



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2023년05월26일
(11) 등록번호 10-2537947
(24) 등록일자 2023년05월24일

(51) 국제특허분류(Int. Cl.)
G06V 10/10 (2023.01) G06N 3/04 (2023.01)
G06N 3/08 (2023.01) G06V 10/40 (2022.01)
(52) CPC특허분류
G06V 10/10 (2022.01)
G06N 3/04 (2023.01)
(21) 출원번호 10-2020-0134493
(22) 출원일자 2020년10월16일
심사청구일자 2020년10월16일
(65) 공개번호 10-2022-0050617
(43) 공개일자 2022년04월25일
(56) 선행기술조사문헌
KR101523740 B1*
KR1020200054367 A*
KR1020200074940 A*
KR1020200077321 A*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
변혜란
서울특별시 서대문구 연세로 50, 제4공학관 810호
기민송
서울특별시 서대문구 연세로 50, 제4공학관 810호
(뒷면에 계속)
(74) 대리인
특허법인(유한)아이시스

전체 청구항 수 : 총 1 항

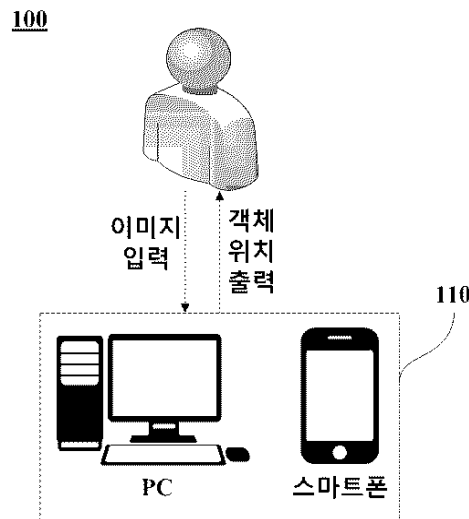
심사관 : 전한철

(54) 발명의 명칭 약한 지도학습 기반 객체 위치 검출 방법 및 장치

(57) 요약

개시된 기술은 약한 지도학습 기반 객체 위치 검출 방법 및 장치에 관한 것으로, 객체 위치 검출 장치가 딥러닝 모델의 제 1 합성곱 레이어를 이용하여 입력된 이미지에 대한 특징 맵(Feature map)을 추출하는 단계; 상기 객체 위치 검출 장치가 상기 특징 맵을 이용하여 제 1 어텐션 맵(Attention map)을 생성하는 단계; 상기 객체 위치 검출 장치가 상기 제 1 어텐션 맵으로부터 드롭된 포어그라운드 마스크(Dropped foreground Mask)를 생성하되, 상기 드롭된 포어그라운드 마스크는 컨트라스트티브 어텐션 로스(Contrastive attention loss)를 기준으로 생성되는 단계; 및 상기 객체 위치 검출 장치가 상기 드롭된 포어그라운드 마스크 또는 상기 제 1 어텐션 맵으로부터 생성된 중요 맵(Importance Map)을 기준으로 객체 위치를 추정하는 단계;를 포함한다. 따라서 객체의 위치 추정에 대한 성능이 저하되는 것을 방지하는 효과가 있다.

대표도 - 도1



(52) CPC특허분류

G06N 3/082 (2023.01)

G06V 10/40 (2023.01)

(72) 발명자

이원영

서울특별시 서대문구 연세로 50, 제4공학관 810호

고성필

서울특별시 서대문구 연세로 50, 제4공학관 810호

이 발명을 지원한 국가연구개발사업

과제고유번호	1711116905
과제번호	2020-0-01361-001
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원(한국연구재단부설)
연구사업명	정보통신방송연구개발사업
연구과제명	인공지능대학원지원(연세대학교)
기 여 율	1/2
과제수행기관명	연세대학교 산학협력단
연구기간	2021.01.01 ~ 2021.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711125843
과제번호	2018-0-00769-004
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원(한국연구재단부설)
연구사업명	정보통신방송연구개발사업
연구과제명	인공지능 시스템을 위한 뉴로모픽 컴퓨팅 SW 플랫폼 기술 개발
기 여 율	1/2
과제수행기관명	한국전자통신연구원
연구기간	2021.01.01 ~ 2021.12.31

명세서

청구범위

청구항 1

객체 위치 검출 장치가 딥러닝 모델의 제 1 합성곱 레이어를 이용하여 입력된 이미지에 대한 특징 맵(Feature map)을 추출하는 단계;

상기 객체 위치 검출 장치가 상기 특징 맵을 이용하여 제 1 어텐션 맵(Attention map)을 생성하는 단계;

상기 객체 위치 검출 장치가 상기 제 1 어텐션 맵에 일정한 임계값을 적용하여 제1 드롭된 포어그라운드 마스크(Dropped foreground Mask)를 생성하는 단계;

상기 객체 위치 검출 장치가 상기 제1 드롭된 포어그라운드 마스크와 상기 특징 맵을 곱한 값을 상기 제1 합성곱 레이어 이후의 레이어인 제2 합성곱 레이어에 입력하는 단계; 및

상기 객체 위치 검출 장치가 상기 제2 합성곱 레이어에서 출력되는 특징 맵으로부터 생성되는 제2 드롭된 포어그라운드 마스크를 기준으로 객체 위치를 추정하는 단계;를 포함하되,

상기 제1 합성곱 레이어는 상기 제1 드롭된 포어그라운드 마스크의 특징과 상기 제1 어텐션 맵에서 추출되는 포어그라운드 마스크(Foreground Mask)의 특징이 최대한 유사해지면서 동시에 상기 제1 드롭된 포어그라운드 마스크와 상기 제1 어텐션 맵에서 추출되는 백그라운드 마스크(Background Mask)의 특징이 최대한 비유사하도록 학습되고,

상기 제1 합성곱 레이어는 상기 제1 어텐션 맵과 상기 제2 합성곱 레이어에서 생성되는 상기 특징 맵에서 생성된 제2 어텐션 맵의 차이가 최대한 줄어들도록 학습되는 객체 위치 검출 방법.

청구항 2

삭제

청구항 3

삭제

청구항 4

삭제

청구항 5

삭제

청구항 6

삭제

청구항 7

삭제

청구항 8

삭제

청구항 9

삭제

청구항 10

삭제

청구항 11

삭제

청구항 12

삭제

청구항 13

삭제

청구항 14

삭제

청구항 15

삭제

청구항 16

삭제

청구항 17

삭제

청구항 18

삭제

청구항 19

삭제

발명의 설명

기술 분야

[0001] 개시된 기술은 약한 지도학습 기반 객체 위치 검출 방법 및 장치에 관한 것이다.

배경 기술

[0002] 지도학습은 합성곱 신경망에 입력되는 이미지에 포함된 객체를 식별하기 위해서 완벽한 정답에 가까운 레이블링이 필요하다. 레이블은 신경망이 객체를 정확하게 분류하기 위해 이용된다. 이와 같이 지도학습은 레이블링에 소요되는 코스트가 높아지는 단점이 존재하므로, 최근에는 비 지도학습이나 이미지 수준의 주석(annotation)만 제공하는 약한 지도학습 기반의 객체 검출 기술도 이용되고 있다.

[0003] 한편, 종래 합성곱 신경망을 이용한 객체 위치 검출 기술에서는 객체에서 가장 강하게 식별되는 영역을 중심으로 객체를 식별하였다. 이에 따라 객체 전체를 찾기보다는 일부에 치중된 식별력을 나타내었다.

[0004] 이러한 문제점을 개선하기 위하여 객체 전체를 찾을 수 있도록 중요한 영역을 가린 상태에서 학습하는 방법들이 개발되었는데 이는 학습모델이 이미지에서 상대적으로 덜 중요한 배경까지 학습하도록 유도하기 때문에 실제 객체의 크기보다 훨씬 큰 박스를 생성하게 되어서 결과적으로 객체의 위치를 식별하는데 정확도가 낮아지는 문제점이 있었다.

