



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0003769
(43) 공개일자 2023년01월06일

(51) 국제특허분류(Int. Cl.)
G06V 10/10 (2023.01) G06F 18/00 (2023.01)
G06N 3/08 (2023.01)
(52) CPC특허분류
G06V 10/10 (2022.01)
G06F 18/22 (2023.01)
(21) 출원번호 10-2021-0085271
(22) 출원일자 2021년06월30일
심사청구일자 2021년06월30일

(71) 출원인
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
변혜란
서울특별시 서대문구 연세로 50, 제4공학관 810호
기민송
서울특별시 서대문구 연세로 50, 제4공학관 810호
(뒷면에 계속)
(74) 대리인
권성현, 유광철, 백두진, 강일신, 김정연

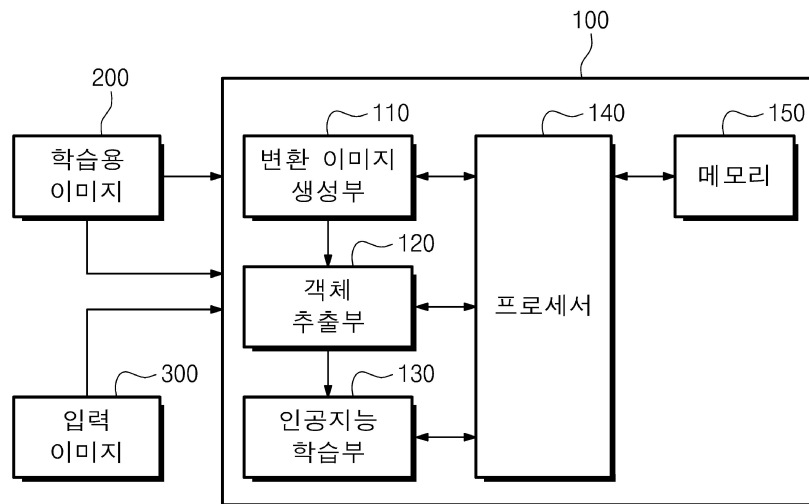
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 객체 위치 검출 장치, 객체 위치 검출 방법 및 객체 위치 검출 방법을 실행시키도록 기록매체에 저장된 컴퓨터 프로그램

(57) 요약

개시된 발명의 일 실시예에 따른 비지도 학습 기반의 객체 위치 검출 방법은, (a) 변환 이미지 생성부에 의하여, 학습용 이미지를 변환하여 변환 이미지를 생성하는 단계; (b) 객체 추출부에 의하여, 상기 학습용 이미지로부터 제1 객체 영역을 추출하고, 상기 변환 이미지로부터 제2 객체 영역을 추출하는 단계; 및 (c) 인공지능 학습부에 의하여, 상기 제1 객체 영역과 상기 제2 객체 영역을 비교하여 인공지능 모델을 학습하는 단계;를 포함할 수 있다.

대표도 - 도1



(52) CPC특허분류

G06N 3/088 (2023.01)

(72) 발명자

어영정

인천광역시 연수구 송도과학로 85, 자유관B 312호

이원영

서울특별시 서대문구 연세로 50, 제4공학관 810호

고성필

서울특별시 서대문구 연세로 50, 제4공학관 810호

이 발명을 지원한 국가연구개발사업

과제고유번호 1711134177

과제번호 2019R1A2C2003760

부처명 과학기술정보통신부

과제관리(전문)기관명 한국연구재단

연구사업명 개인기초연구(과기정통부)(R&D)

연구과제명 특성 정보 자동 생성을 통한 처음 보는 복합카테고리의 이미지와 비디오 생성 및 인식을 위한 제로샷 학습 기술 연구

기 여 율 1/2

과제수행기관명 연세대학교

연구기간 2021.03.01 ~ 2022.02.28

이 발명을 지원한 국가연구개발사업

과제고유번호 1711126082

과제번호 2020-0-01361-002

부처명 과학기술정보통신부

과제관리(전문)기관명 정보통신기획평가원

연구사업명 정보통신방송혁신인재양성(R&D)

연구과제명 인공지능대학원지원(연세대학교)

기 여 율 1/2

과제수행기관명 연세대학교 산학협력단

연구기간 2021.01.01 ~ 2021.12.31

명세서

청구범위

청구항 1

- (a) 변환 이미지 생성부에 의하여, 학습용 이미지를 변환하여 변환 이미지를 생성하는 단계;
- (b) 객체 추출부에 의하여, 상기 학습용 이미지로부터 제1 객체 영역을 추출하고, 상기 변환 이미지로부터 제2 객체 영역을 추출하는 단계; 및
- (c) 인공지능 학습부에 의하여, 상기 제1 객체 영역과 상기 제2 객체 영역을 비교하여 인공지능 모델을 학습하는 단계;를 포함하는 비지도 학습 기반의 객체 위치 검출 방법.

청구항 2

제1항에 있어서,

상기 (c) 단계는:

상기 인공지능 학습부에 의하여, 상기 학습용 이미지에서 상기 제1 객체 영역을 제외한 제1 배경 영역과 상기 변환 이미지에서 상기 제2 객체 영역을 제외한 제2 배경 영역을 비교하여 상기 인공지능 모델을 학습하는 단계;를 포함하는, 비지도 학습 기반의 객체 위치 검출 방법.

청구항 3

제2항에 있어서,

상기 (c) 단계는:

상기 제1 객체 영역과 상기 제2 객체 영역의 유사성을 분석하고, 상기 제1 배경 영역과 상기 제1 객체 영역의 차이를 분석하여 손실 함수를 연산하는 단계; 및

상기 손실 함수가 감소하도록 상기 인공지능 모델을 학습하는 단계;를 포함하는 비지도 학습 기반의 객체 위치 검출 방법.

청구항 4

제1항에 있어서,

상기 (b) 단계는:

(b1) 상기 학습용 이미지를 구성하는 복수개의 채널 중 인식된 객체의 정보가 상대적으로 많이 포함된 채널인 제1 채널에 기초하여 상기 제1 객체 영역을 추출하는 단계; 및

(b2) 상기 변환 이미지를 구성하는 복수개의 채널 중 인식된 객체의 정보가 상대적으로 많이 포함된 채널인 제2 채널에 기초하여 상기 제2 객체 영역을 추출하는 단계;를 포함하는 비지도 학습 기반의 객체 위치 검출 방법.

청구항 5

제4항에 있어서,

상기 (b1) 단계는:

상기 학습용 이미지의 채널을 구성하는 픽셀이 상기 인식된 객체의 정보를 포함하는 픽셀인지 판단하는 단계;

상기 학습용 이미지의 채널에 대하여 상기 객체의 정보를 포함하는 픽셀의 개수를 결정하는 단계; 및

상기 픽셀의 개수에 기초하여 상기 복수개의 학습용 이미지의 채널 중 인식된 객체의 정보가 상대적으로 많이 포함된 채널을 상기 제1 채널로 결정하는 단계;를 포함하고,

상기 (b2) 단계는:

상기 변환 이미지의 채널을 구성하는 픽셀이 상기 인식된 객체의 정보를 포함하는 픽셀인지 판단하는 단계;

상기 변환 이미지의 채널에 대하여 상기 객체의 정보를 포함하는 픽셀의 개수를 결정하는 단계; 및

상기 픽셀의 개수에 기초하여 상기 복수개의 변환 이미지의 채널 중 인식된 객체의 정보가 상대적으로 많이 포함된 채널을 상기 제2 채널로 결정하는 단계;를 포함하는 비지도 학습 기반의 객체 위치 검출 방법.

청구항 6

제1항에 있어서,

객체 추출부에 의하여, 상기 인공지능 모델을 이용하여 입력 이미지로부터 상기 제1 객체 영역을 추출하는 단계; 및

상기 추출된 제1 객체 영역을 기초로 상기 입력 이미지에서의 객체 위치를 결정하는 단계;를 더 포함하는 비지도 학습 기반의 객체 위치 검출 방법.

청구항 7

제1항에 있어서,

상기 (a) 단계는:

상기 변환 이미지 생성부에 의하여, 상기 학습용 이미지를 미리 정해진 각도로 회전시켜 상기 변환 이미지를 생성하는 단계;를 포함하는 비지도 학습 기반의 객체 위치 검출 방법.

청구항 8

제1항에 있어서,

상기 (a) 단계는:

상기 변환 이미지 생성부에 의하여, 상기 학습용 이미지를 확대하거나 축소하여 상기 변환 이미지를 생성하는 단계;를 포함하는 비지도 학습 기반의 객체 위치 검출 방법.

청구항 9

제1항에 있어서,

상기 (a) 단계는:

상기 변환 이미지 생성부에 의하여, 상기 학습용 이미지를 대칭 변환하여 상기 변환 이미지를 생성하는 단계;를 포함하는 비지도 학습 기반의 객체 위치 검출 방법.

청구항 10

제1항에 있어서,

상기 (c) 단계는:

상기 인공지능 학습부에 의하여, 상기 학습용 이미지에서 상기 제1 객체 영역을 제외한 제1 배경 영역과 상기 변환 이미지에서 상기 제2 객체 영역을 제외한 제2 배경 영역을 비교하여 상기 인공지능 모델을 학습하는 단계;

상기 제1 객체 영역과 상기 제2 객체 영역의 유사성을 분석하고, 상기 제1 배경 영역과 상기 제2 배경 영역의 유사성을 분석하여 손실 함수를 연산하는 단계; 및

상기 손실 함수가 감소하도록 상기 인공지능 모델을 학습하는 단계;를 포함하고,

상기 (b) 단계는:

(b1) 상기 학습용 이미지를 구성하는 복수개의 채널 중 인식된 객체의 정보가 상대적으로 많이 포함된 채널인 제1 채널에 기초하여 상기 제1 객체 영역을 추출하는 단계; 및

(b2) 상기 변환 이미지를 구성하는 복수개의 채널 중 인식된 객체의 정보가 상대적으로 많이 포함된 채널인 제2 채널에 기초하여 상기 제2 객체 영역을 추출하는 단계;를 포함하는 비지도 학습 기반의 객체 위치 검출 방법.

청구항 11

제1항 내지 제10항 중 어느 한 항의 비지도 학습 기반의 객체 위치 검출 방법을 실행시키도록 컴퓨터로 판독 가능한 기록매체에 저장된 컴퓨터 프로그램.

청구항 12

학습용 이미지를 변환하여 변환 이미지를 생성하도록 구성되는 변환 이미지 생성부;

상기 학습용 이미지로부터 제1 객체 영역을 추출하고, 상기 변환 이미지로부터 제2 객체 영역을 추출하도록 구성되는 객체 추출부; 및

상기 제1 객체 영역과 상기 제2 객체 영역을 비교하여 인공지능 모델을 학습하도록 구성되는 인공지능 학습부;를 포함하는 비지도 학습 기반의 객체 위치 검출 장치.

청구항 13

제12항에 있어서,

상기 인공지능 학습부는,

상기 학습용 이미지에서 상기 제1 객체 영역을 제외한 제1 배경 영역과 상기 변환 이미지에서 상기 제2 객체 영역을 제외한 제2 배경 영역을 비교하여 상기 인공지능 모델을 학습하도록 구성되는, 비지도 학습 기반의 객체 위치 검출 장치.

