



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2021년08월17일

(11) 등록번호 10-2291111

(24) 등록일자 2021년08월11일

(51) 국제특허분류(Int. Cl.)

G06K 9/46 (2006.01) G06F 17/11 (2006.01)

G06N 3/08 (2006.01) G06T 5/00 (2019.01)

(52) CPC특허분류

G06K 9/46 (2013.01)

G06F 17/11 (2013.01)

(21) 출원번호 10-2020-0011824

(22) 출원일자 2020년01월31일

심사청구일자 2020년01월31일

(65) 공개번호 10-2021-0098082

(43) 공개일자 2021년08월10일

(56) 선행기술조사문헌

Rafael Felix 외 3명, The European Conference on Computer Vision (ECCV), 2018, pp. 21-37 (2018.09.08) 1부.\*

Yongqin Xian 외 3명, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 5542-5551 (2018.06.19) 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)

(뒷면에 계속)

(74) 대리인

특허법인우인

전체 청구항 수 : 총 8 항

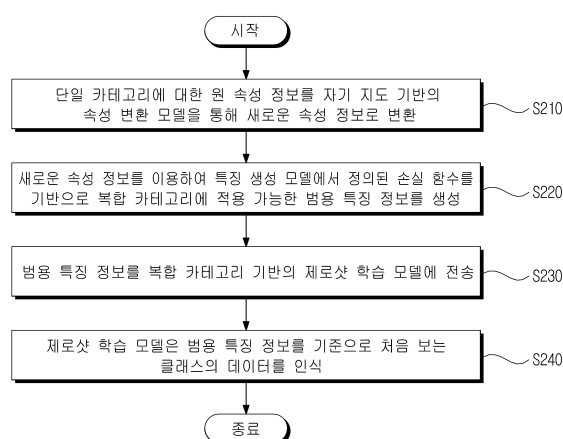
심사관 : 강현일

(54) 발명의 명칭 자기 지도 기반의 제로샷 인식 장치 및 방법

## (57) 요약

본 실시예들은 단일 카테고리에 관한 속성 정보를 변환하고 변환된 속성 정보를 기반으로 적대적 생성 신경망을 통해 범용 특징 정보를 생성하여, 복합 카테고리 데이터 세트에 대하여 제로샷 인식 성능을 향상시킬 수 있는 제로샷 인식 방법 및 장치를 제공한다.

대표도 - 도6



(52) CPC특허분류

G06N 3/08 (2013.01)

G06T 5/002 (2013.01)

(72) 발명자

**이제욱**

서울특별시 서대문구 연세로 50, 제4공학관 810호  
(신촌동, 연세대학교)

**홍기범**

서울특별시 서대문구 연세로 50, 제4공학관 810호  
(신촌동, 연세대학교)

이 발명을 지원한 국가연구개발사업

과제고유번호 2019-11-0366

부처명 과학기술정보통신부

과제관리(전문)기관명 한국연구재단

연구사업명 중견연구자지원사업

연구과제명 특성 정보 자동 생성을 통한 처음 보는 복합카테고리의 이미지와 비디오 생성 및 인  
식을 위한 제로샷 학습 기술 연구(1/3)

기 여 율 1/1

과제수행기관명 연세대학교

연구기간 2019.03.01 ~ 2020.02.29

## 명세서

### 청구범위

#### 청구항 1

컴퓨팅 디바이스에 의한 제로샷 인식 방법에 있어서,

단일 카테고리에 대한 원 속성 정보를 자기 지도 기반의 속성 변환 모델을 통해 새로운 속성 정보로 변환하는 단계;

상기 새로운 속성 정보를 이용하여 특징 생성 모델에서 정의된 손실 함수를 기반으로 복합 카테고리에 적용 가능한 범용 특징 정보를 생성하는 단계; 및

상기 범용 특징 정보를 상기 복합 카테고리 기반의 제로샷 학습 모델에 전송하는 단계를 포함하며,

상기 특징 생성 모델은 랜덤 노이즈 분포와 상기 새로운 속성 정보를 입력받고 상기 범용 특징 정보를 출력하며, 생성 모델과 판별 모델이 상호 작용하는 적대적 생성 신경망에서 정의된 (i) 페널티 가중치를 조건부로 상기 범용 특징 정보에 바서스타인 거리(Wasserstein Distance)를 적용하는 조건부 생성 손실 함수 및 (ii) 자기 지도를 수행하면서 상기 원 속성 정보 또는 상기 새로운 속성 정보에 따른 상기 범용 특징 정보에 대한 상기 판별 모델의 분포를 적용하는 자기 지도 손실 함수를 학습한 결과를 이용하여 상기 범용 특징 정보를 교정하는 것을 특징으로 하는 제로샷 인식 방법.

#### 청구항 2

제1항에 있어서,

상기 원 속성 정보는 복수의 벡터로 입력되고,

상기 속성 변환 모델은 상기 복수의 벡터에서 일부 값을 다른 값으로 변환하는 것을 특징으로 하는 제로샷 인식 방법.

#### 청구항 3

제2항에 있어서,

상기 속성 변환 모델은 상기 복수의 벡터에서 일부 값을 제로 또는 노이즈 값으로 변환하는 것을 특징으로 하는 제로샷 인식 방법.

#### 청구항 4

제2항에 있어서,

상기 속성 변환 모델은 상기 복수의 벡터에서 대응하는 두 개의 값을 상호 치환하여, 상기 원 속성 정보의 매니폴드를 유지하는 것을 특징으로 하는 제로샷 인식 방법.

#### 청구항 5

삭제

#### 청구항 6

삭제

#### 청구항 7

삭제

#### 청구항 8

제1항에 있어서,

슈도 속성 정보를 제로샷 학습 모델에 전송하면,

상기 제로샷 학습 모델은 상기 범용 특징 정보를 기준으로 미사용된 클래스의 데이터를 인식하는 것을 특징으로 하는 제로샷 인식 방법.

