(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0044649
(43) 공개일자 2023년04월04일

(51) 국제특허분류(Int. Cl.)
G06F 18/00 (2023.01) *G06N 3/08* (2023.01)
G06T 5/20 (2006.01) *G06T 7/194* (2017.01)
(52) CPC특허분류
G06V 20/00 (2022.01)
G06N 3/08 (2023.01)
(21) 출원번호 10-2021-0127014
(22) 출원일자 2021년09월27일
심사청구일자 2021년09월27일

(71) 출원인
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
변혜란
서울특별시 서대문구 연세로 50 연세대학교 제4공학관 810호
이원영
서울특별시 서대문구 연세로 50 연세대학교 제 4 공학관 810호
(뒷면에 계속)
(74) 대리인
정부연

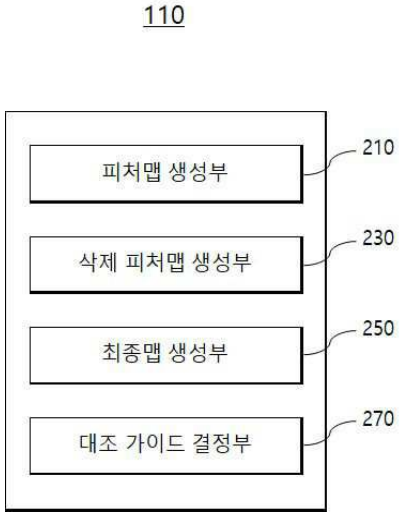
전체 청구항 수 : 총 14 항

(54) 발명의 명칭 **약지도 객체인식 장치 및 방법**

(57) 요 약

약지도 객체인식 장치(Weakly Supervised Object Localization Apparatus)는 입력 이미지에 대한 제1 컨볼루션 연산을 수행하여 피처맵(X)을 생성하는 피처맵 생성부, 상기 피처맵(X)으로 어텐션 맵(A)을 생성하고 상기 어텐션 맵(A)를 통해 상기 입력 이미지에 대한 마스킹 연산을 수행하여 삭제 피처맵(-X)을 생성하는 삭제 피처맵 생성부, 상기 피처맵(X) 및 상기 삭제 피처맵(-X)에 대한 제2 컨볼루션 연산을 수행하여 최종 피처맵(F) 및 최종 삭제 피처맵(-F)을 각각 생성하는 최종맵 생성부, 및 상기 최종 피처맵(F) 및 상기 최종 삭제 피처맵(-F)을 기초로 상기 입력 이미지의 포그라운드 객체에 대한 대조 가이드(contrastive guidance)를 결정하는 대조 가이드 결정부를 포함한다.

대 표 도 - 도2



(52) CPC특허분류
G06T 5/20 (2023.01)
G06T 7/194 (2017.01)

(72) 발명자
기민송
서울특별시 서대문구 연세로 50 연세대학교 제4공학관 810호
이제욱
서울특별시 서대문구 연세로 50 연세대학교 제4공학관 810호
박성호
서울특별시 서대문구 연세로 50 연세대학교 제 4공학관 810호

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 1/3
과제수행기관명 연세대학교 산학협력단
연구기간 2021.01.01 ~ 2021.12.31
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 1/3
과제수행기관명 연세대학교
연구기간 2021.03.01 ~ 2022.02.28
이 발명을 지원한 국가연구개발사업
과제고유번호 1711125843
과제번호 2018-0-00769-004
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 SW컴퓨팅산업원천기술개발(R&D,정보화)
연구과제명 인공지능 시스템을 위한 뉴로모픽 컴퓨팅 SW 플랫폼 기술 개발
기 여 율 1/3
과제수행기관명 한국전자통신연구원
연구기간 2021.01.01 ~ 2021.12.31

명 세 서

청구범위

청구항 1

입력 이미지에 대한 제1 컨볼루션 연산을 수행하여 피처맵(X)을 생성하는 피처맵 생성부;

상기 피처맵(X)으로 어텐션 맵(A)을 생성하고 상기 어텐션 맵(A)를 통해 상기 입력 이미지에 대한 마스킹 연산을 수행하여 삭제 피처맵(-X)을 생성하는 삭제 피처맵 생성부;

상기 피처맵(X) 및 상기 삭제 피처맵(-X)에 대한 제2 컨볼루션 연산을 수행하여 최종 피처맵(F) 및 최종 삭제 피처맵(-F)을 각각 생성하는 최종맵 생성부; 및

상기 최종 피처맵(F) 및 상기 최종 삭제 피처맵(-F)을 기초로 상기 입력 이미지의 포그라운드 객체에 대한 대조 가이드(contrastive guidance)를 결정하는 대조 가이드 결정부를 포함하는 약지도 객체인식 장치(Weakly Supervised Object Localization Apparatus).

청구항 2

제1항에 있어서, 상기 삭제 피처맵 생성부는

상기 피처맵(X)을 채널-와이즈 풀링(channel-wise pooling)을 통해 상기 어텐션 맵(A)을 생성하는 것을 특징으로 하는 약지도 객체인식 장치.

청구항 3

제2항에 있어서, 상기 삭제 피처맵 생성부는

상기 어텐션 맵(A)에서 가장 특징적인 부분에 대한 마스크를 생성하여 상기 입력 이미지에 대한 마스킹 연산을 수행하는 것을 특징으로 하는 약지도 객체인식 장치.

청구항 4

제1항에 있어서, 상기 최종맵 생성부는

상기 최종 피처맵(F) 및 상기 최종 삭제 피처맵(-F)을 통해 상기 입력 이미지에서 서로 다른 지역을 활성화하여 상기 포그라운드 객체를 배경과 멀어지도록 하는 것을 특징으로 하는 약지도 객체인식 장치.

청구항 5

제4항에 있어서, 상기 최종맵 생성부는

상기 최종 피처맵(F) 및 상기 최종 삭제 피처맵(-F)에 대해 채널-와이즈 풀링(channel-wise pooling) 기반의 어텐션 맵(A_F, $-A_F$)을 통해 포그라운드 마스크(Mfg, -Mfg) 및 백그라운드 마스크(Mbg, -Mbg)를 생성하는 것을 특징으로 하는 약지도 객체인식 장치.

청구항 6

제5항에 있어서, 상기 최종맵 생성부는

상기 포그라운드 마스크(Mfg, -Mfg) 및 상기 백그라운드 마스크(Mbg, -Mbg)를 기초로 전경 피처맵(Ffg, -Ffg)

및 배경 피처맵(Fbg, -Fbg)을 생성하는 것을 특징으로 하는 약지도 객체인식 장치.

청구항 7

제6항에 있어서, 상기 최종맵 생성부는

상기 전경 피처맵(Ffg, -Ffg) 및 상기 배경 피처맵(Fbg, -Fbg)을 정규 임베딩 공간에 투영시켜 다차원 피처 벡터(Zfg, Zbg, -Zfg, -Zbg)를 생성하는 것을 특징으로 하는 약지도 객체인식 장치.

청구항 8

제4항에 있어서, 상기 최종맵 생성부는

상기 최종 피처맵(F) 및 상기 최종 삭제 피처맵(-F)에 대해 제3 컨볼루션 연산을 수행하여 키, 쿼리 및 밸류(k, q, v)를 생성하고 상기 키, 쿼리 및 밸류를 가중치 매트릭스(W)로 프로덕트 연산하여 상기 최종 피처맵(F) 및 상기 최종 삭제 피처맵(-F)의 성능을 향상시키는 것을 특징으로 하는 약지도 객체인식 장치.

청구항 9

제1항에 있어서, 상기 대조 가이드 결정부는

상기 최종 피처맵(F) 및 상기 최종 삭제 피처맵(-F)을 기초로 생성된 다차원 피처 벡터(Zfg, Zbg, -Zfg, -Zbg) 중 전경 피처 벡터(Zfg, -Zfg) 각각의 배경 피처 벡터(Zbg, -Zbg) 간의 거리를 확대하면서 전경 피처 백터(Zfg, -Zfg) 간의 거리를 감소시키도록 대조 가이드(contrastive guidance)를 결정하는 것을 특징으로 하는 약지도 객체인식 장치.

청구항 10

입력 이미지에 대한 제1 컨볼루션 연산을 수행하여 피처맵(X)을 생성하는 피처맵 생성단계;

상기 피처맵(X)으로 어텐션 맵(A)을 생성하고 상기 어텐션 맵(A)를 통해 상기 입력 이미지에 대한 마스킹 연산을 수행하여 삭제 피처맵(-X)을 생성하는 삭제 피처맵 생성단계;

상기 피처맵(X) 및 상기 삭제 피처맵(-X)에 대한 제2 컨볼루션 연산을 수행하여 최종 피처맵(F) 및 최종 삭제 피처맵(-F)을 각각 생성하는 최종맵 생성단계; 및

상기 최종 피처맵(F) 및 상기 최종 삭제 피처맵(-F)을 기초로 상기 입력 이미지의 포그라운드 객체에 대한 대조 가이드(contrastive guidance)를 결정하는 대조 가이드 결정단계를 포함하는 약지도 객체인식 방법(Weakly Supervised Object Localization Method).

청구항 11

제10항에 있어서, 상기 삭제 피처맵 생성단계는

상기 피처맵(X)을 채널-와이즈 풀링(channel-wise pooling)을 통해 상기 어텐션 맵(A)을 생성하는 단계; 및

상기 어텐션 맵(A)에서 가장 특징적인 부분에 대한 마스크를 생성하여 상기 입력 이미지에 대한 마스킹 연산을 수행하는 단계를 포함하는 것을 특징으로 하는 약지도 객체인식 방법.

청구항 12

제10항에 있어서, 상기 최종맵 생성단계는

상기 최종 피처맵(F) 및 상기 최종 삭제 피처맵(-F)에 대해 채널-와이즈 풀링(channel-wise pooling) 기반의 어텐션 맵(A_F, $-A_F$)을 통해 포그라운드 마스크(Mfg, -Mfg) 및 백그라운드 마스크(Mbg, -Mbg)를 생성하는 단계;

상기 포그라운드 마스크(Mfg, -Mfg) 및 백그라운드 마스크(Mbg, -Mbg)를 기초로 전경 피처맵(Ffg, -Ffg) 및 배경 피처맵(Fbg, -Fbg)을 생성하는 단계; 및

상기 전경 피처맵(Ffg, -Ffg) 및 배경 피처맵(Fbg, -Fbg)을 정규 임베딩 공간에 투영시켜 다차원 피처 벡터(Zfg, Zbg, -Zfg, -Zbg)를 생성하는 단계를 포함하는 것을 특징으로 하는 약지도 객체인식 방법.