청구항 14

제13항에 있어서,

상기 인공지능 학습부는:

상기 제1 객체 영역과 상기 제2 객체 영역의 유사성을 분석하고, 상기 제1 배경 영역과 상기 제1 객체 영역의 차이를 분석하여 손실 함수를 연산하고; 그리고

상기 손실 함수가 감소하도록 상기 인공지능 모델을 학습하도록 구성되는, 비지도 학습 기반의 객체 위치 검출 장치.

청구항 15

제12항에 있어서,

상기 객체 추출부는:

상기 학습용 이미지를 구성하는 복수개의 채널 중 인식된 객체의 정보가 상대적으로 많이 포함된 채널인 제1 채널에 기초하여 상기 제1 객체 영역을 추출하고; 그리고

상기 변환 이미지를 구성하는 복수개의 채널 중 인식된 객체의 정보가 상대적으로 많이 포함된 채널인 제2 채널에 기초하여 상기 제2 객체 영역을 추출하도록 구성되는, 비지도 학습 기반의 객체 위치 검출 장치.

청구항 16

제15항에 있어서,

상기 객체 추출부는:

상기 학습용 이미지의 채널을 구성하는 픽셀이 상기 인식된 객체의 정보를 포함하는 픽셀인지 판단하고;

상기 학습용 이미지의 채널에 대하여 상기 객체의 정보를 포함하는 픽셀의 개수를 결정하고;

상기 픽셀의 개수에 기초하여 상기 복수개의 학습용 이미지의 채널 중 인식된 객체의 정보가 상대적으로 많이 포함된 채널을 상기 제1 채널로 결정하고;

상기 변환 이미지의 채널을 구성하는 픽셀이 상기 인식된 객체의 정보를 포함하는 픽셀인지 판단하고;

상기 변환 이미지의 채널에 대하여 상기 객체의 정보를 포함하는 픽셀의 개수를 결정하고; 그리고
상기 픽셀의 개수에 기초하여 상기 복수개의 변환 이미지의 채널 중 인식된 객체의 정보가 상대적으로 많이 포함된 채널을 상기 제2 채널로 결정하도록 구성되는, 비지도 학습 기반의 객체 위치 검출 장치.

청구항 17

제12항에 있어서,
프로세서;를 더 포함하고,
상기 객체 추출부는,
상기 인공지능 모델을 이용하여 입력 이미지로부터 상기 제1 객체 영역을 추출하도록 구성되고,
상기 프로세서는,
상기 추출된 제1 객체 영역을 기초로 상기 입력 이미지에서의 객체 위치를 결정하도록 구성되는, 비지도 학습 기반의 객체 위치 검출 장치.

청구항 18

제12항에 있어서,
상기 변환 이미지 생성부는,
상기 학습용 이미지를 미리 정해진 각도로 회전시켜 상기 변환 이미지를 생성하도록 구성되는, 비지도 학습 기반의 객체 위치 검출 장치.

청구항 19

제12항에 있어서,
상기 변환 이미지 생성부는,
상기 학습용 이미지를 확대하거나 축소하여 상기 변환 이미지를 생성하도록 구성되는, 비지도 학습 기반의 객체 위치 검출 장치.

청구항 20

제12항에 있어서,
상기 변환 이미지 생성부는,
상기 학습용 이미지를 대칭 변환하여 상기 변환 이미지를 생성하도록 구성되는, 비지도 학습 기반의 객체 위치 검출 장치.

발명의 설명

기술 분야

[0001] 본 발명은 이미지상 객체의 추가적인 위치 관련 정보에 기초하여 객체의 위치를 검출하는 성능을 개선할 수 있는 비지도 학습 기반의 객체 위치 검출 방법 및 객체 위치 검출 장치에 관한 것이다.

배경 기술

[0002] 기계 학습(Machine Learning)의 한 방법으로서 지도 학습(Supervised Learning)은 이미지에서의 객체 위치를 검출하기 위한 바운딩 박스(bounding box), 세그먼테이션을 사람의 주석 작업에 의하여 컨볼루션 뉴럴 네트워크(CNN: Convolution Neural Network)로 학습하는 방법이다. 그러나, 정확한 주석을 얻기 위해 사람의 큰 노동력을 필요로 한다는 문제가 있다.

[0003] 또다른 기계 학습 방법으로서 비지도 학습(Unsupervised Learning)은 인공 레이블을 분류하기 위하여 분류 손실 함수(classification loss function)를 이용하여 인공지능 모델을 훈련시키는 자가 지도 학습을 적용한다. 그러

나, 비지도 학습은 타겟과 관련된 정보를 인공지능 모델이 학습하기 때문에 배경까지 활성화가 되어 학습의 성능을 낮춘다는 문제가 발생할 수 있다.

- [0004] 또한, 기존의 대조 표현 학습은 이미지를 신경망을 통하여 특징 벡터로 인코딩하는데, 특징 벡터에는 공간 정보가 포함되어 있지 않기 때문에 손실 함수가 객체만을 커버하지 않고 차별적인 부분에 집중하거나 배경을 과도하게 커버한다는 문제가 있다. 따라서, 새로운 대조 손실 함수를 이용하여 보다 효율적인 객체의 위치를 검출할 수 있는 방법이 필요하다.

발명의 내용

해결하려는 과제

- [0005] 본 발명은 학습용 이미지상 객체의 위치 관련 정보를 추가적으로 포함하고 있는 변환 이미지에 기초하여 객체의 전체 영역을 커버할 수 있으며, 학습용 이미지의 차별적인 부분에만 집중하거나 배경 영역을 과도하게 커버하지 않을 수 있어서 객체의 위치를 정확하게 검출할 수 있는 객체 인식 시스템, 객체 인식 방법 및 컴퓨터 프로그램을 제공하는 것이다.
- [0006] 또한, 본 발명은 인공지능 모델에 대한 학습 횟수가 적더라도 종래의 객체 위치 검출 기술보다 이미지 상에서의 객체 위치를 정확하게 검출할 수 있는 비지도 학습 기반의 객체 위치 검출 장치, 객체 위치 검출 방법 및 컴퓨터 프로그램을 제공하는 것이다.

과제의 해결 수단

- [0007] 개시된 발명의 일 측면에 따른 비지도 학습 기반의 객체 위치 검출 방법은, (a) 변환 이미지 생성부에 의하여, 학습용 이미지를 변환하여 변환 이미지를 생성하는 단계; (b) 객체 추출부에 의하여, 상기 학습용 이미지로부터 제1 객체 영역을 추출하고, 상기 변환 이미지로부터 제2 객체 영역을 추출하는 단계; 및 (c) 인공지능 학습부에 의하여, 상기 제1 객체 영역과 상기 제2 객체 영역을 비교하여 인공지능 모델을 학습하는 단계;를 포함할 수 있다.
- [0008] 또한, 상기 (c) 단계는: 상기 인공지능 학습부에 의하여, 상기 학습용 이미지에서 상기 제1 객체 영역을 제외한 제1 배경 영역과 상기 변환 이미지에서 상기 제2 객체 영역을 제외한 제2 배경 영역을 비교하여 상기 인공지능 모델을 학습하는 단계;를 포함할 수 있다.
- [0009] 또한, 상기 (c) 단계는: 상기 제1 객체 영역과 상기 제2 객체 영역의 유사성을 분석하고, 상기 제1 배경 영역과 상기 제1 객체 영역의 차이를 분석하여 손실 함수를 연산하는 단계; 및 상기 손실 함수가 감소하도록 상기 인공지능 모델을 학습하는 단계;를 포함할 수 있다.
- [0010] 또한, 상기 (b) 단계는: (b1) 상기 학습용 이미지를 구성하는 복수개의 채널 중 인식된 객체의 정보가 상대적으로 많이 포함된 채널인 제1 채널에 기초하여 상기 제1 객체 영역을 추출하는 단계; 및 (b2) 상기 변환 이미지를 구성하는 복수개의 채널 중 인식된 객체의 정보가 상대적으로 많이 포함된 채널인 제2 채널에 기초하여 상기 제2 객체 영역을 추출하는 단계;를 포함할 수 있다.
- [0011] 또한, 상기 (b1) 단계는: 상기 학습용 이미지의 채널을 구성하는 픽셀이 상기 인식된 객체의 정보를 포함하는 픽셀인지 판단하는 단계; 상기 학습용 이미지의 채널에 대하여 상기 객체의 정보를 포함하는 픽셀의 개수를 결정하는 단계; 및 상기 픽셀의 개수에 기초하여 상기 복수개의 학습용 이미지의 채널 중 인식된 객체의 정보가 상대적으로 많이 포함된 채널을 상기 제1 채널로 결정하는 단계;를 포함하고, 상기 (b2) 단계는: 상기 변환 이미지의 채널을 구성하는 픽셀이 상기 인식된 객체의 정보를 포함하는 픽셀인지 판단하는 단계; 상기 변환 이미지의 채널에 대하여 상기 객체의 정보를 포함하는 픽셀의 개수를 결정하는 단계; 및 상기 픽셀의 개수에 기초하여 상기 복수개의 변환 이미지의 채널 중 인식된 객체의 정보가 상대적으로 많이 포함된 채널을 상기 제2 채널로 결정하는 단계;를 포함할 수 있다.
- [0012] 또한, 객체 추출부에 의하여, 상기 인공지능 모델을 이용하여 입력 이미지로부터 상기 제1 객체 영역을 추출하는 단계; 및 상기 추출된 제1 객체 영역을 기초로 상기 입력 이미지에서의 객체 위치를 결정하는 단계;를 더 포함할 수 있다.
- [0013] 또한, 상기 (a) 단계는: 상기 변환 이미지 생성부에 의하여, 상기 학습용 이미지를 미리 정해진 각도로 회전시켜 상기 변환 이미지를 생성하는 단계;를 포함할 수 있다.

