L_{G_i}, L_{G_i} = L_{G_{i^{c_{gan}}}} + L_{G_{i^{c_{cls}}}} + L_{G_{i^{u_{gan}}}} + L_{G_{i^{u_{cls}}}}$$

[0049]

[0050] D_{KL} 는 수학식 2의 시맨틱 코드 조절 모델에 해당하고, α 는 KL 손실의 계수이고, k 는 계층의 단계 수이다. 예컨대, α 는 2이고, k 는 3으로 설정될 수 있다.

[0051] 생성 모델의 생성 손실 함수는 (i) 제1 조건부 생성 손실 함수, (ii) 제1 조건부 분류 손실 함수, (iii) 제1 무조건부 생성 손실 함수, 및 (iv) 제1 무조건부 분류 손실 함수가 결합된다.

[0052] 제1 조건부 생성 손실 함수와 제1 조건부 분류 손실 함수는 특성 정보가 조건부로 적용된다.

[0053] (i) 제1 조건부 생성 손실 함수, (ii) 제1 조건부 분류 손실 함수, (iii) 제1 무조건부 생성 손실 함수, 및 (iv) 제1 무조건부 분류 손실 함수는 수학식 4와 같이 표현된다.

수학식 4

$$\begin{aligned} L_{G_{i^{c_{gan}}}} &= \mathbb{E}_{\hat{x}_i \sim p_{G_i}} [\log(1 - D_i(\hat{x}_i, a^y))], \\ L_{G_{i^{c_{cls}}}} &= -\mathbb{E}_{\hat{x}_i \sim p_{G_i}} [\log P(y|\hat{x}_i, a^y)], \\ L_{G_{i^{u_{gan}}}} &= \mathbb{E}_{\hat{x}_i \sim p_{G_i}} [\log(1 - D_i(\hat{x}_i))], \\ L_{G_{i^{u_{cls}}}} &= -\mathbb{E}_{\hat{x}_i \sim p_{G_i}} [\log P(y|\hat{x}_i)] \end{aligned}$$

[0054]

[0055] \hat{x}_i 는 i 번째 스케일에서의 생성 모델의 분산 p_{G_i} 으로부터 추출된다.

[0056] 판별 모델의 판별 손실 함수는 수학식 5와 같이 표현된다.

수학식 5

$$\begin{aligned}
 L_D &= \sum_{i=1}^k L_{D_i}, L_{D_i} = L_{D_{i_{gan}^c}} + L_{D_{i_{cls}^c}} + L_{D_{i_{gan}^u}} + L_{D_{i_{cls}^u}}, \\
 L_{D_{i_{gan}^c}} &= \mathbb{E}_{x_i \sim p_{data_i}} [\log D_i(x_i, a^y)] + \mathbb{E}_{\hat{x}_i \sim p_{G_i}} [\log(1 - D_i(\hat{x}_i, a^y))], \\
 L_{D_{i_{cls}^c}} &= -\mathbb{E}_{x_i \sim p_{data_i}} [\log P(y|x_i, a^y)] - \mathbb{E}_{\hat{x}_i \sim p_{G_i}} [\log P(y|\hat{x}_i, a^y)], \\
 L_{D_{i_{gan}^u}} &= \mathbb{E}_{x_i \sim p_{data_i}} [\log D_i(x_i)] + \mathbb{E}_{\hat{x}_i \sim p_{G_i}} [\log(1 - D_i(\hat{x}_i))], \\
 L_{D_{i_{cls}^u}} &= -\mathbb{E}_{x_i \sim p_{data_i}} [\log P(y|x_i)] - \mathbb{E}_{\hat{x}_i \sim p_{G_i}} [\log P(y|\hat{x}_i)]
 \end{aligned}$$

[0057]

[0058] 판별 모델의 판별 손실 함수는 (i) 제2 조건부 생성 손실 함수, (ii) 제2 조건부 분류 손실 함수, (iii) 제2 무조건부 생성 손실 함수, 및 (iv) 제2 무조건부 분류 손실 함수가 결합된다.

[0059] 제2 조건부 생성 손실 함수와 제2 조건부 분류 손실 함수는 특성 정보가 조건부로 적용된다.