청구항 13

제10항에 있어서, 상기 최종맵 생성단계는

상기 최종 피처맵(F) 및 상기 최종 삭제 피처맵(-F)에 대해 제3 컨볼루션 연산을 수행하여 키(k), 쿼리(q) 및 밸류(v)를 생성하고 상기 키(k), 쿼리(q) 및 밸류(v)를 가중치 매트릭스(W)로 프로덕트 연산하여 상기 최종 피처맵(F) 및 상기 최종 삭제 피처맵(-F)의 성능을 향상시키는 것을 특징으로 하는 약지도 객체인식 방법.

청구항 14

제10항에 있어서, 상기 대조 가이드 결정단계는

상기 최종 피처맵(F) 및 상기 최종 삭제 피처맵(-F)을 기초로 생성된 다차원 피처 벡터(Zfg, Zbg, -Zfg, -Zbg) 중 전경 피처 벡터(Zfg, -Zfg) 각각의 배경 피처 벡터(Zbg, -Zbg) 간의 거리를 확대하면서 전경 피처 벡터(Zfg, -Zfg) 간의 거리를 감소시키도록 대조 가이드(contrastive guidance)를 결정하는 것을 특징으로 하는 약지도 객체인식 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 객체인식 기술에 관한 것으로, 보다 상세하게는 CNN(Convolutional Neural Network) 학습을 기반으로 이미지 내의 객체에 대한 대조 가이드를 통해 정확한 객체의 영역을 검출할 수 있는 약지도 객체인식 장치 및 방법에 관한 것이다.

배 경 기 술

[0003] 객체인식(Object Localization)은 이미지 상의 객체를 식별하기 위해 이미지 내의 단일 객체에 대한 분류 및 위치 검출하는 컴퓨터 비전 기술이다. CNN과 같은 딥러닝 모델은 객체 인식을 위해 해당 객체의 고유의 특징을 자동으로 학습하는데 사용된다.

[0004] 딥러닝을 통한 객체 인식 방법들은 이미 만들어진 데이터 셋과 그 안에 포함되어 있는 객체의 위치에 대한 실제 정보를 같이 학습하는 방법으로 설계되어 있는데, 이런 딥러닝 학습 모델을 완전 지도학습(Fully supervised) 방법이라고 한다. 완전 지도학습 방법을 통한 객체의 위치 검출 방법은 성능이 뛰어나지만 객체의 위치에 대한 실제 정보를 학습 과정에 반드시 포함해야 한다는 단점이 있는데, 이 때문에 시간이 지날수록 다양한 데이터를 학습하면서 객체의 위치에 대한 레이블을 만들어 주어야 하는 데 많은 시간을 소모해야 하는 문제가 있다.

[0005] 그래서, 최근에는 완전 지도학습 방법 외에 약지도 학습(Weakly supervised)의 방법을 통해 다양한 연구를 하고 있다. 약지도 학습이란 학습과정에서 이미지와 그에 대한 클래스 레이블만을 학습시켜 딥러닝 예측 모델을 생성하는 방법이다. 완전 지도학습과 비교하여 약지도 학습은 객체의 실제 위치에 대한 레이블이 필요하지 않기 때문에 많은 인적, 물적 낭비를 줄일 수 있다는 장점이 있다.

[0006] 하지만, 약지도 학습을 통해 객체인식을 하는 기존 방법은 CNN에서의 분류기가 이미지를 어떤 클래스에 속하는

지 분류할 때 가장 특징적인 부분만을 보고 판단하여 이 영역만을 인식하기 때문에 객체 인식 효율이 떨어지는 문제가 있다.

선행기술문헌

특허문헌

[0008] (특허문헌 0001) 한국등록특허 제10-1879207(2018.07.11)호

발명의 내용

해결하려는 과제

[0009] 본 발명의 일 실시예는 CNN(Convolutional Neural Network) 학습을 기반으로 이미지 내의 객체에 대한 대조 가이드를 통해 정확한 객체의 영역을 검출할 수 있는 약지도 객체인식 장치 및 방법을 제공하고자 한다.

[0010] 본 발명의 일 실시예는 객체의 포그라운드 전체를 인식하고 백그라운드를 삭제하여 정확한 객체의 영역을 검출함으로써 객체 인식 효율을 높일 수 있는 약지도 객체인식 장치 및 방법을 제공하고자 한다.

[0011] 본 발명의 일 실시예는 객체 인식 성능을 향상시킬 수 있는 AE(Adversarial erasing) 기반의 새로운 약지도 객체인식(WSOL) 프레임워크를 제안하는 약지도 객체인식 장치 및 방법을 제공하고자 한다.

과제의 해결 수단

[0013] 실시예들 중에서, 약지도 객체인식 장치는 입력 이미지에 대한 제1 컨볼루션 연산을 수행하여 피처맵(X)을 생성하는 피처맵 생성부, 상기 피처맵(X)으로 어텐션 맵(A)을 생성하고 상기 어텐션 맵(A)를 통해 상기 입력 이미지에 대한 마스킹 연산을 수행하여 삭제 피처맵(-X)을 생성하는 삭제 피처맵 생성부, 상기 피처맵(X) 및 상기 삭제 피처맵(-X)에 대한 제2 컨볼루션 연산을 수행하여 최종 피처맵(F) 및 최종 삭제 피처맵(-F)을 각각 생성하는 최종맵 생성부, 및 상기 최종 피처맵(F) 및 상기 최종 삭제 피처맵(-F)을 기초로 상기 입력 이미지의 포그라운드 객체에 대한 대조 가이드(contrastive guidance)를 결정하는 대조 가이드 결정부를 포함한다.

[0014] 상기 삭제 피처맵 생성부는 상기 피처맵(X)을 채널-와이즈 풀링(channel-wise pooling)을 통해 상기 어텐션 맵(A)을 생성할 수 있다.

[0015] 상기 삭제 피처맵 생성부는 상기 어텐션 맵(A)에서 가장 특징적인 부분에 대한 마스크를 생성하여 상기 입력 이미지에 대한 마스킹 연산을 수행할 수 있다.

[0016] 상기 최종맵 생성부는 상기 최종 피처맵(F) 및 상기 최종 삭제 피처맵(-F)을 통해 상기 입력 이미지에서 서로 다른 지역을 활성화하여 상기 포그라운드 객체를 배경과 멀어지도록 할 수 있다.

[0017] 상기 최종맵 생성부는 상기 최종 피처맵(F) 및 상기 최종 삭제 피처맵(-F)에 대해 채널-와이즈 풀링(channel-wise pooling) 기반의 어텐션 맵(A_F, $-A_F$)을 통해 포그라운드 마스크(Mfg, -Mfg) 및 백그라운드 마스크(Mbg, -Mbg)를 생성할 수 있다.

[0018] 상기 최종맵 생성부는 상기 포그라운드 마스크(Mfg, -Mfg) 및 상기 백그라운드 마스크(Mbg, -Mbg)를 기초로 전경 피처맵(Ffg, -Ffg) 및 배경 피처맵(Fbg, -Fbg)을 생성할 수 있다.

[0019] 상기 최종맵 생성부는 상기 전경 피처맵(Ffg, -Ffg) 및 상기 배경 피처맵(Fbg, -Fbg)을 정규 임베딩 공간에 투영시켜 다차원 피처 벡터(Zfg, Zbg, -Zfg, -Zbg)를 생성할 수 있다.

[0020] 상기 최종맵 생성부는 상기 최종 피처맵(F) 및 상기 최종 삭제 피처맵(-F)에 대해 제3 컨볼루션 연산을 수행하여 키, 쿼리 및 밸류(k, q, v)를 생성하고 상기 키, 쿼리 및 밸류를 가중치 매트릭스(W)로 프로덕트 연산하여 상기 최종 피처맵(F) 및 상기 최종 삭제 피처맵(-F)의 성능을 향상시킬 수 있다.

[0021] 상기 대조 가이드 결정부는 상기 최종 피처맵(F) 및 상기 최종 삭제 피처맵(-F)을 기초로 생성된 다차원 피처

벡터(Zfg, Zbg, -Zfg, -Zbg) 중 전경 피처 벡터(Zfg, -Zfg) 각각의 배경 피처 벡터(Zbg, -Zbg) 간의 거리를 확대하면서 전경 피처 백터(Zfg, -Zfg) 간의 거리를 감소시키도록 대조 가이드(contrastive guidance)를 결정할 수 있다.

[0022] 실시예들 중에서, 약지도 객체인식 방법은 입력 이미지에 대한 제1 컨볼루션 연산을 수행하여 피처맵(X)을 생성하는 피처맵 생성단계, 상기 피처맵(X)으로 어텐션 맵(A)을 생성하고 상기 어텐션 맵(A)를 통해 상기 입력 이미지에 대한 마스킹 연산을 수행하여 삭제 피처맵(-X)을 생성하는 삭제 피처맵 생성단계, 상기 피처맵(X) 및 상기 삭제 피처맵(-X)에 대한 제2 컨볼루션 연산을 수행하여 최종 피처맵(F) 및 최종 삭제 피처맵(-F)을 각각 생성하는 최종맵 생성단계, 및 상기 최종 피처맵(F) 및 상기 최종 삭제 피처맵(-F)을 기초로 상기 입력 이미지의 포그라운드 객체에 대한 대조 가이드(contrastive guidance)를 결정하는 대조 가이드 결정단계를 포함한다.

[0023] 상기 삭제 피처맵 생성단계는 상기 피처맵(X)을 채널-와이즈 풀링(channel-wise pooling)을 통해 상기 어텐션 맵(A)을 생성하는 단계, 및 상기 어텐션 맵(A)에서 가장 특징적인 부분에 대한 마스크를 생성하여 상기 입력 이미지에 대한 마스킹 연산을 수행하는 단계를 포함할 수 있다.