- [0014] 또한, 상기 (a) 단계는: 상기 변환 이미지 생성부에 의하여, 상기 학습용 이미지를 확대하거나 축소하여 상기 변환 이미지를 생성하는 단계;를 포함할 수 있다.
- [0015] 또한, 상기 (a) 단계는: 상기 변환 이미지 생성부에 의하여, 상기 학습용 이미지를 대칭 변환하여 상기 변환 이미지를 생성하는 단계;를 포함할 수 있다.
- [0016] 개시된 발명의 일 측면에 따른 컴퓨터 프로그램은, 상기 비지도 학습 기반의 객체 위치 검출 방법을 실행시키도록 컴퓨터로 판독 가능한 기록매체에 저장될 수 있다.
- [0017] 개시된 발명의 일 측면에 따른 비지도 학습 기반의 객체 위치 검출 장치는, 학습용 이미지를 변환하여 변환 이미지를 생성하도록 구성되는 변환 이미지 생성부; 상기 학습용 이미지로부터 제1 객체 영역을 추출하고, 상기 변환 이미지로부터 제2 객체 영역을 추출하도록 구성되는 객체 추출부; 및 상기 제1 객체 영역과 상기 제2 객체 영역을 비교하여 인공지능 모델을 학습하도록 구성되는 인공지능 학습부;를 포함할 수 있다.
- [0018] 또한, 상기 인공지능 학습부는, 상기 학습용 이미지에서 상기 제1 객체 영역을 제외한 제1 배경 영역과 상기 변환 이미지에서 상기 제2 객체 영역을 제외한 제2 배경 영역을 비교하여 상기 인공지능 모델을 학습하도록 구성될 수 있다.
- [0019] 또한, 상기 인공지능 학습부는: 상기 제1 객체 영역과 상기 제2 객체 영역의 유사성을 분석하고, 상기 제1 배경 영역과 상기 제1 객체 영역의 차이를 분석하여 손실 함수를 연산하고; 그리고 상기 손실 함수가 감소하도록 상기 인공지능 모델을 학습하도록 구성될 수 있다.
- [0020] 또한, 상기 객체 추출부는: 상기 학습용 이미지를 구성하는 복수개의 채널 중 인식된 객체의 정보가 상대적으로 많이 포함된 채널인 제1 채널에 기초하여 상기 제1 객체 영역을 추출하고; 그리고 상기 변환 이미지를 구성하는 복수개의 채널 중 인식된 객체의 정보가 상대적으로 많이 포함된 채널인 제2 채널에 기초하여 상기 제2 객체 영역을 추출하도록 구성될 수 있다.
- [0021] 또한, 상기 객체 추출부는: 상기 학습용 이미지의 채널을 구성하는 픽셀이 상기 인식된 객체의 정보를 포함하는 픽셀인지 판단하고; 상기 학습용 이미지의 채널에 대하여 상기 객체의 정보를 포함하는 픽셀의 개수를 결정하고; 상기 픽셀의 개수에 기초하여 상기 복수개의 학습용 이미지의 채널 중 인식된 객체의 정보가 상대적으로 많이 포함된 채널을 상기 제1 채널로 결정하고; 상기 변환 이미지의 채널을 구성하는 픽셀이 상기 인식된 객체의 정보를 포함하는 픽셀인지 판단하고; 상기 변환 이미지의 채널에 대하여 상기 객체의 정보를 포함하는 픽셀의 개수를 결정하고; 그리고 상기 픽셀의 개수에 기초하여 상기 복수개의 변환 이미지의 채널 중 인식된 객체의 정보가 상대적으로 많이 포함된 채널을 상기 제2 채널로 결정하도록 구성될 수 있다.
- [0022] 또한, 프로세서;를 더 포함하고, 상기 객체 추출부는, 상기 인공지능 모델을 이용하여 입력 이미지로부터 상기 제1 객체 영역을 추출하도록 구성되고, 상기 프로세서는, 상기 추출된 제1 객체 영역을 기초로 상기 입력 이미지에서의 객체 위치를 결정하도록 구성될 수 있다.
- [0023] 또한, 상기 변환 이미지 생성부는, 상기 학습용 이미지를 미리 정해진 각도로 회전시켜 상기 변환 이미지를 생성하도록 구성될 수 있다.
- [0024] 또한, 상기 변환 이미지 생성부는, 상기 학습용 이미지를 확대하거나 축소하여 상기 변환 이미지를 생성하도록 구성될 수 있다.
- [0025] 또한, 상기 변환 이미지 생성부는, 상기 학습용 이미지를 대칭 변환하여 상기 변환 이미지를 생성하도록 구성될 수 있다.

발명의 효과

- [0026] 개시된 발명의 일 측면에 따르면, 학습용 이미지 및 변환 이미지에 기초하여 이미지상 객체의 전체 영역을 커버할 수 있고, 학습용 이미지의 차별적인 부분에만 집중하지 않으면서도 배경 영역을 과도하게 커버하지 않도록 인공지능 모델을 학습하여 객체의 위치를 정확하게 검출할 수 있다.
- [0027] 또한, 본 발명의 실시예에 의하면, 인공지능 모델에 대한 학습 횟수가 적더라도 종래의 객체 위치 검출 기술보다 이미지 상에서의 객체 위치를 정확하게 검출하여 객체 위치 검출 성능을 높일 수 있다.

도면의 간단한 설명

- [0028] 도 1은 일 실시예에 따른 객체 위치 검출 장치의 구성도이다.
- 도 2는 일 실시예에 따른 비지도 학습 방법을 설명하기 위한 도면이다.
- 도 3은 일 실시예에 따른 손실 함수를 연산하는 방법을 도시한 도면이다.
- 도 4는 일 실시예에 따라 복수개의 채널 중 일부 채널에 기초하여 객체 영역을 추출하는 것을 설명하기 위한 도면이다.
- 도 5는 일 실시예에 따라 복수개의 채널 중 일부 채널에 기초하여 객체 영역을 추출하는 것을 설명하기 위한 또 다른 도면이다.
- 도 6은 일 실시예에 따라 입력 이미지에서의 객체 위치를 결정하는 과정을 도시한 도면이다.
- 도 7은 일 실시예에 따른 객체 위치 검출 결과와 종래의 객체 위치 검출 결과를 나타낸 도면이다.
- 도 8은 일 실시예에 따른 객체 위치 검출 방법의 순서도이다.
- 도 9는 일 실시예에 따른 객체 위치 검출 방법이 종래의 객체 위치 검출 방법에 비해 개선된 정도를 나타낸 표이다.
- 도 10은 일 실시예에 따라 복수개의 채널 중 일부 채널을 결정하는 기준을 설명하기 위한 도면이다.
- 도 11은 일 실시예에 따라 복수개의 채널 중 일부 채널을 결정하는 기준을 설명하기 위한 또다른 도면이다.
- 도 12는 다양한 입력 이미지 변환에 대하여 일 실시예에 따른 위치 검출 방법이 종래의 위치 검출 방법에 비하여 개선된 정도를 나타낸 표이다.

발명을 실시하기 위한 구체적인 내용

- [0029] 명세서 전체에 걸쳐 동일 참조 부호는 동일 구성요소를 지칭한다. 본 명세서가 실시예들의 모든 요소들을 설명하는 것은 아니며, 개시된 발명이 속하는 기술분야에서 일반적인 내용 또는 실시예들 간에 중복되는 내용은 생략한다. 명세서에서 사용되는 '~부'라는 용어는 소프트웨어 또는 하드웨어로 구현될 수 있으며, 실시예들에 따라 복수의 '~부'가 하나의 구성요소로 구현되거나, 하나의 '~부'가 복수의 구성요소들을 포함하는 것도 가능하다.
- [0030] 또한 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.
- [0031] 제1, 제2 등의 용어는 하나의 구성요소를 다른 구성요소로부터 구별하기 위해 사용되는 것으로, 구성요소가 전술된 용어들에 의해 제한되는 것은 아니다.
- [0032] 단수의 표현은 문맥상 명백하게 예외가 있지 않는 한, 복수의 표현을 포함한다.
- [0033] 각 단계들에 있어 식별부호는 설명의 편의를 위하여 사용되는 것으로 식별부호는 각 단계들의 순서를 설명하는 것이 아니며, 각 단계들은 문맥상 명백하게 특정 순서를 기재하지 않는 이상 명기된 순서와 다르게 실시될 수 있다.
- [0034] 이하 첨부된 도면들을 참고하여 개시된 발명의 작용 원리 및 실시예들에 대해 설명한다.
- [0035] 도 1은 일 실시예에 따른 객체 위치 검출 장치의 구성도이며, 도 2는 일 실시예에 따른 비지도 학습 방법을 설명하기 위한 도면이다.
- [0036] 도 1을 참조하면, 본 발명의 실시예에 따른 객체 검출 장치(100)는 변환 이미지 생성부(110), 객체 추출부(120), 인공지능 학습부(130), 프로세서(140), 메모리(150)를 포함할 수 있다.
- [0037] 객체 위치 검출 장치(100)는 객체를 포함하는 학습용 이미지(200) 또는 입력 이미지(300)에서 객체를 인식하고 이미지상에서 인식된 객체의 위치를 결정할 수 있다.
- [0038] 학습용 이미지(200)는 이미지 데이터의 형태로 객체 검출 장치(100)에 입력될 수 있다. 학습용 이미지(200)는 일 실시예에 따른 객체 검출 방법의 비지도 학습에 이용되는 이미지일 수 있다.
- [0039] 비지도 학습(Unsupervised Learning)이란 기계 학습의 일종으로서, 데이터가 어떻게 구성되어 있는지를 알아내는 학습 중 한 종류일 수 있다. 비지도 학습은 지도 학습(Supervised Learning) 또는 강화 학습(Reinforcement

Learning)과는 달리 입력값에 대한 목표값이 주어지지 않을 수 있다.