선행기술문헌

특허문헌

[0005] (특허문헌 0001) 한국 공개특허 제10-2020-0074940호

발명의 내용

해결하려는 과제

[0006] 개시된 기술은 이미지 내 샘플 대조를 통한 약한 지도학습을 이용하여 객체 위치를 검출하는 방법 및 장치를 제공하는데 있다.

과제의 해결 수단

[0007] 상기의 기술적 과제를 이루기 위하여 개시된 기술의 제 1 측면은 객체 위치 검출 장치가 딥러닝 모델의 제 1 합성곱 레이어를 이용하여 입력된 이미지에 대한 특징 맵(Feature map)을 추출하는 단계, 상기 객체 위치 검출 장치가 상기 특징 맵을 이용하여 제 1 어텐션 맵(Attention map)을 생성하는 단계, 상기 객체 위치 검출 장치가 상기 제 1 어텐션 맵으로부터 드롭된 포어그라운드 마스크(Dropped foreground Mask)를 생성하되, 상기 드롭된 포어그라운드 마스크는 컨트라스티브 어텐션 로스(Contrastive attention loss)를 기준으로 생성되는 단계 및 상기 객체 위치 검출 장치가 상기 드롭된 포어그라운드 마스크 또는 상기 제 1 어텐션 맵으로부터 생성된 중요 맵(Importance Map)을 기준으로 객체 위치를 추정하는 단계를 포함하는 객체 위치 검출 방법을 제공하는데 있다.

[0008] 상기의 기술적 과제를 이루기 위하여 개시된 기술의 제 2 측면은 객체가 포함된 이미지를 입력받는 입력장치, 딥러닝 모델을 저장하는 저장장치 및 상기 딥러닝 모델의 제 1 합성곱 레이어를 이용하여 상기 이미지에 대한 특징 맵(Feature map)을 추출하고, 상기 특징 맵에 어텐션 블록(Attention block)을 설정하여 제 1 어텐션 맵(Attention map)을 생성하고 상기 제 1 어텐션 맵으로부터 드롭된 포어그라운드 마스크(Dropped foreground Mask)를 생성하되, 상기 드롭된 포어그라운드 마스크는 컨트라스티브 어텐션 로스(Contrastive attention loss)를 기준으로 생성되고, 상기 드롭된 포어그라운드 마스크 또는 상기 제 1 어텐션 맵으로부터 생성된 중요 맵(Importance Map)을 기준으로 객체 위치를 추정하는 연산장치를 포함하는 객체 위치 검출 장치를 제공하는데 있다.

발명의 효과

[0009] 개시된 기술의 실시 예들은 다음의 장점들을 포함하는 효과를 가질 수 있다. 다만, 개시된 기술의 실시 예들이 이를 전부 포함하여야 한다는 의미는 아니므로, 개시된 기술의 권리범위는 이에 의하여 제한되는 것으로 이해되어서는 아니 될 것이다.

[0010] 개시된 기술의 일 실시예에 따르면 약한 지도학습 기반 객체 위치 검출 방법 및 장치는 객체의 영역을 나타내는 박스를 생성하되 과도하게 박스가 커지는 것을 방지하는 효과가 있다.

[0011] 또한, 컨트라스티브 어텐션 로스와 포어그라운드 컨시스턴시 로스를 학습에 사용하여 객체 위치 추정 성능이 저하되는 것을 방지하는 효과가 있다.

[0012] 또한, 비 지역적인 어텐션 블록을 적용하여 생성된 어텐션 맵을 기준으로 객체 위치를 추정함으로써 객체의 특정 영역에 치우쳐지지 않고 전체 영역을 커버하는 효과가 있다.

도면의 간단한 설명

[0013] 도 1은 개시된 기술의 일 실시예에 따른 약한 지도학습 기반 객체 위치 검출 시스템을 이용하여 객체 위치를 추정하는 것을 나타낸 도면이다.

도 2는 개시된 기술의 일 실시예에 따라 객체 위치를 추정하는 전체적인 흐름을 나타낸 도면이다.

도 3은 개시된 기술의 일 실시예에 따라 컨트라스티브 어텐션 로스를 계산하는 것을 나타낸 도면이다.

도 4는 개시된 기술의 일 실시예에 따른 객체 위치 검출 장치의 구성을 나타낸 도면이다.

도 5는 개시된 기술의 일 실시예에 따른 객체 위치 검출 방법에 대한 순서도이다.

도 6은 객체 위치 추정에 대한 결과를 비교하는 도면이다.

도 7은 종래의 이미지 데이터셋에 개시된 기술을 적용한 것을 나타낸 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0014] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세한 설명에 상세하게 설명하고자 한다. 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.
- [0015] 제 1, 제 2, A, B 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 해당 구성요소들은 상기 용어들에 의해 한정되지는 않으며, 단지 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제 1 구성요소는 제 2 구성요소로 명명될 수 있고, 유사하게 제 2 구성요소도 제 1 구성요소로 명명될 수 있다. 및/또는 이라는 용어는 복수의 관련된 기재된 항목들의 조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.
- [0016] 본 명세서에서 사용되는 용어에서 단수의 표현은 문맥상 명백하게 다르게 해석되지 않는 한 복수의 표현을 포함하는 것으로 이해되어야 한다. 그리고 "포함한다" 등의 용어는 실시된 특징, 개수, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함을 의미하는 것이지, 하나 또는 그 이상의 다른 특징들이나 개수, 단계 동작 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 배제하지 않는 것으로 이해되어야 한다.
- [0017] 도면에 대한 상세한 설명을 하기에 앞서, 본 명세서에서의 구성부들에 대한 구분은 각 구성부가 담당하는 주기능 별로 구분한 것에 불과함을 명확히 하고자 한다. 즉, 이하에서 설명할 2개 이상의 구성부가 하나의 구성부로 합쳐지거나 또는 하나의 구성부가 보다 세분화된 기능별로 2개 이상으로 분화되어 구비될 수도 있다.
- [0018] 그리고 이하에서 설명할 구성부 각각은 자신이 담당하는 주기능 이외에도 다른 구성부가 담당하는 기능 중 일부 또는 전부의 기능을 추가적으로 수행할 수도 있으며, 구성부 각각이 담당하는 주기능 중 일부 기능이 다른 구성부에 의해 전담되어 수행될 수도 있음은 물론이다. 따라서, 본 명세서를 통해 설명되는 각 구성부들의 존재 여부는 기능적으로 해석되어야 할 것이다.
- [0019] 도 1은 개시된 기술의 일 실시예에 따른 약한 지도학습 기반 객체 위치 검출 시스템을 이용하여 객체 위치를 추정하는 것을 나타낸 도면이다. 도 1을 참조하면 객체 위치 검출 시스템(100)은 사용자의 PC나 스마트폰과 같은 컴퓨팅 장치(110)를 이용하여 사용자가 입력하는 이미지를 수신할 수 있다. 이러한 컴퓨팅 장치는 알고리즘에 따라 연산을 수행 가능한 다른 종류의 장치를 이용할 수도 있다. 컴퓨팅 장치는 사용자의 입력을 수신하는 장치, 연산을 위한 알고리즘, 프로그램 등을 저장하는 장치, 사용자의 입력 및 저장된 알고리즘 또는 프로그램을 이용하여 연산을 수행하는 장치 및 연산 결과를 사용자에게 출력하는 장치를 포함할 수 있다. 입력장치에는 사용자가 입력하는 이미지 및 이미지에 대한 주석이 입력될 수 있다. 일 실시예로, 입력장치는 이미지 파일을 업로드할 수 있는 마우스 및 이미지 파일에 대한 주석을 입력할 수 있는 키보드를 포함할 수 있다. 이러한 입력을 수신하기 위해서 컴퓨팅 장치는 사용자를 위한 소정의 인터페이스를 출력하는 어플리케이션 또는 프로그램을 저장할 수 있다.
- [0020] 한편, 사용자는 컴퓨팅 장치(110)에 데이터를 입력할 수 있다. 사용자가 입력하는 데이터는 기본적으로 객체를 촬영한 이미지이다. 이미지는 사용자가 직접 촬영한 것을 이용할 수도 있고 별도의 데이터셋에서 제공되는 특정 이미지를 사용할 수도 있다. 입력 데이터에는 이러한 이미지와 함께 객체에 대한 주석(annotation)이 포함된다. 여기에서 주석은 전체 레이블링 데이터가 아닌 이미지 레벨의 주석(annotation)만 입력될 수 있다. 즉, 사용자는 객체가 포함된 이미지 및 객체에 대한 주석을 컴퓨팅 장치(110)에 데이터로 입력할 수 있다. 그리고 컴퓨팅 장치는 데이터를 입력받고 저장된 알고리즘에 따라 약한 지도학습을 수행하여 객체의 위치를 추정할 수 있다. 알고리즘에 따라 객체의 위치를 추정하는 것은 컴퓨팅 장치에서 연산을 담당하는 장치를 통해 수행될 수 있으며 연산 결과는 출력 장치를 통해 사용자에게 제공될 수 있다.
- [0021] 한편, 개시된 기술에서는 객체 위치를 보다 정확하게 추정하기 위해서 컨트라스티브 어텐션 로스(Contrastive attention loss)와 포어그라운드 컨시스턴시 로스(Foreground consistency loss)의 2가지 함수를 이용하여 합성곱 레이어를 훈련한다. 먼저 컨트라스티브 어텐션 로스를 계산하기 위해서 이미지에서 추출된 특징 맵(Feature map)을 토대로 어텐션 맵(Attention map)을 생성하고 어텐션 맵에 임계값을 적용하여 복수의 마스크를