[0024] 상기 최종맵 생성단계는 상기 최종 피처맵(F) 및 상기 최종 삭제 피처맵(-F)에 대해 채널-와이즈 풀링(channel-wise pooling) 기반의 어텐션 맵(A_F, $-A_F$)을 통해 포그라운드 마스크(Mfg, -Mfg) 및 백그라운드 마스크(Mbg, -Mbg)를 생성하는 단계, 상기 포그라운드 마스크(Mfg, -Mfg) 및 백그라운드 마스크(Mbg, -Mbg)를 기초로 전경 피처맵(Ffg, -Ffg) 및 배경 피처맵(Fbg, -Fbg)을 생성하는 단계, 및 상기 전경 피처맵(Ffg, -Ffg) 및 배경 피처맵(Fbg, -Fbg)을 정규 임베딩 공간에 투영시켜 다차원 피처 벡터(Zfg, Zbg, -Zfg, -Zbg)를 생성하는 단계를 포함할 수 있다.

[0025] 상기 최종맵 생성단계는 상기 최종 피처맵(F) 및 상기 최종 삭제 피처맵(-F)에 대해 제3 컨볼루션 연산을 수행하여 키(k), 쿼리(q) 및 밸류(v)를 생성하고 상기 키(k), 쿼리(q) 및 밸류(v)를 가중치 매트릭스(W)로 프로덕트 연산하여 상기 최종 피처맵(F) 및 상기 최종 삭제 피처맵(-F)의 성능을 향상시킬 수 있다.

[0026] 상기 대조 가이드 결정단계는 상기 최종 피처맵(F) 및 상기 최종 삭제 피처맵(-F)을 기초로 생성된 다차원 피처 벡터(Zfg, Zbg, -Zfg, -Zbg) 중 전경 피처 벡터(Zfg, -Zfg) 각각의 배경 피처 벡터(Zbg, -Zbg) 간의 거리를 확대하면서 전경 피처 벡터(Zfg, -Zfg) 간의 거리를 감소시키도록 대조 가이드(contrastive guidance)를 결정할 수 있다.

발명의 효과

[0028] 개시된 기술은 다음의 효과를 가질 수 있다. 다만, 특정 실시예가 다음의 효과를 전부 포함하여야 한다거나 다음의 효과만을 포함하여야 한다는 의미는 아니므로, 개시된 기술의 권리범위는 이에 의하여 제한되는 것으로 이해되어서는 아니 될 것이다.

[0029] 본 발명의 일 실시예에 따른 약지도 객체인식 장치 및 방법은 CNN(Convolutional Neural Network) 학습을 기반으로 이미지 내의 객체에 대한 대조 가이드를 통해 정확한 객체의 영역을 검출할 수 있다.

[0030] 본 발명의 일 실시예에 따른 약지도 객체인식 장치 및 방법은 객체의 포그라운드 전체를 인식하고 백그라운드를 삭제하여 정확한 객체의 영역을 검출함으로써 객체 인식 효율을 높일 수 있다.

[0031] 본 발명의 일 실시예에 따른 약지도 객체인식 장치 및 방법은 객체 인식 성능을 향상시킬 수 있는 AE(Adversarial erasing) 기반의 새로운 약지도 객체인식(WSOL) 프레임워크를 제안할 수 있다.

도면의 간단한 설명

[0033] 도 1은 본 발명의 일 실시예에 따른 약지도 객체인식 장치의 시스템 구성을 설명하는 도면이다.

도 2는 도 1의 약지도 객체인식 장치에 있는 프로세스의 기능적 구성을 설명하는 도면이다.

도 3은 도 2의 약지도 객체인식 장치에서 수행되는 약지도 객체인식 과정을 설명하는 순서도이다.

도 4는 본 발명에 따른 약지도 객체인식을 위한 전체 프레임워크를 설명하는 도면이다.

도 5는 원본 분기와 삭제 분기의 피처맵에서 활성화 변화를 나타내는 도면이다.

도 6은 전경 및 배경 피처맵을 정규 임베딩 공간에 투영시켜 다차원 벡터를 생성하는 도면이다.

도 7은 본 발명에 따른 WSOL 프레임워크의 ImageNet 및 CUB-200-2011 데이터셋에 대한 정성적 결과를 나타내는 도면이다.

도 8은 기존 방법 대비 본 발명에 따른 약지도 객체인식 결과를 보여주는 예시도이다.

발명을 실시하기 위한 구체적인 내용

[0034] 본 발명에 관한 설명은 구조적 내지 기능적 설명을 위한 실시예에 불과하므로, 본 발명의 권리범위는 본문에 설명된 실시예에 의하여 제한되는 것으로 해석되어서는 아니 된다. 즉, 실시예는 다양한 변경이 가능하고 여러 가지 형태를 가질 수 있으므로 본 발명의 권리범위는 기술적 사상을 실현할 수 있는 균등물들을 포함하는 것으로 이해되어야 한다. 또한, 본 발명에서 제시된 목적 또는 효과는 특정 실시예가 이를 전부 포함하여야 한다거나 그러한 효과만을 포함하여야 한다는 의미는 아니므로, 본 발명의 권리범위는 이에 의하여 제한되는 것으로 이해되어서는 아니 될 것이다.

[0035] 한편, 본 출원에서 서술되는 용어의 의미는 다음과 같이 이해되어야 할 것이다.

[0036] "제1", "제2" 등의 용어는 하나의 구성요소를 다른 구성요소로부터 구별하기 위한 것으로, 이들 용어들에 의해 권리범위가 한정되어서는 아니 된다. 예를 들어, 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다.

[0037] 어떤 구성요소가 다른 구성요소에 "연결되어"있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결될 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어"있다고 언급된 때에는 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다. 한편, 구성요소들 간의 관계를 설명하는 다른 표현들, 즉 "~사이에"와 "바로 ~사이에" 또는 "~에 이웃하는"과 "~에 직접 이웃하는" 등도 마찬가지로 해석되어야 한다.

[0038] 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한 복수의 표현을 포함하는 것으로 이해되어야 하고, "포함하다"또는 "가지다" 등의 용어는 실시된 특징, 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함을 지정하려는 것이며, 하나 또는 그 이상의 다른 특징이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0039] 각 단계들에 있어 식별부호(예를 들어, a, b, c 등)는 설명의 편의를 위하여 사용되는 것으로 식별부호는 각 단계들의 순서를 설명하는 것이 아니며, 각 단계들은 문맥상 명백하게 특정 순서를 기재하지 않는 이상 명기된 순서와 다르게 일어날 수 있다. 즉, 각 단계들은 명기된 순서와 동일하게 일어날 수도 있고 실질적으로 동시에 수행될 수도 있으며 반대의 순서대로 수행될 수도 있다.

[0040] 본 발명은 컴퓨터가 읽을 수 있는 기록매체에 컴퓨터가 읽을 수 있는 코드로서 구현될 수 있고, 컴퓨터가 읽을 수 있는 기록 매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록 장치를 포함한다. 컴퓨터가 읽을 수 있는 기록 매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피 디스크, 광 데이터 저장 장치 등이 있다. 또한, 컴퓨터가 읽을 수 있는 기록 매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산 방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수 있다.

[0041] 여기서 사용되는 모든 용어들은 다르게 정의되지 않는 한, 본 발명이 속하는 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가진다. 일반적으로 사용되는 사전에 정의되어 있는 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한 이상적이거나 과도하게 형식적인 의미를 지니는 것으로 해석될 수 없다.

[0043] 도 1은 본 발명의 일 실시예에 따른 약지도 객체인식 장치의 시스템 구성을 설명하는 도면이다.

[0044] 도 1을 참조하면, 약지도 객체인식 장치(100)는 프로세서(110), 메모리(130), 사용자 입출력부(150) 및 네트워크 입출력부(170)를 포함하는 컴퓨팅 시스템으로 구성될 수 있다.

[0045] 프로세서(110)는 약지도 객체인식 장치(100)가 동작하는 과정에서의 각 단계들을 처리하는 프로시저를 실행할 수 있고, 그 과정 전반에서 읽혀지거나 작성되는 메모리(130)를 관리할 수 있으며, 메모리(130)에 있는 휘발성

메모리와 비휘발성 메모리 간의 동기화 시간을 스케줄할 수 있다. 프로세서(110)는 약지도 객체인식 장치(100)의 동작 전반을 제어할 수 있고, 메모리(130), 사용자 입출력부(150) 및 네트워크 입출력부(170)와 전기적으로 연결되어 이들 간의 데이터 흐름을 제어할 수 있다. 프로세서(110)는 약지도 객체인식 장치(100)의 CPU(Central Processing Unit)로 구현될 수 있다.

[0046] 메모리(130)는 SSD(Solid State Drive) 또는 HDD(Hard Disk Drive)와 같은 비휘발성 메모리로 구현되어 약지도 객체인식 장치(100)에 필요한 데이터 전반을 저장하는데 사용되는 보조기억장치를 포함할 수 있고, RAM(Random Access Memory)과 같은 휘발성 메모리로 구현된 주기억장치를 포함할 수 있다.

[0047] 사용자 입출력부(150)는 사용자 입력을 수신하기 위한 환경 및 사용자에게 특정 정보를 출력하기 위한 환경을 포함할 수 있다. 예를 들어, 사용자 입출력부(150)는 터치 패드, 터치 스크린, 화상 키보드 또는 포인팅 장치와 같은 어댑터를 포함하는 입력장치 및 모니터 또는 터치스크린과 같은 어댑터를 포함하는 출력장치를 포함할 수 있다. 일 실시예에서, 사용자 입출력부(150)는 원격 접속을 통해 접속되는 컴퓨팅 장치에 해당할 수 있고, 그러한 경우, 약지도 객체인식 장치(100)는 독립적인 서버로서 수행될 수 있다.

[0048] 네트워크 입출력부(170)은 네트워크를 통해 외부 장치 또는 시스템과 연결하기 위한 환경을 포함하고, 예를 들어, LAN(Local Area Network), MAN(Metropolitan Area Network), WAN(Wide Area Network) 및 VAN(Value Added Network) 등의 통신을 위한 어댑터를 포함할 수 있다.

[0050] 도 2는 도 1의 약지도 객체인식 장치에 있는 프로세서의 기능적 구성을 설명하는 도면이다.

[0051] 도 2를 참조하면, 약지도 객체인식 장치(100)는 피처맵 생성부(210), 삭제 피처맵 생성부(230), 최종맵 생성부(250) 및 대조 가이드 결정부(270)를 포함할 수 있고, 이들은 상호 연결될 수 있다.