[0060] $P()$ 는 클래스 레이블 y 에 대해 예측되는 x_i 또는 \hat{x}_i 의 확률이다. 확률은 양립 가능성 점수를 기반으로 산출된다. x_i 는 i 번째 스케일의 진짜 이미지 분산 p_{data_i} 로부터 추출된다. \hat{x}_i 는 i 번째 스케일의 생성 모델의 분산 p_{G_i} 로부터 추출된다. k 는 계층의 단계 수이다.

[0061] 적대적 이미지 생성 모델은 특성 공간 \mathcal{A} 를 임베딩 공간으로 활용하고 클래스 y 의 양립 가능성 점수 $s(y)$ 는 내적으로 정의된다. 특성 공간은 데이터가 매핑된 특정 차원의 임베딩 공간을 의미할 수 있다.

[0062] 양립 가능성 점수는 수학식 6과 같이 표현된다.

수학식 6

$$\begin{aligned}
 s_i^c(y) &= \langle \mathbf{W}_i^c T \phi_i(x_i, a^y), a^y \rangle, \\
 s_i^u(y) &= \langle \mathbf{W}_i^u T \phi_i(x_i), a^y \rangle
 \end{aligned}$$

[0063]

[0064] $\phi_i(x_i, a^y), \phi_i(x_i) \in \mathbb{R}^k$ 는 판별 모델 D_i 에 의해 획득된 이미지 표현 벡터이다. $a^y \in \mathbb{R}^{d_a}$ 는 클래스 y 의 특성 벡터에 대해 주어진 파라미터이다. $\mathbf{W}_i^c \in \mathbb{R}^{k \times d_a}$ 는 조건부 손실을 위한 가중치 매트릭스이고 $\mathbf{W}_i^u \in \mathbb{R}^{k \times d_a}$ 는 무조건부 손실을 위한 가중치 매트릭스이다.

[0065] 가중치 매트릭스는 $\phi_i(x_i, a^y)$ 와 $\phi_i(x_i)$ 를 특성 공간에 매핑한다. 주어진 이미지 x_i 에 대해서 확률 $P(y|x_i, a^y)$ 와 $P(y|x_i)$ 는 학습용 사용된 클래스의 양립 가능성 점수의 소프트맥스 값이다. 확률은

수학식 7과 같이 표현된다.

수학식 7

$$P(y|x_i, a^y) = \frac{\exp(s_i^c(y))}{\sum_{y' \in \mathcal{Y}^s} \exp(s_i^c(y'))},$$

$$P(y|x_i) = \frac{\exp(s_i^u(y))}{\sum_{y' \in \mathcal{Y}^s} \exp(s_i^u(y'))}$$

[0066]

[0067] \hat{x}_i 에 대해서도 수학식 7과 유사하게 확률을 산출할 수 있다.

[0068] 판별 모델의 판별 손실 함수는 생성된 이미지가 실제 이미지에 수렴하도록 (i) 조건부 경사 페널티와 (ii) 무조건부 경사 페널티를 추가로 결합시켜, 생성된 이미지를 정규화한다. 실제 이미지와 생성된 이미지의 분산 유사도가 낮으면 적대적 이미지 생성 모델은 수렴할 수 없다.

[0069] 적대적 이미지 생성 모델은 실제 이미지의 내시 균형(Nash Equilibrium)을 피하기 위해 판별 모델에 경사 페널티를 부과한다.

[0070] i번째 스케일에서 조건부 손실을 위한 정규화 지수인 조건부 경사 페널티와 무조건부 손실을 위한 정규화 지수인 무조건부 경사 페널티는 수학식 8과 같이 표현된다.

수학식 8

$$R_i^c := \frac{\gamma}{2} \mathbb{E}_{x_i \sim p_{data_i}} [\|\nabla D_i(x_i, a^y)\|^2],$$

$$R_i^u := \frac{\gamma}{2} \mathbb{E}_{x_i \sim p_{data_i}} [\|\nabla D_i(x_i)\|^2]$$

[0071]

[0072] 최종적인 판별 모델의 판별 손실 함수는 수학식 9와 같이 표현된다.

수학식 9

$$L_D = \sum_{i=1}^k L_{D_i} + R_i^c + R_i^u$$

[0073]

[0074] 제로샷 이미지 생성 장치가 제로샷 이미지를 생성하면, 제로샷 학습 모델은 생성된 제로샷 이미지와 실제의 이미지를 입력받고 학습하여 처음 보는 클래스를 인식한다.

