



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2022년10월14일
(11) 등록번호 10-2454930
(24) 등록일자 2022년10월11일

- (51) 국제특허분류(Int. Cl.)
G06F 16/583 (2019.01) G06F 40/289 (2020.01)
G06F 40/30 (2020.01) G06V 10/40 (2022.01)
- (52) CPC특허분류
G06F 16/5854 (2019.01)
G06F 40/289 (2020.01)
- (21) 출원번호 10-2020-7008623
- (22) 출원일자(국제) 2019년05월09일
심사청구일자 2020년03월25일
- (85) 번역문제출일자 2020년03월25일
- (65) 공개번호 10-2020-0066617
- (43) 공개일자 2020년06월10일
- (86) 국제출원번호 PCT/CN2019/086274
- (87) 국제공개번호 WO 2020/107813
국제공개일자 2020년06월04일
- (30) 우선권주장
201811459428.7 2018년11월30일 중국(CN)
- (56) 선행기술조사문헌
Anderson, Peter James. "Vision and Language Learning: From Image Captioning and Visual Question Answering towards Embodied Agents." (2018.04.24 공개)
Lin, Yuetan, et al. "Task-driven visual saliency and attention-based visual question answering." arXiv preprint arXiv:1702.06700 (201702.22. 공개)
Chen, Kan, et al. "Abc-cnn: An attention based convolutional neural network for visual question answering." arXiv preprint arXiv:1511.05960 (2015 공개)
- (73) 특허권자
베이징 센스타임 테크놀로지 디벨롭먼트 컴퍼니 리미티드
중국 베이징 100080 하이덴 노스웨스트 4 링 로드 넘버 58 11플로어 룸 1101-1117
- (72) 발명자
리우, 시후이
중국 베이징 100084 하이덴 디스트릭트 중관춘 이스트 로드 1 코트야드 3 빌딩 7층 룸 710-712
샤오, 징
중국 베이징 100084 하이덴 디스트릭트 중관춘 이스트 로드 1 코트야드 3 빌딩 7층 룸 710-712
(뒷면에 계속)
- (74) 대리인
(유)한양특허법인

전체 청구항 수 : 총 17 항

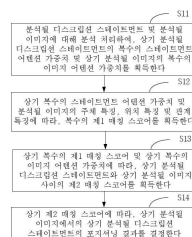
심사관 : 이복현

(54) 발명의 명칭 이미지의 디스크립션 스테이트먼트 포지셔닝 방법 및 장치, 전자 기기 및 저장 매체

(57) 요약

본원 발명은 이미지의 디스크립션 스테이트먼트 포지셔닝 방법 및 장치, 전자 기기 및 저장 매체에 관한 것이다. 상기 방법은, 분석될 디스크립션 스테이트먼트 및 분석될 이미지를 분석하여, 분석될 디스크립션 스테이트먼트의 복수의 스테이트먼트 어텐션 가중치 및 상기 분석될 이미지의 복수의 이미지 어텐션 가중치를 획득하는 단계 복수의 (뒷면에 계속)

대표도



스테이트먼트 어텐션 가중치 및 분석될 이미지의 주체 특징, 위치 특징 및 관계 특징에 따라, 복수의 제1 매칭 스코어를 획득하는 단계 복수의 제1 매칭 스코어 및 복수의 이미지 어텐션 가중치에 따라, 분석될 디스크립션 스테이트먼트와 분석될 이미지 사이의 제2 매칭 스코어를 획득하는 단계 제2 매칭 스코어에 따라, 분석될 이미지에서의 분석될 디스크립션 스테이트먼트의 포지셔닝 결과를 결정하는 단계를 포함한다.

(52) CPC특허분류

G06F 40/30 (2020.01)

G06N 3/0454 (2013.01)

G06N 3/049 (2013.01)

G06N 3/08 (2013.01)

G06V 10/40 (2022.01)

(72) 발명자

왕, 지하오

중국 베이징 100084 하이덴 디스트릭트 중관춘 이스트 로드 1 코트야드 3 빌딩 7층 룸 710-712

리, 홍셴

중국 베이징 100084 하이덴 디스트릭트 중관춘 이스트 로드 1 코트야드 3 빌딩 7층 룸 710-712

왕, 시아오강

중국 베이징 100084 하이덴 디스트릭트 중관춘 이스트 로드 1 코트야드 3 빌딩 7층 룸 710-712

명세서

청구범위

청구항 1

이미지의 디스크립션 스테이트먼트 포지셔닝 방법으로서,

전자 기기에 적용되고,

분석될 디스크립션 스테이트먼트 및 분석될 이미지에 대해 각각 분석 처리하여, 상기 분석될 디스크립션 스테이트먼트의 복수의 스테이트먼트 어텐션 가중치 및 상기 분석될 이미지의 복수의 이미지 어텐션 가중치를 획득하는 단계;