[0052] 피처맵 생성부(210)는 입력 이미지에 대한 제1 컨볼루션 연산을 수행하여 피처맵(X)을 생성할 수 있다. 피처맵 생성부(210)는 컨볼루션 신경망(CNN) 구조로 구현되고 이미지를 통과시켜 각 클래스에 따른 피처맵(X)을 생성할 수 있다.

[0053] CNN 구조는 각 레이어의 입출력 데이터의 형상 유지, 복수의 필터로 이미지의 특징 추출 및 학습, 추출한 이미지의 특징들을 모으고 강화하는 풀링(pooling) 레이어로 일반 인공 신경망보다 적은 학습 파라미터를 갖는다. CNN은 이미지 특징 추출을 위하여 입력 데이터를 필터가 순회하며 컨볼루션을 계산하고, 그 계산 결과를 이용하여 피처맵(Feature map)을 생성한다. 피처맵은 원본 이미지의 위치 정보를 내포할 수 있다.

[0054] 삭제 피처맵 생성부(230)는 어텐션 맵(A)을 생성하고 어텐션 맵(A)을 통해 입력 이미지에 대한 마스킹 연산을 수행하여 삭제 피처맵(-X)을 생성할 수 있다. 삭제 피처맵 생성부(230)는 채널-와이즈 풀링(channel-wise pooling)을 통해 어텐션 맵(A)을 생성할 수 있다. 일 실시예에서, 삭제 피처맵 생성부(230)는 피처맵 생성부(210)의 백본 중간에서 피처맵(X)을 채널-와이즈 풀링을 통해 어텐션 맵(Attention map)(A)을 생성할 수 있다. 삭제 피처맵 생성부(230)는 생성된 어텐션 맵(A)에서 가장 특징적인 부분에 대한 마스크(Mask)를 생성하여 입력 이미지에 대한 마스킹 연산을 수행할 수 있다. 삭제 피처맵 생성부(230)는 임계값(Θ_d)을 어텐션 맵(A)의 가장 높은 값과 곱해 주어 이 값보다 크다면 가장 특징적인 픽셀 부분으로 간주하고 해당 픽셀 부분에 대한 마스크를 생성할 수 있다. 삭제 피처맵 생성부(230)는 마스크를 아래의 수학식 1을 통해 생성할 수 있다.

[0055] [수학식 1]

[0056] $$\mathbf{M_{pix}} = \mathbb{1}[\mathbf{A} > \tau_d], \quad where \quad \tau_d = max(\mathbf{A}) \times \theta_d$$

[0057] 여기에서, Mpix는 마스크에 해당하며, A는 어텐션 맵에 해당한다.

[0058] 삭제 픽처맵 생성부(230)는 어텐션 맵(A)에서 가장 특징적인 픽셀 부분을 영역 단위로 확장하기 위해서 SxS 크기의 커널 사이즈(kernel size)를 가지는 맥스 풀링 레이어(max pooling layer)를 마스크(Mpix)에 곱해주어 영역별 마스크(M)을 생성할 수 있다. 삭제 픽처맵 생성부(230)는 생성된 영역별 마스크(M)를 원래 피처맵(X)에 곱해주어 삭제 피처맵(-X)을 생성할 수 있다.

[0059] 최종맵 생성부(250)는 피처맵(X) 및 삭제 피처맵(-X)에 대한 제2 컨볼루션 연산을 수행하여 최종 피처맵(F) 및 최종 삭제 피처맵(-F)을 각각 생성할 수 있다. 일 실시예에서, 최종맵 생성부(250)는 피처맵(X) 및 삭제 피처맵(-X)을 백본 네트워크의 나머지 컨볼루션 레이어들에 통과시켜 최종 피처맵(F) 및 최종 삭제 피처맵(-F)을 각

각 생성할 수 있다. 최종맵 생성부(250)는 학습이 진행될수록 삭제 피처맵(-X)에서 삭제 영역이 넓어지고 최종 피처맵(F) 및 최종 삭제 피처맵(-F)에서 객체의 전체 영역을 더 활성화시킬 수 있다.

[0060] 최종맵 생성부(250)는 최종 피처맵(F) 및 최종 삭제 피처맵(-F)을 통해 입력 이미지에서 서로 다른 지역을 활성화하여 포그라운드 객체를 배경과 멀어지도록 할 수 있다. 최종맵 생성부(250)는 최종 피처맵(F) 및 최종 삭제 피처맵(-F)에 대해 채널-와이즈 풀링(channel-wise pooling) 기반의 어텐션 맵(A_F, $-A_F$)을 통해 전경 마스크(Mfg, -Mfg) 및 배경 마스크(Mbg, -Mbg)를 생성할 수 있다. 일 실시예에서, 최종맵 생성부(250)는 최종 피처맵(F) 및 최종 삭제 피처맵(-F)에서 채널-와이즈 풀링을 이용하여 어텐션 맵(A_F, $-A_F$)을 생성하고 각각의 어텐션 맵에서 임계값을 이용해 마스크를 생성하여 전경과 배경을 분리할 수 있다. 여기에서, 최종맵 생성부(250)는 생성된 마스크를 최종 피처맵(F)에 곱해주어 전경 피처맵 및 배경 피처맵을 생성할 수 있다. 최종맵 생성부(250)는 포그라운드 마스크 및 백그라운드 마스크를 아래의 수학식 2를 통해 생성할 수 있다.

[0061] [수학식 2]

[0062] $$\mathbf{M_{fg}} = \mathbb{1}[\mathbf{A_F} > \tau_{fg}], \quad \mathbf{M_{bg}} = \mathbb{1}[\mathbf{A_F} < \tau_{bg}]$$

[0063] 여기에서, Mfg는 포그라운드 마스크에 해당하며, Mbg는 백그라운드 마스크에 해당하고, A_F는 채널-와이즈 풀링된 어텐션 맵에 해당한다. τ_{fg} 및 τ_{bg}는 미리 정의된 임계값이다.

[0064] 최종맵 생성부(250)는 포그라운드 마스크(Mfg,-Mfg) 및 백그라운드 마스크(Mbg,-Mbg)를 기초로 전경 피처맵(Ffg,-Ffg) 및 배경 피처맵(Fbg,-Fbg)을 생성할 수 있다. 일 실시예에서, 최종맵 생성부(250)는 전경 피처맵 및 배경 피처맵을 아래의 수학식 3을 통해 생성할 수 있다.

[0065] [수학식 3]

[0066] $$\mathbf{F_{fg}} = \mathbf{F} \odot \mathbf{M_{fg}}, \quad \mathbf{F_{bg}} = \mathbf{F} \odot \mathbf{M_{bg}}$$

[0067] 여기에서, Ffg는 전경 피처맵에 해당하며, Fbg는 배경 피처맵에 해당하고, 피처맵(F)에 각 마스크(M)를 곱하여 생성된다.

[0068] 최종맵 생성부(250)는 전경 피처맵(Ffg, -Ffg) 및 배경 피처맵(Fbg, -Fbg)을 정규 임베딩 공간에 투영시켜 다차원 피처 벡터(Zfg, Zbg, -Zfg, -Zbg)를 생성할 수 있다. 일 실시예에서, 최종맵 생성부(250)는 생성된 전경과 배경의 피처맵들(Ffg, Fbg, -Ffg, -Fbg)을 정규 임베딩 공간에 투영시켜 128차원의 벡터를 생성할 수 있다.

[0069] 최종맵 생성부(250)는 최종 피처맵(F) 및 최종 삭제 피처맵(-F)에 대해 제3 컨볼루션 연산을 수행하여 키(key), 쿼리(query) 및 밸류(value)를 생성하고 키, 쿼리 및 밸류를 가중치 매트릭스(W)로 프로덕트 연산하여 최종 피처맵(F) 및 최종 삭제 피처맵(-F)의 성능을 향상시킬 수 있다. 일 실시예에서, 최종맵 생성부(250)는 가중치 매트릭스(W)를 키(k), 쿼리(q) 간의 화이트닝 도트 프로덕트(whitened dot product) 연산을 통해 생성할 수 있고, 가중치 매트릭스(W)는 아래의 수학식 4로 정의될 수 있다.

[0070] [수학식 4]

[0071] $$\mathbf{W} = \sigma\left((\mathbf{q}_i - \mu_{\mathbf{q}})^T (\mathbf{k}_j - \mu_{\mathbf{k}})\right)$$

[0072] 여기에서, σ는 소프트맥스(softmax) 함수이고, μ는 각 쿼리(q), 키(k) 안의 픽셀(i,j)의 평균값이다.

[0073] 최종맵 생성부(250)에서 최종 생성되는 향상된 피처맵(F')은 아래의 수학식 5를 통해 생성될 수 있다.

[0074] [수학식 5]

[0075] $$\mathbf{F}' = \mathbf{F} \oplus h(\mathbf{v} \otimes \mathbf{W})$$

[0076] 여기에서, h는 배치 정규화를 뒤따르는 1x1 컨볼루션 레이어에 해당한다.

[0077] 대조 가이드 결정부(270)는 최종 피처맵(F) 및 최종 삭제 피처맵(-F)을 기초로 입력 이미지의 전경 객체에 대한 대조 가이드(contrastive guidance)를 결정할 수 있다. 대조 가이드 결정부(270)는 최종 피처맵(F) 및 최종 삭

제 피처맵(-F)을 기초로 생성된 다차원 피처 벡터(Zfg, Zbg, -Zfg, -Zbg) 중 전경 피처 벡터(Zfg, -Zfg) 각각의 배경 피처 벡터(Zbg, -Zbg) 간의 거리를 확대하면서 전경 피처 벡터(Zfg, -Zfg) 간의 거리를 감소시키도록 대조 가이드(contrastive guidance)를 결정할 수 있다. 일 실시예에서, 대조 가이드 결정부(270)는 대조 가이드 손실(contrastive guidance loss)를 이용하여 전경 피처 벡터(Zfg)는 전경 삭제 피처 벡터(-Zfg)와 가까워지도록, 배경 피처 벡터(Zbg)와는 멀어지도록 학습하고, 전경 삭제 피처 벡터(-Zfg)도 마찬가지로 전경 피처 벡터(Zfg)와는 가까워지고 배경 삭제 피처 벡터(-Zbg)와는 멀어지도록 학습할 수 있다. 대조 가이드 손실은 아래의 수학식 6으로 정의될 수 있다.

