(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0028794
(43) 공개일자 2025년03월04일

(51) 국제특허분류(Int. Cl.)
G06N 5/04 (2023.01) *G06N 20/00* (2019.01)
G06N 5/02 (2023.01)
(52) CPC특허분류
G06N 5/04 (2023.01)
G06N 20/00 (2021.08)
(21) 출원번호 10-2023-0109898
(22) 출원일자 2023년08월22일
심사청구일자 2023년08월22일

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
이상근
서울특별시 강남구 선릉로 221(도곡동, 도곡렉슬아파트)
박현태
서울특별시 강동구 상암로 11 (암사동 선사현대아파트) 108동 1601호
(74) 대리인
윤귀상

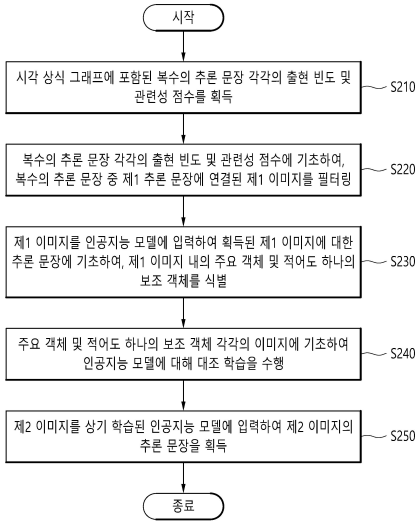
전체 청구항 수 : 총 14 항

(54) 발명의 명칭 **이미지에 대한 다양한 시각 상식 추론 데이터를 생성하는 방법 및 장치**

(57) 요 약

시각 상식 추론 데이터를 생성하는 방법 및 장치가 개시된다. 본 개시의 일 실시예에 따른, 장치에 의해 수행되는, 시각 상식 추론 데이터를 생성하는 방법은, 시각 상식 그래프에 포함된 복수의 추론 문장 각각의 출현 빈도 및 관련성 점수를 획득하는 단계; 상기 복수의 추론 문장 각각의 출현 빈도 및 관련성 점수에 기초하여, 상기 복수의 추론 문장 중 제1 추론 문장에 연결된 제1 이미지를 필터링하는 단계; 상기 제1 이미지를 인공지능 모델에 입력하여 획득된 상기 제1 이미지에 대한 추론 문장에 기초하여, 상기 제1 이미지 내의 주요 객체 및 적어도 하나의 보조 객체를 식별하는 단계; 상기 주요 객체 및 상기 적어도 하나의 보조 객체 각각의 이미지에 기초하여 상기 인공지능 모델에 대해 대조 학습(contrastive learning)을 수행하는 단계; 및 제2 이미지를 상기 학습된 인공지능 모델에 입력하여 상기 제2 이미지의 추론 문장을 획득하는 단계를 포함할 수 있다.

대 표 도 - 도2



(52) CPC특허분류
G06N 5/02 (2023.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711193973
과제번호 2019-0-00079-005
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 정보통신방송혁신인재양성
연구과제명 인공지능대학원지원(고려대학교)
기 여 율 1/2
과제수행기관명 고려대학교산학협력단
연구기간 2023.01.01 ~ 2023.12.31
이 발명을 지원한 국가연구개발사업
과제고유번호 1711188552
과제번호 2021R1A2C3010430
부처명 과학기술정보통신부
과제관리(전문)기관명 한국연구재단
연구사업명 개인기초연구(과기정통부)
연구과제명 미플랫폼: 자기지도학습과 연합학습으로 강화되는 작은 인공지능 플랫폼
기 여 율 1/2
과제수행기관명 고려대학교
연구기간 2023.03.01 ~ 2024.02.29

명 세 서

청구범위

청구항 1

장치에 의해 수행되는, 시각 상식 추론 데이터를 생성하는 방법에 있어서, 상기 방법은:

시각 상식 그래프에 포함된 복수의 추론 문장 각각의 출현 빈도 및 관련성 점수를 획득하는 단계;

상기 복수의 추론 문장 각각의 출현 빈도 및 관련성 점수에 기초하여, 상기 복수의 추론 문장 중 제1 추론 문장에 연결된 제1 이미지를 필터링하는 단계;

상기 제1 이미지를 인공지능 모델에 입력하여 획득된 상기 제1 이미지에 대한 추론 문장에 기초하여, 상기 제1 이미지 내의 주요 객체 및 적어도 하나의 보조 객체를 식별하는 단계;

상기 주요 객체 및 상기 적어도 하나의 보조 객체 각각의 이미지에 기초하여 상기 인공지능 모델에 대해 대조 학습(contrastive learning)을 수행하는 단계; 및

제2 이미지를 상기 학습된 인공지능 모델에 입력하여 상기 제2 이미지의 추론 문장을 획득하는 단계를 포함하는, 방법.

청구항 2

제1항에 있어서,

상기 제1 추론 문장은, 상기 복수의 추론 문장 중 출현 빈도가 제1 임계값을 초과하는 추론 문장인, 방법.

청구항 3

제2항에 있어서,

상기 제1 추론 문장의 출현 빈도는, 상기 제1 추론 문장에 연결된 복수의 이미지의 개수에 기초하고,

상기 제1 추론 문장의 관련성 점수는, 상기 제1 추론 문장에 연결된 복수의 이미지 간의 코사인 유사도의 평균 값인, 방법.

청구항 4

제3항에 있어서,

i) 상기 제1 추론 문장에 연결된 복수의 이미지 중 상기 제1 이미지를 제외한 나머지 이미지 및 ii) 상기 제1 이미지 간의 코사인 유사도 값이 상기 제1 추론 문장에 연결된 복수의 이미지 간의 코사인 유사도의 평균 값보다 낮은, 방법.

청구항 5

제1항에 있어서,

상기 제1 이미지에 대한 추론 문장에 기초하여, 상기 제1 이미지에 포함된 적어도 하나의 객체 별 집중 점수가 획득되고,

상기 적어도 하나의 객체 별 집중 점수에 기초하여 상기 적어도 하나의 객체를 상기 주요 객체 및 상기 적어도 하나의 보조 객체로 분류되는, 방법.

청구항 6

제5항에 있어서,

상기 인공지능 모델은, 상기 대조 학습을 통해, 상기 주요 객체 및 상기 적어도 하나의 보조 객체 간의 관계에 기초하여 상기 제1 이미지에 대한 추론 문장을 출력하도록 학습되는, 방법.

청구항 7

제1항에 있어서,

상기 제2 이미지의 추론 문장에 기초하여 상기 시각 상식 그래프가 업데이트되는, 방법.

청구항 8

시각 상식 추론 데이터를 생성하는 장치에 있어서, 상기 장치는:

메모리; 및

프로세서를 포함하고,

상기 프로세서는,

시각 상식 그래프에 포함된 복수의 추론 문장 각각의 출현 빈도 및 관련성 점수를 획득하고,

상기 복수의 추론 문장 각각의 출현 빈도 및 관련성 점수에 기초하여, 상기 복수의 추론 문장 중 제1 추론 문장에 연결된 제1 이미지를 필터링하는 추론 필터링 모듈;

상기 제1 이미지를 인공지능 모델에 입력하여 획득된 상기 제1 이미지에 대한 추론 문장에 기초하여, 상기 제1 이미지 내의 주요 객체 및 적어도 하나의 보조 객체를 식별하고,

상기 주요 객체 및 상기 적어도 하나의 보조 객체 각각의 이미지에 기초하여 상기 인공지능 모델에 대해 대조 학습(contrastive learning)을 수행하는 반사실적 대조 학습 모듈; 및

제2 이미지를 상기 학습된 인공지능 모델에 입력하여 상기 제2 이미지의 추론 문장을 획득하는 반사실적 추론 생성 모듈을 포함하는, 장치.

청구항 9

제8항에 있어서,

상기 제1 추론 문장은, 상기 복수의 추론 문장 중 출현 빈도가 제1 임계값을 초과하는 추론 문장인, 장치.