상기 복수의 스테이트먼트 어텐션 가중치 및 분석될 이미지의 주체 특징, 위치 특징 및 관계 특징에 따라, 복수의 제1 매칭 스코어를 획득하는 단계 - 상기 분석될 이미지는 복수의 객체를 포함하고, 주체 객체는 상기 복수의 객체 중 어텐션 가중치가 가장 높은 객체이며, 상기 주체 특징은 상기 주체 객체의 특징이고, 상기 위치 특징은 상기 복수의 객체의 위치 특징이며, 상기 관계 특징은 상기 복수의 객체 사이의 관계 특징임 - ;

상기 복수의 제1 매칭 스코어 및 상기 복수의 이미지 어텐션 가중치에 따라, 상기 분석될 디스크립션 스테이트먼트와 상기 분석될 이미지 사이의 제2 매칭 스코어를 획득하는 단계; 및

상기 제2 매칭 스코어에 따라, 상기 분석될 이미지에서의 상기 분석될 디스크립션 스테이트먼트의 포지셔닝 결과를 결정하는 단계를 포함하는 이미지의 디스크립션 스테이트먼트 포지셔닝 방법.

청구항 2

청구항 1에 있어서,

분석될 디스크립션 스테이트먼트 및 분석될 이미지에 대해 각각 분석 처리하여, 상기 분석될 디스크립션 스테이트먼트의 복수의 스테이트먼트 어텐션 가중치 및 상기 분석될 이미지의 복수의 이미지 어텐션 가중치를 획득하는 단계는,

상기 분석될 이미지에 대해 특징 추출을 진행하여, 상기 분석될 이미지의 이미지 특징 벡터를 획득하는 단계;

상기 분석될 디스크립션 스테이트먼트에 대해 특징 추출을 진행하여, 상기 분석될 디스크립션 스테이트먼트의 복수의 워드의 워드 임베딩 벡터를 획득하는 단계; 및

상기 이미지 특징 벡터 및 상기 복수의 워드의 워드 임베딩 벡터에 따라, 상기 분석될 디스크립션 스테이트먼트의 복수의 스테이트먼트 어텐션 가중치 및 상기 분석될 이미지의 복수의 이미지 어텐션 가중치를 획득하는 단계를 포함하는 이미지의 디스크립션 스테이트먼트 포지셔닝 방법.

청구항 3

청구항 1 또는 청구항 2에 있어서,

뉴럴 네트워크를 통해 상기 분석될 디스크립션 스테이트먼트의 복수의 스테이트먼트 어텐션 가중치 및 상기 분석될 이미지의 복수의 이미지 어텐션 가중치를 획득하는 단계를 더 포함하는 이미지의 디스크립션 스테이트먼트 포지셔닝 방법.

청구항 4

청구항 3에 있어서,

상기 복수의 스테이트먼트 어텐션 가중치는 스테이트먼트 주어 가중치, 스테이트먼트 위치 가중치 및 스테이트먼트 관계 가중치를 포함하고,

상기 뉴럴 네트워크는 이미지 어텐션 네트워크를 포함하며,

상기 이미지 어텐션 네트워크는 주체 네트워크, 위치 네트워크 및 관계 네트워크를 포함하고,

상기 복수의 제1 매칭 스코어는 주어 매칭 스코어, 위치 매칭 스코어 및 관계 매칭 스코어를 포함하며,

상기 복수의 스테이트먼트 어텐션 가중치 및 분석될 이미지의 주체 특징, 위치 특징 및 관계 특징에 따라, 복수의 제1 매칭 스코어를 획득하는 단계는,

상기 스테이트먼트 주어 가중치 및 주체 특징을 상기 주체 네트워크에 입력하고 처리를 거쳐, 상기 주어 매칭 스코어를 획득하는 단계;

상기 스테이트먼트 위치 가중치 및 위치 특징을 상기 위치 네트워크에 입력하고 처리를 거쳐, 상기 위치 매칭 스코어를 획득하는 단계; 및

상기 스테이트먼트 관계 가중치 및 관계 특징을 상기 관계 네트워크에 입력하고 처리를 거쳐, 상기 관계 매칭 스코어를 획득하는 단계를 포함하는 이미지의 디스크립션 스테이트먼트 포지셔닝 방법.

