(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0068032
(43) 공개일자 2024년05월17일

(51) 국제특허분류(Int. Cl.)
G06V 20/64 (2022.01) *G01S 17/89* (2020.01)
G06V 10/30 (2022.01) *G06V 10/46* (2022.01)
G06V 10/75 (2022.01) *G06V 10/80* (2022.01)
G06V 20/40 (2022.01) *G06V 20/58* (2022.01)
(52) CPC특허분류
G06V 20/64 (2023.08)
G01S 17/89 (2022.01)
(21) 출원번호 10-2023-0006147
(22) 출원일자 2023년01월16일
심사청구일자 없음
(30) 우선권주장
1020220147333 2022년11월07일 대한민국(KR)

(71) 출원인
삼성전자주식회사
경기도 수원시 영통구 삼성로 129 (매탄동)
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
장수진
경기도 화성시 동탄반석로 70, 435동 1503호 (반송동, 솔빛마을신도브래뉴아파트)
김상필
서울특별시 성북구 안암로 145(안암동5가)
(뒷면에 계속)
(74) 대리인
특허법인 무한

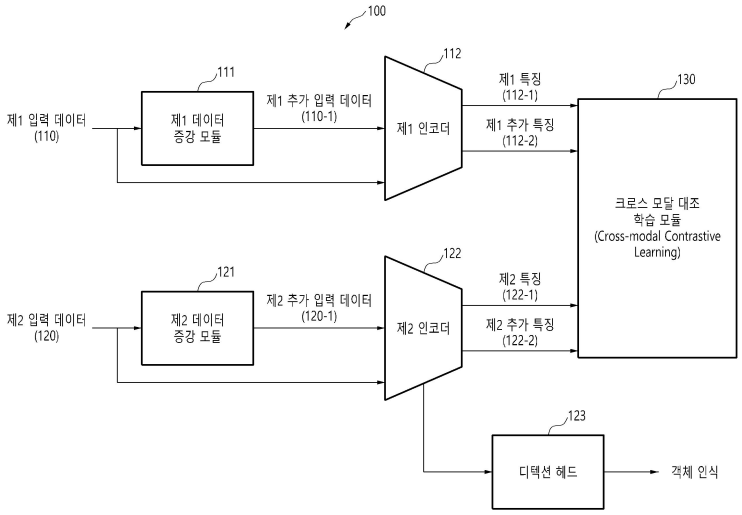
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 **객체 인식기 학습 방법 및 그 장치**

(57) 요 약

아래의 개시는 객체 인식기 학습 장치에 관한 것이다. 객체 인식기 학습 장치는 입력 데이터를 획득 데이터 증강을 통해 추가 입력 데이터를 획득하고, 입력 데이터 및 추가 입력 데이터로부터 특징 벡터들을 추출하여 결정된 손실 함수를 이용하여 대조 학습을 수행할 수 있다. 객체 인식기 학습 장치는 대조 학습 결과를 통해 객체 인식기를 업데이트할 수 있다.

대 표 도



(52) CPC특허분류
G06V 10/30 (2023.08)
G06V 10/467 (2023.08)
G06V 10/754 (2023.08)
G06V 10/811 (2023.08)
G06V 20/41 (2022.01)
G06V 20/58 (2022.01)
G06T 2210/12 (2013.01)

(72) 발명자
김진규
서울특별시 성북구 안암로 145(안암동5가)
노원석
서울특별시 성북구 안암로 145(안암동5가)
장규삼
경기도 성남시 분당구 동판교로 275, 112동 703호 (삼평동, 봇들마을1단지)
이동욱
경기도 용인시 기흥구 서천동로 22, 505동 601호 (서천동, 힐스테이트서천)
지대현
경기도 화성시 동탄문화센터로 38, 414동 1001호 (반송동, 동탄솔빛마을 서해그랑블아파트)

명 세 서

청구범위

청구항 1

대상 객체에 관한 제1 입력 데이터 및 제2 입력 데이터를 획득하는 단계;

상기 제2 입력 데이터에 대하여 데이터 증강을 수행하여 제2 추가 입력 데이터를 획득하는 단계;

상기 제1 입력 데이터를 제1 인코더에 입력하여 공유 임베딩 스페이스에 제1 특징(feature)을 추출하는 단계;

상기 제2 입력 데이터를 제2 인코더에 입력하여 상기 공유 임베딩 스페이스에 제2 특징을 추출하는 단계;

상기 제2 추가 입력 데이터를 상기 제2 인코더에 입력하여 상기 공유 임베딩 스페이스에 제2 추가 특징을 추출하는 단계;

상기 제1 특징, 상기 제2 특징 및 상기 제2 추가 특징에 기초하여 제1 손실 함수를 결정하는 단계;

상기 제2 특징 및 상기 제2 추가 특징에 기초하여 제2 손실 함수를 결정하는 단계; 및

상기 제1 손실 함수 및 상기 제2 손실 함수에 기초하여 상기 제2 인코더의 가중치(weight)를 업데이트하는 단계

를 포함하는

객체 인식기 학습 방법.

청구항 2

제1항에 있어서,

상기 제1 손실 함수를 결정하는 단계는

상기 제1 특징과 상기 제2 특징 및 상기 제2 추가 특징 사이의 제1 긍정/부정 쌍(positive/negative pair) 정보를 획득하는 단계; 및

상기 제1 긍정/부정 쌍 정보에 기초하여 상기 제1 손실 함수를 결정하는 단계

를 포함하는,

객체 인식기 학습 방법.

청구항 3

제2항에 있어서,

상기 제1 긍정/부정 쌍 정보를 획득하는 단계는

상기 제1 특징과 상기 제2 특징 및 상기 제2 추가 특징 사이의 유사도를 획득하는 단계; 및

상기 유사도에 기초하여 상기 제1 긍정/부정 쌍 정보를 획득하는 단계

를 포함하는,

객체 인식기 학습 방법.

청구항 4

제2항에 있어서,

상기 제1 긍정/부정 쌍 정보를 획득하는 단계는

상기 제1 특징, 상기 제2 특징 및 상기 제2 추가 특징 각각에 대응하는 클래스 정보를 획득하는 단계; 및

상기 클래스 정보에 기초하여 상기 제1 긍정/부정 쌍 정보를 획득하는 단계

를 포함하는

객체 인식기 학습 방법.

청구항 5

제1항에 있어서,

상기 제2 손실 함수를 결정하는 단계는

상기 제2 특징과 상기 제2 추가 특징 사이의 제2 긍정/부정 쌍 정보를 획득하는 단계; 및

상기 제2 긍정/부정 쌍 정보에 기초하여 상기 제2 손실 함수를 결정하는 단계

를 포함하는,

객체 인식기 학습 방법.

청구항 6

제5항에 있어서,

상기 제2 긍정/부정 쌍 정보를 획득하는 단계는

상기 제2 특징과 상기 제2 추가 특징 사이의 유사도를 획득하는 단계; 및

상기 유사도에 기초하여 상기 제2 긍정/부정 쌍 정보를 획득하는 단계

를 포함하는,

객체 인식기 학습 방법.

청구항 7

제1항에 있어서,

상기 제1 손실 함수를 결정하는 단계는

상기 제2 특징 및 상기 제2 추가 특징에 적용하기 위한 상기 제1 특징의 의미(semantic) 정보를 추출하는 상기 제1 손실 함수를 결정하는 단계

를 포함하는,

객체 인식기 학습 방법.

청구항 8

제1항에 있어서,

상기 제2 손실 함수를 결정하는 단계는

상기 제1 특징에 포함된 노이즈(noise)를 억제하는 상기 제2 손실 함수를 결정하는, 객체 인식기 학습 방법.

청구항 9

제1항에 있어서,

상기 제1 입력 데이터는

상기 대상 객체에 관한 이미지를 포함하고,

상기 제1 입력 데이터에 대하여 데이터 증강을 수행하여 제1 추가 입력 데이터를 획득하는 단계

를 더 포함하는,

객체 인식기 학습 방법.

청구항 10

제9항에 있어서,

상기 제1 추가 입력 데이터를 획득하는 단계는

상기 제1 입력 데이터에 대하여 랜덤 파라미터 왜곡(Random Parameter Distortion, RPD)을 적용하는 단계

를 포함하는,

객체 인식기 학습 방법.

청구항 11

제10항에 있어서,

상기 랜덤 파라미터 왜곡은

상기 이미지의 스케일, 파라미터 및 바운딩 박스(bounding box) 중 적어도 하나 이상을 임의로 변형하는 방법

을 포함하는,

객체 인식기 학습 방법.

청구항 12

제1항에 있어서,

상기 제2 입력 데이터는

상기 대상 객체에 관한 라이다 포인트 세트(LiDAR point set)를 포함하고,

제2 추가 입력 데이터를 획득하는 단계는

상기 제2 입력 데이터에 대하여 RPS(Random Point Sparsity)를 적용하여 상기 제2 추가 입력 데이터를 획득하는 단계

를 포함하는,

객체 인식기 학습 방법.