청구항 10

제9항에 있어서,

상기 제1 추론 문장의 출현 빈도는, 상기 제1 추론 문장에 연결된 복수의 이미지의 개수에 기초하고,

상기 제1 추론 문장의 관련성 점수는, 상기 제1 추론 문장에 연결된 복수의 이미지 간의 코사인 유사도의 평균 값인, 장치.

청구항 11

제10항에 있어서,

i) 상기 제1 추론 문장에 연결된 복수의 이미지 중 상기 제1 이미지를 제외한 나머지 이미지 및 ii) 상기 제1 이미지 간의 코사인 유사도 값이 상기 제1 추론 문장에 연결된 복수의 이미지 간의 코사인 유사도의 평균 값보다 낮은, 장치.

청구항 12

제8항에 있어서,

상기 제1 이미지에 대한 추론 문장에 기초하여, 상기 제1 이미지에 포함된 적어도 하나의 객체 별 집중 점수가 획득되고,

상기 적어도 하나의 객체 별 집중 점수에 기초하여 상기 적어도 하나의 객체를 상기 주요 객체 및 상기 적어도 하나의 보조 객체로 분류되는, 장치.

청구항 13

제12항에 있어서,

상기 인공지능 모델은, 상기 대조 학습을 통해, 상기 주요 객체 및 상기 적어도 하나의 보조 객체 간의 관계에 기초하여 상기 제1 이미지에 대한 추론 문장을 출력하도록 학습되는, 장치.

청구항 14

제8항에 있어서,

상기 제2 이미지의 추론 문장에 기초하여 상기 시각 상식 그래프가 업데이트되는, 장치.

발명의 설명

기 술 분 야

[0001] 본 개시는 시각-언어 인공지능 모델을 활용하여 이미지에 해당하는 상식 추론 문장을 생성하는 방법 및 장치에 관한 것이다. 보다 상세하게는 본 개시는 이미지 및 해당 이미지를 설명하는 문장을 분석하여 해당 이미지가 내포하고 있는 다양한 이야기에 관한 상식 수준의 다양한 문장을 생성하는 방법 및 장치에 관한 것이다.

배 경 기 술

[0002] 알파고의 등장 이후, 사람 수준에 근접한 인공지능 모델을 설계하고자 하는 연구가 급속도로 진행되고 있다. 일예로, 인공지능 모델이 사람이 일반적으로 지니고 있는 상식 지식을 활용할 수 있게 하는 AI2 (Allen institute for artificial intelligent) 기반의 연구가 진행되고 있다.

[0003] 특히, 언어에 담긴 상식 지식 뿐만 아니라 이미지에 담긴 상식 지식까지 활용하기 위하여, 시각 정보를 추가적으로 활용하는 연구가 진행되고 있다. 예로, 상식 지식을 명시적으로 표현할 수 있도록, Microsoft AI 등 다양한 글로벌 인공지능 연구소에서 시각 상식 인공지능 모델 연구를 수행해왔다.

[0004] 다만, 여전히 최신 모델들은 이미지의 사소한 부분에 집중하지 못한다는 한계가 존재하였으며, 이에 따라, 해당 모델들의 추론 능력은 사람 수준에 미치지 못한다는 문제가 있다.

[0005] 일반적으로 시각 상식 인공지능 모델은 이미지 및 사람이 작성한 이미지에 해당하는 상식 추론 문장을 이용해 학습을 진행하고 있다. 다만, 해당 모델은 이미지를 세부적으로 표현한 상식 추론 문장보다, 표현하지 못한 상식 추론 문장을 더욱 빈번히 학습하여, 서로 다른 이미지에 대해 동일한 문장을 생성하는 경향을 보이고 있다. 반면, 사람은 각 이미지의 사소하지만 중요한 부분을 파악하여 서로 다른 이미지에 대해 서로 다른 문장을 생성한다.

발명의 내용

해결하려는 과제

[0006] 본 개시는 상술된 문제점을 해결하기 위해 안출된 것으로서, 본 개시의 목적은 보다 사람 수준에 근접한 인공지능 모델을 설계하기 위해 이미지를 더욱 세부적으로 묘사하는 서술적인 상식 추론 문장을 생성하는 방법 및 장치를 제공함에 있다.

[0007] 본 개시의 목적들은 이상에서 언급한 목적으로 제한되지 않으며, 언급되지 않은 본 발명의 다른 목적 및 장점들은 하기의 설명에 의해서 이해될 수 있고, 본 개시의 실시예에 의해 보다 분명하게 이해될 것이다. 또한, 본 개시의 목적 및 장점들은 청구범위에 나타낸 수단 및 그 조합에 의해 실현될 수 있음을 쉽게 알 수 있을 것이다.

과제의 해결 수단

[0008] 본 개시의 일 실시예로, 장치에 의해 수행되는, 시각 상식 추론 데이터를 생성하는 방법은, 시각 상식 그래프에 포함된 복수의 추론 문장 각각의 출현 빈도 및 관련성 점수를 획득하는 단계; 상기 복수의 추론 문장 각각의 출현 빈도 및 관련성 점수에 기초하여, 상기 복수의 추론 문장 중 제1 추론 문장에 연결된 제1 이미지를 필터링하는 단계; 상기 제1 이미지를 인공지능 모델에 입력하여 획득된 상기 제1 이미지에 대한 추론 문장에 기초하여, 상기 제1 이미지 내의 주요 객체 및 적어도 하나의 보조 객체를 식별하는 단계; 상기 주요 객체 및 상기 적어도 하나의 보조 객체 각각의 이미지에 기초하여 상기 인공지능 모델에 대해 대조 학습(contrastive learning)을 수행하는 단계; 및 제2 이미지를 상기 학습된 인공지능 모델에 입력하여 상기 제2 이미지의 추론 문장을 획득하는

단계를 포함할 수 있다.

[0009] 그리고, 상기 제1 추론 문장은, 상기 복수의 추론 문장 중 출현 빈도가 제1 임계값을 초과하는 추론 문장일 수 있다.

[0010] 그리고, 상기 제1 추론 문장의 출현 빈도는, 상기 제1 추론 문장에 연결된 복수의 이미지의 개수에 기초하고, 상기 제1 추론 문장의 관련성 점수는, 상기 제1 추론 문장에 연결된 복수의 이미지 간의 코사인 유사도의 평균 값일 수 있다.

[0011] 그리고, i) 상기 제1 추론 문장에 연결된 복수의 이미지 중 상기 제1 이미지를 제외한 나머지 이미지 및 ii) 상기 제1 이미지 간의 코사인 유사도 값이 상기 제1 추론 문장에 연결된 복수의 이미지 간의 코사인 유사도의 평균 값보다 낮을 수 있다.

[0012] 그리고, 상기 제1 이미지에 대한 추론 문장에 기초하여, 상기 제1 이미지에 포함된 적어도 하나의 객체 별 집중 점수가 획득되고, 상기 적어도 하나의 객체 별 집중 점수에 기초하여 상기 적어도 하나의 객체를 상기 주요 객체 및 상기 적어도 하나의 보조 객체로 분류될 수 있다.

[0013] 그리고, 상기 인공지능 모델은, 상기 대조 학습을 통해, 상기 주요 객체 및 상기 적어도 하나의 보조 객체 간의 관계에 기초하여 상기 제1 이미지에 대한 추론 문장을 출력하도록 학습될 수 있다.

[0014] 그리고, 상기 제2 이미지의 추론 문장에 기초하여 상기 시각 상식 그래프가 업데이트될 수 있다.