청구항 5

청구항 4에 있어서,

상기 복수의 이미지 어텐션 가중치는 주체 객체 가중치, 객체 위치 가중치 및 객체 관계 가중치를 포함하고,

상기 복수의 제1 매칭 스코어 및 상기 복수의 이미지 어텐션 가중치에 따라, 상기 분석될 디스크립션 스테이트먼트와 상기 분석될 이미지 사이의 제2 매칭 스코어를 획득하는 단계는,

상기 주체 객체 가중치, 상기 객체 위치 가중치 및 상기 객체 관계 가중치에 따라, 상기 주어 매칭 스코어, 상기 위치 매칭 스코어 및 상기 관계 매칭 스코어에 대해 가중 평균을 진행하여, 상기 제2 매칭 스코어를 결정하는 단계를 포함하는 이미지의 디스크립션 스테이트먼트 포지셔닝 방법.

청구항 6

청구항 1 또는 청구항 2에 있어서,

상기 이미지의 디스크립션 스테이트먼트 포지셔닝 방법은,

상기 분석될 이미지를 특징 추출 네트워크에 입력하고 처리를 거쳐, 상기 주체 특징, 상기 위치 특징 및 상기 관계 특징을 획득하는 단계를 더 포함하는 이미지의 디스크립션 스테이트먼트 포지셔닝 방법.

청구항 7

청구항 1 또는 청구항 2에 있어서,

상기 제2 매칭 스코어에 따라, 상기 분석될 이미지에서의 상기 분석될 디스크립션 스테이트먼트의 포지셔닝 결과를 결정하는 단계는,

상기 제2 매칭 스코어가 기설정 임계값보다 크거나 같은 경우, 상기 주체 객체의 이미지 영역을 상기 분석될 디스크립션 스테이트먼트의 포지셔닝 위치로 결정하는 단계를 포함하는 이미지의 디스크립션 스테이트먼트 포지셔닝 방법.

청구항 8

청구항 3에 있어서,

뉴럴 네트워크를 통해 상기 분석될 디스크립션 스테이트먼트의 복수의 스테이트먼트 어텐션 가중치 및 상기 분석될 이미지의 복수의 이미지 어텐션 가중치를 획득하는 단계 이전에, 상기 이미지의 디스크립션 스테이트먼트 포지셔닝 방법은,

샘플 세트를 이용하여 상기 뉴럴 네트워크를 트레이닝하는 단계를 더 포함하고,

상기 샘플 세트는 복수의 포지티브 샘플 페어 및 복수의 네거티브 샘플 페어를 포함하며,

각각의 포지티브 샘플 페어는 제1 샘플 이미지 및 상기 제1 샘플 이미지의 제1 샘플 디스크립션 스테이트먼트를 포함하고,

각각의 네거티브 샘플 페어는 제1 샘플 이미지 및 상기 제1 샘플 디스크립션 스테이트먼트에서 워드를 제거한

후의 제2 샘플 디스크립션 스테이트먼트를 포함하거나, 또는 제1 샘플 디스크립션 스테이트먼트 및 상기 제1 샘플 이미지에서 이미지 어텐션 가중치가 가장 높은 영역을 제거한 후의 제2 샘플 이미지를 포함하는 이미지의 디스크립션 스테이트먼트 포지셔닝 방법.

청구항 9

청구항 8에 있어서,

상기 뉴럴 네트워크는 언어 어텐션 네트워크를 더 포함하고, 상기 이미지의 디스크립션 스테이트먼트 포지셔닝 방법은,

상기 포지티브 샘플 페어의 제1 샘플 디스크립션 스테이트먼트 및 제1 샘플 이미지를 상기 언어 어텐션 네트워크에 입력하여, 상기 제1 샘플 디스크립션 스테이트먼트의 복수의 워드의 어텐션 가중치를 획득하는 단계;

사전 결정된 식별자를 사용하여 상기 제1 샘플 디스크립션 스테이트먼트 중 어텐션 가중치가 가장 높은 워드를 교체하여, 제2 샘플 디스크립션 스테이트먼트를 획득하는 단계; 및

상기 제1 샘플 이미지 및 상기 제2 샘플 디스크립션 스테이트먼트를 네거티브 샘플 페어로 하는 단계를 더 포함하는 이미지의 디스크립션 스테이트먼트 포지셔닝 방법.