#### 청구항 9

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

상기 프로세서는 단일 카테고리에 대한 원 속성 정보를 자기 지도 기반의 속성 변환 모델을 통해 새로운 속성 정보로 변환하고,

상기 프로세서는 상기 새로운 속성 정보를 이용하여 특징 생성 모델에서 정의된 손실 함수를 기반으로 복합 카테고리에 적용 가능한 범용 특징 정보를 생성하고,

상기 프로세서는 상기 범용 특징 정보를 상기 복합 카테고리 기반의 제로샷 학습 모델에 전송하며,

상기 특징 생성 모델은 랜덤 노이즈 분포와 상기 새로운 속성 정보를 입력받고 상기 범용 특징 정보를 출력하며, 생성 모델과 판별 모델이 상호 작용하는 적대적 생성 신경망에서 정의된 (i) 페널티 가중치를 조건부로 상기 범용 특징 정보에 바스스타인 거리(Wasserstein Distance)를 적용하는 조건부 생성 손실 함수 및 (ii) 자기 지도를 수행하면서 상기 원 속성 정보 또는 상기 새로운 속성 정보에 따른 상기 범용 특징 정보에 대한 상기 판별 모델의 분포를 적용하는 자기 지도 손실 함수를 학습한 결과를 이용하여 상기 범용 특징 정보를 교정하는 것을 특징으로 하는 제로샷 인식 장치.

#### 청구항 10

제9항에 있어서,

상기 원 속성 정보는 복수의 벡터로 입력되고,

상기 속성 변환 모델은 상기 복수의 벡터에서 대응하는 두 개의 값을 상호 치환하여, 상기 원 속성 정보의 매니폴드를 유지하는 것을 특징으로 하는 제로샷 인식 장치.

#### 청구항 11

삭제

#### 청구항 12

삭제

#### 청구항 13

삭제

#### 청구항 14

제9항에 있어서,

상기 프로세서가 슈도 속성 정보를 제로샷 학습 모델에 전송하면,

상기 제로샷 학습 모델은 상기 범용 특징 정보를 기준으로 미사용된 클래스의 데이터를 인식하는 것을 특징으로 하는 제로샷 인식 장치.

### 발명의 설명

### 기술 분야

본 실시예가 속하는 기술 분야는 제로샷 인식 장치 및 방법에 관한 것이다.

[0001]

## 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 속성 정보(Attribute)를 올바르게 추론할 수 있도록 모델을 학습하면, 처음 보는 클래스의 데이터에 대해서도 해당 클래스의 속성 정보를 추론하여 처음 보는 클래스를 인식할 수 있다.
- [0004] 데이터 세트에 정의된 속성 정보(given attribute)는 클래스를 대표하는 주요 특성을 의미하며, 예를 들어 새의 부리 길이, 날개 색상, 몸 크기 등이 있다. 기존 제로샷 학습 기술들은 속성 정보를 핵심 단서로 사용하기 때문에, 데이터 세트에 학습용으로 미리 정의되어 있지 않다면 정확도에 영향을 준다.
- [0005] 사용자가 정의한 속성 정보는 제로샷 학습에서 서로 다른 카테고리를 구분함에 있어서 필수 요소 중 하나이다. 속성 정보를 다양한 카테고리에 대해 새롭게 정의하는 것은 사용자의 노동력과 전문적 지식이 필요하고 많은 시간과 비용이 필요하므로, 사용자가 속성 정보를 전부 설정하는 것은 현실적으로 불가능하다.
- [0006] 기존의 제로샷 학습 모델은 사용자에게 의해 정의되어 있지 않은 개념을 전혀 이해를 할 수 없는 문제가 있다. 예를 들어, 말과 얼룩말을 구분하기 위해서 필요한 "줄무늬"라는 속성 정보가 사용자에게 의해 정의되어 있지 않은 경우, 기존의 제로샷 학습 모델은 말과 얼룩말을 구분하지 못하는 실정이다.
- [0007] 단일 카테고리 데이터 세트는 하나의 카테고리(클래스의 집합)에 한정된 클래스로만 이루어진 데이터 세트이다. 예를 들면 꽃 종류들만 포함된 데이터 세트 또는 새 종류들만 포함된 데이터 세트 등이 있다.
- [0008] 복합 카테고리 데이터 세트는 둘 이상의 카테고리로 이루어져 있으며, 단일화된 속성 정보를 정의하기 어려운 데이터 세트이다. 예를 들면 강아지, 꽃, 자동차, 가구 등을 포함하는 서로 다른 카테고리들로 구성된다.
- [0009] 하나의 카테고리에서 학습된 모델은 다른 카테고리에서 사용할 수 없는 한계가 있다. 속성 정보의 체계가 다르므로, 새 종류로 학습된 제로샷 모델은 꽃 종류의 데이터 세트에 사용할 수 없다. 다른 카테고리를 인식 가능한 제로샷 학습 모델이 없는 실정이다.

## 선행기술문헌

### 특허문헌

- [0010] (특허문헌 0001) US 2018/0197050 (2018.07.12)

## 발명의 내용

### 해결하려는 과제

- [0011] 본 발명의 실시예들은 단일 카테고리에 관한 속성 정보를 변환하고 변환된 속성 정보를 기반으로 적대적 생성 신경망을 통해 범용 특징 정보를 생성하여, 복합 카테고리 데이터 세트에 대하여 제로샷 인식 성능을 향상시키는 데 주된 목적이 있다.
- [0012] 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

- [0013] 본 실시예의 일 측면에 의하면, 컴퓨팅 디바이스에 의한 제로샷 인식 방법에 있어서, 단일 카테고리에 대한 원 속성 정보를 자기 지도 기반의 속성 변환 모델을 통해 새로운 속성 정보로 변환하는 단계, 상기 새로운 속성 정보를 이용하여 특징 생성 모델에서 정의된 손실 함수를 기반으로 복합 카테고리에 적용 가능한 범용 특징 정보를 생성하는 단계, 및 상기 범용 특징 정보를 상기 복합 카테고리 기반의 제로샷 학습 모델에 전송하는 단계를 포함하는 제로샷 인식 방법을 제공한다.
- [0014] 본 실시예의 다른 측면에 의하면, 하나 이상의 프로세서 및 상기 하나 이상의 프로세서에 의해 실행되는 하나

이상의 프로그램을 저장하는 메모리를 포함하는 제로샷 인식 장치에 있어서, 상기 프로세서는 단일 카테고리에 대한 원 속성 정보를 자기 지도 기반의 속성 변환 모델을 통해 새로운 속성 정보로 변환하고, 상기 프로세서는 상기 새로운 속성 정보를 이용하여 특징 생성 모델에서 정의된 손실 함수를 기반으로 복합 카테고리에 적용 가능한 범용 특징 정보를 생성하고, 상기 프로세서는 상기 범용 특징 정보를 상기 복합 카테고리 기반의 제로샷 학습 모델에 전송하는 것을 특징으로 하는 제로샷 인식 장치를 제공한다.