[0075] 제로샷 이미지 생성 장치가 학습용 사용된 데이터 세트 S를 수렴시키면, 생성 모델 G와 처음 보는 클래스 \tilde{y} 의 특성 정보 $a^{\tilde{y}}$ 를 이용하여 생성된 임의의 처음 보는 이미지 \hat{x} 를 생성한다.

[0076] 생성된 처음 보는 데이터 세트 $\hat{\mathcal{U}} = \{(\hat{x}, \tilde{y}, a^{\tilde{y}}) | \hat{x} \in \mathcal{X}^{\hat{u}}, \tilde{y} \in \mathcal{Y}^u, a^{\tilde{y}} \in \mathcal{A}\}$ 에 사용된 데이터 세트 \mathcal{S} 를 결합하여 전체 데이터 세트 $\mathcal{T} = \mathcal{S} \cup \hat{\mathcal{U}}$ 를 생성한다.

[0077] 제로샷 학습 모델은 특성 공간 \mathcal{A} 을 임베딩 공간으로 활용한다. 단일 클래스 y 의 양립 가능성 점수 $s(y)$ 는 수학적 식 10과 같이 내적으로 정의된다.

수학적 식 10

[0078]
$$s(y) = \langle \mathbf{W}^T \phi(x), a^y \rangle$$

[0079] $\phi(x) \in \mathbb{R}^k$ 는 분류 모델을 통해 추출된 이미지 표현 벡터이다. $a^y \in \mathbb{R}^{d_a}$ 는 클래스 y 의 특성 벡터이고, $\mathbf{W} \in \mathbb{R}^{k \times d_a}$ 는 풀리 커넥티드 레이어(Fully Connected Layer)를 갖는 가중치 매트릭스이고, 특성 공간 \mathcal{A} 에 $\phi(x)$ 가 매핑된다.

[0080] 주어진 이미지 x 의 클래스 레이블을 추론하기 위해서, 이미지는 $f_a(x) = \mathbf{W}^T \phi(x)$ 특성 표현에 투영된다. 양립 가능성 점수 $s(y)$ 는 최선의 매칭된 클래스를 선택하도록 사용될 수 있다.

수학적 식 11

[0081]
$$y^* = \arg \max_{y \in \mathcal{Y}^t} (s(y)) = \arg \max_{y \in \mathcal{Y}^t} \langle f_a(x) \rangle$$

[0082] $\mathcal{Y}^t = \mathcal{Y}^s \cup \mathcal{Y}^u$ 로 GZSL를 설정할 수 있다. 가장 높은 양립 가능성 점수를 갖는 y^* 는 예측된 클래스이다.

[0083] 제로샷 이미지 생성 장치는 시맨틱 코드를 이용하여 시맨틱 특징을 조절하고, 계층적으로 연결된 적대적 이미지 생성 모델과 업샘플링 모델을 통하여 처음 보는(Unseen) 클래스의 이미지를 생성함으로써, 사용된(Seen) 클래스 및 처음 보는(Unseen) 클래스 간의 데이터 불균형 문제를 해결하고 제로샷 인식 성능을 향상시킬 수 있다.

[0084] 도 6은 본 발명의 다른 실시예에 따른 제로샷 이미지 생성 방법을 예시한 흐름도이다. 제로샷 이미지 생성 방법은 컴퓨팅 디바이스에 의하여 수행될 수 있으며, 제로샷 이미지 생성 장치와 동일한 방식으로 동작한다.

[0085] 단계 S210에서 프로세서는 시맨틱 특징을 가변한 시맨틱 코드를 생성한다.

[0086] 시맨틱 코드를 생성하는 단계(S210)는, 특성 정보(Attribute)의 평균과 표준 편차를 산출하고, 정규 분포와 조절된 가우시안 분포 간의 엔트로피 변화를 산출하는 시맨틱 코드 조절 모델을 이용하여, 특성 정보에 무작위성을 증가시킨 시맨틱 코드를 생성한다.