생성한다. 어텐션 맵은 특징 맵에 어텐션 블록을 적용하여 생성하는 것으로, 어텐션 블록은 컴퓨팅 장치에 저장된 알고리즘에 따라 다소 다르게 설정될 수 있다. 다만, 개시된 기술에서 언급하는 어텐션 블록은 채널 풀링을 대체하기 위한 것으로 비 지역적인(non-local) 특성을 갖는다. 즉, 공간의 유사성을 고려하여 특징 맵에서 향상된 어텐션 맵을 생성하는 것이다. 특징 맵이 주어지면 어텐션 블록은 이를 채널(Channel), 가로(Height) 및 세로(Weight)로 각각 임베딩하고 처음 두 임베딩 간의 유사성에 의해 가중치가 부여된 세 번째 임베딩의 공간 합계를 출력할 수 있다. 이와 같이 채널 풀링된 결과에 의해 특징 맵으로부터 향상된 어텐션 맵이 정의될 수 있다.

[0022] 한편, 어텐션 맵을 이용하여 포어그라운드(Foreground) 마스크와 백그라운드(Background) 마스크 및 포어그라운드 마스크에서 특정 영역을 가린 형태의 드롭된 포어그라운드 마스크(Dropped foreground mask)의 3가지 마스크를 생성한다. 여기에서 특정 영역이라 함은 이미지의 배경과 전경 중 가장 차별화된 영역을 의미한다. 종래의 신경망 기반 객체 위치 검출 기술에서는 이러한 특정 영역을 기점으로 특징 맵에 대한 바운딩 박스를 생성하기 때문에 객체의 전체 영역을 커버하지 못하거나 실제 객체의 크기보다 훨씬 큰 바운딩 박스를 생성하는 경우가 발생하였다. 따라서 개시된 기술에서는 이러한 문제점을 해소하고자 가장 차별화된 영역을 가려서 생성된 Dropped foreground mask를 이용한다. 컴퓨팅 장치는 각각의 마스크와 입력된 특징을 곱하는 것으로 이미지의 특징을 임베딩할 수 있다.

[0023] 한편, 드롭된 포어그라운드 마스크는 컨트라스티브 어텐션 로스를 기준으로 생성된다. 예컨대, 제 1 어텐션 맵을 특징공간으로 임베딩하는데 있어서 드롭된 포어그라운드 마스크가 포어그라운드 마스크의 임베딩 결과와는 최대한 유사하도록, 백그라운드 마스크의 임베딩 결과와는 최대한 멀어지도록 생성될 수 있다. 드롭된 포어그라운드 마스크는 제 1 어텐션 맵에 임계값을 적용하여 배경과 전경 중 가장 차별화된 영역을 가려서 생성될 수 있으며 생성된 마스크를 통해 임베딩된 특징값이 포어그라운드 마스크와 가까운 공간에 위치할 수 있고 백그라운드 마스크의 특징값과는 다소 떨어진 공간에 위치할 수 있다.

[0024] 한편, 컴퓨팅 장치는 컨트라스티브 어텐션 로스를 복수의 합성곱 레이어 각각에 특징 맵을 입력함으로써 초기 레이어에서 상위 레이어로 순차적으로 수행될 수 있다. 가장 먼저 컨트라스티브 어텐션 로스를 계산하는 초기 레이어에서는 이미지에서 추출된 특징 맵을 그대로 입력값으로 이용하나, 이후의 상위 레이어에서는 앞선 레이어에서 생성된 Dropped foreground mask와 중요 맵(Importance map) 중 랜덤하게 선택된 하나를 특징 맵에 곱한 것을 입력값으로 이용한다. 여기에서 Dropped foreground mask는 상술한 바와 같이 어텐션 맵에 임계값(Threshold)를 적용함으로써 생성되는 것이며, 중요 맵은 어텐션 맵에 시그모이드 활성화를 적용함으로써 생성된다.

[0025] 즉, 이후의 레이어에서는 Dropped foreground mask 또는 중요 맵이 가중치로 적용된 특징 맵을 입력받고 앞선 레이어와 동일하게 컨트라스티브 어텐션 로스를 수행하는 것이다. 여기에서 Dropped foreground mask 또는 중요 맵은 pixel-wise multiplication이 된다. pixel-wise multiplication은 채널 차원에서 브로드캐스팅하는 element-wise multiplication을 의미한다. 이 경우, 이후의 레이어에는 입력값에 가중치가 적용되었으므로 앞선 레이어에서 생성한 어텐션 맵과 차이가 있는 어텐션 맵이 생성된다. 이에 따라 두 어텐션 맵 간의 차이가 적어지도록 학습하기 위해서 포어그라운드 컨시스턴시 로스를 적용한다. 설명을 위해 예시적으로 2개의 레이어에 대해서만 언급하였으나 실제로 합성곱 신경망의 레이어는 다수개가 존재하므로 첫 번째 레이어와 마지막 레이어 간의 어텐션 맵이 서로 유사해지도록 포어그라운드 컨시스턴시 로스를 계산할 수 있다.

[0026] 도 2는 개시된 기술의 일 실시예에 따라 객체 위치를 추정하는 전체적인 흐름을 나타낸 도면이다. 도 2를 참조하면 딥러닝 모델(200)에 포함된 복수개의 합성곱 레이어에서 각각 컨트라스티브 어텐션 로스(220)가 수행되며 첫 번째 레이어(201)의 어텐션 맵(210)과 마지막 레이어(202)의 어텐션 맵이 유사한 값을 갖도록 포어그라운드 컨시스턴시 로스(230)가 수행된다.