- [0040] 변환 이미지 생성부(110)는 학습용 이미지(200)를 변환하여 변환 이미지(400)를 생성하도록 구성될 수 있다.
- [0041] 이때, 변환 이미지 생성부(110)는 학습용 이미지(200)를 미리 정해진 각도로 회전시켜 변환 이미지(400)를 생성하도록 구성될 수 있다.
- [0042] 예를 들어, 변환 이미지 생성부(110)는 학습용 이미지(200)를 시계 방향으로 90도 회전시켜 변환 이미지(400)를 생성하거나, 학습용 이미지(200)를 반시계 방향으로 90도 회전시켜 변환 이미지(400)를 생성할 수 있으나, 학습용 이미지(200)를 회전시키는 각도가 90도에 한정되는 것은 아니다.
- [0043] 변환 이미지 생성부(110)는 학습용 이미지(200)를 확대하거나 축소하여 변환 이미지(400)를 생성하도록 구성될 수 있다.
- [0044] 예를 들어, 변환 이미지 생성부(110)는 학습용 이미지(200)의 크기를 2배 확대하여 변환 이미지(400)를 생성하거나, 학습용 이미지(200)의 크기를 절반으로 축소하여 변환 이미지(400)를 생성할 수 있으나, 학습용 이미지(200)를 확대 또는 축소시키는 스케일(scale)이 이에 한정되는 것은 아니다.
- [0045] 변환 이미지 생성부(110)는 학습용 이미지(200)를 대칭 변환하여 변환 이미지(400)를 생성하도록 구성될 수 있다.
- [0046] 전술한 바와 같이 변환 이미지 생성부(110)는 학습용 이미지(200)를 회전, 확대, 축소 및 대칭 변환의 방법으로 변환하여 변환 이미지(400)를 생성할 수 있으나, 학습용 이미지(200)를 변환시켜 변환 이미지(400)를 생성하는 방법이 이에 한정되는 것은 아니며, 다양한 변환 방법이 이용될 수 있다.
- [0047] 객체 추출부(120)는 학습용 이미지(200)로부터 제1 객체 영역(201)을 추출하고, 변환 이미지(400)로부터 제2 객체 영역(401)을 추출하도록 구성될 수 있다.
- [0048] 객체 추출부(120)는 딥러닝 기반의 객체 검출 기술에 기초하여 이미지로부터 객체 영역을 인식하는 인코더일 수 있다.
- [0049] 딥러닝 기반의 객체 검출 기술은 이미지로부터 추출되는 객체 영역의 특징(feature)을 데이터를 기반으로 학습한다. 이때, 이미지로부터 특징(feature)을 추출하는 방식을 학습하기 위해 여러 단계의 컨볼루션 계층(convolution layer)을 쌓은 CNN(Convolutional Neural Networks) 구조가 활용될 수 있으나 이에 한정되는 것은 아니다.
- [0050] 도 2를 참조하면, 인공지능 학습부(130)는 제1 객체 영역(201)과 제2 객체 영역(401)을 비교하여 인공지능 모델을 학습하도록 구성될 수 있다.
- [0051] 즉, 인공지능 학습부(130)는 변환되기 전의 학습용 이미지(200)에서 추출된 객체 영역과 변환된 후의 변환 이미지(400)에서 추출된 객체 영역을 비교하여 인공지능 모델을 학습할 수 있다.
- [0052] 객체 추출부(120)는 학습용 이미지(200)와 변환 이미지(400)에 대해서 각각 개별적으로 객체를 추출할 수 있다. 따라서, 제1 객체 영역(201)과 제2 객체 영역(401)이 반드시 일치하는 것은 아닐 수 있다. 하지만, 제2 객체 영역(401)은 학습용 이미지(200)가 일정한 방식의 변환 과정을 거쳐 생성된 변환 이미지(400)로부터 추출된 것이므로 제1 객체 영역(201)과 어느정도 일치하는 경향을 가질 수 있다. 구체적으로, 제2 객체 영역(401)을 역변환하여 생성된 영역은 제1 객체 영역(201)과 어느정도 일치하는 형태의 객체 영역일 수 있다.
- [0053] 예를 들어, 변환 이미지(400)가 학습용 이미지(200)를 시계 방향으로 90도 회전시켜 생성된 이미지일 경우, 제2 객체 영역(401)을 반시계 방향으로 90도 회전시켜 생성된 영역은 제1 객체 영역(201)과 어느정도 일치하는 형태의 객체 영역일 수 있다.
- [0054] 만약 객체 위치 검출 장치(100)의 성능이 뛰어나다면 변환되기 전의 이미지인 학습용 이미지(200)에서 객체를 인식하든, 변환된 후의 이미지인 변환 이미지(400)에서 객체를 인식하든 거의 일치하는 영역을 제1 객체 영역(201) 및 제2 객체 영역(401)으로 추출할 것이다. 따라서 일 실시예에 따른 인공지능 학습부(130)는 학습을 반복할수록 제1 객체 영역(201)과 제2 객체 영역(401)이 서로 유사한 영역이 되도록 인공지능 모델을 학습할 수 있다.
- [0055] 인공지능 학습부(130)는 제1 배경 영역(202)과 제2 배경 영역(402)을 비교하여 인공지능 모델을 학습하도록 구성될 수 있다.

- [0056] 제1 배경 영역(202)은 학습용 이미지(200)에서 제1 객체 영역(201)이 제외된 영역일 수 있다. 또한, 제2 배경 영역(402)은 변환 이미지(400)에서 제2 객체 영역(401)이 제외된 영역일 수 있다.
- [0057] 즉, 인공지능 학습부(130)는 변환되기 전의 학습용 이미지(200)에서 추출된 배경 영역과 변환된 후의 변환 이미지(400)에서 추출된 배경 영역을 비교하여 인공지능 모델을 학습할 수 있다.
- [0058] 제2 배경 영역(402)은 학습용 이미지(200)가 일정한 방식의 변환 과정을 거쳐 생성된 변환 이미지(400)로부터 추출된 것이므로 제1 배경 영역(202)과 어느정도 일치하는 경향을 가질 수 있다. 구체적으로, 제2 배경 영역(402)을 역변환하여 생성된 영역은 제1 배경 영역(202)과 어느정도 일치하는 형태의 객체 영역일 수 있다.
- [0059] 예를 들어, 변환 이미지(400)가 학습용 이미지(200)를 시계 방향으로 90도 회전시켜 생성된 이미지일 경우, 제2 배경 영역(402)을 반시계 방향으로 90도 회전시켜 생성된 영역은 제1 배경 영역(202)과 어느정도 일치하는 형태의 객체 영역일 수 있다.
- [0060] 만약 객체 위치 검출 장치(100)의 성능이 뛰어나다면 거의 일치하는 영역을 제1 배경 영역(202) 및 제2 배경 영역(402)으로 추출할 것이다. 따라서 일 실시예에 따른 인공지능 학습부(130)는 학습을 반복할수록 제1 배경 영역(202)과 제2 배경 영역(402)이 서로 유사한 영역이 되도록 인공지능 모델을 학습할 수 있다.
- [0061] 인공지능 학습부(130)는 학습용 데이터를 통해, 학습용 이미지(200)의 특징을 입력 변수로 설정하고 이미지에서의 객체의 위치를 출력 변수로 설정하여 인공지능 모델을 학습하도록 구성될 수 있다.
- [0062] 학습용 이미지(200)의 특징은 해당 학습용 이미지(200)에 대한 다양한 특징을 나타내는 정보일 수 있다. 예를 들어, 학습용 이미지(200)의 특징은 해당 학습용 이미지(200)의 각 픽셀 단위에서의 색상, 명도 등에 대한 정보일 수 있다.
- [0063] 인공지능 모델을 학습하는 것은 기계 학습(Machine Learning) 방식을 통한 학습일 수 있다. 기계 학습이란 다수의 파라미터로 구성된 모델을 이용하며, 주어진 데이터로 파라미터를 최적화하는 것을 의미할 수 있다. 인공지능 학습부(130)는 입력에 따라 인공지능 모델을 통해 최종적으로 나온 결과물(output)인 출력들을 이용하여 인공지능 모델을 학습할 수 있다. 인공지능 모델은 메모리(150)에 저장될 수 있다.
- [0064] 변환 이미지 생성부(110), 객체 추출부(120), 인공지능 학습부(130)는 객체 위치 검출 장치(100)에 포함된 복수개의 프로세서(140) 중 어느 하나의 프로세서(140)를 포함할 수 있다. 또한, 지금까지 설명된 본 발명의 실시예 및 앞으로 설명할 실시예에 따른 객체 인식 방법은, 프로세서(140)에 의해 구동될 수 있는 프로그램의 형태로 구현될 수 있다.
- [0065] 여기서 프로그램은, 프로그램 명령, 데이터 파일 및 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 프로그램은 기계어 코드나 고급 언어 코드를 이용하여 설계 및 제작된 것일 수 있다. 프로그램은 상술한 부호 수정을 위한 방법을 구현하기 위하여 특별히 설계된 것일 수도 있고, 컴퓨터 소프트웨어 분야에서 통상의 기술자에게 기 공지되어 사용 가능한 각종 함수나 정의를 이용하여 구현된 것일 수도 있다. 전술한 정보 표시 방법을 구현하기 위한 프로그램은, 프로세서(140)에 의해 판독 가능한 기록매체에 기록될 수 있다. 이때, 기록매체는 메모리(150)일 수 있다.
- [0066] 메모리(150)는 전술한 동작 및 후술하는 동작을 수행하는 프로그램을 저장할 수 있으며, 메모리(150)는 저장된 프로그램을 실행시킬 수 있다. 프로세서(140)와 메모리(150)가 복수인 경우에, 이들이 하나의 칩에 집적되는 것도 가능하고, 물리적으로 분리된 위치에 마련되는 것도 가능하다. 메모리(150)는 데이터를 일시적으로 기억하기 위한 S램(Static Random Access Memory, S-RAM), D램(Dynamic Random Access Memory) 등의 휘발성 메모리를 포함할 수 있다. 또한, 메모리(150)는 제어 프로그램 및 제어 데이터를 장기간 저장하기 위한 롬(Read Only Memory), 이피롬(Erasable Programmable Read Only Memory: EPROM), 이이피롬(Electrically Erasable Programmable Read Only Memory: EEPROM) 등의 비휘발성 메모리를 포함할 수 있다.
- [0067] 프로세서(140)는 각종 논리 회로와 연산 회로를 포함할 수 있으며, 메모리(150)로부터 제공된 프로그램에 따라 데이터를 처리하고, 처리 결과에 따라 제어 신호를 생성할 수 있다.
- [0068] 도 3은 일 실시예에 따른 손실 함수를 연산하는 방법을 도시한 도면이다.
- [0069] 도 3을 참조하면, 인공지능 학습부(130)는 제1 객체 영역(201)과 제2 객체 영역(401)의 유사성을 분석하고, 제1 배경 영역(202)과 제1 객체 영역(201)의 차이를 분석하여 손실 함수를 연산할 수 있다.

[0070] [방정식 1]

$$\mathbf{M}_{fg} = \mathbb{1}[(1 - \mathbf{A}) > \theta_{bg}]$$

reverse attention map

[0071]

[0072] 구체적으로, [방정식 1]을 참조하면, \mathbf{M}_{fg} 는 이미지의 배경 영역을 산출하기 위하여 객체 영역을 마킹하는데 이용되는 인자이고, \mathbf{A} 는 attention map으로서 이미지에서 각 픽셀별로 객체 영역으로 인식된 정도를 나타내는 임베딩 값일 수 있다. \mathbf{M}_{fg} 는 각 픽셀에 대해서 1에서 \mathbf{A} 를 차감한 값($1-\mathbf{A}$)이 미리 정해진 임계값(θ_{bg})을 초과하면 1로 설정될 수 있으며, 각 픽셀에 대해서 1에서 \mathbf{A} 를 차감한 값($1-\mathbf{A}$)이 미리 정해진 임계값(θ_{bg}) 이하이면 0으로 설정될 수 있다.

[0073] [방정식 2]

$$\mathbf{A}_{bg} = (1 - \mathbf{A}) \odot \mathbf{M}_{fg}$$

[0074]

[0075] [방정식 2]를 참조하면, \mathbf{A}_{bg} 는 background attention map으로서 객체 영역이 마킹이 된 상태의 이미지에 대응되는 임베딩 인자일 수 있다. \mathbf{A}_{bg} 는 각 픽셀에 대해서 1에서 \mathbf{A} 를 차감한 값($1-\mathbf{A}$)에 각 픽셀 별 \mathbf{M}_{fg} 값을 곱하여 산출할 수 있다. 결과적으로 \mathbf{A}_{bg} 는 이미지에서 객체 영역이 제외되어 배경 영역을 나타내는 맵일 수 있다.

[0076] [방정식 3]

$$\mathcal{L}_{pec} = \mathbb{E}_{\mathbf{x}} [[d(\mathbf{A}_{tf2orig}, \mathbf{A}_{orig}) - d(\mathbf{A}_{tf2orig}, \mathbf{A}_{bg}) + m]_+ +$$

$$[(d(\mathbf{A}_{orig2tf}, \mathbf{A}_{tf}) - d(\mathbf{A}_{orig2tf}, \mathbf{A}_{bg}) + m)_+].$$

[0077]

[0078] [방정식 3]을 참조하면, \mathcal{L}_{pec} 는 일 실시예에 따른 손실함수일 수 있다. 구체적으로, \mathcal{L}_{pec} 는 트리플렛 손실(Triplet Loss)을 나타내는 함수일 수 있다.