[0087] 단계 S220에서 프로세서는 시맨틱 코드를 입력받아 계층적으로 연결된 적대적 이미지 생성 모델과 업샘플링 모델을 갖는 제로샷 이미지 생성 모델을 통하여 제로샷 이미지를 생성한다. 업샘플링 모델은 이미지의 열과 행에 픽셀을 추가하여 상기 이미지의 크기를 증가시킨다.

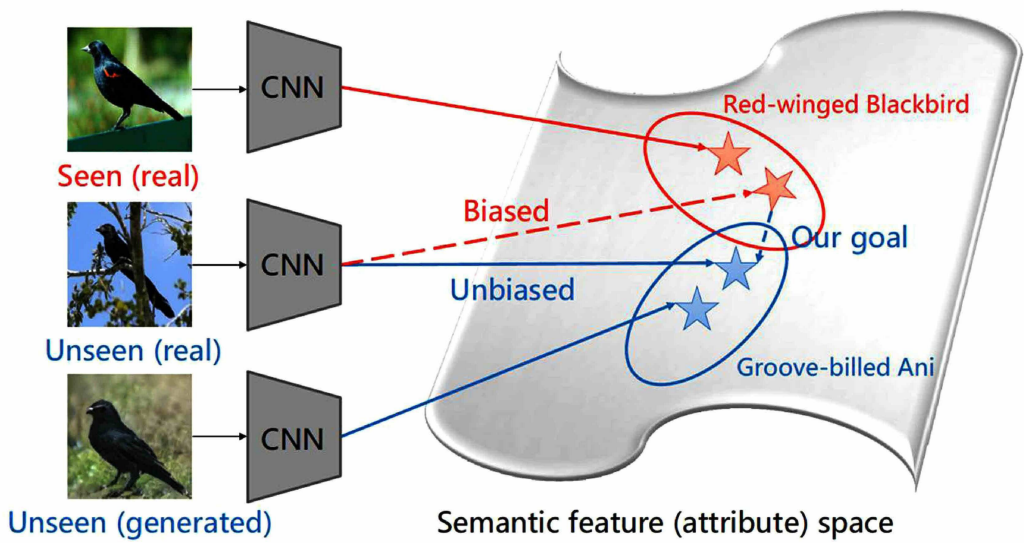
[0088] 제로샷 이미지 생성 모델은 시맨틱 코드와 랜덤 노이즈 분포를 입력받고 제로샷 이미지를 출력하며, 제1 적대적 이미지 생성 모델에 제1 업샘플링 모델이 연결되고, 제1 업샘플링 모델에 제2 적대적 이미지 생성 모델이 연결되고, 제2 적대적 이미지 생성 모델에 제2 업샘플링 모델이 연결되고, 제2 업샘플링 모델에 제3 적대적 이미지

생성 모델이 연결되며, 제1 적대적 이미지 생성 모델, 제2 적대적 이미지 생성 모델, 및 제3 적대적 이미지 생성 모델을 순차적으로 거치면서 저해상도의 이미지를 고해상도의 이미지로 변환한다.