[0027] 먼저 첫 번째 레이어(201)에서 입력된 이미지에 대한 특징 맵을 추출할 수 있다. 특징 맵의 추출은 종래의 합성곱 신경망에서 수행되는 것과 유사한 방식으로 처리된다. 예컨대, 특징 맵에 비 지역적 특성을 갖는 어텐션 블록을 적용하여 어텐션 맵을 생성할 수 있다. 그리고 생성된 어텐션 맵을 기준으로 드롭된 포어그라운드 마스크를 생성하고 어텐션 맵과 드롭된 포어그라운드 마스크를 이용하여 객체 위치를 추정하는 연산을 수행하게 된다. 객체의 위치 추정에 대한 결과는 컨트라스티브 어텐션 로스를 통해 조정될 수 있다. 예컨대, 딥러닝 모델이 객체 위치를 추정하기 위해서 컨트라스티브 어텐션 로스 값이 작아지도록 학습될 수 있다.

[0028] 한편, 컨트라스티브 어텐션 로스를 기준으로 생성되는 드롭된 포어그라운드 마스크(Dropped foreground mask)는 이하의 수학식 1에 따라 계산될 수 있다.

수학식 1

$$\mathbf{M}_{\text{dfg}} = \mathbb{1}[\mathbf{A} < \theta_{\text{fg}}] \wedge \mathbb{1}[\mathbf{A} > \theta_{\text{bg}}]$$

[0029]

[0031] 여기에서 \mathbf{A} 는 어텐션 맵을 의미하며, θ_{fg} 는 사전에 정의된 전경(Foreground) 영역이고, θ_{bg} 는 사전에 정의된 배경(Background) 영역을 의미한다. 즉, 드롭된 포어그라운드 마스크는 어텐션 맵에서 가장 차별화된 영역을 제거하여 생성되며, 임베딩된 특징이 포어그라운드 마스크의 임베딩 결과와는 유사해지고, 백그라운드 마스크의 임베딩 결과는 서로 멀어지도록 생성될 수 있다.

[0032] 한편, 아래의 수학식 2를 참조하여 컨트라스티브 어텐션 로스를 계산할 수 있다.

수학식 2

$$\mathcal{L}_{\text{ca}} = \mathbb{E}_{\mathbf{x}}[(d(\mathbf{z}_{\text{dfg}}, \mathbf{z}_{\text{fg}}) - d(\mathbf{z}_{\text{dfg}}, \mathbf{z}_{\text{bg}}) + m)_{+}]$$

[0033]

[0035] 여기에서 \mathbf{z}_{dfg} 는 드롭된 포어그라운드 마스크를 통해 임베딩된 특징이고, \mathbf{z}_{fg} 는 포어그라운드 마스크를 통해 임베딩된 특징이고, \mathbf{z}_{bg} 는 백그라운드 마스크를 통해 임베딩된 특징을 의미한다. 그리고 $d(\mathbf{z}_{\text{dfg}}, \mathbf{z}_{\text{fg}})$ 는 제 1 마스크 및 제 2 마스크의 인스턴스의 배열이고, $d(\mathbf{z}_{\text{dfg}}, \mathbf{z}_{\text{bg}})$ 는 제 1 마스크 및 제 3 마스크의 인스턴스의 배열이고, m 은 마진이다. 즉, 제 1 마스크를 통해 출력된 특징과 제 2 마스크를 통해 출력된 특징의 인스턴스는 서로 유사하고, 제 1 마스크의 특징과 제 3 마스크의 특징의 인스턴스는 서로 상이하게 출력된다.

[0036] 컨트라스티브 어텐션 로스 함수는 쿼리가 동등한 인스턴스와 유사하고 쿼리가 상이한 인스턴스 간에는 유사하지 않을 때 값이 적게 도출되는 함수이다. 즉, 드롭된 포어그라운드 마스크는 컨트라스티브 어텐션 로스의 계산 결과를 작아지도록 하는 마스크를 의미한다. 컨트라스티브 어텐션 로스는 어텐션 맵에 배경을 포함하면 비 유사성으로 인한 불이익이 발생하기 때문에 경계에 도달할 때까지 어텐션 맵을 가이드할 수 있다. 따라서, 삼중항 손실(Triplet loss)와 같이 값이 샘플을 추출할 필요가 없고 값이 큰 네거티브 샘플을 관리하지 않고 제 1 마스크 내지 제 3 마스크를 통해 출력되는 특징을 토대로 포지티브 샘플과 네거티브 샘플로 간주할 수 있다.

[0037] 한편, 첫 번째 레이어는 다음 레이어의 입력값을 생성하기 위해서 특징 맵에 가중치를 적용한다. 여기에서 특징 맵에 대한 가중치로 Dropped foreground mask가 적용될 수도 있고 중요 맵이 적용될 수도 있다. 이는 딥러닝 모델에서 랜덤하게 결정된 어느 하나를 특징 맵에 곱하는 것으로 다음 레이어의 입력값을 계산할 수 있다.

[0038] 한편, 두 번째 레이어는 가중치가 적용된 특징 맵을 입력값으로 하여 첫 번째 레이어와 동일한 방식으로 연산을 수행한다. 그리고 세 번째 레이어는 두 번째 레이어에서 계산된 가중치가 적용된 특징 맵을 입력값으로 하여 마찬가지로 동일하게 연산을 수행한다. 이러한 방식에 따라 마지막 레이어(202)까지 연산을 순차적으로 진행할 수 있다. 마지막 레이어까지 연산이 종료되면 첫 번째 레이어의 어텐션 맵과 마지막 레이어의 어텐션 맵을 비교할 수 있다. 레이어의 개수에 따라 첫 번째 레이어와 마지막 레이어에서 각각 생성된 어텐션 맵의 차이가 커질 수 있다. 객체 위치 검출 장치는 두 어텐션 맵이 서로 유사해지도록 딥러닝 모델을 학습할 수 있다. 예컨대, 첫 번째 레이어의 배경 활성화(Activation)를 마지막 레이어를 기준으로 감소시킴으로써 두 어텐션 맵 간의 차이를 줄이는 포어그라운드 컨시스턴시 로스를 계산할 수 있다.

[0039] 한편, 아래의 수학식 3을 참조하여 포어그라운드 컨시스턴시 로스를 계산할 수 있다.

수학식 3

$$\mathcal{L}_{fc} = \|A_i - A_j\|_2^2$$

[0040]

[0042] 여기서 A_i 는 i번째 어텐션 맵이고, A_j 는 j번째 어텐션 맵을 의미한다. 즉, i번째 어텐션 맵이 초기 레이어에서 출력되는 어텐션 맵이고 j번째 어텐션 맵이 이후의 레이어를 통해 출력되는 어텐션 맵이다. 어텐션 맵은 대략적으로 모든 위치(Location)의 활성화를 나타낸 맵이며, 초기 레이어는 객체의 전체 범위 대신 가장자리나 모서리와 같은 국부적인 영역을 구별하는 것에서 더 활성화될 수 있다. 이러한 문제점을 해소하기 위해서 초기 레이어의 어텐션 맵과 이후 레이어의 어텐션 맵이 일관성을 유지할 수 있도록 포어그라운드 컨시스턴시 로스를 계산함으로써 두 어텐션 맵이 유사해지도록 학습할 수 있다.

[0043]

한편, 객체 위치 검출 장치는 이하의 수학식 4에 따라 딥러닝 모델의 전체 로스를 계산할 수 있다.