[0078] [수학식 6]

[0079]
$$\mathcal{L}_{cg} = \{\max\left[\|(\mathbf{z}_{fg} - \bar{\mathbf{z}}_{fg})\|^2 - \|(\mathbf{z}_{fg} - \mathbf{z}_{bg})\|^2 + m, 0\right] + \max\left[\|(\bar{\mathbf{z}}_{fg} - \mathbf{z}_{fg})\|^2 - \|(\bar{\mathbf{z}}_{fg} - \bar{\mathbf{z}}_{bg})\|^2 + m, 0\right]\}$$

[0080] 여기에서, $\mathcal{L}_{cg}$ 는 대조 가이드 손실에 해당하며, m은 마진에 해당한다.

[0081] 대조 가이드 손실은 대상 객체에서 보완적으로 발견된 영역을 활용하여 4중 관계(원본 및 삭제된 분기의 전경 및 배경 피처맵)를 최적화할 수 있다. 따라서, 전체 객체를 올바른 범위로 발견하도록 가이드할 수 있다.

[0083] 도 3은 도 2의 약지도 객체인식 장치에서 수행되는 약지도 객체인식 과정을 설명하는 순서도이다.

[0084] 도 3을 참조하면, 약지도 객체인식 장치(100)는 피처맵 생성부(210)를 통해 입력 이미지에 대한 제1 컨볼루션 연산을 수행하여 피처맵(X)을 생성할 수 있다(단계 S310). 약지도 객체인식 장치(100)는 삭제 피처맵 생성부(230)를 통해 피처맵(X)으로 어텐션 맵(A)을 생성하고 어텐션 맵(A)을 통해 입력 이미지에 대한 마스킹 연산을 수행하여 삭제 피처맵(-X)을 생성할 수 있다(단계 S330). 약지도 객체인식 장치(100)는 최종맵 생성부(250)를 통해 피처맵(X) 및 삭제 피처맵(-X)에 대한 제2 컨볼루션 연산을 수행하여 최종 피처맵(F) 및 최종 삭제 피처맵(-F)을 각각 생성할 수 있다(단계 S350). 약지도 객체인식 장치(100)는 대조 가이드 결정부(270)를 통해 최종 피처맵(F) 및 최종 삭제 피처맵(-F)을 기초로 입력 이미지의 전경 객체에 대한 대조 가이드(contrastive guidance)를 결정할 수 있다(단계 S370).

[0086] 도 4는 본 발명에 따른 약지도 객체인식을 위한 전체 프레임워크를 설명하는 도면이다.

[0087] 본 발명은 약지도 객체인식을 위한 기존 방법들, CAM(class activation mapping)이나 HaS(Hide-and Seek), ACoL(Adversarial Complementary Leraning), ADL(Attention-based Dropout Layer), EIL(Erasing Integrated Leasning) 등의 AE(adversarial erasing) 방법에서의 성능 저하를 극복하기 위해 AE 기반의 새로운 WSOL(Weakly Supervised Object Localization) 프레임워크를 제안한다.

[0088] 도 4에서 볼 수 있듯이, 본 발명에 따른 WSOL 프레임워크(400)는 SRD(Scheduled Region Drop)구성(410), CG(Contrastive Guidance)구성(430), PNL(Pairwise Non-Local Block)구성(450)의 세 가지 핵심 요소로 이루어지며, 분류 네트워크를 활용하고 클래스 레이블만 사용하여 대조 가이드 손실과 분류 손실로 훈련한다.

[0089] SRD 구성(410)은 영역-레벨에 원본 피처맵에서 가장 구별되는 부분을 점진적으로 삭제하여 덜 유익한 영역을 효과적으로 발견하도록 네트워크를 촉진한다. SRD 구성(410)은 삭제 분기(Erased branch)의 입력이 되는 삭제 피처맵(-X)을 생성한다. 이 분기는 원래 분기(Original branch)의 가중치를 공유한다. 네트워크는 원본 및 삭제 피처맵(X,-X)을 동시에 피드-포워드하고 최종 피처맵(F,-F)을 출력하여 보완 영역을 탐색한다.

[0090] CG 구성(430)은 이중-분기의 전경 피처들이 함께 당겨지면서 각 배경 피처에서 멀어지도록 권장한다. 이것은 모델이 배경과 구별되는 전경의 표현을 학습하도록 하여 활성화가 배경으로 확장되는 것을 방지한다.

[0091] 또한 PNL 구성(450)은 피처맵의 픽셀 간의 관계를 학습하여 네트워크를 가속화하여 가장 독특한 영역의 다른 관련 부분을 발견한다. PNL 구성(450)은 픽셀 관계 간의 컨텍스트 정보(contextual information)를 학습하여 향상된 피처맵을 생성한다. 향상된 피처맵은 대조 손실을 계산하기 위한 대조 가이드에 대한 입력으로 제공된다.

대조 가이드 손실 $\mathcal{L}_{cg}$ 은 활성화 맵을 배경으로 확산시키지 않고 전체 객체 영역을 탐색하도록 네트워크를 가

이드한다.

[0092] 본 발명에 따른 WSOL 프레임워크(400)를 이루는 각 구성에 대해 이하에서 좀더 구체적으로 설명한다.

[0093] SRD(Scheduled Region Drop)

[0094] 적대적 삭제를 사용하는 기존의 WSOL 방법은 픽셀 수준에서 가장 구별되는 부분을 삭제하여 삭제 피처맵을 생성한다. 그러나 픽셀 수준의 드롭 만 사용하여 가장 유익한 부분에 인접한 픽셀을 완전히 제거하는 것은 어렵다. 이러한 나머지 정보 픽셀은 삭제된 분기가 보완 영역(즉, 대상 객체의 덜 구별되는 부분)을 발견하는 것을 방해한다. 차별화된 영역을 보다 효과적으로 제거하기 위해 영역별 삭제 전략을 제안한다.

[0095] 먼저, 채널-와이즈 풀링을 통해 원본 피처맵(X)의 어텐션 맵(A$\in^{\mathbb{R}^{1\times H\times W}}$)을 얻는다. 그런 다음, 다음과 같이 픽셀 수준 이진 마스크(Mpix$\in^{R^{1\times H\times W}}$)를 생성한다.

[0096] $$\mathbf{M_{pix}} = \mathbb{1}[\mathbf{A} > \tau_{\mathrm{d}}], \quad where \quad \tau_{\mathrm{d}} = max(\mathbf{A}) \times \theta_{\mathrm{d}}$$

[0097] τ_d는 어텐션 맵(A)의 가장 높은 값과 미리 정의된 드롭 임계값(Θ_d)의 곱으로 나타낸다.

[0098] Mpix의 각 픽셀을 S x S 제곱 영역의 크기로 확장하여 영역 드롭 마스크(M)를 생성한다. 구체적으로, (S,S)의 커널 크기를 갖는 최대 풀링 계층을 Mpix에 적용한다.

[0099] 마지막으로, 삭제 피처맵(-X)은 피처맵(X)과 마스크(M) 사이의 스페셜-와이즈 곱(spatial-wise multiplication)에 의해 생성된다. 피처맵(^{X})과 삭제 피처맵($^{\bar{X}}$) 모두 가중치를 공유하는 네트워크의 이후 레이어에 동시에 공급된다. 또한 고정된 드롭 임계값($^{\theta_d}$)이 불안정한 성능을 유발함을 관찰했다. 삭제 분기는 넓은 범위에서 가장 구별되는 부분을 버리기 때문에(즉, 지역 수준의 하락) 초기 훈련 단계에서 분류하는 데 어려움을 겪는다. 이 문제를 해결하기 위해 감소 임계값을 1에서 $^{\theta_d}$까지 선형적으로 감소시켜 훈련 시작 시 이중 분기 간의 불일치를 줄인다. 전반적으로, SRD 구성(410)은 도 5에서와 같이 삭제 영역을 점진적으로 늘리고 덜 구별되는 영역으로 활성화를 성공적으로 확장한다.

[0101] 도 5는 원본 분기와 삭제 분기의 피처맵에서 활성화 변화를 나타내는 도면이다.

[0102] 도 5에서, 학습이 진행될수록 삭제 피처맵($^{\bar{X}}$)에서 삭제영역이 넓어지고 최종 피처맵(F) 및 최종 삭제 피처맵($^{\bar{F}}$)에서 객체의 전체 영역을 더 활성화시킨다.

[0104] CG(Contrastive Guidance)

[0105] 대조 학습은 긍정적인 쌍을 끌어들이고 부정적인 쌍을 밀어냄으로써 의미 있는 표현을 학습하는 것을 목표로 한다. 마찬가지로 이 대조 학습 개념을 사용하기 위해 도 6에서와 같이 전경을 양의 쌍으로, 배경을 음의 쌍으로 구성한다.

[0107] 도 6은 전경 및 배경 피처맵을 정규 임베딩 공간에 투영시켜 다차원 벡터를 생성하는 도면이다.

[0108] 도 6에서, 최종 피처맵(^{F}, $^{\bar{F}}$)의 전경과 배경은 각각 원래 피처맵(^{X}) 및 삭제 피처맵($^{\bar{X}}$)이 있는 이중 분기에서 인코딩된다. 채널-와이즈 풀링을 통해 생성된 어텝션 맵(AF)의 강도를 임계값으로 지정하여 전경 및 배경 마스크($^{M_{fg}}$, $^{M_{bg}}$)를 생성한다. 그런 다음 각 마스크를 곱한 전경 및 배경 피처맵($^{F_{fg}}$, $^{F_{bg}}$)을 생성한다.

[0109] $$\mathbf{M_{fg}} = \mathbb{1}[\mathbf{A_F} > \tau_{\mathrm{fg}}], \quad \mathbf{M_{bg}} = \mathbb{1}[\mathbf{A_F} < \tau_{\mathrm{bg}}]$$

[0110] $$\mathbf{F}_{\mathrm{fg}}=\mathbf{F}\odot\mathbf{M}_{\mathrm{fg}},\quad \mathbf{F}_{\mathrm{bg}}=\mathbf{F}\odot\mathbf{M}_{\mathrm{bg}}$$

[0111] 여기서 τ_{fg} 및 τ_{bg}는 미리 정의된 임계값이다. 각 전경 및 배경 피처맵은 투영 헤드를 사용하여 정규화된 임베딩 공간에 투영된다. ReLU(Rectified Linear Unit) 활성화가 있는 2개의 1x1 컨볼루션 레이어로 구성되며 각각의 128차원 피처 벡터(z_{fg}, z_{bg}, $\bar{z}_{fg}$, $\bar{z}_{bg}$)를 출력한다. 공식적으로 대조 가이드 손실은 다음과 같이 제공됩니다.