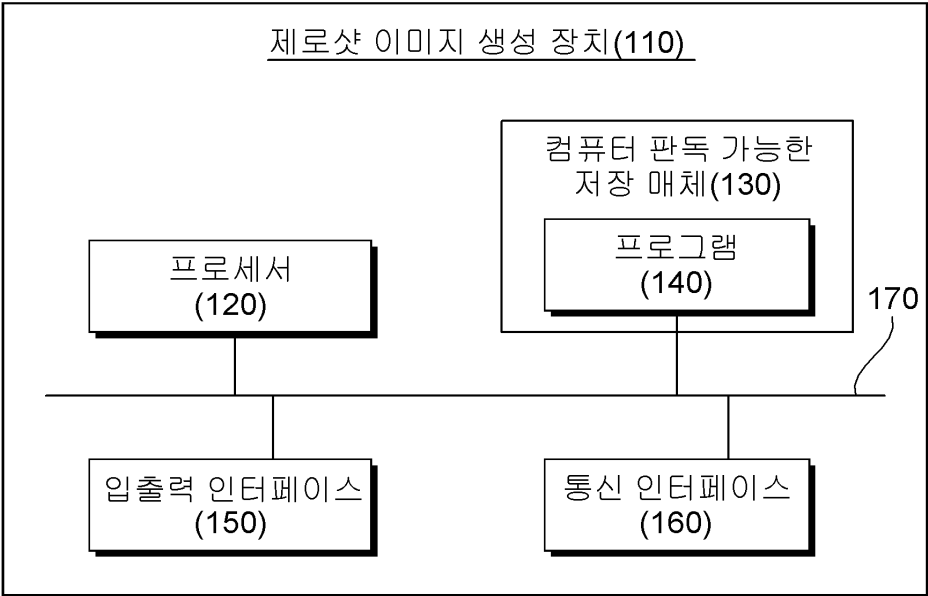
- [0089] 적대적 이미지 생성 모델은 생성 모델과 판별 모델이 상호 작용하는 적대적 생성 신경망에서 생성 모델의 생성 손실 함수와 판별 모델의 판별 손실 함수를 최적화하여, 처음 보는 클래스의 이미지를 생성한다.
- [0090] 생성 모델의 생성 손실 함수는 (i) 제1 조건부 생성 손실 함수, (ii) 제1 조건부 분류 손실 함수, (iii) 제1 무조건부 생성 손실 함수, 및 (iv) 제1 무조건부 분류 손실 함수가 결합되며, 제1 조건부 생성 손실 함수와 제1 조건부 분류 손실 함수는 특성 정보가 적용된다.
- [0091] 판별 모델의 판별 손실 함수는 (i) 제2 조건부 생성 손실 함수, (ii) 제2 조건부 분류 손실 함수, (iii) 제2 무조건부 생성 손실 함수, 및 (iv) 제2 무조건부 분류 손실 함수가 결합되며, 제2 조건부 생성 손실 함수와 제2 조건부 분류 손실 함수는 특성 정보가 적용된다.
- [0092] 판별 모델의 판별 손실 함수는 생성된 이미지가 실제 이미지에 수렴하도록 (i) 조건부 경사 페널티와 (ii) 무조건부 경사 페널티를 추가로 결합시켜, 생성된 이미지를 정규화한다.
- [0093] 제로샷 이미지를 생성하면, 제로샷 학습 모델은 생성된 제로샷 이미지와 실제의 이미지를 입력받고 학습하여 처음 보는 클래스를 인식한다.
- [0094] 도 7 및 도 8은 본 발명의 실시예들에 따른 시뮬레이션 결과를 예시한 도면이다. 도 7은 본 실시예들에 의해 생성된 설명 가능한 처음 보는 이미지의 예시이고, 도 8은 두 개의 특성 정보를 보간(Interpolation)하여 생성된 처음 보는 이미지의 예시이다. 도 8의 (a)는 외양 변화이고, 도 8의 (b)는 색깔 변화이고, 도 8의 (c)는 외양과 색깔 변화이다.
- [0095] 도 7와 도 8에 도시된 바와 같이, 생성된 처음 보는 클래스의 이미지는 실제 이미지와 매우 유사한 것을 쉽게 파악할 수 있다.
- [0096] 제로샷 이미지 생성 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.
- [0097] 제로샷 이미지 생성 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 모듈 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.
- [0098] 도 6에서는 각각의 과정을 순차적으로 실행하는 것으로 기재하고 있으나 이는 예시적으로 설명한 것에 불과하고, 이 분야의 기술자라면 본 발명의 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 도 6에 기재된 순서를 변경하여 실행하거나 또는 하나 이상의 과정을 병렬적으로 실행하거나 다른 과정을 추가하는 것으로 다양하게 수정 및 변형하여 적용 가능할 것이다.
- [0099] 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.
- [0100] 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