수학식 4

$$\mathcal{L}_{total} = \mathcal{L}_{cls} + \mathcal{L}_{ca} + \mathcal{L}_{fc}$$

[0044]

[0046] 여기서 \mathcal{L}_{cls} 는 종래의 클래스 로스를 의미하고, \mathcal{L}_{ca} 는 컨트라스티브 어텐션 로스를 의미하고, \mathcal{L}_{fc} 는 포어그라운드 컨시스턴시 로스를 의미한다. 클래스 로스는 네트워크의 마지막에 GAP(Global average pooling) 레이어를 구성함으로써 소프트맥스 출력값을 생성하고 one-hot ground truth label이 주어지면 계산할 수 있으며, 앞서 언급한 수학식 1을 통해 컨트라스티브 어텐션 로스를 계산하고 수학식 2를 통해 포어그라운드 컨시스턴시 로스를 계산할 수 있다.

[0047]

객체 위치 검출 장치는 이와 같이 컨트라스티브 어텐션 로스와 포어그라운드 컨시스턴시 로스를 계산하고 이미지 내 클래스 분류를 위한 클래스피케이션 로스(Classification loss)와 합산하여 딥러닝 모델의 전체 로스를 계산할 수 있다. 객체 위치 검출 장치는 컨트라스티브 어텐션 로스 값이 작아지도록 드롭된 포어그라운드 마스크를 생성하고, 레이어 별 어텐션 맵이 유사해지도록 포어그라운드 컨시스턴시 로스가 작아지도록 딥러닝 모델을 학습할 수 있다. 즉, 전체 로스가 작아지도록 학습과정을 반복하여 객체 전체를 커버하는 바운딩 박스를 생성하되 박스의 크기가 과도하게 커지는 것을 방지할 수 있다. 따라서, 종래의 계층적 방식으로 학습된 모델에서 성능 저하가 발생하는 것과는 다르게 객체 전체에 해당하는 영역과 유사한 박스를 형성함으로써 객체 위치 추정 성능을 높일 수 있다.

[0048]

도 3은 개시된 기술의 일 실시예에 따라 컨트라스티브 어텐션 로스를 계산하는 것을 나타낸 도면이다. 도 3을 참조하면 각 레이어에서는 어텐션 맵을 이용하여 3개의 마스크를 생성할 수 있다. 종래와 같이 이미지의 배경에 해당하는 영역을 제거하여 포어그라운드 마스크를 생성하거나 이미지의 전경에 해당하는 영역을 제거하여 백그라운드 마스크를 생성할 수 있다. 그리고 앞서 도 1 및 도 2를 통해 설명한 방식으로 드롭된 포어그라운드 마스크(310)를 생성할 수 있다. 드롭된 포어그라운드 마스크(310)는 제 1 어텐션 맵에 임계값(310a)을 적용하여 생성된다. 여기에서 임계값(310a)은 이미지의 배경과 전경 중에서 가장 차별된 특징이 나타나는 영역일 수 있다. 가령 객체가 동물인 경우에는 동물의 얼굴이 차별된 영역으로 간주될 수 있다.

[0049]

드롭된 포어그라운드 마스크(310)와 포어그라운드 마스크는 학습이 진행되면서 차별된 영역에서 객체 영역 내 덜 중요한 부분까지 학습하도록 유도될 수 있다. 그리고 백그라운드 마스크는 배경 쪽으로 차별된 영역이 발생하도록 유도될 수 있다. 이러한 과정에 따라 드롭된 포어그라운드 마스크와 포어그라운드 마스크의 임베딩 결과는 서로 유사해지도록 학습되고, 드롭된 포어그라운드 마스크와 백그라운드 마스크의 임베딩 결과는 서로 멀어지도록 학습될 수 있다. 즉, triplet loss 형태를 이용하여 학습을 진행할 수 있다.

[0050]

3개의 마스크에 각각 입력된 제 1 어텐션 맵은 특징공간(310b)으로 임베딩된다. 특징공간 내에서 드롭된 포어그라운드 마스크의 특징과 포어그라운드 마스크의 특징은 서로 최대한 유사하게 표현되고, 드롭된 포어그라운드 마스크의 특징과 백그라운드 마스크의 특징은 최대한 비유사하게 표현된다. 예컨대, 드롭된 포어그라운드 마스크

크의 특징과 포어그라운드 마스크의 특징은 특징공간 내 인접하는 위치에 표현될 수 있고, 드롭된 포어그라운드 마스크의 특징과 백그라운드 마스크의 특징은 최대한 멀리 떨어지도록 표현될 수 있다.