### 발명의 효과

- [0015] 이상에서 설명한 바와 같이 본 발명의 실시예들에 의하면, 단일 카테고리에 관한 속성 정보를 변환하고 변환된 속성 정보를 기반으로 적대적 생성 신경망을 통해 범용 특징 정보를 생성하여, 복합 카테고리 데이터 세트에 대하여 제로샷 인식 성능을 향상시킬 수 있는 효과가 있다.
- [0016] 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

- [0017] 도 1은 기존의 제로샷 학습 모델을 예시한 도면이다.
- 도 2는 단일 카테고리 및 복합 카테고리를 예시한 도면이다.
- 도 3은 본 발명의 일 실시예에 따른 제로샷 인식 장치를 예시한 블록도이다.
- 도 4는 본 발명의 일 실시예에 따른 제로샷 인식 장치의 속성 변환 모델을 예시한 도면이다.
- 도 5는 본 발명의 일 실시예에 따른 제로샷 인식 장치의 특징 생성 모델을 예시한 도면이다.
- 도 6은 본 발명의 다른 실시예에 따른 제로샷 인식 방법을 예시한 흐름도이다.
- 도 7 및 도 8은 본 발명의 실시예들에 따른 시뮬레이션 결과를 예시한 도면이다.

### 발명을 실시하기 위한 구체적인 내용

- [0018] 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.
- [0019] 도 1은 기존의 제로샷 학습 모델을 예시한 도면이다.
- [0020] 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 있으나, 기존의 제로샷 학습 역시 속성 정보가 정의되지 않으면 새로운 클래스를 인식할 수 없는 문제가 있다.
- [0021] 기존의 제로샷 학습은 속성 정보가 정의된 데이터 세트에서만 학습이 가능하다. 하나의 카테고리로 학습된 모델은 다른 카테고리에 사용할 수 없는 한계를 갖는다. 예를 들어, 새 종류들로 학습된 제로샷 모델은 꽃 종류들의 데이터 세트에 사용할 수 없다. 정의된 속성 정보의 체계가 다르고, 서로 다른 카테고리를 인식할 수 있는 제로샷 학습 기술이 없기 때문이다.
- [0022] 기존의 제로샷 학습 모델을 예시한 도 1을 참조하면, 기존의 제로샷 학습은 학습에 사용한 클래스(seen class)의 데이터로부터 해당 클래스와 관련된 속성 정보(attribute)를 올바르게 추론할 수 있도록 모델을 학습하면, 처음 보는 클래스(Unseen Class)의 데이터에 대해서도 해당 클래스의 속성 정보를 추론한 후 처음 보는 클래스를 인식한다.
- [0023] 학습에 사용한 클래스(Seen Class)는 속성 정보를 학습하는데 활용되는 학습용 클래스이고, 처음 보는 클래스(Unseen Class)는 학습에서 전혀 사용되지 않은 테스트용 제로샷 클래스이다. 단일 카테고리 데이터 세트는 하나의 카테고리(클래스의 집합)에 한정된 클래스로만 이루어진 데이터 세트이다. 예를 들면, 꽃 종류들만 포함된 데이터 세트, 새 종류들만 포함된 데이터 세트 등이 있다.
- [0024] 데이터 세트에 정의된 속성 정보(Given 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 속성 정보를 핵심 단서로 사용한다. 데이터 세트에 학습용으로 미리 정의되어 있는 경우 제로샷 학

습시 정확도가 낮아지는 문제가 있다.

- [0025] 데이터 세트에 정의되지 않은 속성 정보(Not Given Attribute)는 기존 데이터 세트에 정의되어 있지 않은 속성 정보를 의미한다.
- [0026] 도 2는 단일 카테고리 및 복합 카테고리를 예시한 도면이다.
- [0027] 기존의 제로샷 학습 기술은 단일 카테고리 데이터 세트에서 연구되며, 복합 카테고리 데이터 세트에서는 제로샷 기술의 분류 정확도가 높지 않다.
- [0028] 단일 카테고리 데이터 세트는 하나의 카테고리(클래스의 집합)에 한정된 클래스로만 이루어진 데이터 세트이다. 예를 들면 꽃 종류들만 포함된 데이터 세트 또는 새 종류들만 포함된 데이터 세트 등이 있다.
- [0029] 복합 카테고리 데이터 세트는 둘 이상의 카테고리로 이루어져 있으며, 단일화된 속성 정보를 정의하기 어려운 데이터 세트이다. 예를 들면 강아지, 꽃, 자동차, 가구 등을 포함하는 서로 다른 카테고리들로 구성된다.
- [0030] 하나의 카테고리에서 학습된 모델은 다른 카테고리에서 사용할 수 없는 한계가 있다. 속성 정보의 체계가 다르므로, 새 종류로 학습된 제로샷 모델은 꽃 종류의 데이터 세트에 사용할 수 없다. 다른 카테고리를 인식 가능한 제로샷 학습 모델이 없는 실정이다.
- [0031] 본 실시예에서는 자기 지도 기반으로 속성 정보를 변환하고, 변환된 속성 정보를 기반으로 범용 특징 정보를 생성하여 기존 제로샷 학습 방식의 한계를 극복한다. 본 실시예는 범용 특징 정보를 통해 복합 카테고리 데이터 세트에 포함된 객체들이나 학습에서 사용되지 않은 객체들을 인식할 수 있다.
- [0032] 도 3은 본 발명의 일 실시예에 따른 제로샷 인식 장치를 예시한 블록도이다.
- [0033] 제로샷 인식 장치(110)는 적어도 하나의 프로세서(120), 컴퓨터 판독 가능한 저장매체(130) 및 통신 버스(170)를 포함한다.
- [0034] 프로세서(120)는 제로샷 인식 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 인식 장치(110)로 하여금 예시적인 실시예에 따른 동작들을 수행하도록 구성될 수 있다.
- [0035] 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 제로샷 인식 장치(110)에 의해 액세스되고 원하는 정보를 저장할 수 있는 다른 형태의 저장 매체, 또는 이들의 적합한 조합일 수 있다.
- [0036] 통신 버스(170)는 프로세서(120), 컴퓨터 판독 가능한 저장 매체(140)를 포함하여 제로샷 인식 장치(110)의 다른 다양한 컴포넌트들을 상호 연결한다.
- [0037] 제로샷 인식 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 인식 장치(110)의 다른 컴포넌트들에 연결될 수 있다.
- [0038] 제로샷 인식 장치(110)는 단일 카테고리에 대한 원 속성 정보를 자기 지도 기반의 속성 변환 모델을 통해 새로운 속성 정보로 변환하고, 새로운 속성 정보를 이용하여 특징 생성 모델에서 정의된 손실 함수를 기반으로 복합 카테고리에 적용 가능한 범용 특징 정보를 생성하고, 범용 특징 정보를 복합 카테고리 기반의 제로샷 학습 모델에 전송하고, 제로샷 학습 모델을 통해 범용 특징 정보를 기준으로 미사용된 클래스의 데이터를 인식한다.
- [0039] 도 4는 본 발명의 일 실시예에 따른 제로샷 인식 장치의 속성 변환 모델을 예시한 도면이다.
- [0040] 속성 변환 모델은 입력된 속성 정보를 벡터 레벨에서 변환하는 모델이다. 복수의 샘플이 벡터 레벨로 입력되어 임베딩 공간을 구성한다.
- [0041] 도 4의 (a)를 참조하면 원 속성 정보는 복수의 벡터로 입력된다. 속성 변환 모델은 복수의 벡터에서 일부 값을