청구항 13

제12항에 있어서,

상기 RPS는

상기 라이다 포인트 세트에 보간법(interpolation)을 적용하는 방법

을 포함하는,

객체 인식기 학습 방법.

청구항 14

대상 객체에 관한 제1 입력 데이터 및 제2 입력 데이터를 획득하고, 상기 제2 입력 데이터에 대하여 데이터 증강을 수행하여 제2 추가 입력 데이터를 획득하는 데이터 증강 모듈;

상기 제1 입력 데이터에서 공유 임베딩 스페이스에 제1 특징(feature)을 추출하고, 상기 제2 입력 데이터에서 상기 공유 임베딩 스페이스에 제2 특징을 추출하고, 상기 제2 추가 입력 데이터에서 상기 공유 임베딩 스페이스에 제2 추가 특징을 추출하는 인코딩 모듈; 및

상기 제1 특징, 상기 제2 특징 및 상기 제2 추가 특징에 기초하여 제1 손실 함수를 결정하고, 상기 제2 특징 및 상기 제2 추가 특징에 기초하여 제2 손실 함수를 결정하고, 상기 제1 손실 함수 및 상기 제2 손실 함수에 기초하여 상기 제2 인코더의 가중치(weight)를 업데이트하는 크로스 모달 대조 학습 모듈

을 포함하는, 객체 인식기 학습 장치.

청구항 15

제14항에 있어서,

상기 크로스 모달 대조 학습 모듈은

상기 제1 특징과 상기 제2 특징 및 상기 제2 추가 특징 사이의 제1 긍정/부정 쌍(positive/negative pair) 정보를 획득하고,

상기 제1 긍정/부정 쌍 정보에 기초하여 상기 제1 손실 함수를 결정하는,

객체 인식기 학습 장치.

청구항 16

제15항에 있어서,

상기 크로스 모달 대조 학습 모듈은

상기 제1 특징과 상기 제2 특징 및 상기 제2 추가 특징 사이의 유사도를 획득하고,

상기 유사도에 기초하여 상기 제1 긍정/부정 쌍 정보를 획득하는,

객체 인식기 학습 장치.

청구항 17

제16항에 있어서,

상기 크로스 모달 대조 학습 모듈은

상기 제1 특징, 상기 제2 특징 및 상기 제2 추가 특징 각각에 대응하는 클래스 정보를 획득하고,

상기 클래스 정보에 기초하여 상기 제1 긍정/부정 쌍 정보를 획득하는,

객체 인식기 학습 방법.

청구항 18

제14항에 있어서,

상기 크로스 모달 대조 학습 모듈은

상기 제2 특징과 상기 제2 추가 특징 사이의 제2 긍정/부정 쌍 정보를 획득하고,

상기 제2 긍정/부정 쌍 정보에 기초하여 상기 제2 손실 함수를 결정하는,

객체 인식기 학습 방법.

청구항 19

제18항에 있어서,

상기 크로스 모달 대조 학습 모듈은

상기 제2 특징과 상기 제2 추가 특징 사이의 유사도를 획득하고,

상기 유사도에 기초하여 상기 제2 긍정/부정 쌍 정보를 획득하는,

객체 인식기 학습 방법.

청구항 20

대상 객체에 관한 타겟 라이다 포인트 세트(LiDAR point set)를 센싱하는 라이다 센서; 및

프로세서

를 포함하고,

상기 프로세서는

상기 타겟 라이다 포인트 세트를 인공 신경망 모델에 입력하여, 상기 대상 객체에 대응하는 특징 벡터를 획득하고,

상기 특징 벡터를 상기 인공 신경망 모델의 디텍션 헤드(detection head)에 입력하여 상기 대상 객체를 추정하고,

상기 인공 신경망 모델은

상기 타겟 라이다 포인트 세트와 상이한 도메인을 갖는 소스 라이다 포인트 세트와 이미지 데이터에 기초하여 학습되는, 전자 장치.

발명의 설명

기 술 분 야

[0001] 아래의 개시는 객체 인식기 학습 방법 및 장치에 관한 것이다.

배 경 기 술

[0002] 최근 첨단 운전자 지원 시스템(Advanced Driver Assistance Systems, ADAS)에서 삼차원 객체 인식(3D Objection Detection) 기술이 활용되고 있다. 삼차원 객체 인식 기술은Front Facing Cameras, Multi-Cameras, Surround View Monitor(SVM)등에 사용될 수 있다. 또한, 삼차원 객체 인식 기술은 자율 주행 차량 주변 객체의 정확한 위치(차량 기준 객체의 3차원 위치) 및 분류 정보(예: 보행자, 차량, 신호등)를 인식하고 최적의 주행 경로를 판단하기 위한 인식기능으로 안전하고 효과적인 자율주행 알고리즘 개발에 필수적인 기술일 수 있다.

발명의 내용

해결하려는 과제

과제의 해결 수단

[0004] 일 실시예에 따른 객체 인식기 학습 방법은 대상 객체에 관한 제1 입력 데이터 및 제2 입력 데이터를 획득하는 단계, 상기 제2 입력 데이터에 대하여 데이터 증강을 수행하여 제2 추가 입력 데이터를 획득하는 단계, 상기 제1 입력 데이터를 제1 인코더에 입력하여 공유 임베딩 스페이스에 제1 특징(feature)을 추출하는 단계, 상기 제2 입력 데이터를 제2 인코더에 입력하여 상기 공유 임베딩 스페이스에 제2 특징을 추출하는 단계, 상기 제2 추가 입력 데이터를 상기 제2 인코더에 입력하여 상기 공유 임베딩 스페이스에 제2 추가 특징을 추출하는 단계, 상기 제1 특징, 상기 제2 특징 및 상기 제2 추가 특징에 기초하여 제1 손실 함수를 결정하는 단계, 상기 제2 특징 및 상기 제2 추가 특징에 기초하여 제2 손실 함수를 결정하는 단계 및 상기 제1 손실 함수 및 상기 제2 손실 함수에 기초하여 상기 제2 인코더의 가중치(weight)를 업데이트하는 단계를 포함할 수 있다.

[0005] 상기 제1 손실 함수를 결정하는 단계는 상기 제1 특징과 상기 제2 특징 및 상기 제2 추가 특징 사이의 제1 긍정/부정 쌍(positive/negative pair) 정보를 획득하는 단계 및 상기 제1 긍정/부정 쌍 정보에 기초하여 상기 제1 손실 함수를 결정하는 단계를 포함할 수 있다.

[0006] 상기 제1 긍정/부정 쌍 정보를 획득하는 단계는 상기 제1 특징과 상기 제2 특징 및 상기 제2 추가 특징 사이의 유사도를 획득하는 단계 및 상기 유사도에 기초하여 상기 제1 긍정/부정 쌍 정보를 획득하는 단계를 포함할 수 있다.

[0007] 상기 제1 긍정/부정 쌍 정보를 획득하는 단계는 상기 제1 특징, 상기 제2 특징 및 상기 제2 추가 특징 각각에 대응하는 클래스 정보를 획득하는 단계 및 상기 클래스 정보에 기초하여 상기 제1 긍정/부정 쌍 정보를 획득하는 단계를 포함할 수 있다.

[0008] 상기 제2 손실 함수를 결정하는 단계는 상기 제2 특징과 상기 제2 추가 특징 사이의 제2 긍정/부정 쌍 정보를 획득하는 단계 및 상기 제2 긍정/부정 쌍 정보에 기초하여 상기 제2 손실 함수를 결정하는 단계를 포함할 수 있다.

[0009] 상기 제2 긍정/부정 쌍 정보를 획득하는 단계는 상기 제2 특징과 상기 제2 추가 특징 사이의 유사도를 획득하는 단계 및 상기 유사도에 기초하여 상기 제2 긍정/부정 쌍 정보를 획득하는 단계를 포함할 수 있다.

[0010] 상기 제1 손실 함수를 결정하는 단계는 상기 제2 특징 및 상기 제2 추가 특징에 적용하기 위한 상기 제1 특징의 의미(semantic) 정보를 추출하는 상기 제1 손실 함수를 결정하는 단계를 포함할 수 있다.

[0011] 상기 제2 손실 함수를 결정하는 단계는 상기 제1 특징에 포함된 노이즈(noise)를 억제하는 상기 제2 손실 함수를 결정할 수 있다.

[0012] 상기 제1 입력 데이터는 상기 대상 객체에 관한 이미지를 포함하고, 상기 제1 입력 데이터에 대하여 데이터 증강을 수행하여 제1 추가 입력 데이터를 획득하는 단계를 더 포함할 수 있다.

[0013] 상기 제1 추가 입력 데이터를 획득하는 단계는 상기 제1 입력 데이터에 대하여 랜덤 파라미터 왜곡(Random Parameter Distortion, RPD)을 적용하는 단계를 포함할 수 있다.