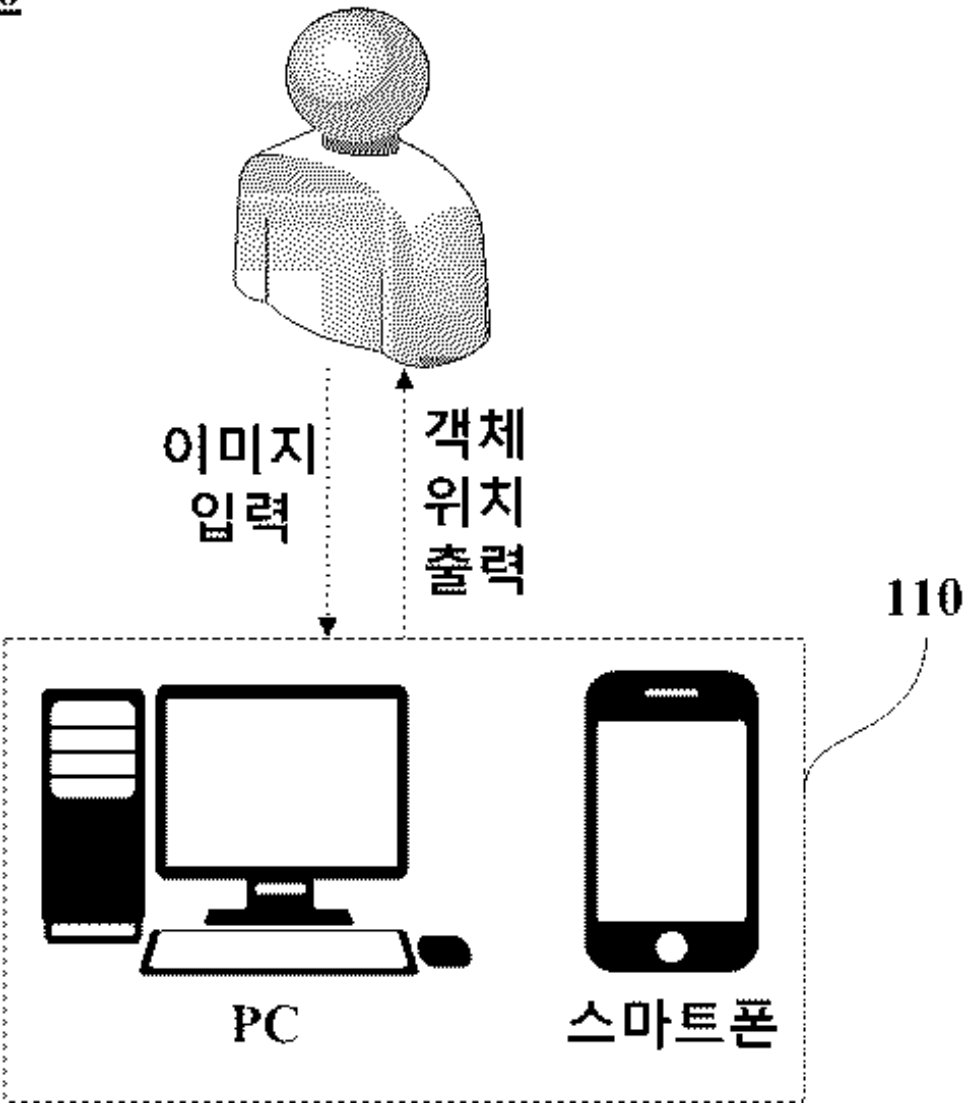
- [0051] 도 4는 개시된 기술의 일 실시예에 따른 객체 위치 검출 장치의 구성을 나타낸 도면이다. 도 4를 참조하면 객체 위치 검출 장치(400)는 입력장치(410), 저장장치(420) 및 연산장치(430)를 포함한다. 객체 위치 검출 장치(400)는 내부에 특정 계산을 수행하도록 프로그램 또는 알고리즘을 탑재하며 입력된 데이터를 토대로 계산을 수행하여 특정한 값의 데이터를 출력하는 컴퓨팅 장치일 수 있다. 예컨대, PC나 스마트폰과 같은 장치를 이용할 수 있다.
- [0052] 입력장치(410)는 객체가 포함된 이미지를 입력받는다. 입력장치는 객체 위치 검출 장치에 탑재되거나 연결될 수 있으며 사용자가 입력하는 이미지를 데이터로 전달받을 수 있다. 예컨대, 키보드 또는 마우스를 입력장치로 이용할 수 있다.
- [0053] 저장장치(420)는 이미지에 대한 객체 위치를 검출하도록 학습된 딥러닝 모델이 저장된다. 저장장치는 객체 위치 검출 장치의 메모리에 해당되며 연산장치의 제어에 따라 저장된 딥러닝 모델이 호출될 수 있다.
- [0054] 연산장치(430)는 딥러닝 모델의 제 1 합성곱 레이어를 이용하여 이미지에 대한 특징 맵(Feature map)을 추출한다. 그리고, 특징 맵에 어텐션 블록(Attention block)을 설정하여 제 1 어텐션 맵(Attention map)을 생성한다. 그리고 제 1 어텐션 맵으로부터 드롭된 포어그라운드 마스크(Dropped foreground Mask)를 생성한다. 앞서 설명한 바와 같이 드롭된 포어그라운드 마스크는 컨트라스티브 어텐션 로스(Contrastive attention loss)를 기준으로 생성되는 것이다. 연산장치(430)는 드롭된 포어그라운드 마스크 또는 제 1 어텐션 맵으로부터 생성된 중요 맵(Importance Map)을 기준으로 객체 위치를 추정한다. 연산장치는 컨트라스티브 어텐션 로스 및 포어그라운드 컨시스턴시 로스를 계산할 수 있는 성능의 CPU 또는 프로세서일 수 있다. 연산장치는 입력장치에서 전달되는 입력 데이터 및 저장장치에 저장된 딥러닝 모델을 이용하여 객체 위치 검출을 위한 연산을 수행하여 출력 데이터를 생성할 수 있다. 여기에서 입력 데이터는 객체가 포함된 이미지 및 주석이고 출력되는 데이터는 이미지에 포함된 객체의 위치 내지는 좌표일 수 있다.
- [0055] 한편, 필요에 따라 연산장치(430)에서 처리된 결과를 출력할 수 있는 출력장치를 더 구비할 수 있다. 출력장치는 연산장치에서 전달되는 객체 위치 검출 결과를 출력하는 디스플레이일 수 있다. 출력장치는 객체 위치 검출 장치와 하나로 연결된 디스플레이일 수도 있고 별도로 연결된 모니터일 수도 있다.
- [0056] 도 5는 개시된 기술의 일 실시예에 따른 객체 위치 검출 방법에 대한 순서도이다. 도 5를 참조하면 객체 위치 검출 장치는 이하의 순서에 따라 객체 위치 검출 방법을 수행할 수 있다.
- [0057] 첫번째로 510 단계에 따라 객체 위치 검출 장치가 제 1 합성곱 레이어를 이용하여 입력된 이미지에 대한 특징 맵(Feature map)을 추출한다. 제 1 합성곱 레이어는 복수의 합성곱 레이어들 중 초기의 레이어를 의미한다. 객체 위치 검출 장치는 사용자가 입력하는 이미지를 수신하여 제 1 합성곱 레이어에 입력할 수 있다. 그리고 제 1 합성곱 레이어는 이미지에 포함된 객체에 대한 특징 맵을 추출할 수 있다.
- [0058] 두 번째로 520 단계에서 객체 위치 검출 장치는 특징 맵을 이용하여 제 1 어텐션 맵(Attention map)을 생성한다. 객체 위치 검출 장치는 특징 맵에 어텐션 블록을 설정하여 제 1 합성곱 레이어에 대한 제 1 어텐션 맵을 생성한다.
- [0059] 세 번째로, 530 단계에서 객체 위치 검출 장치는 제 1 어텐션 맵을 기준으로 드롭된 포어그라운드 마스크를 생성한다. 드롭된 포어그라운드 마스크는 컨트라스티브 어텐션 로스(Contrastive attention loss)에 대한 결과값을 작아지도록 하는 가중치 역할을 수행하는 것을 의미한다. 객체 위치 검출 장치는 입력된 이미지로부터 생성한 제 1 어텐션 맵을 기준으로 3개의 마스크를 생성할 수 있다. 일 실시예로, 배경을 가려서 생성한 포어그라운드 마스크, 전경을 가려서 생성한 백그라운드 마스크 및 전경과 배경 중 가장 차별된 영역을 가려서 생성한 드롭된 포어그라운드 마스크를 생성할 수 있다. 그리고 드롭된 포어그라운드 마스크가 포어그라운드 마스크의 임베딩 결과와 가까운 공간에 위치하면서 백그라운드 마스크의 임베딩 결과와는 멀어지도록 드롭된 포어그라운드 마스크를 생성할 수 있다.
- [0060] 다음으로 540 단계에서 드롭된 포어그라운드 마스크 또는 중요 맵을 기준으로 객체의 위치를 추정한다. 객체의 위치는 객체를 중심으로 생성된 바운딩 박스의 위치를 토대로 계산될 수 있다. 다만 객체 위치 검출 장치는 객체의 특정 영역에 치우쳐지지 않고 객체 전체를 모두 포함하면서 과도하게 크게 생성되지 않도록 드롭된 포어그라운드 마스크를 이용할 수 있다.

- [0061] 한편, 이와 같이 제 1 합성곱 레이어에서의 연산이 종료되면 객체 위치 검출 장치는 550 단계에 따라 제 2 합성곱 레이어에 가중치가 적용된 객체 이미지를 입력한다. 가중치가 적용된 객체 이미지는 제 1 어텐션 맵에 Dropped foreground 마스크 및 중요맵(Importance map) 중 랜덤하게 선택된 하나를 가중치로 적용하여 생성된다.
- [0062] 560 단계에서 객체 위치 검출 장치는 제 2 합성곱 레이어를 이용하여 제 2 어텐션 맵을 생성한다. 제 2 합성곱 레이어에 가중치가 적용된 이미지를 입력하고 제 1 합성곱 레이어와 동일하게 이미지에 어텐션 블록을 적용하여 제 2 어텐션 맵을 생성할 수 있다.
- [0063] 570 단계에서 객체 위치 검출 장치는 제 1 어텐션 맵과 제 2 어텐션 맵이 서로 유사해지도록 딥러닝 모델을 학습한다. 일 실시예로, 포어그라운드 컨시스턴시 로스를 계산하여 제 1 합성곱 레이어에서 출력된 제 1 어텐션 맵과 제 2 합성곱 레이어에서 출력된 제 2 어텐션 맵이 서로 비슷한 값이되도록 학습할 수 있다.
- [0064] 도 6은 객체 위치 추정에 대한 결과를 비교하는 도면이다. 도 6을 참조하면 녹색 박스가 예측된 값이고 붉은색 박스가 실제 객체의 영역에 대한 값을 나타낸다. 종래의 ADL(Autonomos Deep Learning)과 SPG(Self Producted Guidance)과 개시된 기술의 일 실시예에 따른 객체 위치 검출 기술을 비교한 결과 ADL에서는 덜 차별적인 부분에서 더 많이 활성화하려는 시도로 인하여 녹색 박스의 크기가 과도하게 확장되는 것을 확인할 수 있다. 그리고 SPG에서는 배경을 억제하려고 하지만 객체 영역을 과도한 크기로 평가하고 있어서 일부 녹색 박스의 크기가 여전히 크게 설정되는 것을 확인할 수 있다. 반면, 개시된 객체 위치 검출 기술에서는 배경 영역으로 과도하게 확장되지 않으면서 객체의 전체 범위를 포함할 수 있도록 붉은색 박스와 유사한 형태로 녹색 박스가 형성되는 것을 확인할 수 있다.
- [0065] 도 7은 종래의 이미지 테이터셋에 개시된 기술을 적용한 것을 나타낸 도면이다. 도 7을 참조하면 붉은색으로 표시된 영역이 중요도가 높은 영역이고 녹색으로 표시된 영역이 중요도가 낮은 영역을 의미한다. 중요도가 높은 영역은 객체가 포함된 영역을 의미하며 중요도가 낮은 영역은 배경을 의미한다. 종래의 ImageNet과 CUB의 데이터셋을 통해 테스트를 수행한 결과 개시된 기술이 적용된 모델에 의해 생성된 녹색 박스와 붉은색 박스가 유사하게 형성되는 것을 확인할 수 있다. 따라서, 비 지역적인 어텐션 맵을 기준으로 객체 위치를 추정함으로써 특정 영역에 치우쳐지지 않고 전체 영역을 커버하는 효과가 있다.
- [0066] 개시된 기술의 일 실시예에 따른 약한 지도학습 기반 객체 위치 검출 방법 및 장치는 이해를 돕기 위하여 도면에 도시된 실시 예를 참고로 설명되었으나, 이는 예시적인 것에 불과하며, 당해 분야에서 통상적 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시 예가 가능하다는 점을 이해할 것이다. 따라서, 개시된 기술의 진정한 기술적 보호범위는 첨부된 특허청구범위에 의해 정해져야 할 것이다.