[0015] 그리고, 본 개시의 일 실시예로, 시각 상식 추론 데이터를 생성하는 장치는, 메모리; 및 프로세서를 포함하고, 상기 프로세서는, 시각 상식 그래프에 포함된 복수의 추론 문장 각각의 출현 빈도 및 관련성 점수를 획득하고, 상기 복수의 추론 문장 각각의 출현 빈도 및 관련성 점수에 기초하여, 상기 복수의 추론 문장 중 제1 추론 문장에 연결된 제1 이미지를 필터링하는 추론 필터링 모듈; 상기 제1 이미지를 인공지능 모델에 입력하여 획득된 상기 제1 이미지에 대한 추론 문장에 기초하여, 상기 제1 이미지 내의 주요 객체 및 적어도 하나의 보조 객체를 식별하고, 상기 주요 객체 및 상기 적어도 하나의 보조 객체 각각의 이미지에 기초하여 상기 인공지능 모델에 대해 대조 학습(contrastive learning)을 수행하는 반사실적 대조 학습 모듈; 및 제2 이미지를 상기 학습된 인공지능 모델에 입력하여 상기 제2 이미지의 추론 문장을 획득하는 반사실적 추론 생성 모듈을 포함할 수 있다.

[0016] 상기 과제의 해결 수단은, 본 개시의 특징을 모두 열거한 것은 아니다. 본 개시의 다양한 특징과 그에 따른 장점과 효과는 아래의 구체적인 실시예를 참조하여 보다 상세하게 이해될 수 있을 것이다.

발명의 효과

[0017] 본 개시의 다양한 실시예에 의해, 인공지능 모델은 입력 이미지의 세부적인 부분에 집중함으로써 입력 이미지 및 다른 이미지들 간의 차이점을 보다 명확히 나타내는 추론 문장을 출력하도록 학습될 수 있다.

[0018] 또한, 본 개시의 다양한 실시예에 의해, 인공지능 모델은 향상된 이미지 이해를 바탕으로 사람 수준에 근접한 추론 동작을 수행하도록 학습될 수 있다.

[0019] 또한, 본 개시의 다양한 실시예에 의해 학습된 인공지능 모델이 자율주행 모듈 등에 부착되었을 때, 기존 시스템 대비 보다 향상된 성능이 제공될 수 있다.

[0020] 본 개시의 효과들은 이상에서 언급된 효과로 제한되지 않으며, 언급되지 않은 또 다른 효과들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[0021] 도 1a는 시각 상식 추론 데이터를 생성하는 장치의 구성을 도시한 블록도이다.

도 1b는 본 개시에 적용될 수 있는 시각 상식 그래프를 예시한다.

도 2는 본 개시의 일 실시예에 따른, 이미지에 대한 다양한 시각 상식 추론 데이터를 생성하는 방법을 설명하기 위한 순서도이다.

도 3은 본 개시의 일 실시예에 따른, 이미지에 대한 다양한 시각 상식 추론 데이터를 생성하는 방법을 설명하기 위한 블록도이다.

도 4a, 도 4b, 및 도 4c는 본 개시의 일 실시예에 따른, 이미지에 대한 다양한 시각 상식 추론 데이터를 생성하

는 방법을 설명하기 위한 도면이다.

도 5는 본 개시의 일 실시예에 따른, 이미지에 대해 생성된 다양한 시각 상식 추론 데이터를 예시한다.

발명을 실시하기 위한 구체적인 내용

[0022] 본 개시의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나, 본 개시는 이하에서 개시되는 실시예들에 제한되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있으며, 단지 본 실시예들은 본 개시의 개시가 완전하도록 하고, 본 개시가 속하는 기술 분야의 통상의 기술자에게 본 개시의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 개시는 청구항의 범주에 의해 정의될 뿐이다.

[0023] 본 명세서에서 사용된 용어는 실시예들을 설명하기 위한 것이며 본 개시를 제한하고자 하는 것은 아니다. 본 명세서에서, 단수형은 문구에서 특별히 언급하지 않는 한 복수형도 포함한다. 명세서에서 사용되는 "포함한다(comprises)" 및/또는 "포함하는(comprising)"은 언급된 구성요소 외에 하나 이상의 다른 구성요소의 존재 또는 추가를 배제하지 않는다.

[0024] 명세서 전체에 걸쳐 동일한 도면 부호는 동일한 구성 요소를 지칭하며, "및/또는"은 언급된 구성요소들의 각각 및 하나 이상의 모든 조합을 포함한다. 비록 "제1", "제2" 등이 다양한 구성요소들을 서술하기 위해서 사용되나, 이들 구성요소들은 이들 용어에 의해 제한되지 않음은 물론이다. 이들 용어들은 단지 하나의 구성요소를 다른 구성요소와 구별하기 위하여 사용하는 것이다. 따라서, 이하에서 언급되는 제1 구성요소는 본 개시의 기술적 사상 내에서 제2 구성요소일 수도 있음은 물론이다.

[0025] 다른 정의가 없다면, 본 명세서에서 사용되는 모든 용어(기술 및 과학적 용어를 포함)는 본 개시가 속하는 기술 분야의 통상의 기술자에게 공통적으로 이해될 수 있는 의미로 사용될 수 있을 것이다. 또한, 일반적으로 사용되는 사전에 정의되어 있는 용어들은 명백하게 특별히 정의되어 있지 않는 한 이상적으로 또는 과도하게 해석되지 않는다.

[0026] 공간적으로 상대적인 용어인 "아래(below)", "아래(beneath)", "하부(lower)", "위(above)", "상부(upper)" 등은 도면에 도시되어 있는 바와 같이 하나의 구성요소와 다른 구성요소과의 상관관계를 용이하게 기술하기 위해 사용될 수 있다. 공간적으로 상대적인 용어는 도면에 도시되어 있는 방향에 더하여 사용시 또는 동작 시 구성요소들의 서로 다른 방향을 포함하는 용어로 이해되어야 한다.

[0027] 예를 들어, 도면에 도시되어 있는 구성요소를 뒤집을 경우, 다른 구성요소의 "아래(below)"또는 "아래(beneath)"로 기술된 구성요소는 다른 구성요소의 "위(above)"에 놓여질 수 있다. 따라서, 예시적인 용어인 "아래"는 아래와 위의 방향을 모두 포함할 수 있다. 구성요소는 다른 방향으로도 배향될 수 있으며, 이에 따라 공간적으로 상대적인 용어들은 배향에 따라 해석될 수 있다.

[0028] 이하에서는 이미지에 대한 다양한 시각 상식 추론 데이터를 생성하는 방법 및 장치에 대해 구체적으로 설명하도록 한다.

[0030] 상술한 바와 같이, 이미지에 기초하여 상식 문장을 출력하도록 학습된 기존의 인공지능 모델은 이미지의 세부적인 부분을 명확히 검출 및 식별하지 못한다는 한계가 존재하였다. 이에 따라, 기존의 인공지능 모델은 여러 이미지들의 세부적인 차이를 인식하지 못하고 동일한 문장 데이터(즉, 동일한 추론 결과)를 출력하는 경우가 빈번하였다.

[0031] 예를 들어, 기존의 인공지능 모델은 두 사람이 단순히 얘기하고 있는 이미지 및 두 사람이 운동하고 있는 이미지를 구분하지 못하고, "talk to person" 추론 문장을 생성하도록 학습될 수 있다.

[0032] 이는 기존의 인공지능 모델이 세부적인 차이의 구분 없이 포괄적인 추론 결과 데이터를 출력하도록 더욱 빈번히 학습했기 때문이다. 이를 해결하기 위하여, 이하에서는 인공지능 모델이 이미지의 세부적인 부분을 묘사하고 있는 부분에 보다 집중하도록 인공지능 모델을 학습시키는 방법 등을 설명하도록 한다.

[0034] 본 개시의 일 실시예로, 도 1a는 시각 상식 추론 데이터를 생성하는 장치의 구성을 도시한 블록도이다. 도 1a에서 설명하는 장치는 스마트 폰, 태블릿 PC, 웨어러블 장치, 노트북 등을 포함할 수 있으나 이에 제한되는 것은

아니다.

[0035] 도 1a에 도시된 바와 같이, 장치(10)는 메모리(11), 통신 모듈(12), 디스플레이(13), 및 프로세서(14)를 포함할 수 있다. 다만, 이에 국한되는 것은 아니며, 장치(10)는 필요한 동작에 따라 당업자 관점에서 자명한 범위 내에서 소프트웨어 및 하드웨어 구성이 수정/추가/생략될 수 있다.