[0079] 트리플렛 손실은 앵커(anchor)를 positive input 및 negative input과 비교하는 인공 신경 네트워크에 대한 손실 함수(loss function)일 수 있다. 이때 앵커 인풋(anchor input)과 positive input 사이의 거리(distance)는 최소화되어야 하며, 앵커 인풋(anchor input)과 negative input과의 거리(distance)는 최대가 되도록 하는 것이 바람직할 수 있다. 트리플렛 손실은 워드 임베딩(word embeddings), 벡터 학습, 행렬 학습과 같은 임베딩 학습의 목적으로 유사성을 학습하는데 이용될 수 있다. 트리플렛 손실 함수(Triplet Loss Function)는 [방정식 3]과 같은 형태의 유클리디안 거리 함수(Euclidean Distance Function)의 형태를 할 수 있다.

[0080] 학습용 이미지(200)에 대한 임베딩 값인 \mathbf{A}_{orig} 및 변환 이미지(400)에 대한 임베딩 값인 \mathbf{A}_{tf} 는 positive input에 해당하는 임베딩 값일 수 있다.

[0081] $\mathbf{A}_{orig2tf}$ 는 \mathbf{A}_{orig} 에 대하여 학습용 이미지(200)에 가해진 변환과 동일한 변환을 했을 때 생성되는 임베딩 값일 수 있다. $\mathbf{A}_{tf2orig}$ 는 \mathbf{A}_{tf} 에 대하여 학습용 이미지(200)에 가해진 변환의 역변환을 했을 때 생성되는 임베딩 값일 수 있다. 이러한 $\mathbf{A}_{tf2orig}$ 및 $\mathbf{A}_{orig2tf}$ 는 앵커 인풋(anchor input)에 해당하는 임베딩 값일 수 있다. \mathbf{A}_{bg} 는 negative input에 해당하는 임베딩 값일 수 있다.

[0082] 인공지능 학습부(130)는 손실 함수가 감소하도록 인공지능 모델을 학습할 수 있다. 즉, 인공지능 학습부(130)는 일 실시예에 따른 트리플렛 손실 함수(\mathcal{L}_{pec})가 최소화되도록 학습을 수행할 수 있다.

[0083] 즉, 인공지능 학습부(130)는 $\mathbf{A}_{tf2orig}$ 와 \mathbf{A}_{orig} 사이의 거리(distance)가 최소가 되고, $\mathbf{A}_{tf2orig}$ 와 \mathbf{A}_{bg} 와의 거리(distance)는 최대가 되도록 학습을 반복하여 수행할 수 있다. 또한, 인공지능 학습부(130)는 $\mathbf{A}_{orig2tf}$ 와 \mathbf{A}_{tf} 사이의 거리(distance)가 최소가 되고, $\mathbf{A}_{orig2tf}$ 와 \mathbf{A}_{bg} 와의 거리(distance)는 최대가 되도록 학습을 반복하여 수행할 수 있다.

[0084] 결과적으로, 인공지능 학습부(130)는 앵커 인풋(anchor input)과 positive input 사이의 거리(distance)는 최소화하고, 앵커 인풋(anchor input)과 negative input과의 거리(distance)를 최대화하여 손실 함수가 감소하

도록 인공지능 모델을 학습할 수 있다.

- [0085] 도 4는 일 실시예에 따라 복수개의 채널 중 일부 채널에 기초하여 객체 영역을 추출하는 것을 설명하기 위한 도면이며, 도 5는 일 실시예에 따라 복수개의 채널 중 일부 채널에 기초하여 객체 영역을 추출하는 것을 설명하기 위한 또다른 도면이다.
- [0086] 도 4를 참조하면, 기존의 객체 위치 검출 방법은 학습용 이미지(200)를 구성하는 복수개의 채널 전부에 대해서 풀링(Pooling)을 수행하여 객체 영역을 추출했다. 즉, 종래의 방법은 각 채널에 포함된 객체의 정보와 상관없이 단순히 각 채널들에 대한 객체 추출 결과를 평균 내어 최종적으로 객체 영역을 추출하였다.
- [0087] 이러한 기존의 방식은 객체의 정보가 다른 채널에 비하여 다소 적게 포함된 채널에 대해서도 다른 채널과 동일한 비중을 두어 객체 영역을 추출하게 되므로 객체의 주변 배경에 대해서도 잘못된 인식을 수행할 수 있다는 문제가 있었다.
- [0088] 도 4 및 도 5를 참조하면, 일 실시예에 따른 객체 추출부(120)는 학습용 이미지(200)를 구성하는 복수개의 채널들 중 일부 채널인 제1 채널에 기초하여 제1 객체 영역(201)을 추출할 수 있다. 제1 채널은 복수개의 채널들 중 인식된 객체의 정보가 상대적으로 많이 포함된 채널일 수 있다.
- [0089] 또한, 객체 추출부(120)는 변환 이미지(400)를 구성하는 복수개의 채널들 중 일부 채널인 제2 채널에 기초하여 제2 객체 영역(401)을 추출할 수 있다. 제2 채널은 복수개의 채널들 중 인식된 객체의 정보가 상대적으로 많이 포함된 채널일 수 있다.
- [0090] 즉, 객체 추출부(120)는 복수개의 채널들 중 인식된 객체의 정보가 상대적으로 많이 포함된 채널에 대해 상대적으로 많은 가중치를 부여하고, 인식된 객체의 정보가 상대적으로 적게 포함된 채널에 대해 상대적으로 적은 가중치를 부여하여 평균을 내는 방식으로 객체의 영역을 추출할 수 있다.
- [0091] 결과적으로, 일 실시예에 따른 객체 위치 검출 방법은 객체의 정보가 상대적으로 많이 포함된 채널에 보다 높은 가중치를 두어 객체 영역을 추출할 수 있어서, 종래의 기술에 비하여 주변 배경을 객체로 인식하는 오류가 적게 발생할 수 있다는 효과가 있다.
- [0092] 학습용 이미지(200) 및 변환 이미지(400)는 복수개의 픽셀로 구성될 수 있다. 이때 일부 픽셀은 이미지에 포함된 객체의 정보를 포함하는 픽셀이고, 일부 픽셀은 객체의 정보를 전혀 포함하지 않은 주변 배경에 대한 픽셀일 수 있다.
- [0093] 객체 추출부(120)는 학습용 이미지(200)의 채널을 구성하는 픽셀이 인식된 객체의 정보를 포함하는 픽셀인지 판단하고, 해당 학습용 이미지(200)의 채널에 대하여 객체의 정보를 포함하는 픽셀의 개수를 결정할 수 있다.
- [0094] 객체 추출부(120)는 픽셀의 개수에 기초하여 복수개의 학습용 이미지(200)의 채널 중 인식된 객체의 정보가 상대적으로 많이 포함된 채널을 제1 채널로 결정할 수 있다.
- [0095] 이때, 제1 채널을 결정하는 방식은 복수개의 채널들 중 객체의 정보를 포함하는 픽셀의 개수가 많은 상위 몇 퍼센트의 채널들을 제1 채널로 결정하는 방식일 수 있다.
- [0096] 예를 들어, 채널의 개수가 4이고, 미리 설정된 비율 값이 50% 라면 객체 추출부(120)는 4개의 채널 중 객체의 정보를 포함하는 픽셀의 개수가 가장 많은 채널과 두번째로 많은 채널, 즉 2개의 채널을 제1 채널로 결정할 수 있다.
- [0097] 한편, 전술한 방식은 학습용 이미지(200)에서 객체를 추출하는 과정만이 아니라 변환 이미지(400)에서 객체를 추출하는 과정에서도 사용될 수 있다.
- [0098] 객체 추출부(120)는 변환 이미지(400)의 채널을 구성하는 픽셀이 인식된 객체의 정보를 포함하는 픽셀인지 판단하고, 변환 이미지(400)의 채널에 대하여 객체의 정보를 포함하는 픽셀의 개수를 결정할 수 있다.
- [0099] 또한, 객체 추출부(120)는 픽셀의 개수에 기초하여 복수개의 변환 이미지(400)의 채널 중 인식된 객체의 정보가 상대적으로 많이 포함된 채널을 제2 채널로 결정할 수 있다.
- [0100] 한편, 제1 채널 및 제2 채널을 결정하는 방식이 반드시 복수개의 채널들 중 객체의 정보를 포함하는 픽셀의 개수가 많은 상위 몇 퍼센트의 채널들을 제1 채널 및 제2 채널로 결정하는 방식으로 한정되는 것은 아니며, 복수개의 채널들 중 인식된 객체의 정보가 상대적으로 많이 포함된 채널에 대해 상대적으로 많은 가중치를 부여하여 평균을 내는 방식으로 객체의 영역을 추출할 수 있다면 어떠한 방식이 사용되더라도 상관없다.