다른 값으로 변환한다.

- [0042] 도 4의 (b)를 참조하면 속성 변환 모델은 복수의 벡터에서 일부 값을 제로(0)로 변환할 수 있다.
- [0043] 도 4의 (c)를 참조하면 속성 변환 모델은 복수의 벡터에서 일부 값을 노이즈 값으로 변환할 수 있다. 노이즈 값으로는 가우시안 노이즈가 적용될 수 있다.
- [0044] 도 4의 (d)를 참조하면 속성 변환 모델은 복수의 벡터에서 대응하는 두 개의 값을 상호 치환한다. 동일한 열에 대응하는 다른 속성 정보 값을 열 단위로 치환한다. 제로 또는 노이즈와 달리 실제 값을 사용하므로, 원 속성 정보의 매니폴드를 유지할 수 있다. 학습 분야에서 매니폴드는 데이터가 있는 공간으로, 고차원에서 저차원으로 표현된 점들의 연결 또는 그룹을 의미한다.
- [0045] 도 5는 본 발명의 일 실시예에 따른 제로샷 인식 장치의 특징 생성 모델을 예시한 도면이다.
- [0046] 사람이 정의한 속성 정보 기반의 데이터 세트에 국한되지 않는 범용적인 제로샷 학습을 위해, 원 속성 정보와 변경된 속성 정보를 이용하여 특징 생성 모델을 학습한다.
- [0047] 특징 생성 모델은 랜덤 노이즈 분포와 새로운 속성 정보를 입력받고 범용 특징 정보를 출력하며, 생성 모델과 판별 모델이 상호 작용하는 적대적 생성 신경망에서 (i) 조건부 생성 손실 함수 및 (ii) 자기 지도 손실 함수를 최적화하여 상기 범용 특징 정보를 교정한다.
- [0048] 조건부 생성 손실 함수( $L_{gan}$ )는 수학식 1과 같이 표현되고, 자기 지도 손실 함수( $L_{ss}$ )는 수학식 2와 같이 표현된다.

### 수학식 1

$$L_{gan} = E_{x \sim P_{data}} [D(x, a^y)] - E_{\hat{x} \sim P_g} [D(\hat{x}, a^y)] - \lambda E_{\bar{x} \sim P_{\bar{x}}} [(\|\nabla_{\bar{x}} D(\bar{x}, a^y)\|_2 - 1)^2]$$

- [0049]
- [0050] 조건부 생성 손실 함수는 페널티 가중치를 조건부로 범용 특징 정보에 바서스타인 거리(Wasserstein Distance)를 적용한다.

[0051]  $x$ 는 클래스  $y$ 의 CNN 특징 정보를 의미한다.  $\hat{x} = G(a^y, z)$ 는 생성 모델의 분포  $P_g$ 로부터 클래스  $y$ 의 생성된 특징이다.  $z$ 는 특징 생성 모델의 자유도를 높이는 노이즈 벡터이다.  $\bar{x} = \alpha x + (1 - \alpha)\hat{x}$ 는  $\alpha \sim U(0, 1)$  조건에 따른 처음 보는 클래스의 CNN 특징 정보이다.  $\lambda$ 는 페널티 가중치 파라미터이다. 로그 우도(Log Likelihood) 대신에 바서스타인 거리를 적용하고, 경사 페널티를 추가로 포함한다. 바서스타인 거리는 두 확률분포의 연관성을 측정하여 그 거리의 기대값이 가장 작을 때의 거리를 의미한다. 결합 확률분포는 두 분포가 동시에 일어날 때의 사건에 대한 확률분포를 의미한다.

### 수학식 2

$$L_{ss} = -E_{x \sim P_{data}} E_{s \sim \mathcal{S}} [\log D(S = s | x^s)] - E_{\hat{x} \sim P_g} E_{s \sim \mathcal{S}} [\log D(S = s | \hat{x}^s)]$$

- [0052]
- [0053] 자기 지도 손실 함수는 자기 지도를 수행하면서 원 속성 정보( $s=0$ ) 또는 새로운 속성 정보( $s=1$ )에 따른 상기 범용 특징 정보에 대한 판별 모델의 분포( $D(S | x^s)$ )를 적용한다. 속성 정보에 대한 치환 유무를 분별하는 목적 함수이다. 속성 변환 모델은 특정 확률 상수  $p$ 에 의해 열 단위로 속성 값을 섞는 함수이며, 매니폴드를 그대로 유지하면서 이상치(outlier) 또는 노이즈에 강인하다.



## 수학식 3

$$\min_G \max_D L_{total} = L_{gan} + L_{ss}$$

[0054]

[0055] 모든 손실은 이상치와 노이즈에 강인한 특징을 생성하게 하여 복합 카테고리 데이터 세트에 적용 가능하다.

[0056] 프로세서가 슈도 속성 정보를 제로샷 학습 모델에 전송하면, 제로샷 학습 모델은 범용 특징 정보를 기준으로 미 사용된 클래스의 데이터를 인식한다.