[0014] 상기 랜덤 파라미터 왜곡은 상기 이미지의 스케일, 파라미터 및 바운딩 박스(bounding box) 중 적어도 하나 이상을 임의로 변형하는 방법을 포함할 수 있다.

[0015] 상기 제2 입력 데이터는 상기 대상 객체에 관한 라이다 포인트 세트(LiDAR point set)를 포함하고, 제2 추가 입력 데이터를 획득하는 단계는 상기 제2 입력 데이터에 대하여 RPS(Random Point Sparsity)를 적용하여 상기 제2 추가 입력 데이터를 획득하는 단계를 포함할 수 있다.

[0016] 상기 RPS는 상기 라이다 포인트 세트에 보간법(interpolation)을 적용하는 방법을 포함할 수 있다.

[0017] 일 실시예에 따른 객체 인식기 학습 장치는 대상 객체에 관한 제1 입력 데이터 및 제2 입력 데이터를 획득하고,

상기 제2 입력 데이터에 대하여 데이터 증강을 수행하여 제2 추가 입력 데이터를 획득하는 데이터 증강 모듈, 상기 제1 입력 데이터에서 공유 임베딩 스페이스에 제1 특징(feature)을 추출하고, 상기 제2 입력 데이터에서 상기 공유 임베딩 스페이스에 제2 특징을 추출하고, 상기 제2 추가 입력 데이터에서 상기 공유 임베딩 스페이스에 제2 추가 특징을 추출하는 인코딩 모듈 및 상기 제1 특징, 상기 제2 특징 및 상기 제2 추가 특징에 기초하여 제1 손실 함수를 결정하고, 상기 제2 특징 및 상기 제2 추가 특징에 기초하여 제2 손실 함수를 결정하고, 상기 제1 손실 함수 및 상기 제2 손실 함수에 기초하여 상기 제2 인코더의 가중치(weight)를 업데이트하는 크로스 모달 대조 학습 모듈을 포함할 수 있다.

[0018] 상기 크로스 모달 대조 학습 모듈은 상기 제1 특징과 상기 제2 특징 및 상기 제2 추가 특징 사이의 제1 긍정/부정 쌍(positive/negative pair) 정보를 획득하고, 상기 제1 긍정/부정 쌍 정보에 기초하여 상기 제1 손실 함수를 결정할 수 있다.

[0019] 상기 크로스 모달 대조 학습 모듈은 상기 제1 특징과 상기 제2 특징 및 상기 제2 추가 특징 사이의 유사도를 획득하고, 상기 유사도에 기초하여 상기 제1 긍정/부정 쌍 정보를 획득할 수 있다.

[0020] 상기 크로스 모달 대조 학습 모듈은 상기 제1 특징, 상기 제2 특징 및 상기 제2 추가 특징 각각에 대응하는 클래스 정보를 획득하고, 상기 클래스 정보에 기초하여 상기 제1 긍정/부정 쌍 정보를 획득할 수 있다.

[0021] 상기 크로스 모달 대조 학습 모듈은 상기 제2 특징과 상기 제2 추가 특징 사이의 제2 긍정/부정 쌍 정보를 획득하고, 상기 제2 긍정/부정 쌍 정보에 기초하여 상기 제2 손실 함수를 결정할 수 있다.

[0022] 상기 크로스 모달 대조 학습 모듈은 상기 제2 특징과 상기 제2 추가 특징 사이의 유사도를 획득하고, 상기 유사도에 기초하여 상기 제2 긍정/부정 쌍 정보를 획득할 수 있다.

[0023] 일 실시예에 따른 전자 장치는 대상 객체에 관한 타겟 라이다 포인트 세트(LiDAR point set)를 센싱하는 라이다 센서 및 프로세서를 포함하고, 상기 프로세서는 상기 타겟 라이다 포인트 세트를 인공 신경망 모델에 입력하여, 상기 대상 객체에 대응하는 특징 벡터를 획득하고, 상기 특징 벡터를 상기 인공 신경망 모델의 디텍션 헤드(detection head)에 입력하여 상기 대상 객체를 추정하고, 상기 인공 신경망 모델은 상기 타겟 라이다 포인트 세트와 상이한 도메인을 갖는 소스 라이다 포인트 세트와 이미지 데이터에 기초하여 학습될 수 있다.

발명의 효과

도면의 간단한 설명

[0026] 도 1은 일 실시예에 따른 객체 인식기 학습 장치의 동작 과정을 개략적으로 도시한 것이다.

도 2는 일 실시예에 따른 객체 인식기 학습 장치를 개략적으로 도시한 것이다.

도 3은 일 실시예에 따른 객체 인식기 학습 장치의 동작을 설명하기 위한 순서도이다.

도 4a는 일 실시예에 따른 특징 벡터들을 간략하게 도시한 것이다.

도 4b는 일 실시예에 따른 인트라(intra) 모달 대조 학습을 간략하게 도시한 것이다.

도 4c는 일 실시예에 따른 인터(inter) 모달 대조 학습을 간략하게 도시한 것이다.

발명을 실시하기 위한 구체적인 내용

[0027] 실시예들에 대한 특정한 구조적 또는 기능적 설명들은 단지 예시를 위한 목적으로 개시된 것으로서, 다양한 형태로 변경되어 구현될 수 있다. 따라서, 실제 구현되는 형태는 개시된 특정 실시예로만 한정되는 것이 아니며, 본 명세서의 범위는 실시예들로 설명한 기술적 사상에 포함되는 변경, 균등물, 또는 대체물을 포함한다.

[0028] 제1 또는 제2 등의 용어를 다양한 구성요소들을 설명하는데 사용될 수 있지만, 이런 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 해석되어야 한다. 예를 들어, 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소는 제1 구성요소로도 명명될 수 있다.

[0029] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다.

[0030] 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 설명된 특징, 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함으로 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0031] 본 문서에서, "A 또는 B", "A 및 B 중 적어도 하나", "A 또는 B 중 적어도 하나", "A, B 또는 C", "A, B 및 C 중 적어도 하나", 및 "A, B, 또는 C 중 적어도 하나"와 같은 문구들 각각은 그 문구들 중 해당하는 문구에 함께 나열된 항목들 중 어느 하나, 또는 그들의 모든 가능한 조합을 포함할 수 있다.

[0032] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 해당 기술분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가진다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 갖는 것으로 해석되어야 하며, 본 명세서에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

[0033] 실시예들은 퍼스널 컴퓨터, 랩톱 컴퓨터, 태블릿 컴퓨터, 스마트 폰, 텔레비전, 스마트 가전 기기, 지능형 자동차, 키오스크, 웨어러블 장치 등 다양한 형태의 제품으로 구현될 수 있다. 이하, 실시예들을 첨부된 도면들을 참조하여 상세하게 설명한다. 첨부 도면을 참조하여 설명함에 있어, 도면 부호에 관계없이 동일한 구성 요소는 동일한 참조 부호를 부여하고, 이에 대한 중복되는 설명은 생략하기로 한다.

[0035] 도 1은 일 실시예에 따른 객체 인식기 학습 장치의 동작 과정을 개략적으로 도시한 것이다.

[0036] 도 1의 하나 이상의 블록들 및 블록들의 조합은 특정 기능을 수행하는 특수 목적 하드웨어 기반 컴퓨터, 또는 특수 목적 하드웨어 및 컴퓨터 명령들의 조합에 의해 구현될 수 있다.

[0037] 삼차원에서 객체의 위치와 카테고리를 분류하는 삼차원 객체 탐지 기술은 최근 고정밀 LiDAR 센서를 활용하여 인상적인 성능을 산출하고 있으며, 특히 자율주행 분야에서 핵심적인 역할을 하고 있다. 하지만, 데이터 수집에 방대한 자원이 요구될 수 있고, 대부분의 데이터셋은 편중된 구조(예: LiDAR 해상도, 지역 및 날씨)를 지니고 있으며, 편중된 환경은 네트워크가 특정 도메인(domain)에 매몰되도록 유도할 수 있다. 전술한 문제들을 해결하기 위해 UDA(Unsupervised Domain Adaptation)는 학습하지 않은 타 도메인에서도 일관된 성능을 유지하는 것을 목표로, 레이블링(labeling)된 소스 데이터셋에서 학습한 모델을 레이블링되지 않은 타겟에서 성능을 검증할 수 있다.

[0038] LiDAR 기반 삼차원 객체 탐지는 LiDAR 센서로부터 360도의 서라운드 뷰 포인트 클라우드(Surround View Point Cloud)를 수집하여, Voxel 또는 Pillar 형태로 변환하여 인코더(Encoder)에 입력하는 기술일 수 있다. 인코더에서 출력된 각각의 특징(feature)들은 디텍션 헤드(Detection Head)에서 디코딩(decoding) 과정을 거쳐 객체의 위치 정보들을 담고 있는 바운딩 박스(Bounding Box)들과 클래스(class)를 예측할 수 있다.