- [0101] 도 6은 일 실시예에 따라 입력 이미지에서의 객체 위치를 결정하는 과정을 도시한 도면이다.
- [0102] 도 6을 참조하면, 객체 추출부(120)는 인공지능 모델을 이용하여 입력 이미지(300)로부터 제1 객체 영역(201)을 추출하도록 구성될 수 있다. 즉, 객체 추출부(120)는 학습용 이미지(200)에 기초하여 학습된 인공지능 모델을 통하여 실제로 사용자가 객체의 위치를 알고자 하는 입력 이미지(300)로부터 객체 영역을 추출할 수 있다.
- [0103] 프로세서(140)는 추출된 제1 객체 영역(201)을 기초로 입력 이미지(300)에서의 객체 위치를 결정할 수 있다. 이때, 프로세서(140)는 제1 객체 영역(201)을 기초로 입력 이미지(300) 상에 바운딩 박스(Bounding box)를 생성하고, 바운딩 박스의 좌표에 기초하여 입력 이미지(300)에서의 객체 위치를 결정할 수 있다.
- [0104] 도 7은 일 실시예에 따른 객체 위치 검출 결과와 종래의 객체 위치 검출 결과를 나타낸 도면이다.
- [0105] 도 7을 참조하면, 인공지능 모델을 학습하지 않은 상태(0 epoch), 인공지능 모델을 10번 학습한 상태(10 epoch) 및 인공지능 모델을 40번 학습한 상태(40 epoch) 별로 객체 위치 검출을 수행했을 때 종래의 객체 위치 검출 방법(Baseline) 및 본 발명의 객체 위치 검출 방법(Ours)에 따른 객체 위치 검출 결과를 확인할 수 있다.
- [0106] 적색 박스(red box)는 미리 정해진 객체 위치의 정답(ground truth)을 나타낸 바운딩 박스일 수 있다.
- [0107] 녹색 박스(green box)는 종래의 객체 위치 검출 방법 및 본 발명의 객체 위치 검출 방법에 따라 검출된 객체의 위치를 나타낸 바운딩 박스일 수 있다.
- [0108] 도 7을 참조하면, 인공지능 모델을 학습하지 않은 상태에서도 본 발명의 객체 위치 검출 방법(Ours)이 종래의 방법(Baseline)보다 정답에 가까운 객체 위치를 검출해낸 것을 확인할 수 있다. 즉, 본 발명의 객체 위치 검출 방법(Ours)은 객체의 주변 배경을 객체로 잘못 인식하는 오류가 덜 일어나는 것을 확인할 수 있다.
- [0109] 이는 본 발명의 객체 위치 검출 방법의 경우 종래의 방법과 달리 단순히 각 채널들에 대한 객체 추출 결과를 평균 내어 최종적으로 객체 영역을 추출한 것이 아니라 객체의 정보가 상대적으로 많이 포함된 채널에 보다 높은 가중치를 두어 객체 영역을 추출하였기 때문에 나타난 결과일 수 있다.
- [0110] 인공지능 모델을 10번 학습한 상태 및 인공지능 모델을 40번 학습한 상태에서는 본 발명의 객체 위치 검출 방법(Ours)이 정답에 가까운 객체 위치를 보다 정확하게 검출한 것을 확인할 수 있다.
- [0111] 이상에서 설명된 구성요소들의 성능에 대응하여 적어도 하나의 구성요소가 추가되거나 삭제될 수 있다. 또한, 구성요소들의 상호 위치는 시스템의 성능 또는 구조에 대응하여 변경될 수 있다는 것은 당해 기술 분야에서 통상의 지식을 가진 자에게 용이하게 이해될 것이다.
- [0112] 도 8은 일 실시예에 따른 객체 위치 검출 방법의 순서도이다. 이는 본 발명의 목적을 달성하기 위한 바람직한 실시예일 뿐이며, 필요에 따라 일부 구성이 추가되거나 삭제될 수 있음은 물론이다.
- [0113] 도 8을 참조하면, 변환 이미지 생성부(110)는 학습용 이미지(200)를 변환하여 변환 이미지(400)를 생성할 수 있다(1001). 이때, 변환 이미지 생성부(110)는 학습용 이미지(200)를 미리 정해진 각도로 회전시키거나, 확대하거나, 축소하거나, 대칭 변환하는 방식 중 하나의 방식으로 변환 이미지(400)를 생성할 수 있다.
- [0114] 객체 추출부(120)는 학습용 이미지(200)를 구성하는 복수개의 채널 중 인식된 객체의 정보가 상대적으로 많이 포함된 채널인 제1 채널에 기초하여 제1 객체 영역(201)을 추출할 수 있다(1002).
- [0115] 이때, 객체 추출부(120)는 학습용 이미지(200)의 채널을 구성하는 픽셀이 인식된 객체의 정보를 포함하는 픽셀인지 판단하고, 학습용 이미지(200)의 채널에 대하여 객체의 정보를 포함하는 픽셀의 개수를 결정하고, 픽셀의 개수에 기초하여 복수개의 학습용 이미지(200)의 채널 중 인식된 객체의 정보가 상대적으로 많이 포함된 채널을 제1 채널로 결정할 수 있다.
- [0116] 객체 추출부(120)는 변환 이미지(400)를 구성하는 복수개의 채널 중 인식된 객체의 정보가 상대적으로 많이 포함된 채널인 제2 채널에 기초하여 제2 객체 영역(401)을 추출할 수 있다(1003).
- [0117] 이때, 객체 추출부(120)는 변환 이미지(400)의 채널을 구성하는 픽셀이 인식된 객체의 정보를 포함하는 픽셀인지 판단하고, 변환 이미지(400)의 채널에 대하여 객체의 정보를 포함하는 픽셀의 개수를 결정하고, 픽셀의 개수에 기초하여 복수개의 변환 이미지(400)의 채널 중 인식된 객체의 정보가 상대적으로 많이 포함된 채널을 제2 채널로 결정할 수 있다.
- [0118] 인공지능 학습부(130)는 제1 객체 영역(201)과 제2 객체 영역(401)의 유사성을 분석할 수 있다(1004). 또한, 인

공지능 학습부(130)는 제1 배경 영역(202)과 제1 객체 영역(201)의 차이를 분석할 수 있다(1005).

- [0119] 인공지능 학습부(130)는 전술한 분석 결과에 기초하여 손실 함수를 연산할 수 있다(1006). 이때, 인공지능 학습부가 연산하는 손실 함수는 트리플렛 손실 함수일 수 있다.
- [0120] 인공지능 학습부(130)는 손실 함수가 감소하도록 인공지능 모델을 학습할 수 있다(1007). 이때, 인공지능 학습부(130)는 트리플렛 손실 함수가 감소하도록 인공지능 모델을 반복하여 학습할 수 있다.
- [0121] 본 발명의 실시예에 따른 비지도 학습 기반의 객체 위치 검출 방법의 성능을 검증하기 위하여, 객체 위치 검출 실험을 진행하였다.
- [0122] 도 9는 일 실시예에 따른 객체 위치 검출 방법이 종래의 객체 위치 검출 방법에 비해 개선된 정도를 나타낸 표이다.
- [0123] 도 9를 참조하면, 입력 이미지(300)에 대하여 회전 변환을 수행하여 객체의 위치를 검출했을 때, 종래의 방법에 사용되는 손실 함수(\mathcal{L}_{cls})만을 연산하고 손실 함수를 감소시키는 방식(Baseline)보다 본 발명에 따른 트리플렛 손실 함수(\mathcal{L}_{pec})를 연산하고 트리플렛 손실 함수를 감소시키는 방식 및 학습용 이미지(200)의 채널 별로 가중치를 다르게 부여하는 방식(\mathbf{A}_i)을 모두 사용하는 방법(Ours(full))이 객체의 위치를 더 잘 검출하는 것을 확인할 수 있다.
- [0124] 구체적으로, 객체 위치의 정답(ground truth)을 나타내는 바운딩 박스와 검출된 객체의 위치를 나타내는 바운딩 박스가 겹치는 비율이 30% 이상인 경우는 종래의 방법(Baseline)은 총 실험 횟수 중 96.75%였으나, 본 발명의 경우 총 실험 횟수 중 97.30% 임을 확인할 수 있다.
- [0125] 또한, 객체 위치의 정답(ground truth)을 나타내는 바운딩 박스와 검출된 객체의 위치를 나타내는 바운딩 박스가 겹치는 비율이 70% 이상인 경우는 종래의 방법(Baseline)은 총 실험 횟수 중 28.66%에 불과하지만, 본 발명의 경우 총 실험 횟수 중 38.64%인 것을 확인할 수 있다.
- [0126] 도 9를 참조하면, 단순히 본 발명에 따른 트리플렛 손실 함수(\mathcal{L}_{pec})를 연산하는 방법만 사용하는 객체 위치 검출 방법보다 종래의 방법에 사용되는 손실 함수(\mathcal{L}_{cls})를 연산하는 방법 및 학습용 이미지(200)의 채널 별로 가중치를 다르게 부여하는 방법(\mathbf{A}_i)을 모두 중첩적으로 이용하는 객체 위치 검출 방법의 성능이 뛰어난 것을 확인할 수 있다.
- [0127] 도 10은 일 실시예에 따라 복수개의 채널 중 일부 채널을 결정하는 기준을 설명하기 위한 도면이며, 도 11은 일 실시예에 따라 복수개의 채널 중 일부 채널을 결정하는 기준을 설명하기 위한 또다른 도면이다.
- [0128] 도 10 및 도 11을 참조하면, 복수개의 채널들 중에서, 객체의 정보를 포함하는 픽셀의 개수가 많은 상위 몇 퍼센트의 채널들을 제1 채널 및 제2 채널로 결정할지 미리 설정할 수 있다.
- [0129] 이때, 실험 결과에 의하면 객체의 정보를 포함하는 픽셀의 개수가 많은 상위 70%의 채널들이 제1 채널 및 제2 채널로 결정되도록 설정하는 것이 가장 효과적으로 객체의 위치를 검출하는 것을 확인할 수 있다.
- [0130] 도 12는 다양한 입력 이미지 변환에 대하여 일 실시예에 따른 위치 검출 방법이 종래의 위치 검출 방법에 비하여 개선된 정도를 나타낸 표이다.
- [0131] 도 12를 참조하면, 변환 이미지(400)를 생성할 때 이용되는 변환 방법에 관계없이, 본 발명에 따른 객체 위치 검출 방법(Ours)이 종래의 방법(Base)보다 객체의 위치를 검출하는 성능이 뛰어난 것을 확인할 수 있다.
- [0132] 구체적으로, 회전 변환(Rotation), 확대 또는 축소하는 변환(Scale), 픽셀 단위로 이동시키는 변환(Translation), 가로 대칭 변환(Hflip), 세로 대칭 변환(Vflip) 등의 변환 방법 중 어느 방법을 이용하여 변환 이미지(400)를 생성하더라도 본 발명에 따른 객체 위치 검출 방법(Ours)이 종래의 방법(Base)보다 객체의 위치를 검출하는 성능이 뛰어난 것을 확인할 수 있다.
- [0133] 이상에서와 같이 첨부된 도면을 참조하여 개시된 실시예들을 설명하였다. 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고도, 개시된 실시예들과 다른 형태로 본 발명이 실시될 수 있음을 이해할 것이다. 개시된 실시예들은 예시적인 것이며, 한정적으로 해석되어서는

는 안 된다.

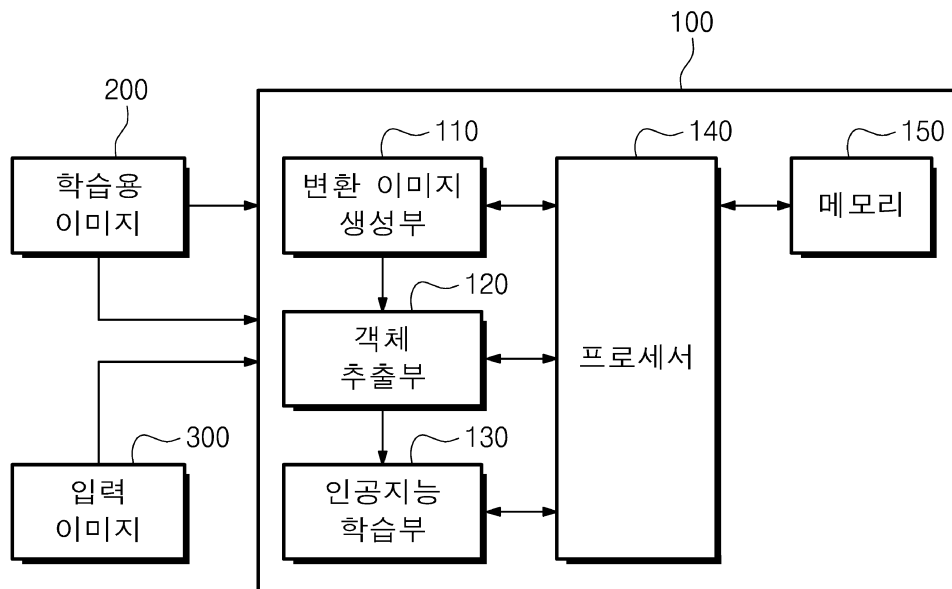
부호의 설명

[0134]