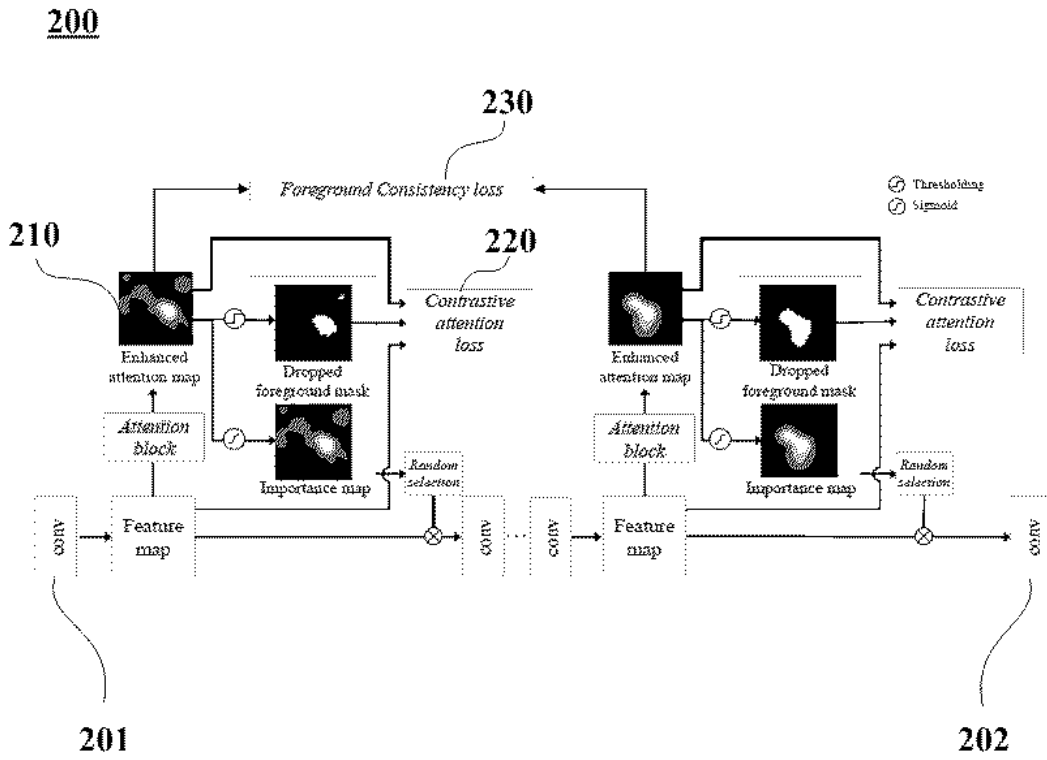
도면

도면1

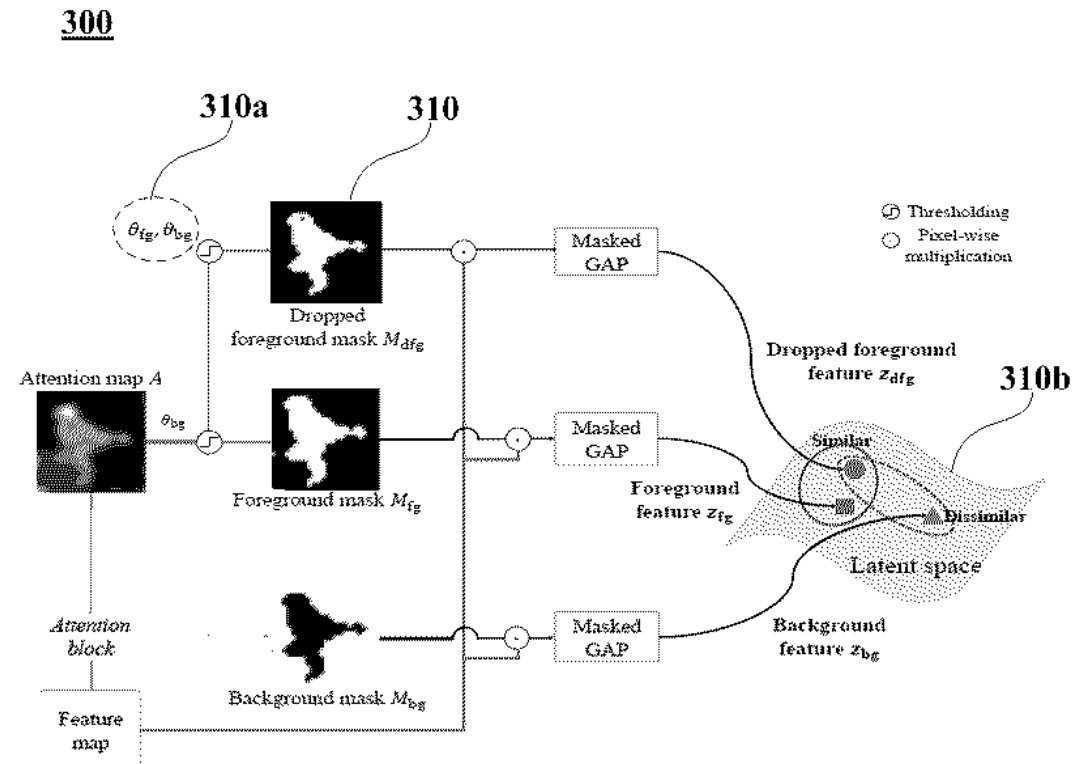
100



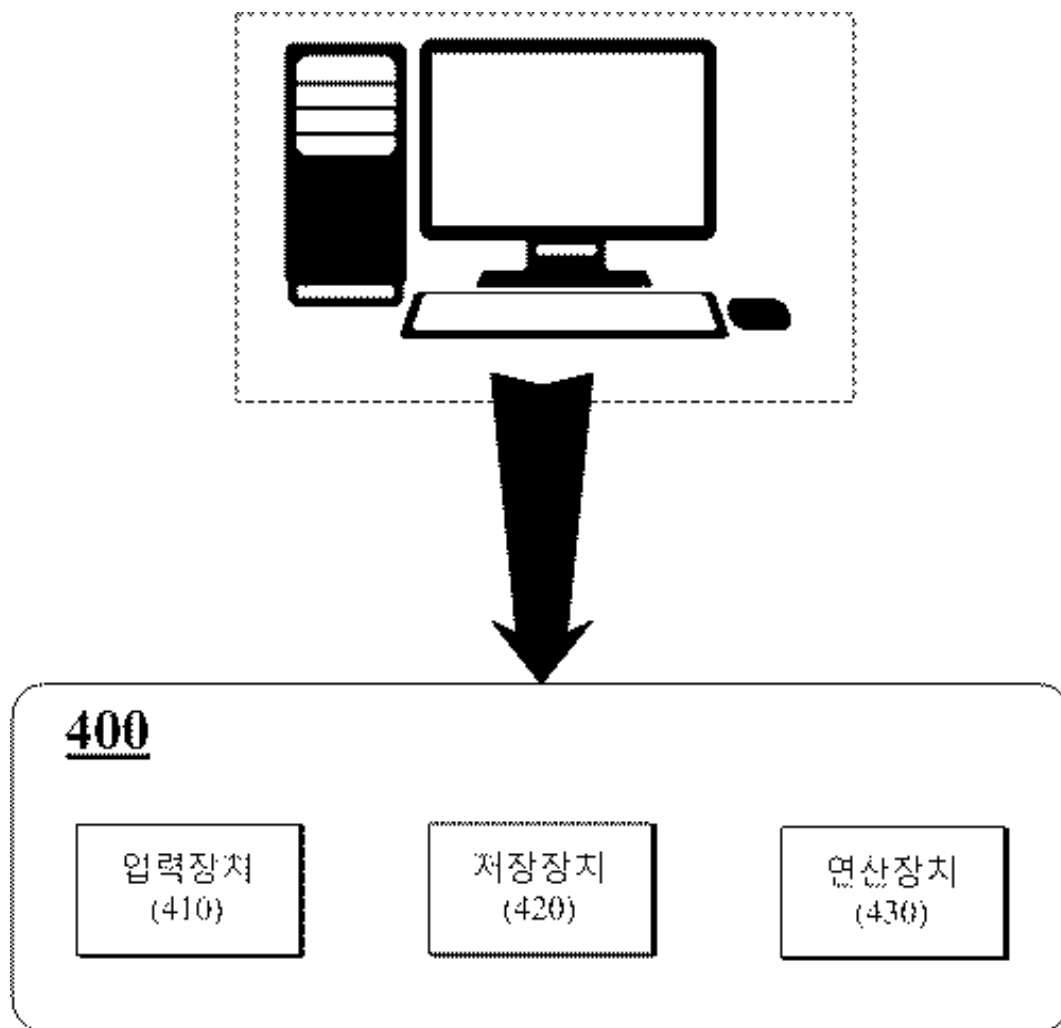