[0112] $$\begin{aligned}\mathcal{L}_{cg}=\{&\max\left[\|(\mathbf{z}_{\mathrm{fg}}-\bar{\mathbf{z}}_{\mathrm{fg}})\|^2-\|(\mathbf{z}_{\mathrm{fg}}-\mathbf{z}_{\mathrm{bg}})\|^2+m,\,0\right]\\&+\max\left[\|(\bar{\mathbf{z}}_{\mathrm{fg}}-\mathbf{z}_{\mathrm{fg}})\|^2-\|(\bar{\mathbf{z}}_{\mathrm{fg}}-\bar{\mathbf{z}}_{\mathrm{bg}})\|^2+m,\,0\right]\}\end{aligned}$$

[0113] 여기서 m은 마진을 나타낸다. 손실 함수는 배경 사이의 거리를 확대하면서 z_{bg}, $\bar{z}_{fg}$의 표현 사이의 거리를 줄이도록 권장한다. 대상 객체의 전체 범위 내에서 다양한 보완 전경을 마이닝할 수 있다.

[0115] PNL(Pairwise Non-Local Block)

[0116] 본 발명에서는 PNL 구성(450)을 사용하여 최종 피처맵(F, $\bar{F}$)에서 대상 객체 영역에 관한 픽셀-별 관계를 강화한다. 대조 가이드 및 분류기에 제공되는 향상된 피처맵을 생성한다. 피처맵 $F\in R^{C\times H\times W}$는 쿼리, 키 및 밸류를 각각 나타내는 $\{q,k,v\}\in R^{C'\times H\times W}$에 3개의 1x1 컨볼루션 레이어로 투영된다. 가중치 매트릭스 $W\in R^{HW\times HW}$는 q,k의 화이트닝 내적 연산에 의해 얻은 각 픽셀 간의 유사도를 나타낸다.

[0117] $$\mathbf{W}=\sigma\left((\mathbf{q}_i-\mu_{\mathrm{q}})^T(\mathbf{k}_j-\mu_{\mathrm{k}})\right)$$

[0118] 여기서 σ는 softmax 함수이고 μ_q,μ_k는 각각 q,k의 각 픽셀 i,j의 공간 평균 값입니다. 그런 다음 향상된 피처맵 F'은 다음과 같이 생성된다.

[0119] $$\mathbf{F}'=\mathbf{F}\oplus h(\mathbf{v}\otimes\mathbf{W})$$

[0120] 여기서, h(·)는 배치 정규화가 뒤따르는 1x1 컨볼루션 레이어를 나타낸다.

[0121] PNL 구성(450)은 쿼리와 키 픽셀 간의 정규화된 차이를 최적화하여 클래스별 영역의 유사성을 고려하여 참석할 장소를 학습한다. 따라서 분류기 및 대조 가이드에 유익한 단서를 제공한다.

[0122] 이하에서, 기존 방법을 훨씬 능가하는 광범위한 실험을 통해 본 발명에서 제안한 새로운 WSOL 프레임워크(400)의 효율성을 확인한다.

[0124] 실험

[0125] 1) 실험 설정

[0126] 데이터셋(Datasets)

[0127] CUB-200-2011, ImageNet의 두 가지 벤치마크에서 제안된 방법을 평가하며, 훈련을 위해 이미지 수준 레이블만 제공된다. CUB-200-2011은 훈련 세트를 위한 5,994개의 이미지와 테스트 세트를 위한 5,794개의 이미지로 구성된 200종의 새(bird)를 포함한다. ImageNet에는 각각 훈련 및 검증 세트에 대한 120만 및 50,000개의 이미지가 포함된 1,000개의 클래스가 있다.

[0128] **평가 지표**

[0129] Top-1 localization(*Top-1 Loc*), GT-known localization(*GT-Loc*) 및 MaxBoxAccV2를 활용하여 방법을 평가한다. *Top-1 Loc*은 IoU 0.5 이상의 경계 상자를 포함하는 올바르게 분류된 이미지의 비율을 정답 값(ground truth)과 함께 나타낸다. *GT-Loc*은 IoU가 50%보다 크면 예측된 상자가 올바른 것으로 간주되는 비율을 측정한다. MaxBoxAccV2는 경계 상자를 생성하기 위한 최적의 임계값을 검색하여 세 가지 IoU 기준(0.3, 0.5, 0.7)에서 위치 식별 성능을 평균화한다.

[0130] **구현 세부 정보**

[0131] VGG16, InceptionV3, ResNet50의 세 가지 백본 네트워크로 방법을 구축한다. 모든 네트워크는 ImageNet 사전 훈련된 가중치를 로드하여 훈련을 시작한다. 본 발명의 PNL과 CG는 분류기 앞에 삽입된다. 드롭 임계값 θ_d를 CUB 데이터 세트의 경우 0.8, ImageNet 데이터 세트의 경우 0.9로 설정했다. 전경 τ_{fg} 및 배경 τ_{bg}의 임계값은 VGG16의 경우 0.9, 0.8로 설정된다. 보완 영역을 추출하기 위해 마지막 드롭 임계값과 함께 예정된 영역 드롭만 활용한다.

[0132] **2) 절제 연구**

[0133] 제안된 구성 요소에 대한 절제 연구는 CUB-200-2011 데이터셋에서 VGG16으로 수행된다.

[0134] **제안된 각 구성 요소의 효과**

[0135] 전체 대상 객체를 위치 식별하기 위해 세 가지 구성 요소를 제안한다. 아래 표 1은 프레임워크에서 개별 요소의 효율성을 보여준다.

[0136] [표 1]

[0137]

Methods	SRD	CG	PNL	MaxBoxAccV2 (%) 0.3	0.5	0.7	Avg	Top-1 Loc (%)
Ours	✓	✓	✓	**99.00**	**88.63**	**53.88**	**80.50**	**65.60**
– SRD	✗	✓	✓	98.65	86.05	46.84	77.18	64.22
– CG	✓	✗	✓	98.29	83.07	41.58	74.31	62.67
– PNL	✓	✓	✗	98.58	86.78	47.26	77.54	63.98

[0138] 대조 가이드(CG)가 없는 경우에는 전체 설정보다 MaxBoxAccV2 측면에서 6.19% 낮은 성능을 달성하고 특히 IoU 0.7에서 12.30% 저하된다. 전체 객체를 위치 식별하기 위해 네트워크에 주어진 이미지의 배경 영역에 대한 가이드를 제공하는 것이 필요한다. 삭제 피처맵 생성(SRD)은 또한 성능을 3.32% 향상시킨다. 프레임워크의 PNL을 제외하고는 성능이 2.96% 감소하며 두 요소에 비해 성능 저하가 가장 적다. 그 결과 모든 구성요소를 사용할 때 최고의 성능을 보여준다.

[0139] **SRD의 위치 및 크기**

[0140] 먼저 삭제 위치가 성능에 미치는 영향을 분석한다. 아래 표 2와 같이, conv4_3 레이어 뒤에 SRD를 삽입할 때 가장 좋은 성능을 보인다. 그러나 초기 레이어(pool2, pool3)에 위치한 SRD의 경우 성능이 약간 저하된다. 이전 연구에서 논의한 바와 같이 이전 레이어가 일반 기능을 추출하고 피처맵에서 로컬로 구별되는 부분(예: 가장자리, 모서리)을 활성화하기 때문이다.

[0141] [표 2]

[0142]

Location	MaxBoxAccV2 (%)	Top-1 Loc (%)
conv4_3	**80.50**	**65.60**
pool3	79.84	64.91
pool2	78.91	64.89

[0143] 또한, 아래 표 3에서 삭제된 영역의 블록 크기에 따른 성능을 조사하였다. 드롭 임계값을 0.8로, 블록 크기를 3으로 설정하여 최상의 성능을 보였다. 원본 피처맵에서 과도한 정보를 지우기 때문에 성능이 저하된다.

[0144] [표 3]

		block_size			
		1	3	5	7
θ_{drop}	0.8	77.5 / 64.4	**80.5 / 65.6**	77.3 / 64.1	68.2 / 55.3
	0.6	78.3 / 64.7	80.1 / 64.3	76.9 / 60.1	71.8 / 52.8
	0.4	79.3 / 64.9	78.9 / 62.3	69.8 / 52.2	56.6 / 38.8

[0145]

[0146] 본 발명의 SRD는 삭제 영역을 점차적으로 증가시키지만, 삭제된 분기는 대상 객체에 대한 충분한 단서 없이 대조 가이드 손실 및 분류 손실을 최적화하는 데 어려움을 겪을 것이라고 생각한다.

[0147] **기존 대조 손실 및 당사 CG 손실과의 비교**

[0148] 아래 표 4는 CG 손실을 기존의 대조 손실(즉, *InfoNCE* 손실)로 대체한 결과를 보여준다.

[0149] [표 4]

Methods	MaxBoxAccV2 (%)				Top-1 Loc (%)
	0.3	0.5	0.7	Avg	
Ours (w/o CG)	98.29	83.07	41.58	74.31	62.67
Ours (w InfoNCE)	98.44	86.38	48.88	77.90	63.46
Ours†	98.79	87.50	50.19	78.89	64.21
Ours	**99.00**	**88.63**	**53.88**	**80.50**	**65.60**

[0150]

[0151] 실험결과, 본 발명의 방법은 *InfoNCE* 손실을 사용하더라도 7.7%의 큰 마진으로 기존의 WSOL 성능을 여전히 능가하는 것으로 나타났다. 그러나 IoU 0.7에서 본 발명의 w/CG(마지막 행) 보다 훨씬 열등한다. 또한 대조 가이드 손실이 없는 본 발명의 성능은 IoU 0.7에서 심각하게 저하된다. 이는 본 발명의 대조 가이드 손실이 전체 객체를 잘 커버하기 위해 기존의 대조 손실보다 네트워크에 적절한 가이드를 제공함을 나타낸다. 또한 대조 학습(세 번째 행)에서 이중 분기의 효과도 검증한다. Ours쪌는 원본 피처맵의 배경만 음수 샘플로 사용한다. 삭제 피처맵의 배경을 버리면 성능이 떨어지는 것을 보여준다. 결과적으로 삭제 피처맵의 배경은 대상 객체의 경계 내에서 활성화를 확장하여 덜 구별되는 부분을 찾는 데 중요한 역할을 한다.