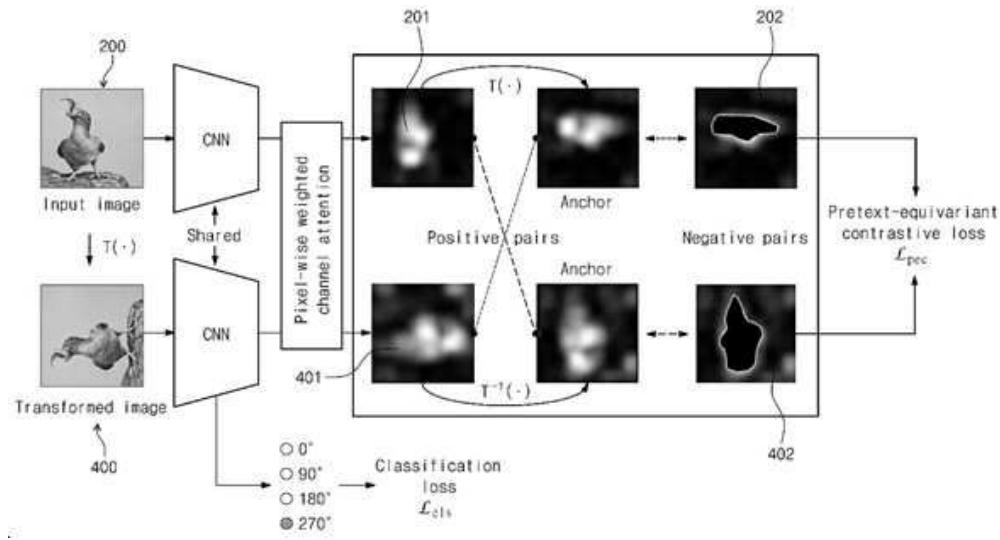
- 100: 객체 위치 검출 장치
- 110: 변환 이미지 생성부
- 120: 객체 추출부
- 130: 인공지능 학습부
- 140: 프로세서
- 150: 메모리
- 200: 학습용 이미지
- 201: 제1 객체 영역
- 202: 제1 배경 영역
- 300: 입력 이미지
- 400: 변환 이미지
- 401: 제2 객체 영역
- 402: 제2 배경 영역