[0057] 임베딩 공간으로서 사용자에게 의해 정의된 속성 공간과 클래스 y에 대한 양립 가능성 점수 s(y)를 사용한다.

## 수학식 4

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

[0058]

[0059] W는 풀리 커넥티드 레이어(Fully Connected Layer)를 갖는 가중치 매트릭스이다. 제로샷 학습 모델은 주어진 특징 x에 대해 클래스 레벨을 예측하기 위해서, 특징은 속성 표현에 투영될 수 있다. 양립 가능성 점수 s(y)는 최선의 매칭된 클래스  $y^*$ 를 선택하도록 사용될 수 있다.

## 수학식 5

$$y^* = \arg \max_{y \in \mathcal{Y}^t} (s(y)) = \arg \max_{y \in \mathcal{Y}^t} \langle f(x), a^y \rangle$$

[0060]

[0061]  $\mathcal{Y}^t = \mathcal{Y}^s \cup \mathcal{Y}^u$  조건에서, 가장 높은 양립 가능성 점수를 가는  $y^*$ 는 예측된 클래스이다.

$\mathcal{Y}^t = \mathcal{Y}^u$  로 ZSL(Zero Shot Learning)을 설정할 수 있고,  $\mathcal{Y}^t = \mathcal{Y}^s \cup \mathcal{Y}^u$  로 GZSL(Generalized Zero Shot Learning)를 설정할 수 있다. GZSL는 테스트 단계에서 처음 보는 클래스( $\mathcal{Y}^u$ , Unseen Class)와 사용된 클래스( $\mathcal{Y}^s$ , Seen Class)를 동시에 분류한다.

[0062] 제로샷 학습 모델은 특징 추출 모델을 포함할 수 있으며, 특징 추출 모델은 CNN(Convolutional Neural Network)으로 구현될 수 있다. 특징 추출 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.

[0063] 도 6은 본 발명의 다른 실시예에 따른 제로샷 인식 방법을 예시한 흐름도이다. 제로샷 인식 방법은 컴퓨팅 디바이스에 의하여 수행될 수 있으며, 제로샷 인식 장치와 동일한 방식으로 동작한다.

[0064] 단계 S210에서 프로세서는 단일 카테고리에 대한 원 속성 정보를 자기 지도 기반의 속성 변환 모델을 통해 새로운 속성 정보로 변환한다. 원 속성 정보는 복수의 벡터로 입력된다. 속성 변환 모델은 복수의 벡터에서 일부 값을 다른 값으로 변환한다. 속성 변환 모델은 복수의 벡터에서 일부 값을 제로 또는 노이즈 값으로 변환할 수 있다. 속성 변환 모델은 복수의 벡터에서 대응하는 두 개의 값을 상호 치환하여, 원 속성 정보의 매니폴드를 유지할 수 있다.

[0065] 단계 S220에서 프로세서는 새로운 속성 정보를 이용하여 특징 생성 모델에서 정의된 손실 함수를 기반으로 복합 카테고리에 적용 가능한 범용 특징 정보를 생성한다. 특징 생성 모델은 랜덤 노이즈 분포와 상기 새로운 속성 정보를 입력받고 상기 범용 특징 정보를 출력하며, 생성 모델과 판별 모델이 상호 작용하는 적대적 생성 신경망

에서 (i) 조건부 생성 손실 함수 및 (ii) 자기 지도 손실 함수를 최적화하여 상기 범용 특징 정보를 교정한다. 조건부 생성 손실 함수는 페널티 가중치를 조건부로 범용 특징 정보에 바서스타인 거리(Wasserstein Distance)를 적용할 수 있다. 자기 지도 손실 함수는 자기 지도를 수행하면서 원 속성 정보 또는 새로운 속성 정보에 따른 범용 특징 정보에 대한 판별 모델의 분포를 적용할 수 있다.

[0066] 단계 S220에서 프로세서는 범용 특징 정보를 복합 카테고리 기반의 제로샷 학습 모델에 전송한다.

[0067] 단계 S220에서 프로세서는 슈도 속성 정보를 제로샷 학습 모델에 전송하면, 제로샷 학습 모델은 범용 특징 정보를 기준으로 미사용된 클래스의 데이터를 인식할 수 있다.

[0068] 도 7 및 도 8은 본 발명의 실시예들에 따른 시뮬레이션 결과를 예시한 도면이다.

[0069] 자동 생성된 처음 보는 클래스의 데이터 및 속성 정보에 대한 정량 평가를 위하여 제로샷 학습에서 널리 사용되는 Average Per-Class Top-1 Accuracy(%) 측정 방법을 이용하였다. Average Per-Class Top-1 Accuracy(%)는 각 클래스 별로 인식 정확도(%)의 평균값을 계산하여 최종 제로샷 성능 평가의 척도로 사용한다.

[0070] 복합 카테고리 제로샷 데이터 세트인 ImageNet 21K를 사용하였고, 테스트 클래스의 집합을 총 9 가지를 사용하여 평가하였다. 2H는 전체 21K 클래스 중에서 WordNet hierarchy에서 학습 클래스로부터 2칸 이내로 떨어진 모든 클래스들의 집합이다. 3H는 전체 21K 클래스 중에서 WordNet hierarchy에서 학습 클래스로부터 3칸 이내로 떨어진 모든 클래스들의 집합이다. M500는 전체 21K 클래스 중에서 가장 데이터 양이 많은 500개의 클래스들의 집합이다. M1K는 전체 21K 클래스 중에서 가장 데이터 양이 많은 1,000개의 클래스들의 집합이다. M5K는 전체 21K 클래스 중에서 가장 데이터 양이 많은 5,000개의 클래스들의 집합이다. L500는 전체 21K 클래스 중에서 가장 데이터 양이 적은 500개의 클래스들의 집합이다. L1K는 전체 21K 클래스 중에서 가장 데이터 양이 적은 1,000개의 클래스들의 집합이다. L5K는 전체 21K 클래스 중에서 가장 데이터 양이 적은 5,000개의 클래스들의 집합이다. A1120K는 전체 21K 클래스를 모두 사용한다.