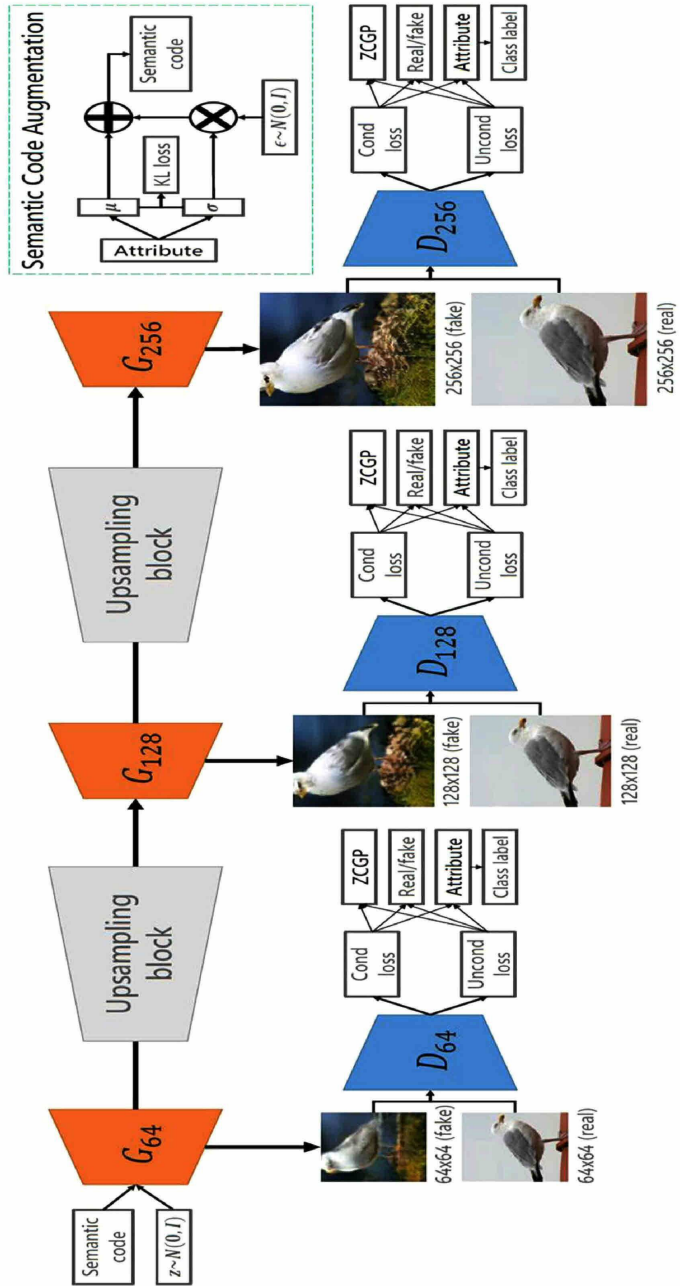
도면1



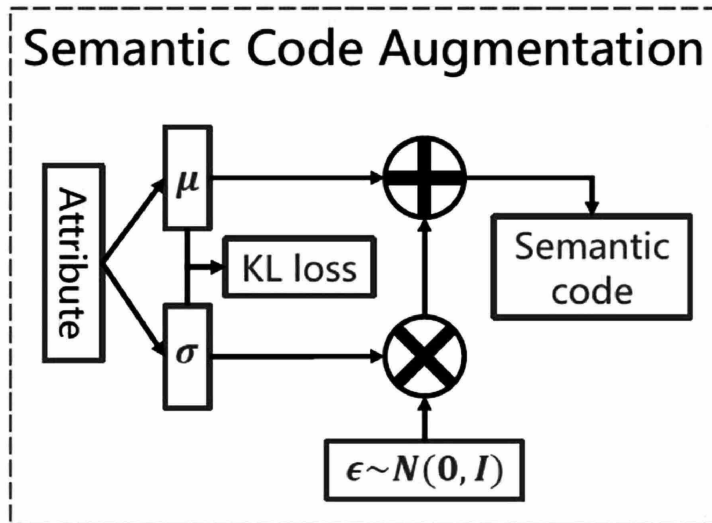
도면2



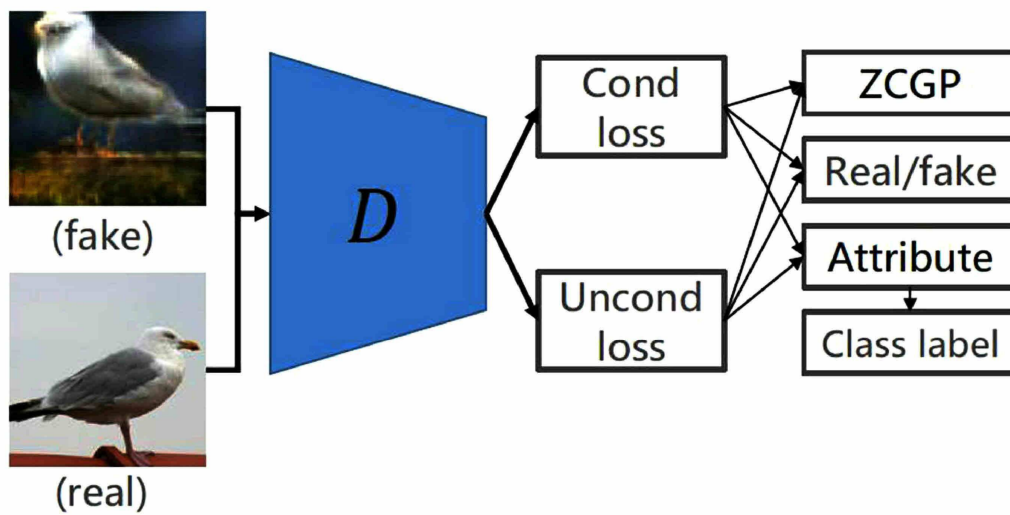