[0036] 메모리(11)는 본 장치(10)의 다양한 기능을 지원하는 데이터와, 제어부의 동작을 위한 프로그램을 저장할 수 있고, 입/출력되는 데이터들을 저장할 있고, 본 장치에서 구동되는 다수의 응용 프로그램(application program 또는 애플리케이션(application)), 본 장치의 동작을 위한 데이터들, 명령어들을 저장할 수 있다. 이러한 응용 프로그램 중 적어도 일부는, 무선 통신을 통해 외부 장치로부터 다운로드 될 수 있다.

[0037] 이러한, 메모리(11)는 플래시 메모리 타입(flash memory type), 하드디스크 타입(hard disk type), SSD 타입(Solid State Disk type), SDD 타입(Silicon Disk Drive type), 멀티미디어 카드 마이크로 타입(multimedia card micro type), 카드 타입의 메모리(예를 들어 SD 또는 XD 메모리 등), 램(random access memory; RAM), SRAM(static random access memory), 롬(read-only memory; ROM), EEPROM(electrically erasable programmable read-only memory), PROM(programmable read-only memory), 자기 메모리, 자기 디스크 및 광디스크 중 적어도 하나의 타입의 저장매체를 포함할 수 있다. 또한, 메모리(11)는 본 장치와는 분리되어 있으나, 유선 또는 무선으로 연결된 데이터베이스가 될 수도 있다.

[0038] 통신 모듈(12)은 외부 장치(예로, 복수의 노드에 대응되는 복수의 센서)와 통신을 가능하게 하는 하나 이상의 구성 요소를 포함할 수 있다. 예를 들어, 통신 모듈(12)은 무선통신 모듈, 유선통신 모듈 또는 위치정보 모듈 중 적어도 하나를 포함할 수 있다.

[0039] 여기서, 무선 통신 모듈은 와이파이(Wi-Fi) 모듈, 와이브로(Wireless broadband) 모듈 외에도, GSM(global System for Mobile Communication), CDMA(Code Division Multiple Access), WCDMA(Wideband Code Division Multiple Access), UMTS(universal mobile telecommunications system), TDMA(Time Division Multiple Access), LTE(Long Term Evolution), 4G, 5G, 6G 등 다양한 무선 통신 방식을 지원하는 무선 통신 모듈을 포함할 수 있다.

[0040] 디스플레이(13)는 본 장치(10)에서 처리되는 정보(예를 들어, 복수의 센서로부터 수신된 다변량 시계열 데이터 등)를 표시(출력)한다. 예를 들어, 디스플레이는 본 장치(10)에서 구동되는 응용 프로그램(일 예로, 어플리케이션)의 실행화면 정보, 또는 이러한 실행화면 정보에 따른 UI(User Interface), GUI(Graphic User Interface) 정보를 표시할 수 있다.

[0041] 프로세서(14)는 본 장치(10) 내의 구성요소들의 동작을 제어하기 위한 알고리즘 또는 알고리즘을 재현한 프로그램에 대한 데이터를 저장하는 메모리, 및 메모리에 저장된 데이터를 이용하여 전술한 동작을 수행하는 적어도 하나의 프로세서(미도시)로 구현될 수 있다. 이때, 메모리와 프로세서는 각각 별개의 칩으로 구현될 수 있다. 또는, 메모리와 프로세서는 단일 칩으로 구현될 수도 있다.

[0042] 즉, 프로세서(14)는 본 장치(10) 내의 구성 요소들을 제어함으로써 장치의 전반적인 동작 및 기능을 제어할 수 있다. 프로세서는 이하의 도면에서 설명되는 본 개시에 따른 다양한 실시 예들을 본 장치 상에서 구현하기 위하여, 위에서 살펴본 구성요소들을 중 어느 하나 또는 복수를 조합하여 제어할 수 있다.

[0043] 프로세서(14)는 추론 필터링 모듈(14-1), 반사실적 대조 학습 모듈(14-2), 및 반사실적 추론 생성 모듈(14-3)을 포함할 수 있다. 추론 필터링 모듈(14-1)은 일정 기준에 따라 시각 상식 그래프에 포함된 적어도 하나의 추론 문장에 연결된 이미지 중 특정 이미지를 필터링(즉, 적어도 하나의 추론 문장과 특정 이미지 간의 연결을 끊는 동작)을 수행할 수 있다.

[0044] 여기서, 적어도 하나의 추론 문장에 특정 이미지가 연결되어 있다는 것은 특정 이미지의 시각적 특성에 기초한 추론 문장이 해당 특정 이미지에 라벨링(labeling)되어 있음을 의미할 수 있다.

[0045] 일 예로, 도 1b에 도시된 바와 같이, 특정 이미지(20)에 복수의 추론 문장이 연결되어 있을 수 있다. 시각 상식 그래프는 복수의 이미지 각각에 복수의 추론 문장을 연결한 그래프를 의미할 수 있다.

[0046] 필터링 모듈(14-1)은 시각 상식 그래프에 포함된 복수의 추론 문장 각각의 출현 빈도 및 관련성 점수를 획득할 수 있다. 복수의 추론 문장 각각의 출현 빈도는 복수의 추론 문장에 연결된 복수의 이미지의 개수에 기초할 수 있다. 복수의 추론 문장 각각의 관련성 점수는 복수의 추론 문장 각각에 연결된 복수의 이미지 간의 코사인 유

사도의 평균 값일 수 있다.

[0047] 추론 필터링 모듈(14-1)은 복수의 추론 문장 각각의 출현 빈도 및 관련성 점수에 기초하여 복수의 추론 문장에 연결된 적어도 하나의 이미지를 필터링할 수 있다.

[0048] 반사실적 대조 학습 모듈(14-2)은 특정 이미지를 인공지능 모델에 입력하여 획득된 특정 이미지에 대한 추론 문장에 기초하여, 특정 이미지 내의 주요 객체 및 적어도 하나의 보조 객체를 식별할 수 있다.

[0049] 반사실적 대조 학습 모듈(14-2)은 주요 객체의 이미지(즉, 사실 이미지) 및 적어도 하나의 보조 객체 이미지(즉, 반사실 이미지)에 기초하여 인공지능 모델에 대해 대조 학습(contrastive learning)을 수행할 수 있다. 즉, 반사실적 대조 학습 모듈(14-2)은 인공지능 모델이 사실 이미지 및 적어도 하나의 반사실 이미지의 관계를 학습하도록 제어할 수 있다.

[0050] 반사실적 대조 추론 모듈(14-3)은 (시각 상식 그래프에 포함된) 이미지를 학습된 인공지능 모델에 입력하여 해당 이미지의 추론 문장을 획득할 수 있다. 학습된 인공지능 모델은 이미지에 포함된 시각적 세부 사항에 집중할 수 있다. 따라서, 반사실적 대조 추론 모듈(14-3)은 학습된 인공지능 모델을 통해 해당 이미지에 포함된 객체 간의 관계를 잘 반영한 추론 문장을 획득할 수 있다.

[0052] 도 2는 본 개시의 일 실시예에 따른, 이미지에 대한 다양한 시각 상식 추론 데이터를 생성하는 방법을 설명하기 위한 순서도이다.

[0053] 장치는 시각 상식 그래프에 포함된 복수의 추론 문장 각각의 출현 빈도 및 관련성 점수를 획득할 수 있다(S210).

[0054] 복수의 추론 문장 각각의 출현 빈도는 복수의 추론 문장에 연결된 복수의 이미지의 개수에 기초할 수 있다. 복수의 추론 문장 각각의 관련성 점수는 복수의 추론 문장 각각에 연결된 복수의 이미지 간의 코사인 유사도의 평균 값일 수 있다.

[0055] 장치는 복수의 추론 문장 각각의 출현 빈도 및 관련성 점수에 기초하여, 복수의 추론 문장 중 제1 추론 문장에 연결된 제1 이미지를 필터링할 수 있다(S220).