[0039] 도 1을 참조하면, 일 실시예에 따른 객체 인식기 학습 장치(100)는 소스 도메인(101)(Source domain)에서 레이블링(labeling) 된 데이터셋에 기초하여 학습된 인공 신경망 모델을 레이블링되지 않은 데이터셋을 학습하는 타겟 도메인(102)(Target domain)의 인공 신경망 모델에 적용할 수 있다.

[0040] 객체 인식기 학습 장치(100)는 이미지(I^s)를 입력으로 수신할 수 있다. 이미지는 객체 인식기 학습 장치(100)에서 랜덤 파라미터 왜곡(Random Parameter Distortion, RPD) 방법을 통해 추가 이미지로 변형될 수 있다. 이미지 인코더(Image Encoder)는 이미지의 특징(F_I^s) 및 추가 이미지의 특징을 공유 임베딩 스페이스(shared embedding space)로 추출할 수 있다.

[0041] 객체 인식기 학습 장치(100)는 소스 도메인의 라이다 포인트 클라우드(LiDAR Point Cloud)(이하, 라이다 포인트

세트(LiDAR Point set)라고 한다)(P^s)를 입력으로 수신할 수 있다. 라이다 포인트 세트는 객체 인식기 학습 장치(100)에서 RPS(Random Point Sparsity) 방법을 통해 추가 라이다 포인트 세트($\hat{P}^s$)로 변형될 수 있다. 라이다 인코더는 라이다 포인트 세트의 특징(F_P^s) 및 추가 라이다 포인트 세트($F_{\hat{P}}^s$)의 특징을 공유 임베딩 스페이스로 추출할 수 있다.

[0042] 객체 인식기 학습 장치(100)는 크로스 모달 대조 학습을 통해 상기 특징들을 비교 학습할 수 있다. 객체 인식기 학습 장치(100)는 비교 학습 결과에 기초하여, 라이다 인코더 및 디텍션 헤드(Detection Head)의 인공 신경망 모델을 업데이트 할 수 있다. 디텍션 헤드는 레이블링(Labeling)된 소스 도메인의 라이다 포인트 세트를 학습하여 클래스(L_{cls}) 정보 및 바운딩 박스(L_{box}) 정보를 추출할 수 있다.

[0043] 객체 인식기 학습 장치(100)는 소스 도메인에서 업데이트된 인공 신경망 모델을 타겟 도메인의 라이다 인코더 및 디텍션 헤드에 적용하여 타겟 도메인의 객체 추정 인식 성능이 향상될 수 있도록 할 수 있다.

[0044] 이하 도 2 내지 도 4c에서 객체 인식기 학습 장치(100)의 동작 방법을 상세히 설명한다.

[0046] 도 2는 일 실시예에 따른 객체 인식기 학습 장치를 개략적으로 도시한 것이다.

[0047] 도 2의 하나 이상의 블록들 및 블록들의 조합은 특정 기능을 수행하는 특수 목적 하드웨어 기반 컴퓨터, 또는 특수 목적 하드웨어 및 컴퓨터 명령들의 조합에 의해 구현될 수 있다.

[0048] 도 1을 참조한 설명은 도 2를 참조한 설명에도 동일하게 적용될 수 있고, 중복되는 내용은 생략될 수 있다.

[0049] 도 2를 참조하면, 일 실시예에 따른 객체 인식기 학습 장치(100)는 데이터 증강(Data Augmentation) 모듈, 인코더(Encoder) 및 크로스 모달 대조 학습(Cross-modal Contrastive Learning) 모듈을 포함할 수 있다. 데이터 증강 모듈은 제1 데이터 증강 모듈(110-1) 및 제2 데이터 증강 모듈(120-1)을 포함할 수 있다. 인코더는 제1 인코더(111) 및 제2 인코더(121)를 포함할 수 있다.

[0050] "모듈"은, 일체로 구성된 부품의 최소 단위 또는 그 일부가 될 수 있다. "모듈"은 하나 또는 그 이상의 기능을 수행하는 최소 단위 또는 그 일부가 될 수도 있다. "모듈"은 기계적으로 또는 전자적으로 구현될 수 있다. 예를 들면,"모듈"은, 알려졌거나 앞으로 개발될, 어떤 동작들을 수행하는 ASIC(application-specific integrated circuit) 칩, FPGAs(field-programmable gate arrays) 또는 프로그램 가능 논리 장치(programmable-logic device) 중 적어도 하나를 포함할 수 있다.

[0051] 일 실시예에 따른 제1 입력 데이터(110)는 대상 객체에 관한 이미지(I^s)를 포함할 수 있다. 이미지(예: 제1 입력 데이터(110))의 도메인 변경에서는 카메라의 고유한(intrinsic) 특성 또는 외적인(extrinsic) 요인으로 인해 파라미터 편차가 발생할 수 있다. 파라미터 편차가 발생하면, 삼차원 공간에서 객체의 위치를 예측할 때 기하학적(geometric) 오차가 발생될 수 있다.

[0052] 제1 데이터 증강 모듈(111)은 제1 입력 데이터(110)에 대하여 데이터 증강(data argumentation)을 수행하여 제1 추가 입력 데이터(110-1)를 획득할 수 있다. 제1 데이터 증강 모듈(111)은 데이터 증강 방법으로 랜덤 파라미터 왜곡(Random Parameter Distortion, RPD)을 적용할 수 있다. 랜덤 파라미터 왜곡은 이미지의 스케일, 파라미터 및 바운딩 박스(bounding box) 중 적어도 하나 이상을 임의로 변형하는 방법이다. 아래 수학식 1과 같은 방법으로 랜덤 파라미터 왜곡을 적용할 수 있다.

수학식 1

[0053]
$$t_n = \begin{pmatrix} \text{focal}_x & 0 & \text{center}_u & 0 \\ 0 & \text{focal}_y & \text{center}_v & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

[0054] 수학식 1의 행렬에서, $T = \{t_1, t_2, t_3, \cdots, t_n\}$는 LiDAR 좌표를 이미지에 투영시켜주는 변형 행렬들이다. 'focal'은 초점 거리를 나타내며, 'center'는 카메라의 중심 좌표를 의미한다.

수학식 2

[0055] $\tilde{T} = \tilde{K} \cdot T$

[0056] 수학식 2는 수학식 1에 표현된 변형 행렬을 동차 좌표 (Homogeneous Coordinate)로 표현된 scale factor $K = [k_1, k_2, \ldots, k_N]$ 와 곱한 수식을 표현한다. 스케일 팩터 K를 이용하여, 제1 데이터 증강 모듈(111)은 입력된 이미지($I \in \mathbb{R}^{H \times W \times 3}$)를 변형된 이미지($\hat{I} \in \mathbb{R}^{KH \times HW \times 3}$)로 생성할 수 있다.

[0057] 일 실시예에 따른 제1 인코더(112)는 제1 입력 데이터(110)를 수신할 수 있다. 제1 인코더(112)는 수신한 제1 입력 데이터(110)에서 제1 특징(feature)(F_I^s)(112-1)를 추출할 수 있다. 제1 인코더(112)는 제1 추가 입력 데이터(110-1)를 수신할 수 있다. 제1 인코더(112)는 수신한 제1 추가 입력 데이터(110-1)에서 제1 추가 특징(112-2)을 추출할 수 있다.

[0058] 일 실시예에 따른 제2 입력 데이터(120)는 대상 객체에 관한 라이다 포인트 세트(LiDAR point set)를 포함할 수 있다. 라이다 기반의 삼차원 객체 인식은 특정 도메인(예: LiDAR beam, 기후 및 지역)과 다른 도메인으로 변경되는 상황에서 일관된 성능(예: 해상도)이 산출되지 않을 수 있다. 전술한 문제를 해결하기 위해, 라이다 포인트 세트에 RPS(Random Point Sparsity)를 적용하여 제2 추가 입력 데이터(120-1)를 획득할 수 있다.

[0059] 제2 데이터 증강 모듈(121)은 제2 입력 데이터(120)에 RPS를 적용하여 제2 추가 입력 데이터(120-1)를 획득할 수 있다. RPS는 데이터셋 별 라이다의 해상도 차이를 보간법(interpolation)을 적용하여 무작위로 라이다 포인트 세트의 밀집도를 조절하는 방법이다. 보간법이란 알고 있는 데이터 값들을 이용하여 모르는 값을 추정하는 방법의 한 종류이다. 예를 들어, 실변수 x 의 함수 f(x)의 모양은 미지이나, 어떤 간격(예: 등간격 또는 부등간격)을 가지는 2개 이상인 변수의 값 x_i(i=1,2,⋯,n)에 대한 함수값 $f(x_i)$가 알려져 있을 경우, 그 사이의 임의의 x에 대한 함수값을 추정하는 것을 말한다. 실험이나 관측에 의하여 얻은 관측값으로부터 관측하지 않은 점에서의 값을 추정하는 경우나 로그표등의 함수표에서 표에 없는 함수값을 구하는 등의 경우에 이용된다.