도면2



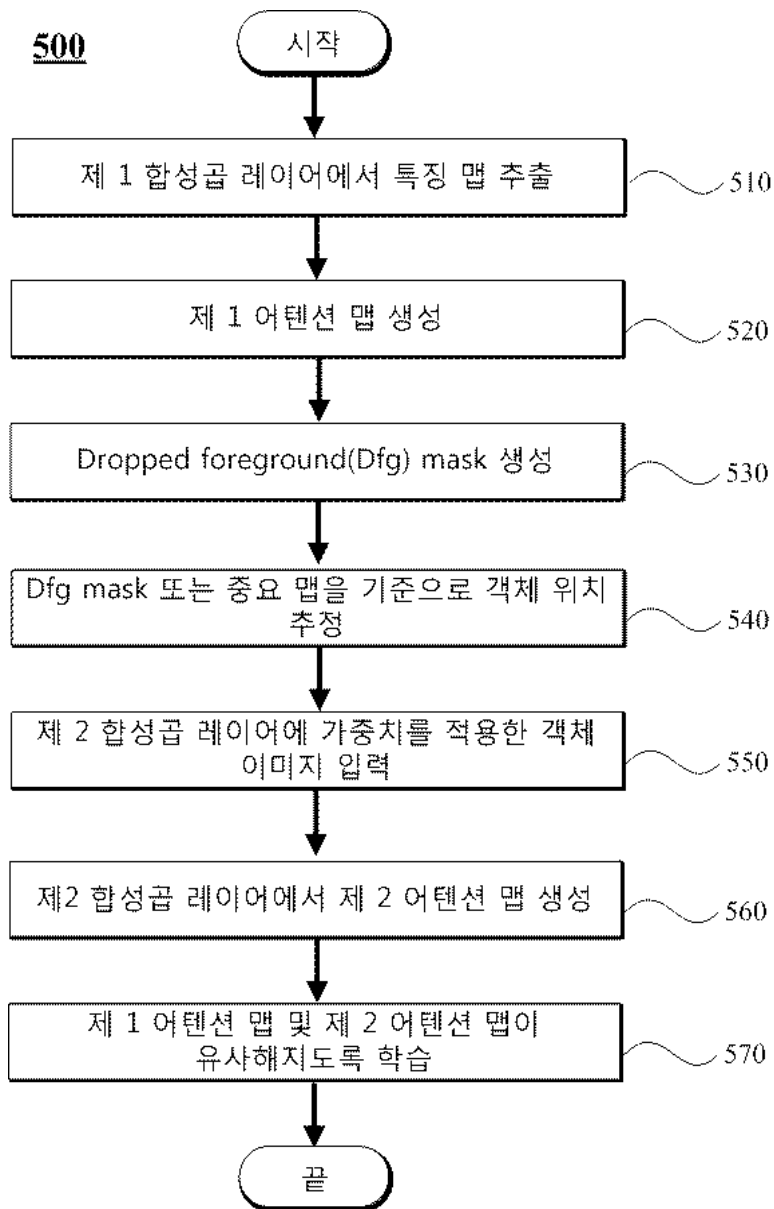
도면3



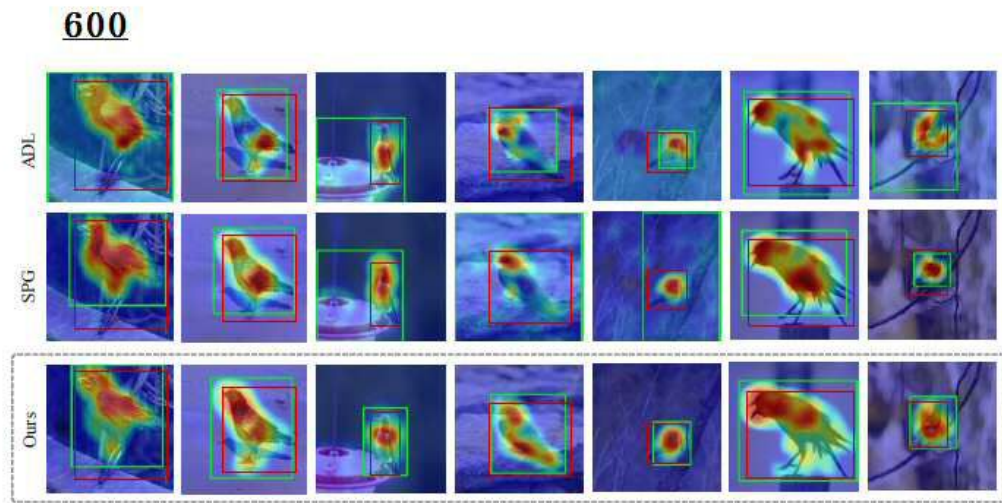
도면4



도면5



도면6



도면7