[0056] 이 때, 제1 추론 문장은 복수의 추론 문장 중 출현 빈도가 제1 임계값을 초과하는 추론 문장일 수 있다. 추론 문장의 출현 빈도가 클수록, 해당 추론 문장이 다양한 이미지에 연결되어 있음을 의미할 수 있다. 장치는 출현 빈도가 제1 임계값을 초과하는 추론 문장을 필터링의 대상이 되는 추론 문장으로 식별할 수 있다.

[0057] 상술한 바와 같이, 제1 추론 문장의 출현 빈도는 제1 추론 문장에 연결된 복수의 이미지의 개수에 기초할 수 있다. 제1 추론 문장의 관련성 점수는 제1 추론 문장에 연결된 복수의 이미지 간의 코사인 유사도의 평균 값일 수 있다.

[0058] 제1 추론 문장에 연결된 복수의 이미지 중 제1 이미지를 제외한 나머지 이미지 및 ii) 제1 이미지 간의 코사인 유사도 값이 제1 추론 문장에 연결된 복수의 이미지 간의 코사인 유사도의 평균 값보다 낮을 경우, 장치는 제1 추론 문장에 연결된 제1 이미지를 필터링할 수 있다.

[0059] 제1 이미지를 인공지능 모델에 입력하여 획득된 제1 이미지에 대한 추론 문장에 기초하여, 장치는 제1 이미지 내의 주요 객체 및 적어도 하나의 보조 객체를 식별할 수 있다(S230).

[0060] 구체적으로, 장치는 제1 이미지에 대한 추론 문장에 기초하여 제1 이미지에 포함된 적어도 하나의 객체 별 집중 점수를 획득할 수 있다. 장치는 적어도 하나의 객체 별 집중 점수에 기초하여 적어도 하나의 객체를 주요 객체 및 적어도 하나의 보조 객체로 분류할 수 있다.

[0061] 일 예로, 제1 이미지에 대한 추론 문장 중 제1 일정 비율 이상의 문장이 제1 객체에 관한 것인 경우, 장치는 제1 객체를 주요 객체로 식별할 수 있다. 제1 이미지에 대한 추론 문장 중 제2 일정 비율 이하의 문장이 제2 객체에 관한 것인 경우, 장치는 제2 객체를 보조 객체로 식별할 수 있다.

[0062] 장치는 주요 객체 및 보조 객체 각각의 이미지에 기초하여 인공지능 모델에 대해 대조 학습을 수행할 수 있다(S240).

[0063] 장치는 주요 객체의 이미지(즉, 사실 이미지) 및 적어도 하나의 보조 객체의 이미지(즉, 반사실 이미지)에 대한 대조 학습을 수행할 수 있도록 인공지능 모델을 제어할 수 있다. 인공지능 모델은 대조 학습을 통해 주요 객체

및 적어도 하나의 부가 객체 간의 관계에 기초하여 제1 이미지에 대한 추론 문장을 출력하도록 학습될 수 있다.

[0064] 장치는 제2 이미지를 학습된 인공지능 모델에 입력하여 제2 이미지의 추론 문장을 획득할 수 있다(S250).

[0065] 학습된 인공지능 모델은 이미지에 포함된 시각적 세부 사항에 집중할 수 있다. 따라서, 장치는 학습된 인공지능 모델에 제2 이미지를 입력함으로써 제2 이미지에 포함된 객체 간의 관계를 잘 반영한 추론 문장을 획득할 수 있다.

[0066] 장치는 학습된 인공지능 모델을 통해 획득된 추론 문장을 이용하여 시각 상식 그래프를 업데이트할 수 있다.

[0068] 도 3은 본 개시의 일 실시예에 따른, 이미지에 대한 다양한 시각 상식 추론 데이터를 생성하는 방법을 설명하기 위한 블록도이다.

[0069] 본 개시는 인공지능 모델이 학습 과정에서 이미지의 세부적인 부분에 집중할 수 있도록 하는 방법에 관한 것이다. 인공지능 모델이 이미지의 세부적인 부분을 집중할 수 있도록 하는 방법은 포괄적인 추론 필터링 단계, 반사실적 대조 학습 단계 및 반사실적 추론 생성 단계를 포함할 수 있다.

[0070] 즉, 장치는 추론 필터링 모듈(14-1), 반사실적 대조학습 모듈(14-2), 및 반사실적 추론 생성 모듈(14-3)을 통해 인공지능 모델이 이미지의 세부적인 부분을 집중할 수 있도록 제어할 수 있다.

[0071] 구체적으로, 포괄적인 추론 필터링 단계에서, 장치는 추론 필터링 모듈(14-1)을 통해 시각 상식 그래프(100)에 포함되어 있는 포괄적인 추론 데이터(예로, 추론 문장) 중 이미지와 관련이 상대적으로 적은 추론 데이터를 필터링할 수 있다. 이에 따라, 인공지능 모델의 이미지 이해 및 이미지 기반 학습을 방해할 수 있는 요인을 제거할 수 있다.

[0072] 일 예로, 장치는 (시각 상식 그래프(100)에 포함된) 추론 문장의 출현 빈도 및 추론 문장의 관련성 점수에 기초하여 해당 추론 문장이 얼마나 포괄적인지 여부를 식별할 수 있다.

[0073] 추론 문장의 출현 빈도는 해당 문장에 연결되어 있는 이미지의 개수에 기초하여 산출될 수 있다. 일 예로, 장치는 추론 필터링 모듈(14-1)을 통해 보다 많은 이미지에 연결되어 있는 추론 문장을 보다 포괄적인 추론 문장으로 판단할 수 있다. 그리고, 장치는 해당 추론 문장 및 더 많은 이미지 간의 연결을 필터링할 수 있다.

[0074] 그리고, 추론 문장의 관련성 점수는 해당 추론 문장에 연결된 이미지들 간의 코사인 유사도의 평균에 기초하여 산출될 수 있다. 즉, 해당 추론 문장에 연결된 이미지들 간의 코사인 유사도의 평균이 낮을수록, 해당 추론 문장이 보다 많은 다른 이미지(즉, 서로 상이한 이미지)에 연결되어 있음을 의미할 수 있다.

[0075] 장치는 추론 필터링 모듈(14-1)을 통해 수학식 1을 이용하여 관련성 점수를 획득할 수 있다.

수학식 1

[0076]
$$R(s) = \frac{1}{f(s)^2} \times \sum_{i=1}^{f(s)} \sum_{j=1}^{f(s)} \cos(I_i^s, I_j^s)$$

[0077] 수학식 1에서 s는 추론 문장을 의미하고, f(s)는 출현 빈도를 의미하고, cos(I, I)는 두 이미지 간의 코사인 유사도를 의미할 수 있다. 코사인 유사도를 산출하기 위하여, 장치는 이미지의 표현을 추출하는 과정에서 CLIP, ResNet, VGG-net 등의 모델을 이용할 수 있다.

[0078] 즉, 장치는 추론 필터링 모듈(14-1)을 통해 추론 문장의 출현 빈도 및 추론 문장의 관련성 점수에 기초하여 각 추론 문장 별로 몇 개의 이미지와의 연결을 필터링할지를 나타내는 N(s) 값을 수학식 2 및 수학식 3을 통해 결정할 수 있다. 장치는 추론 필터링 모듈(14-1)을 통해 추론 문장 별로 이미지 간의 유사도의 평균이 낮은 순으로 필터링을 진행할 수 있다.

수학식 2

[0079] $$P(s) = 1 - \sqrt{\frac{t \times R(s)}{f(s)}}$$

수학식 3

[0080] $$N(s) = P(s) \times f(s)$$

[0081] 수학식 2에서 t는 실험적으로 결정된 임계값(예로, 10)을 나타낸다. 본 발명을 위해 10을 사용했다. 장치는 시각 상식 그래프(100)에 포함된 데이터를 추론 필터링 모듈(14-1)을 통해 필터링하고, 필터링된 데이터에 기초하여 인공지능 모델을 학습시킬 수 있다.

[0082] 장치는 반사실적 대조 학습 모듈(14-2)을 통해 인공지능 모델(예로, 시각 상식 모델(500))을 학습시킬 수 있다.