[0071] 본 실시예에 따른 제로샷 인식 장치는 복합 카테고리 데이터 세트의 일종인 ImageNet에 대해서 최신 제로샷 기술의 Top-1 분류 정확도가 가장 우수한 결과를 보여준다. 본 실시예에 따른 제로샷 인식 장치는 카테고리의 종류와 특성 정보 범위에 제한을 두지 않으며, 기존 기술의 한계를 극복 가능하다.

[0072] 제로샷 인식 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.

[0073] 제로샷 인식 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.

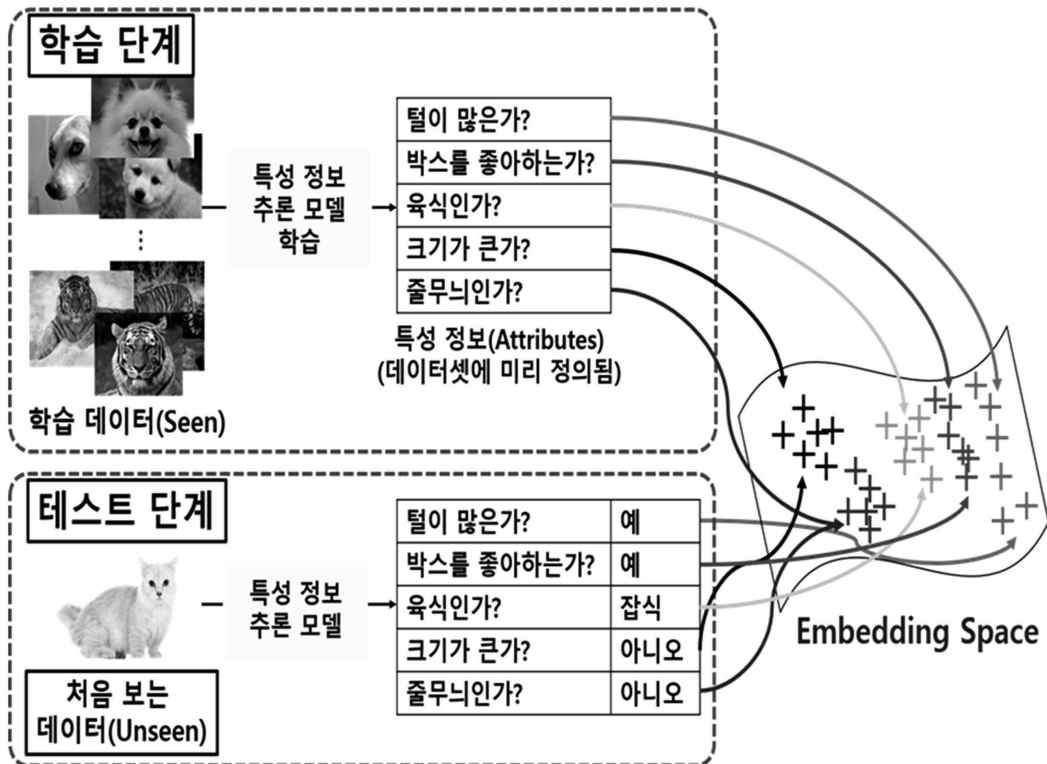
[0074] 도 5 및 도 6에서는 각각의 과정을 순차적으로 실행하는 것으로 기재하고 있으나 이는 예시적으로 설명한 것에 불과하고, 이 분야의 기술자라면 본 발명의 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 도 5 및 도 6에 기재된 순서를 변경하여 실행하거나 또는 하나 이상의 과정을 병렬적으로 실행하거나 다른 과정을 추가하는 것으로 다양하게 수정 및 변형하여 적용 가능할 것이다.

[0075] 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.

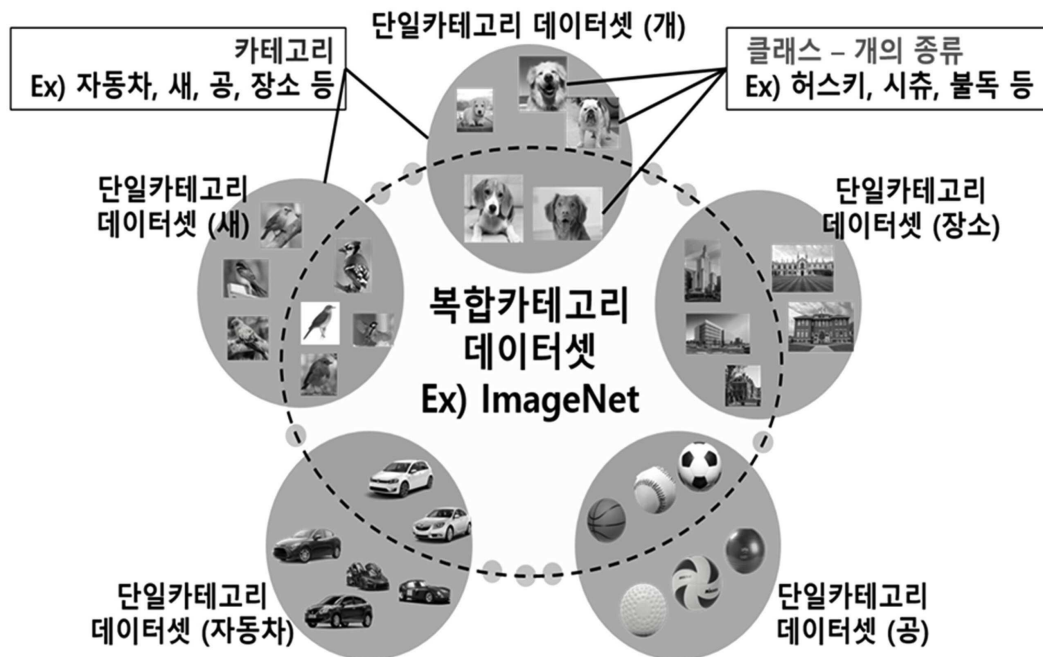
[0076] 본 실시예들은 본 실시예의 기술 사상을 설명하기 위한 것이고, 이러한 실시예에 의하여 본 실시예의 기술 사상의 범위가 한정되는 것은 아니다. 본 실시예의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 실시예의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