도면

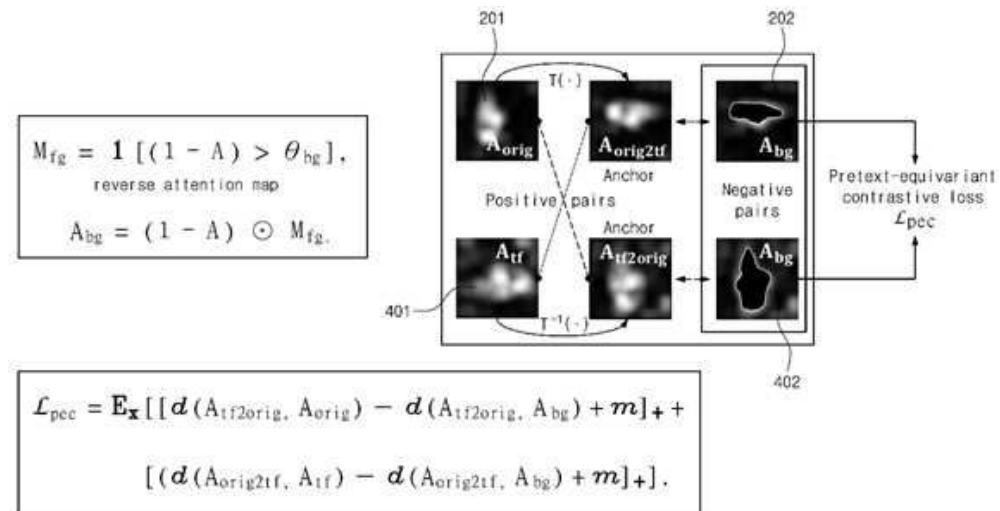
도면1



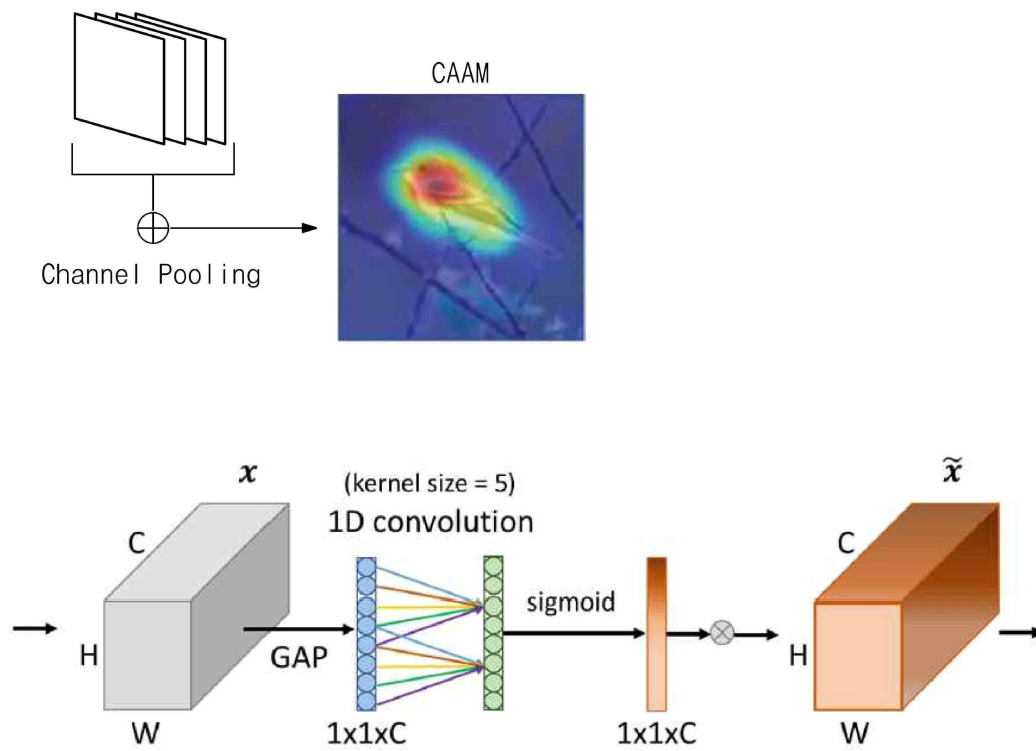
도면2



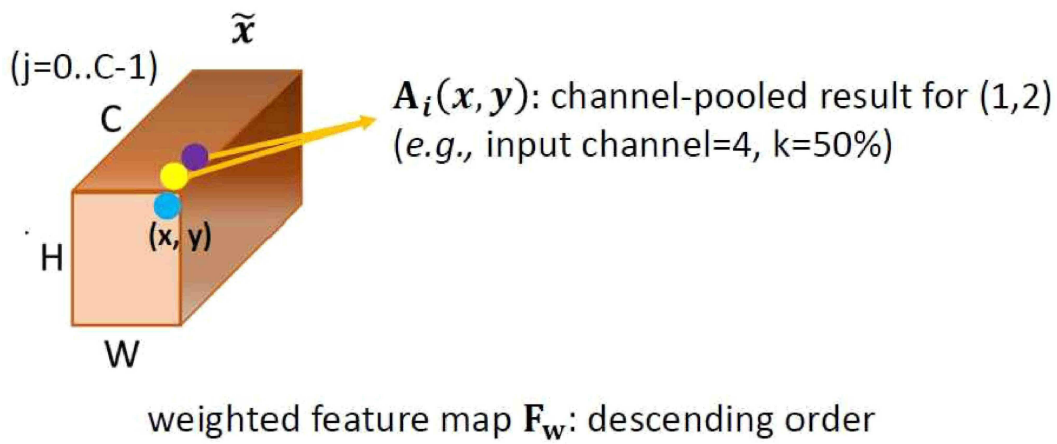
도면3



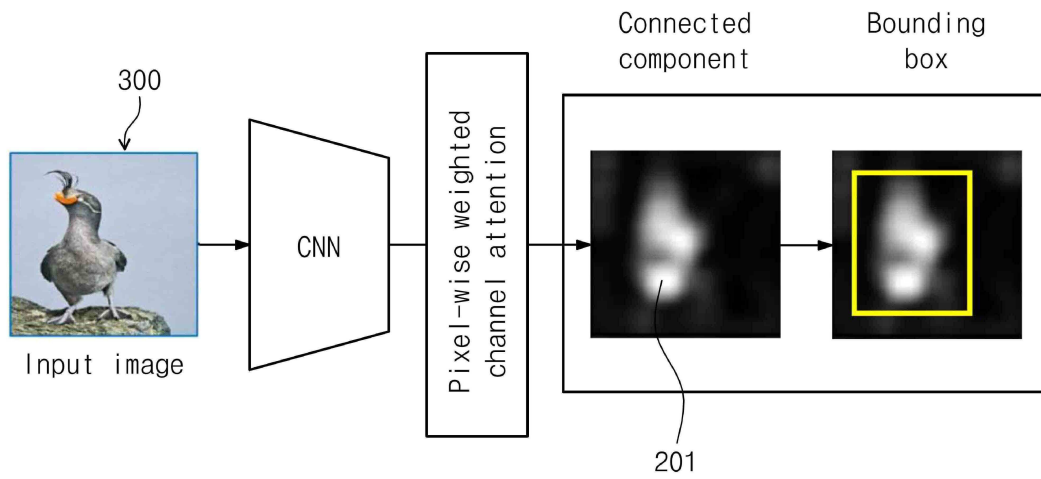
도면4



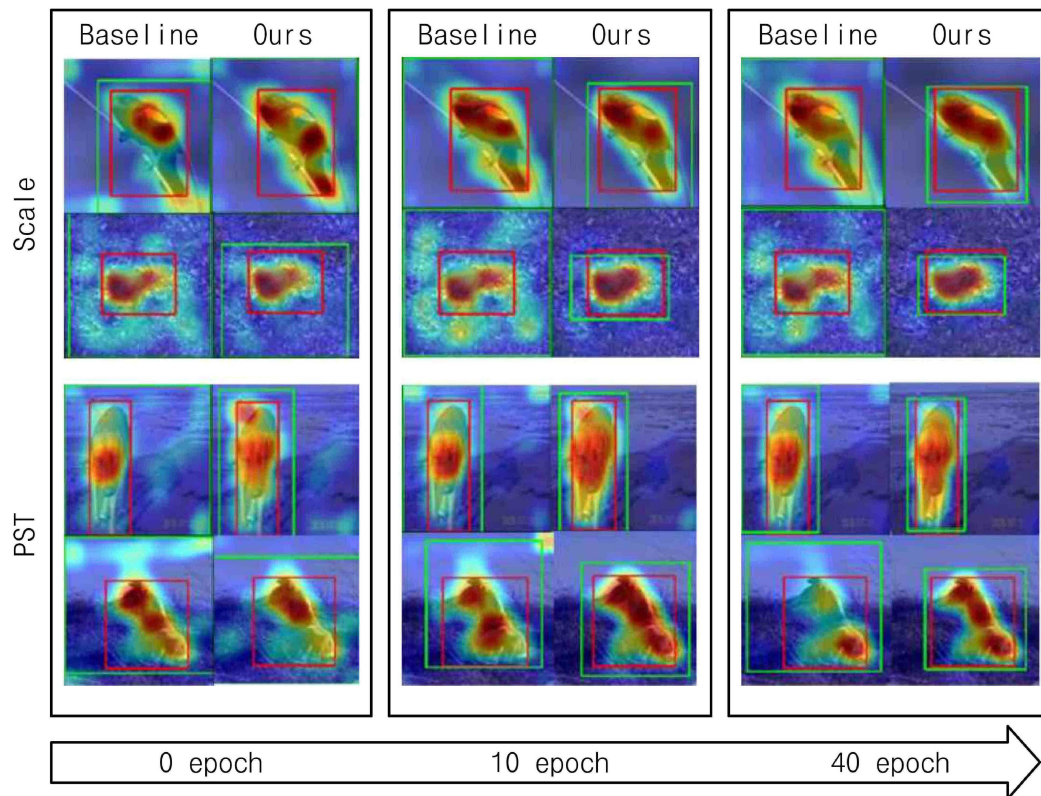
도면5



도면6

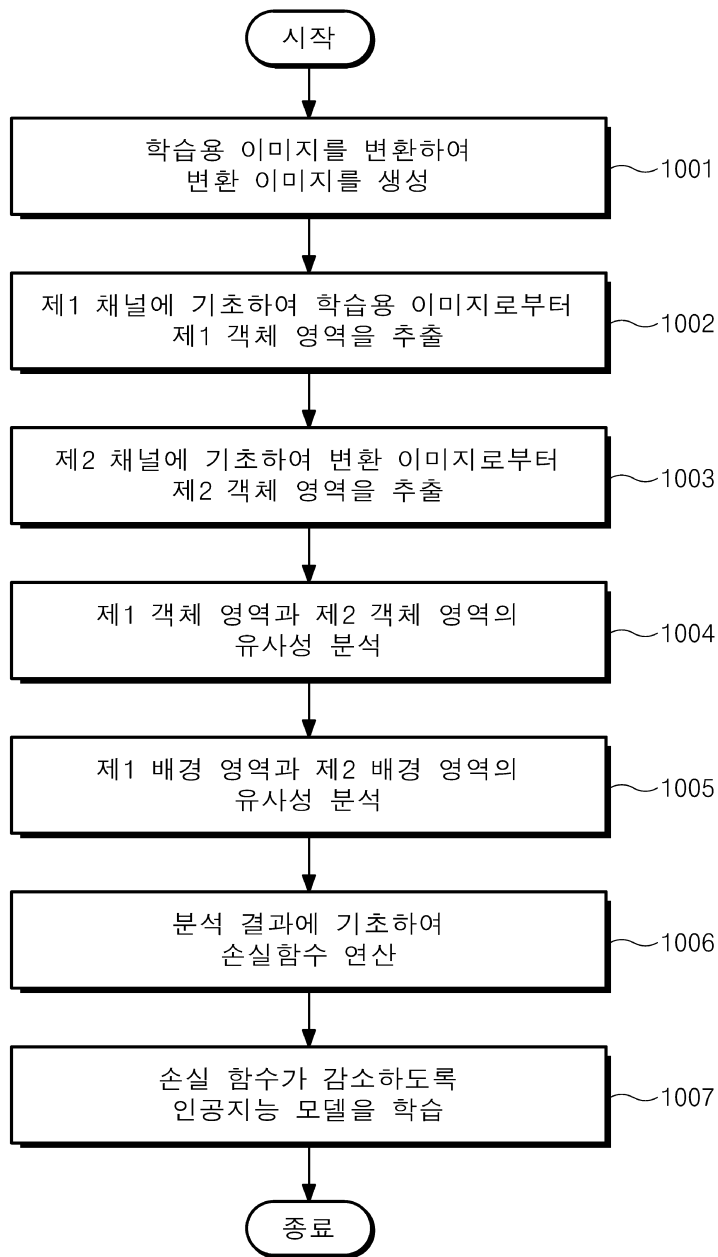


도면7



- red box: ground truth
- green box: predicted results *PST: Point Symmetric Transformation

도면8

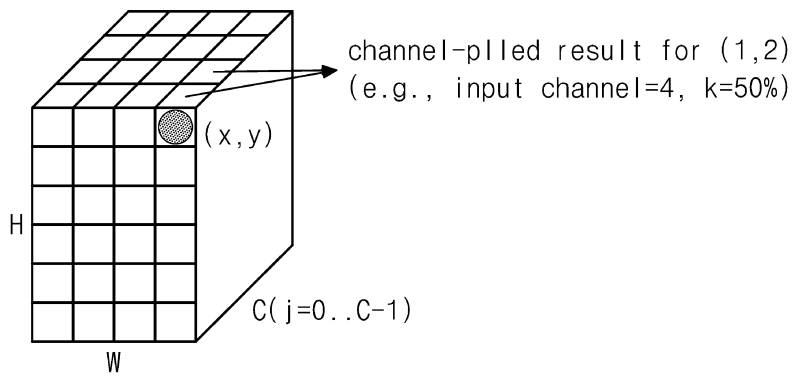
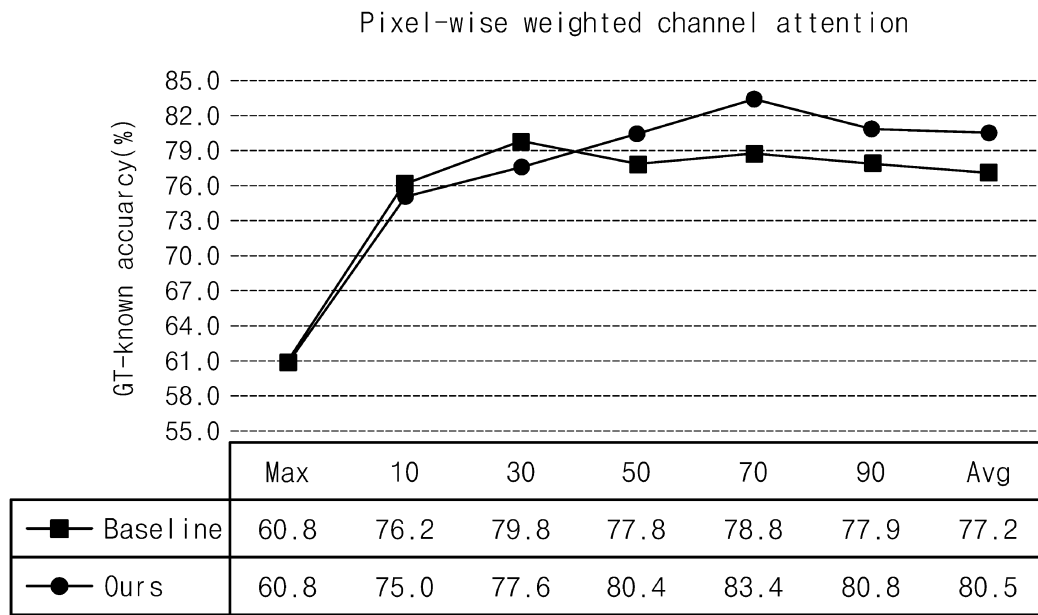


도면9

Method	CUB-200-2011[1]			Localization accuracy@IoU(%)			
	\mathcal{L}_{cls}	\mathcal{L}_{pec}	A_i	30	50	70	Mean
Baseline	✓	✗	✗	96.75	77.21	28.66	67.54
Ours(full)	✓	✓	✓	97.30	83.65	38.64	73.20
$-\mathcal{L}_{cls}$	✗	✓	✓	94.73	77.25	40.24	70.74(-2.46)
$-\mathcal{L}_{pec}$	✓	✗	✓	96.65	78.97	32.70	69.44(-3.76)
$-A_i$	✓	✓	✗	96.73	79.49	32.27	69.50(-3.70)

- Baseline: rotation prediction task with \mathcal{L}_{cls}
- \mathcal{L}_{cls} : classification loss
- \mathcal{L}_{pec} : pretext-equivariant contrastive loss
- A_i : pixel-wise weighted channel attention

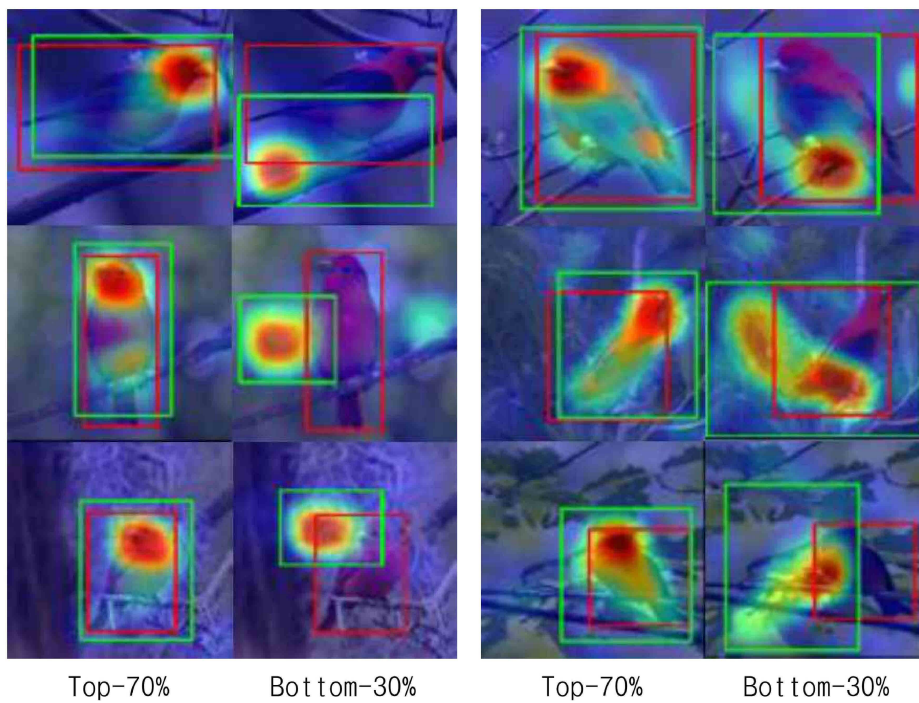
도면10



weighted feature map F_w : descending order

도면11

Method	Localization accuracy@IoU(%)			
	30	50	70	Mean
Baseline	96.75	77.21	28.66	67.54
Ours(top-70%)	97.30	83.65	38.64	73.20
Ours(bottom-30%)	89.04	57.90	17.74	54.89



- red box: ground truth
- green box: predicted results

도면12

Tasks	CUB-200-2011[1]		Stanford Cars		FGVC-Aircraft		Stanford Dogs	
	Base	Ours	Base	Ours	Base	Ours	Base	Ours
Rotation[14]	77.21	<u>83.65</u>	88.69	<u>97.41</u>	91.86	<u>97.14</u>	77.79	82.03
Scale	44.66	<u>83.62</u>	77.41	91.69	65.70	<u>96.69</u>	67.30	<u>82.63</u>
Translation	22.98	83.05	61.37	<u>97.40</u>	47.13	96.36	58.05	66.72
Hflip	7.76	75.12	91.415	96.81	87.15	94.83	78.18	80.88
Vflip	75.94	80.51	91.96	96.08	92.85	94.77	77.07	80.44
PST[7]	48.92	76.80	92.89	95.68	92.41	96.51	67.91	<u>86.64</u>