[0083] 반사실적 대조 학습 모듈(14-2)을 통해 시각 상식 모델(500)은 이미지에 포함된 적어도 하나의 객체 중 어떤 객체가 해당 이미지에 대응되는 문장 데이터와 관련도가 높은지를 식별할 수 있다.

[0084] 세부적으로, 인공지능 모델은 객체 별로 분리된 이미지, 이미지의 설명, 추론 유형을 입력받아 추론 문장을 생성하도록 학습될 수 있다.

[0085] 이 때, 인공지능 모델의 집중 점수가 높은 순으로 객체를 나열하여 절반으로 분리될 수 있다. 집중 점수가 일정 값 이상인 이미지를 사실 이미지로 칭하고, 집중 점수가 일정 값 미만인 이미지를 반사실 이미지라 칭한다.

[0086] 이후 인공지능 모델은 추가적으로 원본 이미지와 사실이미지, 반사실 이미지를 대조함으로써 그 관계를 학습하게 될 수 있으며, 이에 따라 중요 객체에 대한 인공지능 모델의 집중 점수가 올라갈 수 있다.

[0087] 반사실적 추론 생성 모듈(14-3)은 학습이 종료된 후 반사실적 대조 학습과 유사하게 원본 이미지로부터 사실 이미지를 분리하여 사실 이미지만을 이용해 추론 문장을 인공지능 모델을 통해 생성할 수 있다.

[0088] 이에 따라, 인공지능 모델은 중요 객체만을 이용해 추론 문장을 생성하게 된다. 인공지능 모델은 사실 이미지를 분리할 수 있도록 가상의 정답을 우선적으로 생성할 수 있으며, 이를 정답으로써 사용할 수 있다.

[0089] 결과적으로 인공지능 모델은 이미지에 세부적인 부분에 집중할 수 있게 되고, 이로부터 얻은 정보를 통해 이미지의 풍부한 스토리를 서술적으로 묘사하는 추론 문장이 생성될 수 있다.

[0090] 도 4a 내지 도 4c는 본 개시의 일 실시예에 따른, 이미지에 대한 다양한 시각 상식 추론 데이터를 생성하는 방법을 설명하기 위한 도면이다.

[0091] 도 4a에 도시된 바와 같이, 시각 상식 그래프 상에는 "talk to person"이라는 추론 문장에 복수의 이미지(410-1, 410-2, 410-3, 420)가 연결될 수 있다.

[0092] 해당 추론 문장에 연결된 복수의 이미지 중 제1 이미지(420)를 제외한 나머지 이미지(410-1, 410-2, 410-3) 및 ii) 제1 이미지(420) 간의 코사인 유사도 값이 해당 추론 문장에 연결된 복수의 이미지(410-1, 410-2, 410-3, 420-1) 간의 코사인 유사도의 평균 값(430)보다 낮을 경우, 장치는 제1 이미지(420)를 필터링할 수 있다.

[0093] 도 4b에 도시된 바와 같이, 장치는 특정 이미지를 인공지능 모델(예로, 시각 및 언어 변환기(vision and language transformer)에 입력하여 획득된 제1 이미지(420)에 대한 추론 문장에 기초하여, 제1 이미지(420) 내의 주요 객체 및 적어도 하나의 부가 객체를 식별할 수 있다.

[0094] 장치는 주요 객체의 이미지(440) 및 부가 객체의 이미지(450)에 기초하여 인공지능 모델에 대해 대조 학습(contrastive learning)을 수행할 수 있다.

[0095] 도 4c에 도시된 바와 같이, 장치는 제1 이미지(420)를 학습된 인공지능 모델에 입력하여 제1 이미지(420)에 대응되는 추론 문장(예로, "lift weight")을 획득할 수 있다.

[0096] 도 5는 본 개시의 일 실시예에 따른, 이미지에 대해 생성된 다양한 시각 상식 추론 데이터를 예시한다.

[0097] 도 5의 (a)에 도시된 바와 같이, 장치는 제2 이미지(510)를 학습된 인공지능 모델에 입력하여 제2 이미지(510)에 대응되는 추론 문장(예로, "Person 1 needed to be brought the meal")을 획득할 수 있다. 대조 학습을 수행하지 않은 인공지능 모델은 동일 이미지(510)를 통해 "Person 1 needed to feel sick"이라는 문장을 출력할 수 있다. 즉, 대조 학습을 수행한 인공지능 모델은 주요 객체(person 1)과 다른 객체(예로, meal) 간의 관계에 기초한 추론 문장을 출력할 수 있다.

[0098] 도 5의 (b)에 도시된 바와 같이, 장치는 제2 이미지(520)를 학습된 인공지능 모델에 입력하여 제2 이미지(520)에 대응되는 추론 문장(예로, "Person 1 wanted to keep warm while she says")을 획득할 수 있다. 대조 학습을 수행하지 않은 인공지능 모델은 동일 이미지(520)를 통해 "Person wanted to have some rest"이라는 문장을 출력할 수 있다. 즉, 대조 학습을 수행한 인공지능 모델은 주요 객체(person 1)과 다른 객체(예로, meal) 간의 관계에 기초한 추론 문장을 출력할 수 있다.

[0099] 도 5의 (c)에 도시된 바와 같이, 장치는 제2 이미지(530)를 학습된 인공지능 모델에 입력하여 제2 이미지(530)에 대응되는 추론 문장(예로, "Person 1 will most likely thank Person 2 for taking care of her")을 획득할 수 있다. 대조 학습을 수행하지 않은 인공지능 모델은 동일 이미지(530)를 통해 "Person 1 will most likely get more warm"이라는 문장을 출력할 수 있다. 즉, 대조 학습을 수행한 인공지능 모델은 주요 객체(person 1)과 다른 객체(예로, person 2) 간의 관계에 기초한 추론 문장을 출력할 수 있다.

[0101] 이상, 본 개시에서 설명되는 실시예와 첨부된 도면은 본 개시에 포함되는 기술적 사상의 일부를 예시적으로 설명하는 것에 불과하다. 따라서, 본 명세서에 개시된 실시예들은 본 개시의 기술적 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이므로, 이러한 실시예에 의하여 본 발명의 기술 사상의 범위가 한정되는 것이 아님은 자명하다.

[0102] 본 개시의 명세서 및 도면에 포함된 기술적 사상의 범위 내에서 당업자가 용이하게 유추할 수 있는 변형 예와 구체적인 실시 예는 모두 본 개시의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