[0060] 일 실시예에 따른 제2 인코더(122)는 제2 입력 데이터(120)를 수신할 수 있다. 제2 인코더(122)는 수신한 제2 입력 데이터(120)에서 제2 특징(feature)(F_P^s)(122-1)를 추출할 수 있다. 제2 인코더(122)는 제2 추가 입력 데이터(120-1)를 수신할 수 있다. 제2 인코더(122)는 수신한 제2 추가 입력 데이터(120-1)에서 제2 추가 특징($F_{\hat{P}}^s$)(122-2)을 추출할 수 있다.

[0061] 객체 인식기 학습 장치(100)는 제1 특징(111-1), 제1 추가 특징(111-2), 제2 특징(121-1) 및 제2 추가 특징(121-2)를 공유 임베딩 스페이스(shared embedding space)에 전달할 수 있다. 즉, 제1 인코더(112) 및 제2 인코더(122)는 공유 임베딩 스페이스에 상기 특징들을 추출할 수 있다.

[0062] 일 실시예에 따른 크로스 모달 대조 학습 모듈(130)은 공유 임베딩 스페이스에 전달된 상기 특징들에 기초하여, 대조 학습을 수행할 수 있다. 크로스 모달 대조 학습 모듈(130)은 인트라 모달(Intra-modal) 대조 학습 및 인터 모달(Inter-modal) 대조 학습을 수행할 수 있다. 후술하는 도 3 내지 도 4c에서 대조 학습 방법에 대해 상세히 설명된다.

[0064] 도 3은 일 실시예에 따른 객체 인식기 학습 장치의 동작을 설명하기 위한 순서도이다.

[0065] 도 3의 동작은 도시된 순서 및 방식으로 수행될 수 있지만, 도시된 실시예의 사상 및 범위를 벗어나지 않으면서 일부 동작의 순서가 변경되거나 일부 동작이 생략될 수 있다. 도 5에 도시된 다수의 동작은 병렬로 또는 동시에 수행될 수 있다.

[0066] 도 1내지 도 2를 참조한 설명은 도 3에도 적용될 수 있고, 중복되는 내용은 생략될 수 있다.

[0067] 설명의 편의를 위해, 단계들(310 내지 350)은 도 2에 도시된 객체 인식기 학습 장치(100)를 사용하여 수행되는 것으로 기술된다. 그러나 이 단계들(310 내지 350)은 어떤 다른 적절한 전자 기기를 통해, 그리고 어떤 적절한 시스템 내에서도 사용될 수 있을 것이다.

[0068] 단계(310)에서, 일 실시예에 따른 객체 인식기 학습 장치(100)는 대상 객체에 관한 제1 입력 데이터(110) 및 제2 입력 데이터(120)를 획득할 수 있다. 제1 입력 데이터(110) 대상 객체에 관한 이미지를 포함할 수 있다. 제2 입력 데이터(120)는 대상 객체에 관한 라이다 포인트 세트(LiDAR Point set)를 포함할 수 있다.

[0069] 단계(320)에서, 일 실시예에 따른 객체 인식기 학습 장치(100)는 제2 입력 데이터(120)에 대하여 데이터 증강을 수행하여 제2 추가 입력 데이터(120-1)를 획득할 수 있다.

[0070] 객체 인식기 학습 장치(100)의 제2 데이터 증강 모듈(121)은 제2 입력 데이터(120)에 대하여 RPS를 적용하여 제2 추가 입력 데이터(120-1)를 획득할 수 있다. RPS는 라이다 포인트 세트에 보간법을 적용하는 방법일 수 있다.

[0071] 객체 인식기 학습 장치(100)의 제1 데이터 증강 모듈(111)은 제1 입력 데이터(110)에 대하여 데이터 증강을 수행하여 제1 추가 입력 데이터(110-1)을 획득할 수 있다. 제1 데이터 증강 모듈(111)은 제1 입력 데이터(110)에 대하여 랜덤 파라미터 왜곡을 적용할 수 있다. 랜덤 파라미터 왜곡은 이미지의 스케일, 파라미터 및 바운딩 박스 중 적어도 하나 이상을 임의로 변형하는 방법일 수 있다.

[0072] 단계(330)에서, 일 실시예에 따른 객체 인식기 학습 장치(100)는 제1 입력 데이터(110), 제2 입력 데이터(120) 및 제2 추가 입력 데이터(120-1)에 대한 각각의 특징들을 공유 임베딩 스페이스에 추출할 수 있다.

[0073] 단계(331)에서, 일 실시예에 따른 객체 인식기 학습 장치(100)는 제1 입력 데이터(110) 제1 인코더(112)에 입력하여 공유 임베딩 스페이스에 제1 특징(112-1)를 추출할 수 있다.

[0074] 단계(332)에서, 일 실시예에 따른 객체 인식기 학습 장치(100)는 제2 입력 데이터를 제2 인코더(122)에 입력하여 공유 임베딩 스페이스에 제2 특징(122-1)을 추출할 수 있다.

[0075] 단계(333)에서, 일 실시예에 따른 객체 인식기 학습 장치(100)는 제 2 추가 입력 데이터(120-1)를 제2 인코더(122)에 입력하여 공유 임베딩 스페이스에 제2 추가 특징(122-2)을 추출할 수 있다.

[0076] 단계(340)에서, 일 실시예에 따른 객체 인식기 학습 장치(100)는 크로스 모달 대조 학습을 이용하여, 제1 손실 함수 및 제2 손실 함수를 결정할 수 있다.

[0077] 단계(341)에서, 일 실시예에 따른 객체 인식기 학습 장치(100)는 제1 특징(112-1), 제2 특징(122-1) 및 제2 추가 특징(122-2)에 기초하여, 제1 손실 함수를 결정할 수 있다.

[0078] 크로스 모달 대조 학습 모듈(130)은 제2 특징(122-1) 및 제2 추가 특징(122-2)에 적용하기 위한 제1 특징(112-1)의 의미(semantic) 정보를 추출하는 제1 손실 함수를 결정할 수 있다.

[0079] 크로스 모달 대조 학습 모듈(130)은 제1 특징(112-1)과 제2 특징(122-1) 및 제2 추가 특징(122-2) 사이의 제1

긍정/부정 쌍(positive/negative pair) 정보를 획득할 수 있다. 크로스 모달 대조 학습 모듈(130)은 제1 긍정/부정 쌍 정보에 기초하여, 제1 손실 함수를 결정할 수 있다.

[0080] 크로스 모달 대조 학습 모듈(130)은 제1 특징(112-1)과 제2 특징(122-1) 및 제2 추가 특징(122-2) 사이의 유사도를 획득할 수 있다. 크로스 모달 대조 학습 모듈(130)은 유사도에 기초하여 제1 긍정/부정 쌍 정보를 획득할 수 있다.

[0081] 크로스 모달 대조 학습 모듈(130)은 제1 특징(112-1), 제2 특징(122-1) 및 제2 추가 특징(122-2) 각각에 대응하는 클래스 정보를 획득할 수 있다. 크로스 모달 대조 학습 모듈(130)은 클래스 정보에 기초하여 제1 긍정/부정 쌍 정보를 획득할 수 있다.

[0082] 단계(342)에서, 일 실시예에 따른 객체 인식기 학습 장치는 제2 특징(122-1) 및 제2 추가 특징(122-2)에 기초하여 제2 손실 함수를 결정할 수 있다.

[0083] 크로스 모달 대조 학습 모듈(130)은 제1 특징(112-1)에 포함된 노이즈(noise)를 억제하는 제2 손실 함수를 결정할 수 있다.

[0084] 크로스 모달 대조 학습 모듈(130)은 제2 특징(122-1)과 제2 추가 특징(122-2) 사이의 제2 긍정/부정 쌍 정보를 획득할 수 있다. 크로스 모달 대조 학습 모듈(130)은 제2 긍정/부정 쌍 정보에 기초하여 제2 손실 함수를 결정할 수 있다.

[0085] 크로스 모달 대조 학습 모듈(130)은 제2 특징(122-1)과 제2 추가 특징(122-2) 사이의 유사도를 획득할 수 있다. 크로스 모달 대조 학습 모듈(130)은 유사도에 기초하여, 제2 긍정/부정 쌍 정보를 획득할 수 있다.

[0086] 크로스 모달 대조 학습 모듈(130)은 제1 특징(112-1), 제2 특징(122-1) 및 제2 추가 특징(122-2)에 대하여 수행한 동작을 제1 특징(112-1)을 제1 추가 특징(112-2)로 대체하여 수행할 수 있다.

[0087] 아래에서 도 4를 참조하여 크로스 모달 대조 학습 모듈(130)의 동작을 상세히 설명한다.

[0088] 단계(350)에서, 일 실시예에 따른 객체 인식기 학습 장치는 제1 손실 함수 및 제2 손실 함수에 기초하여, 제2 인코더의 가중치를 업데이트할 수 있다.