도면3



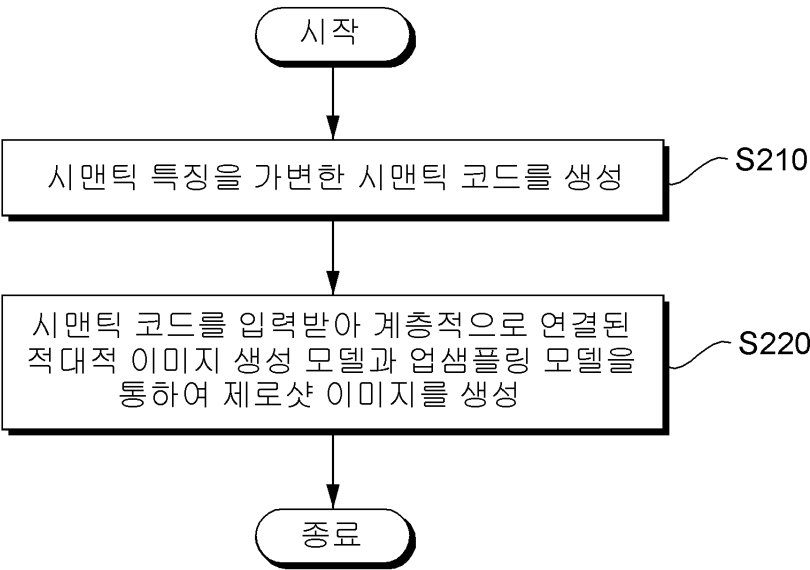
도면4



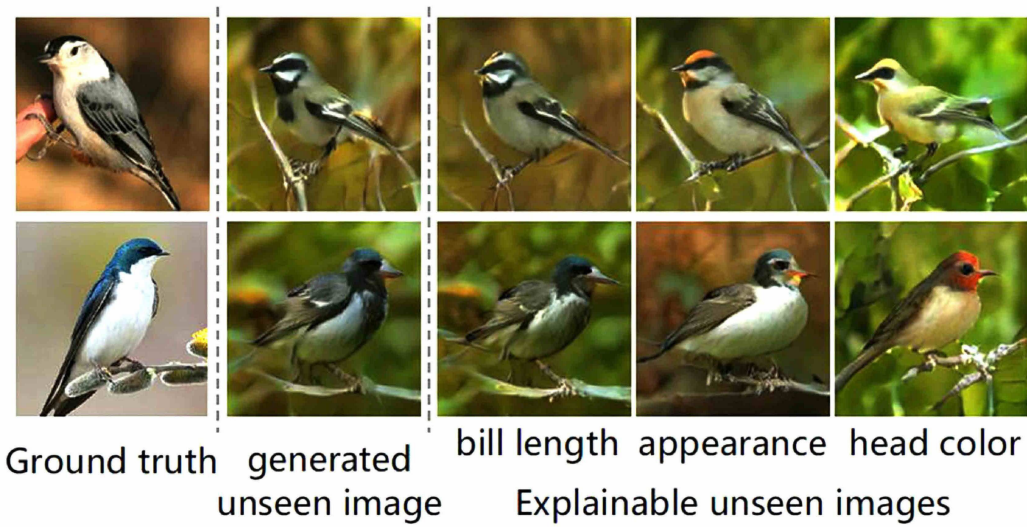
도면5



도면6



도면7



도면8