도면

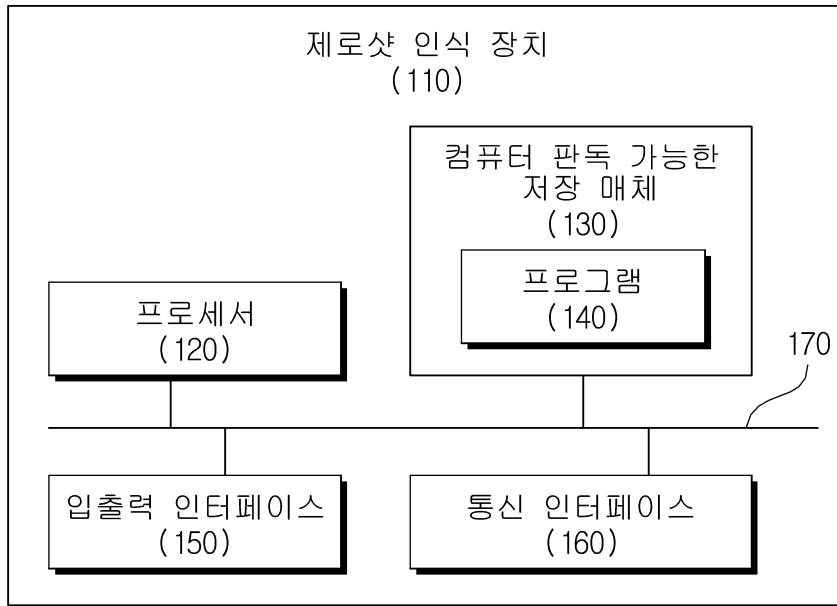
도면1



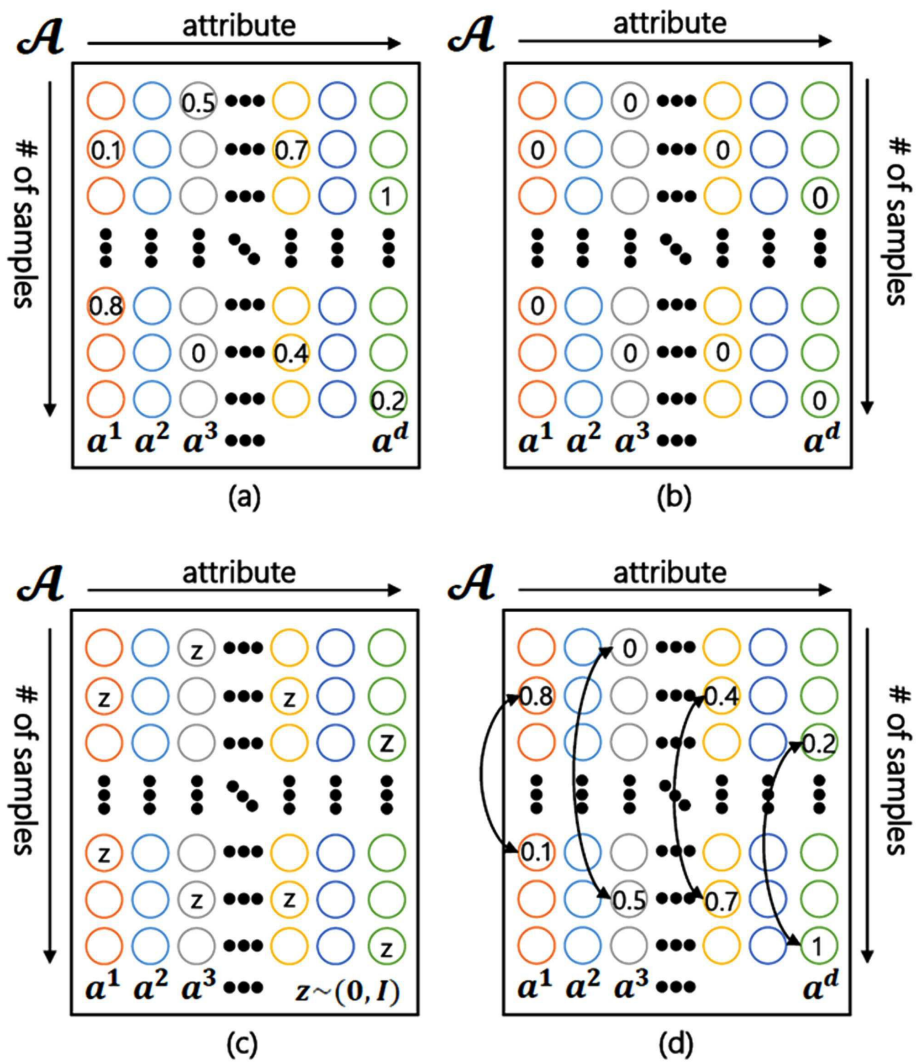
도면2



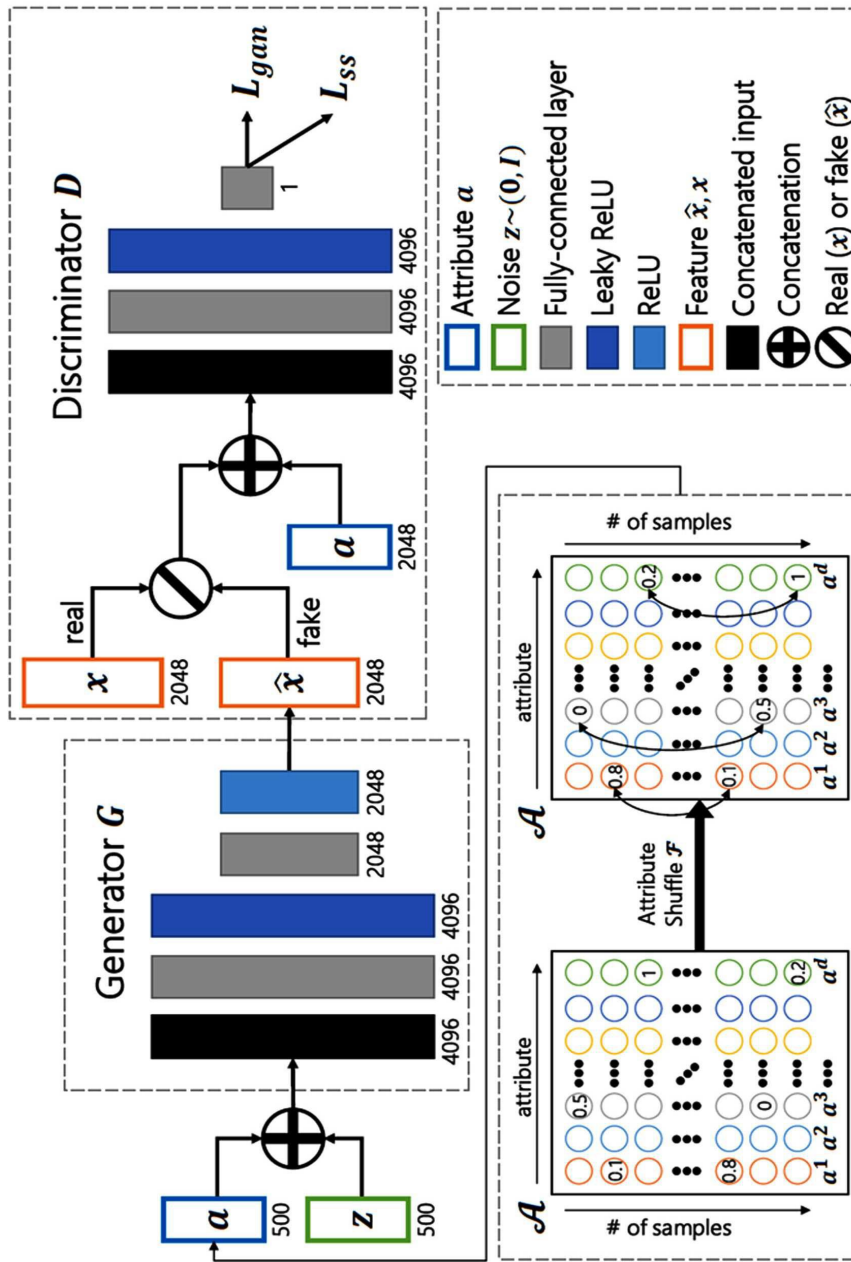
도면3



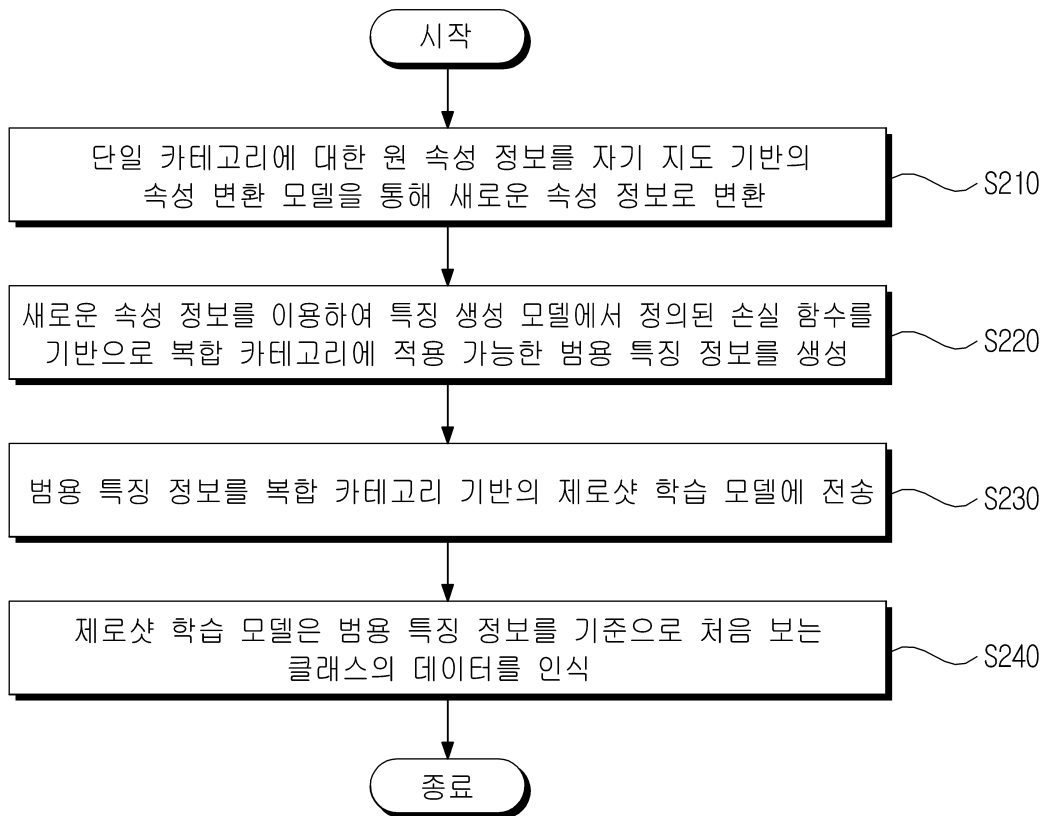
도면4



도면5



도면6





도면7

Method	Hierarchy		Most Populated			Least Populated			All 20K
	2H	3H	500	1K	5K	500	1K	5K	
CONSE [Norouzi <i>et al.</i> , 2014]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SYNC [Changpinyo <i>et al.</i> , 2016]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ESZSL [Romera-Paredes and Torr, 2015]	1.3	0.5	1.7	1.5	0.8	0.6	0.4	0.3	0.3