[0090] 도 4a는 일 실시예에 따른 특징 벡터들을 간략하게 도시한 것이다.

[0091] 도 4b는 일 실시예에 따른 인트라(intra) 모달 대조 학습을 간략하게 도시한 것이다.

[0092] 도 4c는 일 실시예에 따른 인터(inter) 모달 대조 학습을 간략하게 도시한 것이다.

[0093] 도 1 내지 도 3을 참조한 설명은, 도 4a 내지 도 4c에도 동일하게 적용될 수 있고, 중복되는 내용은 생략될 수 있다.

[0094] 도 4a를 참조하면, 제1 입력 데이터(110)의 제1 특징(112-1)은 복수의 제1 특징 벡터들(410) (F_I^s)을 포함할 수 있다. 제2 입력 데이터(120)의 제2 특징(122-1)은 복수의 제2 특징 벡터들(420) (F_P^s)을 포함할 수 있다. 제2 추가 입력 데이터(120-1)의 제2 추가 특징(122-2)은 복수의 제2 추가 특징 벡터들(430) $(F_{\hat{P}}^s)$을 포함할 수 있다.

[0095] 도 4b를 참조하면, 일 실시예에 따른 크로스 모달 대조 학습 모듈(130)은 피벗 기반 인트라 모달 대조 학습(Pivot-based Intra-modal Contrastive Learning)(이하, 인트라 모달 대조 학습 이라고 한다)을 수행할 수 있다. 인트라 모달 대조 학습은 이미지의 특징 벡터를 라이다 포인트 세트에 융합시키기 위한 학습 방법일 수 있다. 라이다 포인트 세트가 이미지의 풍부한 기하학적(rich-geometric) 의미(semantic) 정보를 포함할 수 있도록 하는 대조 학습 방법일 수 있다. 제1 특징 벡터들(410), 제2 특징 벡터들(420) 및 제2 추가 특징 벡터들(430)에는 현지화(localization) 및 객체 클래스(class) 정보가 포함될 수 있다. 인트라 모달 대조 학습은 지역 기하학적(regional) 정보와 의미론적(semantic) 정보를 연결하는 방법일 수 있다.

[0096] 인트라 모달 대조 학습은 제1 특징 벡터들(410)을 기준(pivot)으로 인스턴스(instance) 수준에서 대조 학습을

수행하는 것일 수 있다. 인트라 모달 대조 학습에서, 크로스 모달 대조 학습 모듈(130)은 유사도에 기초하여 제1 긍정/부정 쌍을 행렬 $S = \underset{1 \leqslant j \leqslant N_s}{\arg\max}\, \psi(F_i^P, F_j^I), 1 \leqslant i \leqslant N_s$ 와 같이 획득할 수 있다. N_s는 소스 도메인의 총 샘플의 수, $\psi(x,y) = \frac{x \cdot y}{\|x\| \|y\|}$ 는 모달 내 특징에 걸친 코사인 유사성(cosine similarty)를 계산할 수 있다.

[0097] 크로스 모달 대조 학습 모듈(130)은 클래스 정보에 기초하여 제1 긍정/부정 쌍을 획득할 수 있다.

[0098] 예를 들어, 크로스 모달 대조 학습 모듈(130)은 제1 특징 벡터(411) 및 제1 특징 벡터(412)를 기준으로 클래스와의 제1 긍정/부정 쌍을 획득할 수 있다. 제2 특징 벡터(421)와 제2 추가 특징 벡터(431)가 A class를 형성하고, 제2 특징 벡터(422)와 제2 추가 특징 벡터(432)가 B class를 형성할 수 있다. 크로스 모달 대조 학습 모듈(130)은 제1 특징 벡터(411)과 유사한 A class를 제1 긍정 쌍으로 구성하고, 제1 특징 벡터(411)과 비유사한 B class를 제1 부정 쌍으로 구성할 수 있다. 반대로, 크로스 모달 대조 학습 모듈(130)은 제1 특징 벡터(412)와 비유사한 A class를 제1 부정 쌍으로 구성하고, 제1 특징 벡터(412)와 유사한 B class를 제1 긍정 쌍으로 구성할 수 있다. 크로스 모달 대조 학습 모듈(130)은 클래스에 제1 긍정/부정 쌍을 상술한 유사도에 기초한 제1 긍정/부정 쌍에 적용하여 제1 긍정/부정 쌍을 업데이트할 수 있다. 업데이트된 제1 긍정/부정 쌍의 행렬은 $S = S \times C_F$ 로 표현될 수 있다. C_F는 클래스 쌍 정보를 포함하는 행렬로서, 동일한 클래스의 특징 벡터에 대해서는 1의 값을 갖고 다른 클래스의 특징 벡터에 대해서는 0의 값을 가질 수 있다.

[0099] 크로스 모달 대조 학습 모듈(130)은 풍부한 의미 정보를 포함하는 이미지의 정보를 가지는 제1 특징 벡터들(410)을 피벗(pivot)으로 설정하고 제2 특징 벡터들(420) 및 제2 추가 특징 벡터들(430) 각각과 순차적으로 대조 학습을 할 수 있다. 즉, 공유 임베딩 스페이스에서 각 모달리티로부터 동일한 클래스의 객체들은 유사한 특징 벡터들을 가질 수 있다. 크로스 모달 대조 학습 모듈(130)을 유사한 특징을 기반으로한 제1 긍정/부정 쌍을 반영하여, 수학식 3과 같은 제1 손실 함수를 결정할 수 있다.

수학식 3

[0100]
$$\mathcal{L}_{intra} = -\sum_{(i,j)\in S_+} \log \frac{\exp(\boldsymbol{f}_i^{I\top} \cdot \boldsymbol{f}_j^P/\tau)}{\sum_{(i,k)\in S_-} \exp(\boldsymbol{f}_i^{I\top} \cdot \boldsymbol{f}_k^P/\tau)}$$

[0101] f^I는 이미지의 객체 샘플(제1 특징 벡터들(410))이고, f^P는 라이다 포인트 세트(예: 포인트 클라우드)의 객체 특징(제2 특징 벡터들(420))이고, τ는 스케일링 계수이다. 크로스 모달 대조 학습 모듈(130)은 L_{intra}를 최소화 함으로써 삼차원 라이다 포인트 세트의 특징 벡터들이 이미지 특징 벡터들의 풍부한 의미론적 정보를 포함하도록 할 수 있다. 결과적으로, 인트라 모달 대조 학습은 모달리티간의 불일치를 감소시킬 수 있다.

[0102] 도 4c를 참조하면, 크로스 모달 대조 학습 모듈(130)은 도메인 적응형 인터 모달 대조 학습(Domain Adaptive Inter-modal Contrastive Learning)(이하, 인터 모달 대조 학습이라고 함)을 수행할 수 있다.

[0103] 인트라 모달 대조 학습은 멀티 모달리티 환경에서 이미지의 의미(semantic) 정보 이외에도 기하학적 노이즈가 발생할 수 있다.

[0104] 인터 모달 대조 학습은 라이다 포인트 세트에 인스턴스(instance) 수준의 대조 학습을 할 수 있다. 소스 도메인의 라이다 센서의 라이다 포인트 세트 (P)가 주어지면 데이터 증강을 수행하여 추가 라이다 포인트 세트 $(\hat{P})$를 획득할 수 있다. 크로스 모달 대조 학습 모듈(130)은 라이다 포인트 세트 (P)와 추가 라이다 포인트 세트 $(\hat{P})$에

서 추출한 제2 특징(122-1) 및 제2 추가 특징(122-1) 사이의 유사도-우선순위 기준에 따라 제2 긍정/부정 쌍을 획득할 수 있다. 인터 모달 대조 학습은 인트라 모달 대조 학습과는 달리 클래스 정보를 고려하지 않고 객체의 특징 간 유사성에 중점을 두고 대조 학습을 할 수 있다.

[0105] 예를 들어, 크로스 모달 대조 학습 모듈(130)은 제2 특징 벡터(421)를 기준으로 제2 추가 특징 벡터(431)은 유사한 긍정 쌍으로, 제2 특징 벡터(422)는 비유사한 부정 쌍으로 대조 학습할 수 있다. 다른 예로, 크로스 모달 대조 학습 모듈(130)은 제2 추가 특징 벡터(432)를 기준으로 제2 추가 특징 벡터(431)은 비유사한 부정 쌍으로, 제2 특징 벡터(422)는 유사한 긍정 쌍으로 대조 학습 할 수 있다.

[0106] 크로스 모달 대조 학습 모듈(130)은 획득한 제2 긍정/부정 쌍을 적용하여 대조 정렬에 대해 수학식 4와 같은 제2 손실 함수를 결정할 수 있다.