부호의 설명

[0103] 10 : 장치

11 : 메모리

12 : 통신 모듈

13 : 디스플레이

14 : 프로세서

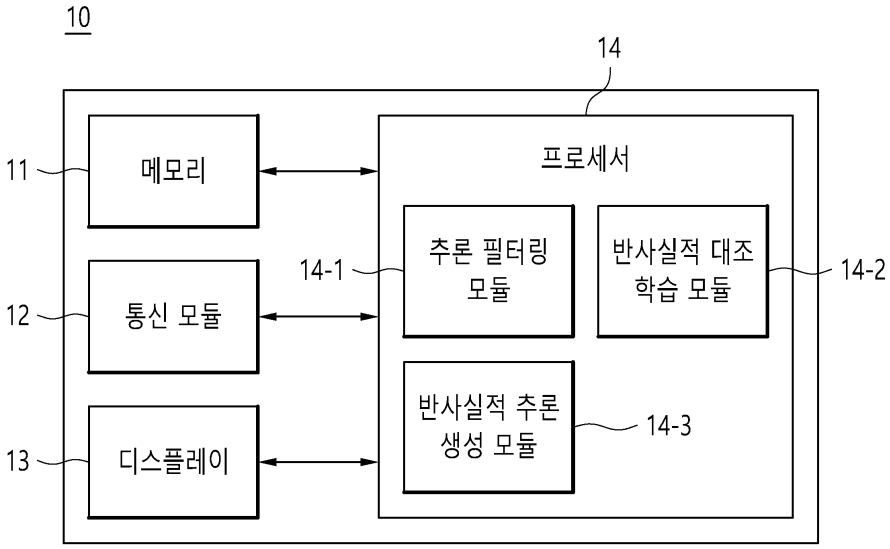
14-1 : 추론 필터링 모듈

14-2 : 반사실적 대조 학습 모듈

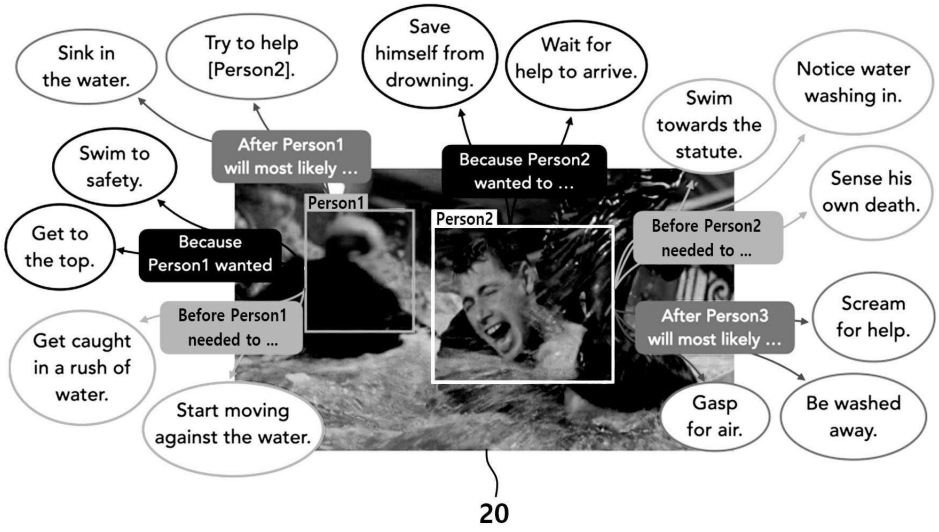
14-3 : 반사실적 추론 생성 모듈

도면

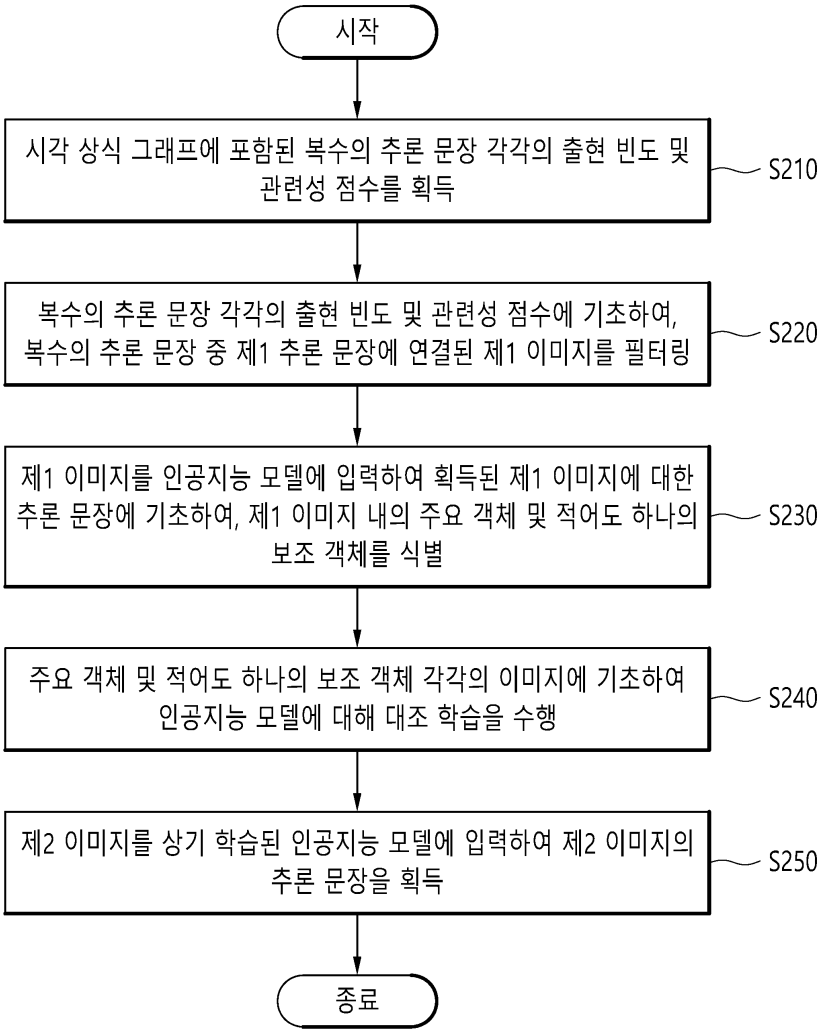
도면1a


10
14
11
메모리
프로세서
추론 필터링 모듈
반사실적 대조 학습 모듈
14-1
14-2
12
통신 모듈
반사실적 추론 생성 모듈
14-3
13
디스플레이

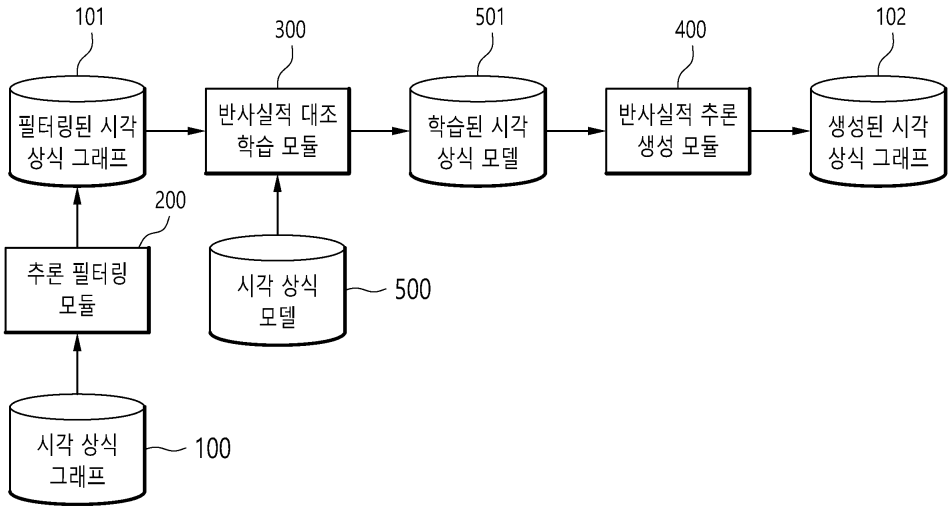
도면1b


Sink in the water.
Try to help [Person2].
Save himself from drowning.
Wait for help to arrive.
Notice water washing in.
Swim towards the statute.
Swim to safety.
After Person1 will most likely ...
Because Person2 wanted to ...
Sense his own death.
Person1
Person2
Get to the top.
Because Person1 wanted
Before Person2 needed to ...
Before Person1 needed to ...
After Person3 will most likely ...
Scream for help.
Get caught in a rush of water.
Start moving against the water.
Gasp for air.
Be washed away.
20