[0152] **3) 최신 방법과의 비교**

[0153] MaxBoxAccV2, *GT-known Loc* 및 *Top-1 Loc* 측면에서 CUB-200-2011 및 ImageNet 데이터 세트에 대한 WSOL 최신 방법과 본 발명의 방법을 비교한다.

[0154] **MaxBoxAcv2.** 아래 표 5에서 본 발명의 방법은 3개의 백본에 대한 *MaxBoxAccV2* 측면에서 CUB 및 ImageNet 데이터 세트의 다른 모든 방법보다 성능이 뛰어난다.

[0155] [표 5]

Methods	CUB-200-2011				ImageNet			
	VGG	Inc	Res	Avg	VGG	Inc	Res	Avg
CAM [38]	63.7	56.7	63.0	61.1	60.0	63.4	63.7	62.4
HaS [26]	63.7	53.4	64.7	60.6	60.6	63.7	63.4	62.6
ACoL [36]	57.4	56.2	66.5	60.0	57.4	63.7	62.3	61.2
SPG [37]	56.3	55.9	60.4	57.5	59.9	63.3	63.3	62.2
ADL [6]	66.3	58.8	58.4	61.1	59.8	61.4	63.7	61.7
CutMix [35]	62.3	57.5	62.8	60.8	59.4	63.9	63.3	62.2
InCA [14]	66.7	60.3	63.2	63.4	61.3	62.8	65.1	63.1
MinMaxCAM [29]	70.2	-	68.0	-	62.2	-	**65.7**	-
Ours	**80.5**	**75.8**	**73.3**	**76.5**	**65.3**	**64.8**	65.5	**64.7**

[0156]

[0157] 본 발명은 CUB(+13.1%)와 ImageNet(+1.6%)에서 눈에 띄는 개선을 달성했다. 특히, 본 발명의 방법은 CUB-InceptionV3의 InCA보다 15.5%, ImageNet-VGG16의 MinMaxCAM 보다 3.1% 향상되었다.

[0158] **GT-known Loc 및 Top-1 Loc.** 아래 표 6은 기존 매트릭스를 사용한 정량적 결과를 보여준다.

[0159] [표 6]

Methods	Backbone	CUB-200-2011		ImageNet	
		GT-Loc	Top-1 Loc	GT-Loc	Top-1 Loc
CAM [38]	VGG16	56.00	44.15	57.72	42.80
ACoL [36]	VGG16	54.10	45.92	62.96	45.83
ADL [6]	VGG16	75.41	52.36	-	44.92
MEIL [19]	VGG16	-	57.46	-	46.81
RCAM [2]	VGG16	80.72	61.30	61.69	44.69
GCNet [18]	VGG16	81.10	63.24	-	-
Ours	VGG16	**88.54**	**65.60**	**65.04**	**48.01**
CAM [38]	InceptionV3	55.10	43.70	62.68	46.30
SPG [37]	InceptionV3	-	46.64	64.69	48.60
DANet [33]	InceptionV3	67.70	52.52	-	47.53
RCAM [2]	GoogLeNet	65.10	51.05	62.76	47.70
GCNet [18]	InceptionV3	75.30	58.58	-	49.10
Ours	InceptionV3	**87.95**	**64.72**	**66.86**	**50.63**
CAM [38]	ResNet50	-	49.41	51.86	38.99
CutMix [35]	ResNet50	-	54.80	-	47.30
ADL [6]	ResNet50-SE	-	62.29	-	48.53
RCAM [2]	ResNet50-SE	74.51	58.39	64.40	51.96
Ours	ResNet50	**85.17**	**69.71**	**66.46**	**52.59**

[0160]

[0161] CUB 및 ImageNet 데이터셋 모두에서 본 발명의 방법은 *GT-Loc, Top-1 Loc*에 관한 최첨단 성능을 달성한다.

[0162] **4) 정성적 결과**

[0163] 도 7은 본 발명에 따른 WSOL 프레임워크의 ImageNet 및 CUB-200-2011 데이터셋에 대한 정성적 결과를 나타내는 도면으로, 실측 상자는 빨간색으로, 예측 상자는 녹색으로 표시하였다.

[0164] 도 7에서, 본 발명의 방법은 전체 객체를 올바르게 위치 식별하고 실제와 비교하여 엄격한 경계 상자를 출력한다. 훈련 단계에서 SRD 및 대조 가이드 손실을 사용하여 배경 영역을 제한한다. 따라서 본 발명의 방법은 덜 구별되는 부분으로 확산될 뿐만 아니라 배경에서 활성화를 억제한다.

[0166] 도 8은 기존 방법 대비 본 발명에 따른 약지도 객체인식 결과를 보여주는 예시도이다.

[0167] 도 8에 보여진 바와 같이, 이미지에 대한 약지도 객체인식의 기존 방법들(ACoL, EIL)은 객체의 가장 특징적인 부분의 영역만을 인식하거나 객체의 더 넓은 영역을 인식하기 위해 가장 특징적인 부분을 지운 후 학습하더라도 배경까지 인식하여 너무 넓은 부분을 인식하기 때문에 객체 인식의 정확도가 떨어진다. 반면, 본 발명의 방법(Ours)은 가장 특징적인 부분을 학습이 진행됨에 따라 점점 넓은 영역을 지우는 SRD(Scheduled Region Drop), 이중-분기에서 추출한 최종 피처맵 및 최종 삭제 피처맵에서 서로 다른 지역을 활성화하여 각 피처맵의 전경과 배경을 나누어 전경끼리는 비슷하도록 하고 배경과는 멀어지게 하여 전경과 배경의 피처를 학습하는 CG(Contrastive Guidance), CG 및 분류기에 더 향상된 피처맵을 제공하는 PNL(Pairwise Non-Local) 블록을 통해 객체의 전경 전체를 인식하고 배경을 억제하여 정확한 객체의 영역을 검출할 수 있다.

[0169] 상기에서는 본 출원의 바람직한 실시예를 참조하여 설명하였지만, 해당 기술 분야의 숙련된 통상의 기술자는 하기의 특허 청구의 범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 출원을 다양하게 수정 및 변경시킬 수 있음을 이해할 수 있을 것이다.