수학식 4

[0107]
$$\mathcal{L}_{inter} = -\sum_{(i,j)\in S_+} \log \frac{\exp(\boldsymbol{f}_i^{P\top} \cdot \boldsymbol{f}_j^{\hat{P}}/\tau)}{\sum_{(i,k)\in S_-} \exp(\boldsymbol{f}_i^{P\top} \cdot \boldsymbol{f}_k^{\hat{P}}/\tau)}$$

[0109] 수학식 4에서, f^{P}는 오리지날 라이다 포인트 세트들의 객체 샘플들이고, $f^{\hat{P}}$ 는 추가 라이다 포인트 세트들의 객체 샘플들이고, τ는 스케일링 계수이다. 제2 손실 함수 L_{inter}를 최소화함으로써, 객체 인식기는 타겟 도메인을 간접적으로 경험하여 도메인 변경에서 발생할 수 있는 문제를 감소시킬 수 있다.

[0110] 크로스 모달 대조 학습 모듈(130)은 결정된 제1 손실 함수 및 제2 손실 함수에 기초하여, 수학식 5와 같은 대조 손실 함수(L_{cont})를 결정할 수 있다.

수학식 5

[0111]
$$L_{cont} = \lambda_{intra} L_{intra} + \lambda_{inter} L_{inter}$$

[0112] 수학식 5를 참조하면, 제2 손실 함수(L_{inter})는 객체 간 클래스 별 상호 작용을 무시(ignore)하는 반면, 제1 손실 함수(L_{intra})는 데이터셋의 불균형(imbalance) 클래스 문제로 인해 임베딩 벡터가 동일한 클래스 내에서 비차별적(non-discriminative) 객체 기능을 갖도록 구동(driving)하는 것을 방지(prevent)할 수 있다.

[0113] 크로스 모달 대조 학습 모듈(130)은 대조 손실 함수(L_{cont})외에도 삼차원 바운딩 박스 및 클래스와 관련된(associate) 전통적인 레버리지 손실 함수(L_{det})를 이용할 수 있다. 결론적으로, 크로스 모달 대조 학습 모듈은 대조 손실 함수(L_{cont}) 및 레버리지 손실 함수(L_{det})에 기초하여 총 손실 함수 (L_{total})를 결정할 수 있다.

크로스 모달 대조 학습 모듈(130)은 아래 수학식 6과 같이 총 손실 함수 (L_{total})를 이용하여, 대조 학습을 수행할 수 있다.

수학식 6

[0114] $$L_{total} = \lambda_{det} L_{det} + L_{cont}$$

[0115] 수학식 6에서, λ_{det}는 그리드 서치(grid serach)에서 L_{det}의 강도(strengh)를 핸들링(handling)하기 위해 파생(derive)된 하이퍼파라미터이다.

[0116] 이상에서 설명된 실시예들은 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치, 방법 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0117] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치, 또는 전송되는 신호 파(signal wave)에 영구적으로, 또는 일시적으로 구체화(embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

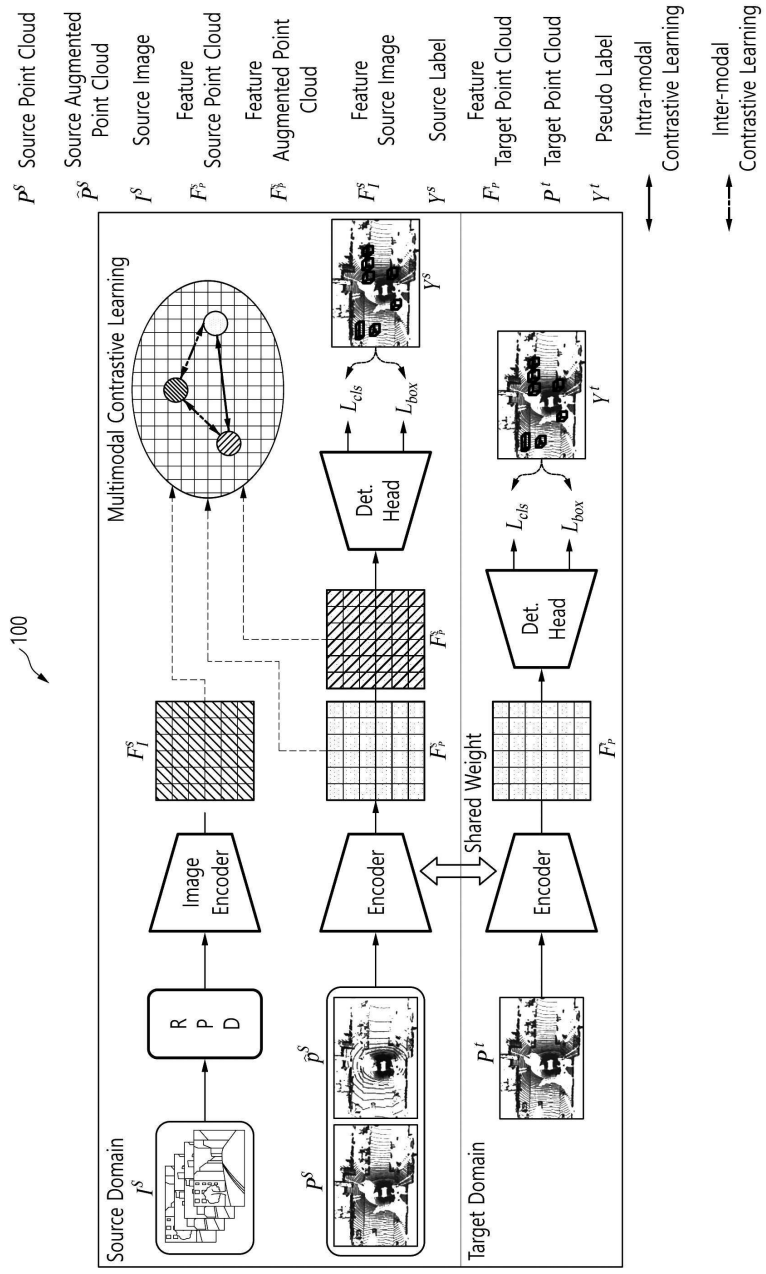
[0118] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), 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.

[0119] 이상과 같이 실시예들이 비록 한정된 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기를 기초로 다양한 기술적 수정 및 변형을 적용할 수 있다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

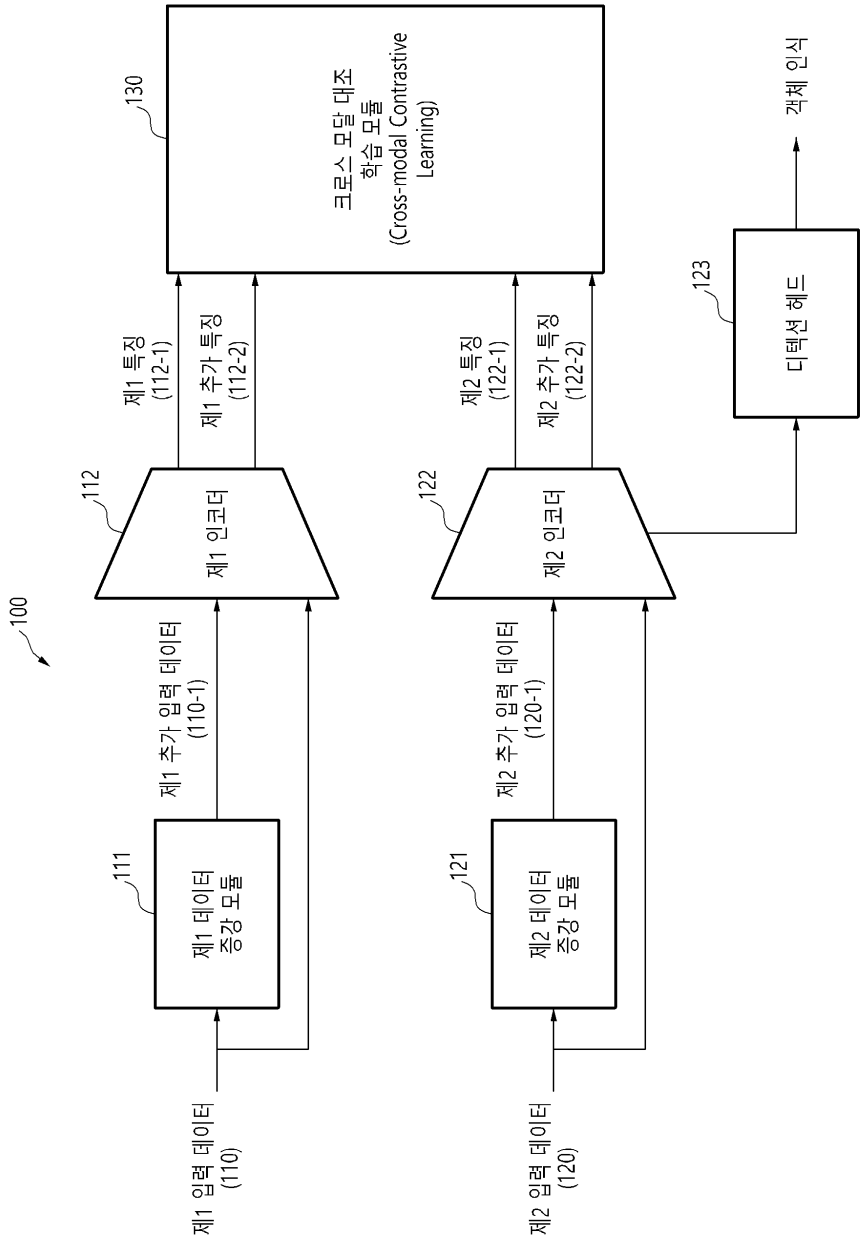
[0120] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