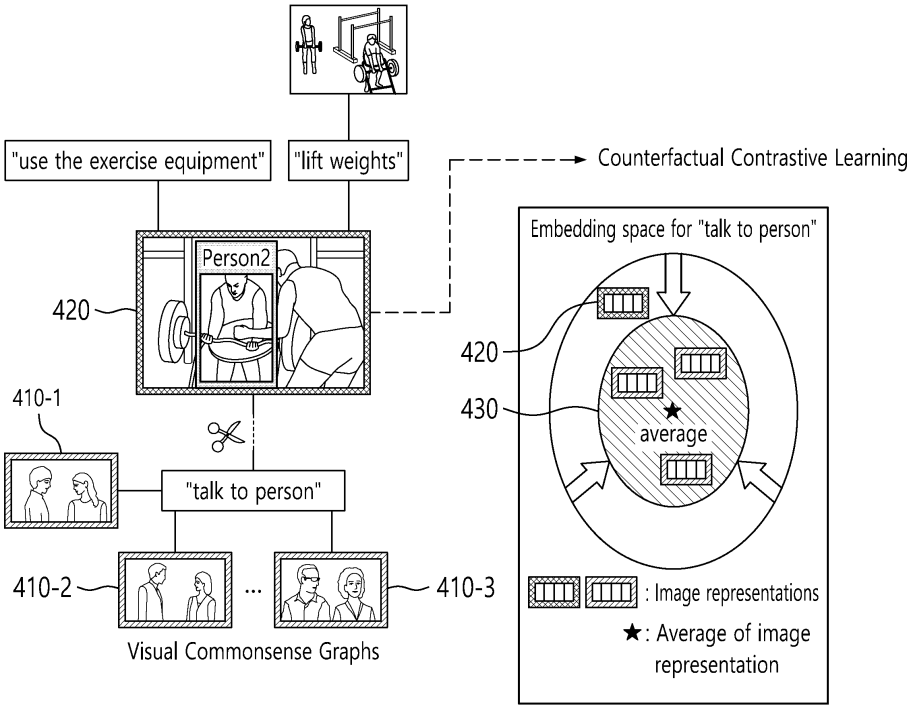
도면2


시작
시각 상식 그래프에 포함된 복수의 추론 문장 각각의 출현 빈도 및 관련성 점수를 획득
S210
복수의 추론 문장 각각의 출현 빈도 및 관련성 점수에 기초하여, 복수의 추론 문장 중 제1 추론 문장에 연결된 제1 이미지를 필터링
S220
제1 이미지를 인공지능 모델에 입력하여 획득된 제1 이미지에 대한 추론 문장에 기초하여, 제1 이미지 내의 주요 객체 및 적어도 하나의 보조 객체를 식별
S230
주요 객체 및 적어도 하나의 보조 객체 각각의 이미지에 기초하여 인공지능 모델에 대해 대조 학습을 수행
S240
제2 이미지를 상기 학습된 인공지능 모델에 입력하여 제2 이미지의 추론 문장을 획득
S250
종료

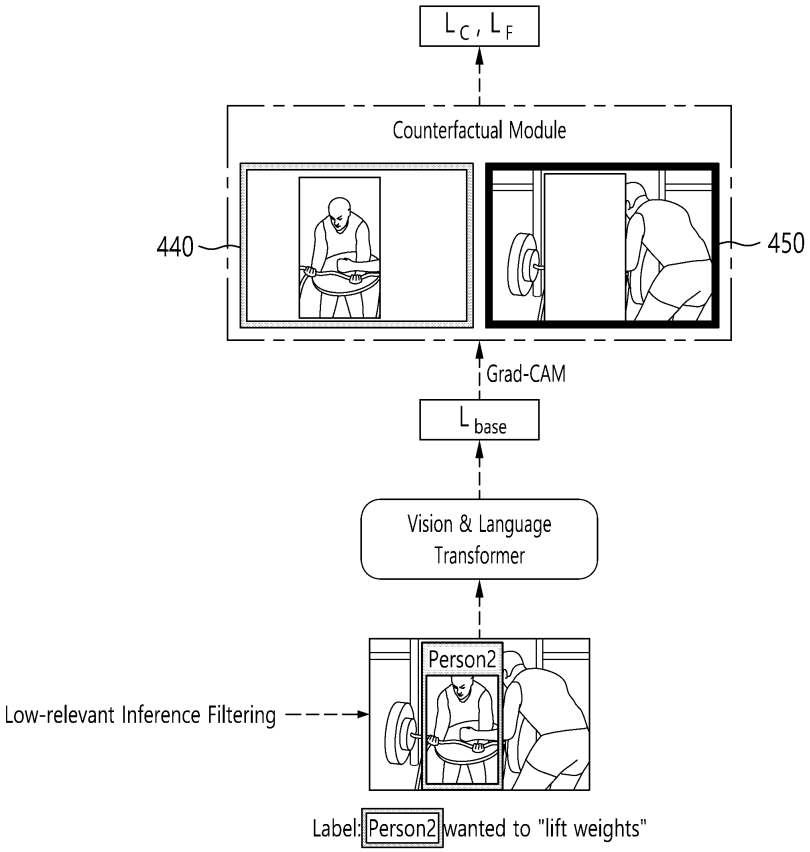
도면3


101
필터링된 시각 상식 그래프
300
반사실적 대조 학습 모듈
501
학습된 시각 상식 모델
400
반사실적 추론 생성 모듈
102
생성된 시각 상식 그래프
200
추론 필터링 모듈
시각 상식 모델
500
시각 상식 그래프
100

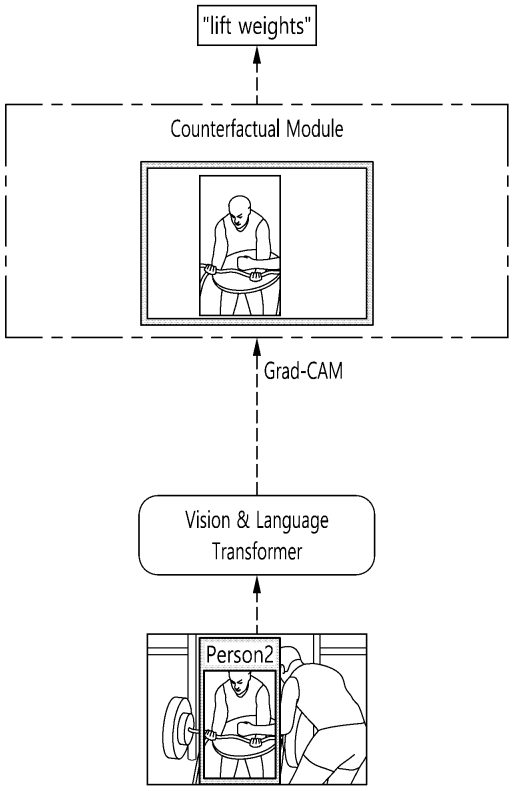
도면4a


"use the exercise equipment"
"lift weights"
Counterfactual Contrastive Learning
Person2
420
410-1
"talk to person"
410-2
410-3
Visual Commonsense Graphs
Embedding space for "talk to person"
420
430
average
: Image representations
★: Average of image representation

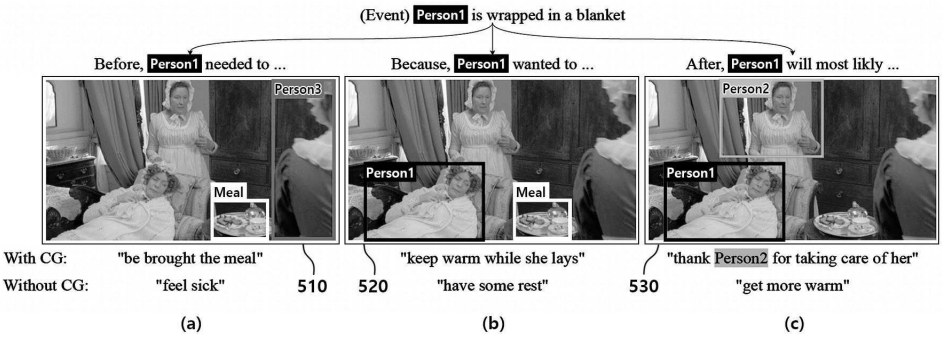
도면4b


L_C, L_F
Counterfactual Module
440
450
Grad-CAM
L_{base}
Vision & Language Transformer
Person2
Low-relevant Inference Filtering
Label: Person2 wanted to "lift weights"

도면4c


"lift weights"
Counterfactual Module
Grad-CAM
Vision & Language Transformer
Person2

도면5


(Event) Person1 is wrapped in a blanket
Before, Person1 needed to ...
Because, Person1 wanted to ...
After, Person1 will most likly ...
Person3
Meal
Person1
Meal
Person2
Person1
With CG: "be brought the meal"
Without CG: "feel sick"
510
"keep warm while she lays"
"have some rest"
520
"thank Person2 for taking care of her"
"get more warm"
530
(a)
(b)
(c)