부호의 설명

[0171] 100: 약지도 객체인식 장치

110: 프로세서 130: 메모리

150: 사용자 입출력부 170: 네트워크 입출력부

210: 피처맵 생성부 230: 삭제 피처맵 생성부

250: 최종맵 생성부 270: 대조 가이드 결정부

400: 본 발명에서 제안한 WSOL 프레임워크

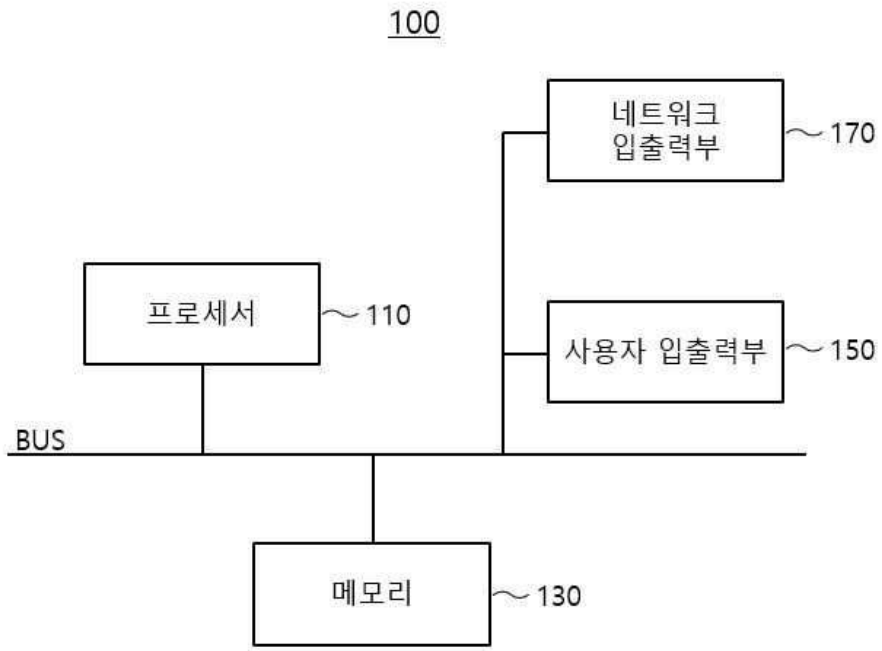
410: SRD(Scheduled region drop) 구성

430: CG(Contrastive guidance) 구성

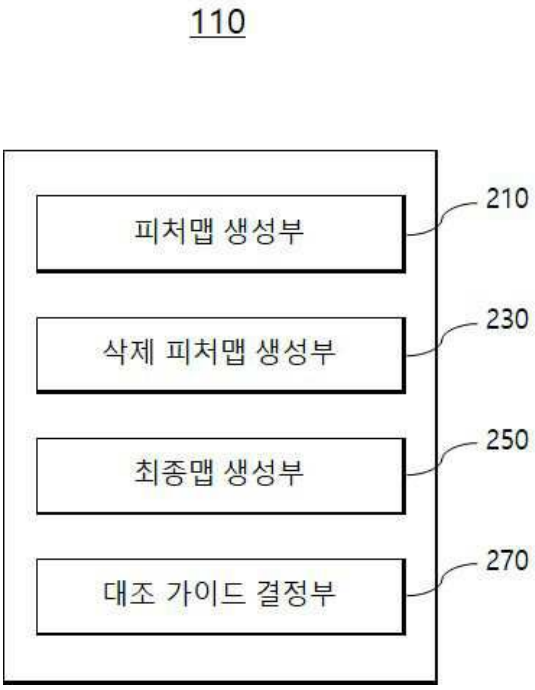
450: PNL(Pair-wise Non-Local) 구성

도면

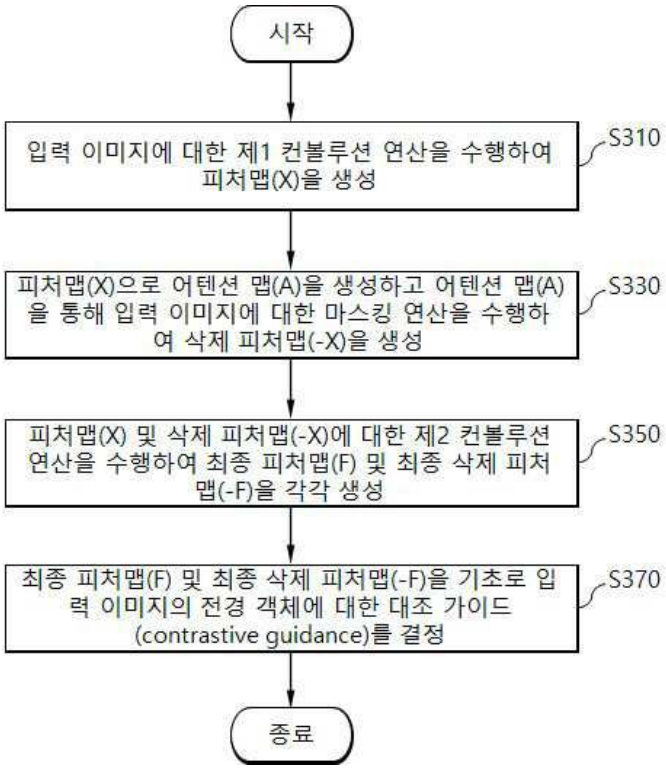
도면1



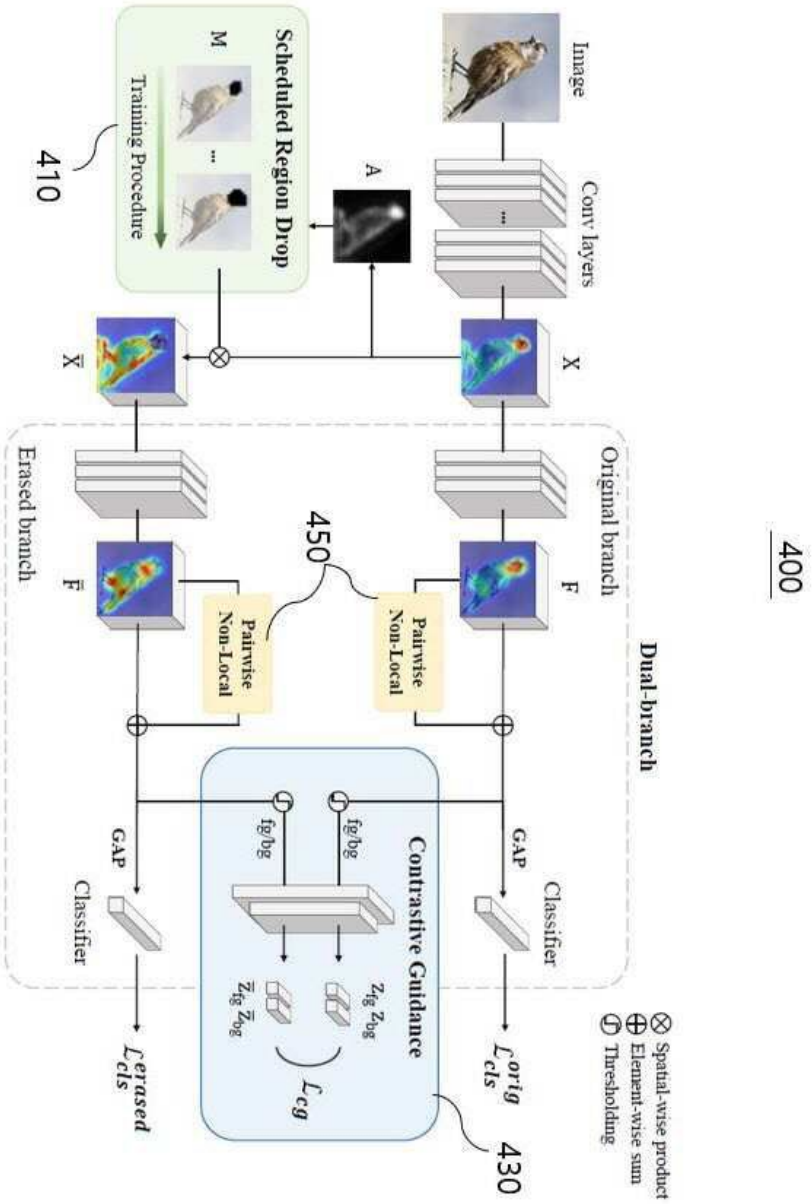
도면2


110
피처맵 생성부
210
삭제 피처맵 생성부
230
최종맵 생성부
250
대조 가이드 결정부
270

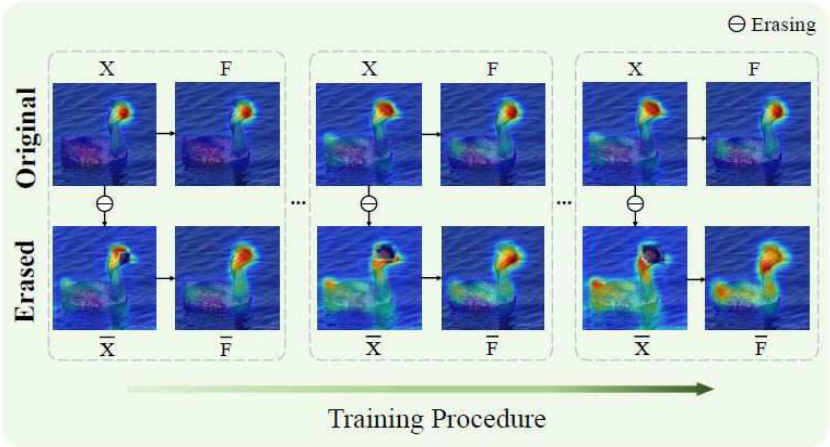
도면3


시작
입력 이미지에 대한 제1 컨볼루션 연산을 수행하여 피처맵(X)을 생성
S310
피처맵(X)으로 어텐션 맵(A)을 생성하고 어텐션 맵(A)을 통해 입력 이미지에 대한 마스킹 연산을 수행하여 삭제 피처맵(-X)을 생성
S330
피처맵(X) 및 삭제 피처맵(-X)에 대한 제2 컨볼루션 연산을 수행하여 최종 피처맵(F) 및 최종 삭제 피처맵(-F)을 각각 생성
S350
최종 피처맵(F) 및 최종 삭제 피처맵(-F)을 기초로 입력 이미지의 전경 객체에 대한 대조 가이드(contrastive guidance)를 결정
S370
종료

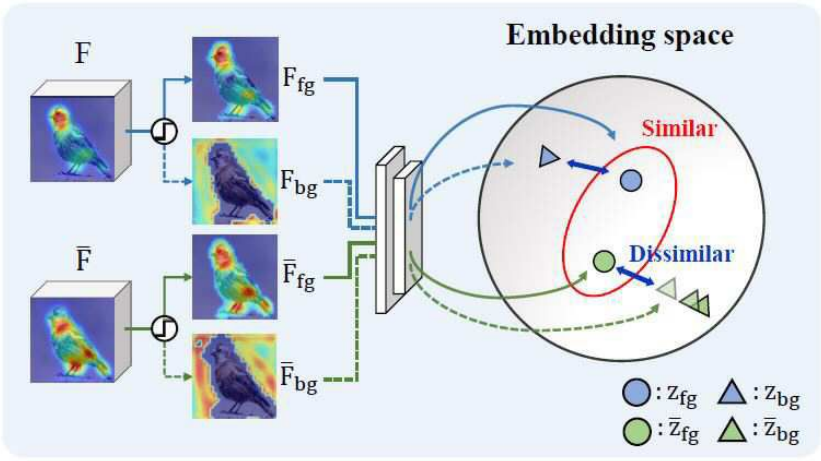
도면4


400
Image
Conv layers
X
A
Scheduled Region Drop
M
Training Procedure
410
$\bar{X}$
Dual-branch
Original branch
Erased branch
F
$\bar{F}$
450
Pairwise Non-Local
Pairwise Non-Local
GAP
GAP
Classifier
Classifier
fg/bg
fg/bg
Contrastive Guidance
z_{fg} z_{bg}
$\bar{z}_{fg}$ $\bar{z}_{bg}$
$\mathcal{L}_{cg}$
$\mathcal{L}_{cls}^{orig}$
$\mathcal{L}_{cls}^{erased}$
430
Spatial-wise product
Element-wise sum
Thresholding

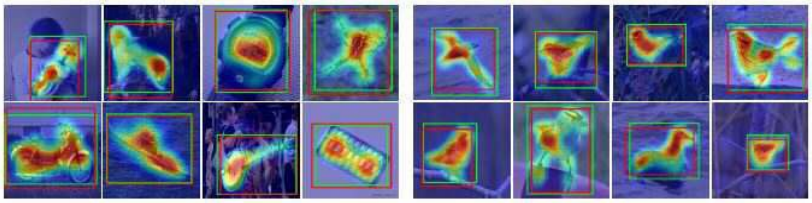
도면5


Erasing
Original
Erased
X
F
$\bar{X}$
$\bar{F}$
Training Procedure

도면6


F
F_{fg}
F_{bg}
$\bar{F}$
$\bar{F}_{fg}$
$\bar{F}_{bg}$
Embedding space
Similar
Dissimilar
: z_{fg}
: z_{bg}
: $\bar{z}_{fg}$
: $\bar{z}_{bg}$

도면7


(a) ImageNet
(b) CUB-200-2011

도면8

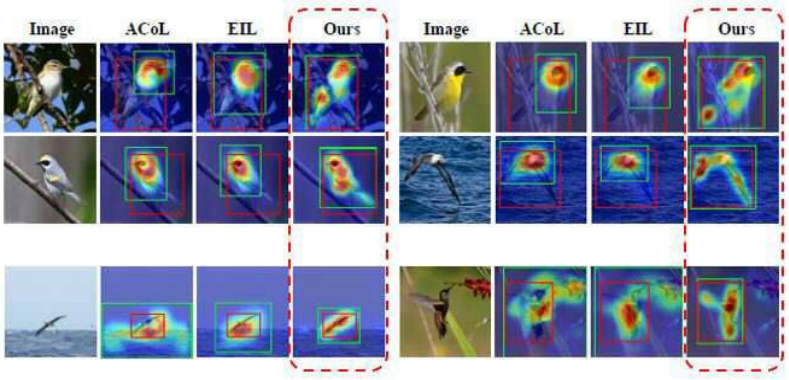


Image
ACoL
EIL
Ours
Image
ACoL
EIL
Ours