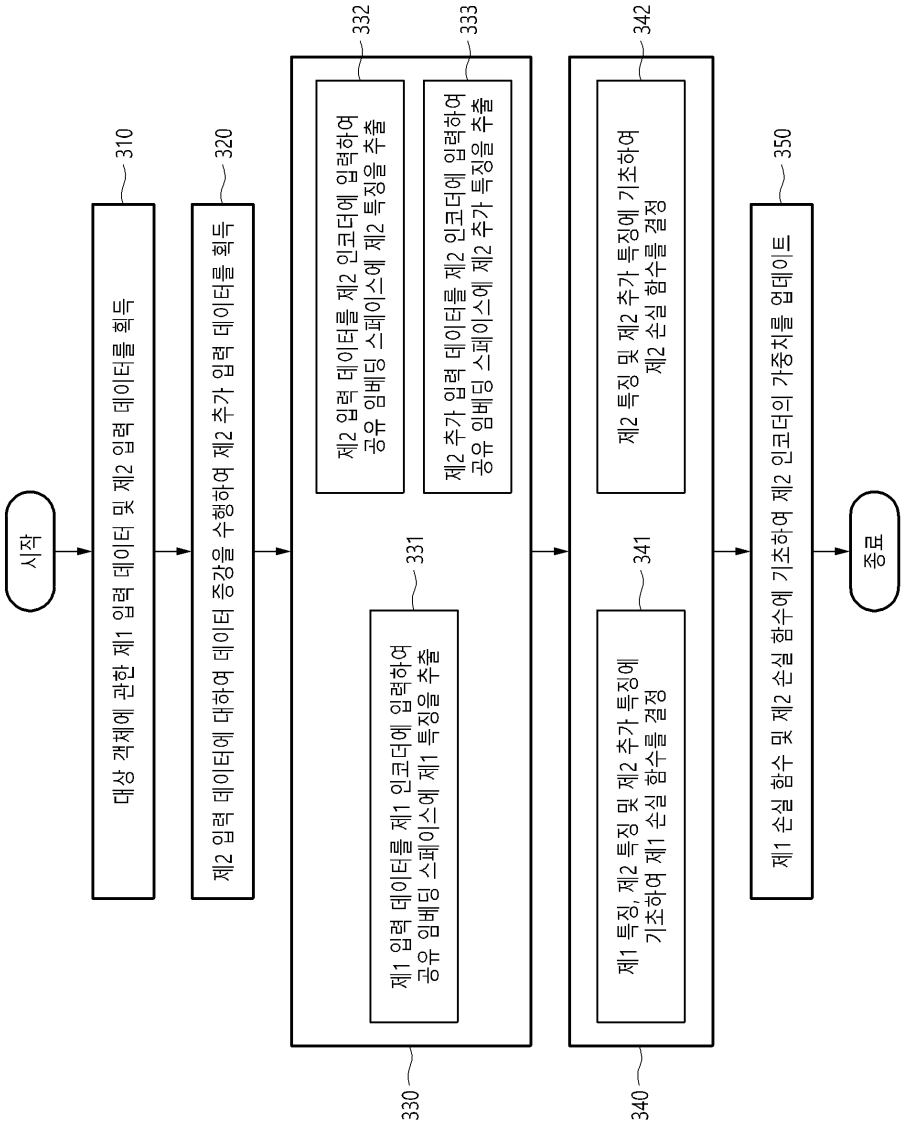
도면1


100
Source Point Cloud
Source Augmented Point Cloud
Source Image
Feature Source Point Cloud
Feature Augmented Point Cloud
Feature Source Image
Source Label
Feature Target Point Cloud
Target Point Cloud
Pseudo Label
Intra-modal Contrastive Learning
Inter-modal Contrastive Learning
Multimodal Contrastive Learning
Source Domain
Target Domain
Shared Weight
R P D
Image Encoder
Encoder
Encoder
Det. Head
Det. Head

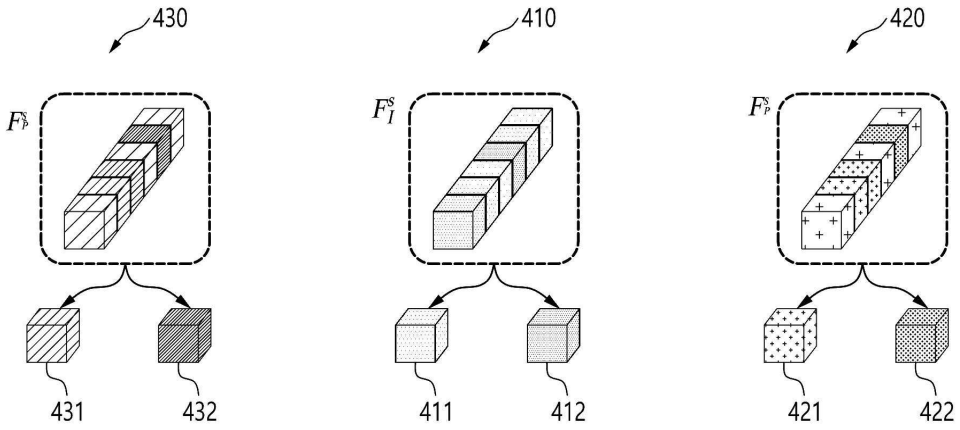
도면2


100
제1 입력 데이터
(110)
제2 입력 데이터
(120)
111
제1 데이터
증강 모듈
121
제2 데이터
증강 모듈
제1 추가 입력 데이터
(110-1)
제2 추가 입력 데이터
(120-1)
112
제1 인코더
122
제2 인코더
제1 특징
(112-1)
제1 추가 특징
(112-2)
제2 특징
(122-1)
제2 추가 특징
(122-2)
130
크로스 모달 대조
학습 모듈
(Cross-modal Contrastive
Learning)
123
디텍션 헤드
객체 인식

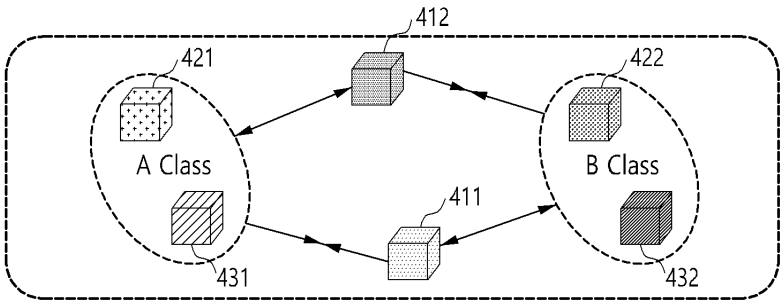
도면3


시작
대상 객체에 관한 제1 입력 데이터 및 제2 입력 데이터를 획득
310
제2 입력 데이터에 대하여 데이터 증강을 수행하여 제2 추가 입력 데이터를 획득
320
330
제1 입력 데이터를 제1 인코더에 입력하여 공유 임베딩 스페이스에 제1 특징을 추출
331
제2 입력 데이터를 제2 인코더에 입력하여 공유 임베딩 스페이스에 제2 특징을 추출
332
제2 추가 입력 데이터를 제2 인코더에 입력하여 공유 임베딩 스페이스에 제2 추가 특징을 추출
333
340
제1 특징, 제2 특징 및 제2 추가 특징에 기초하여 제1 손실 함수를 결정
341
제2 특징 및 제2 추가 특징에 기초하여 제2 손실 함수를 결정
342
제1 손실 함수 및 제2 손실 함수에 기초하여 제2 인코더의 가중치를 업데이트
350
종료

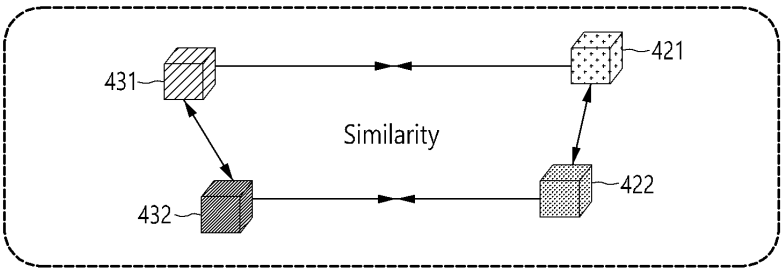
도면4a


430
410
420
431
432
411
412
421
422

도면4b


412
421
422
A Class
B Class
411
431
432

도면4c


431
421
Similarity
432
422