f-CLSWGAN [Xian <i>et al.</i> , 2018b]	4.4	1.3	4.7	4.1	1.8	3.3	2.5	0.8	0.5
LATEM [Xian <i>et al.</i> , 2016]	2.0	0.8	2.6	2.1	1.0	1.0	0.6	0.3	0.3
DEVISE [Frome <i>et al.</i> , 2013]	2.2	0.8	2.9	2.3	1.1	1.6	1.3	0.4	0.3
SJE [Akata <i>et al.</i> , 2015]	1.8	0.7	2.4	1.9	1.0	1.7	1.4	0.4	0.3
ALE [Akata <i>et al.</i> , 2016]	2.2	0.8	2.9	2.3	1.1	1.9	1.2	0.4	0.3
CMT [Socher <i>et al.</i> , 2013]	1.1	0.3	1.6	1.2	0.6	0.8	0.4	0.2	0.2
SAE [Kodirov <i>et al.</i> , 2017]	0.8	0.3	1.7	1.3	0.5	0.1	0.1	0.0	0.2
Ours	<b>5.1</b>	<b>1.5</b>	<b>12.1</b>	<b>6.2</b>	<b>2.3</b>	<b>5.4</b>	<b>3.2</b>	<b>1.1</b>	<b>0.7</b>



도면8