청구항 10

청구항 8에 있어서,

상기 이미지의 디스크립션 스테이트먼트 포지셔닝 방법은,

상기 포지티브 샘플 페어의 제1 샘플 디스크립션 스테이트먼트 및 제1 샘플 이미지를 이미지 어텐션 네트워크에 입력하여, 상기 제1 샘플 이미지의 어텐션 가중치를 획득하는 단계;

상기 제1 샘플 이미지 중 어텐션 가중치가 가장 높은 이미지 영역을 제거하여, 제2 샘플 이미지를 획득하는 단계; 및

상기 제2 샘플 이미지 및 상기 제1 샘플 디스크립션 스테이트먼트를 네거티브 샘플 페어로 하는 단계를 더 포함하는 이미지의 디스크립션 스테이트먼트 포지셔닝 방법.

청구항 11

청구항 8에 있어서,

샘플 세트를 이용하여 상기 뉴럴 네트워크를 트레이닝하는 단계는,

상기 뉴럴 네트워크의 제1 손실 및 제2 손실에 따라, 상기 뉴럴 네트워크의 전체 손실을 결정하는 단계; 및

상기 전체 손실에 따라, 상기 뉴럴 네트워크를 트레이닝하는 단계를 포함하는 이미지의 디스크립션 스테이트먼트 포지셔닝 방법.

청구항 12

청구항 11에 있어서,

상기 뉴럴 네트워크의 제1 손실 및 제2 손실에 따라, 상기 뉴럴 네트워크의 전체 손실을 결정하는 단계 이전에, 상기 이미지의 디스크립션 스테이트먼트 포지셔닝 방법은,

상기 제1 손실을 획득하는 단계를 더 포함하고,

상기 제1 손실을 획득하는 단계는,

동일한 포지티브 샘플 페어의 제1 샘플 이미지 및 제1 샘플 디스크립션 스테이트먼트를 상기 뉴럴 네트워크에 입력하고 처리를 거쳐, 제1 트레이닝 스코어를 획득하는 단계;

상이한 포지티브 샘플 페어의 제1 샘플 이미지 및 제1 샘플 디스크립션 스테이트먼트를 상기 뉴럴 네트워크에 입력하고 처리를 거쳐, 제2 트레이닝 스코어를 획득하는 단계; 및

복수의 제1 트레이닝 스코어 및 복수의 제2 트레이닝 스코어에 따라, 제1 손실을 획득하는 단계를 포함하는 이

미지의 디스크립션 스테이트먼트 포지셔닝 방법.

청구항 13

청구항 11에 있어서,

상기 뉴럴 네트워크의 제1 손실 및 제2 손실에 따라, 상기 뉴럴 네트워크의 전체 손실을 결정하는 단계 이전에, 상기 이미지의 디스크립션 스테이트먼트 포지셔닝 방법은,

상기 제2 손실을 획득하는 단계를 더 포함하고,

상기 제2 손실을 획득하는 단계는,

동일한 네거티브 샘플 페어의 제2 샘플 이미지 및 제1 샘플 디스크립션 스테이트먼트를 상기 뉴럴 네트워크에 입력하고 처리를 거쳐, 제3 트레이닝 스코어를 획득하는 단계;

상이한 네거티브 샘플 페어의 제2 샘플 이미지 및 제1 샘플 디스크립션 스테이트먼트를 상기 뉴럴 네트워크에 입력하고 처리를 거쳐, 제4 트레이닝 스코어를 획득하는 단계;

동일한 네거티브 샘플 페어의 제1 샘플 이미지 및 제2 샘플 디스크립션 스테이트먼트를 상기 뉴럴 네트워크에 입력하고 처리를 거쳐, 제5 트레이닝 스코어를 획득하는 단계;

상이한 네거티브 샘플 페어의 제1 샘플 이미지 및 제2 샘플 디스크립션 스테이트먼트를 상기 뉴럴 네트워크에 입력하고 처리를 거쳐, 제6 트레이닝 스코어를 획득하는 단계; 및

복수의 제3 트레이닝 스코어, 복수의 제4 트레이닝 스코어, 복수의 제5 트레이닝 스코어 및 복수의 제6 트레이닝 스코어에 따라, 제2 손실을 획득하는 단계를 포함하는 이미지의 디스크립션 스테이트먼트 포지셔닝 방법.

청구항 14

청구항 11에 있어서,

상기 뉴럴 네트워크의 제1 손실 및 제2 손실에 따라, 상기 뉴럴 네트워크의 전체 손실을 결정하는 단계는,

상기 제1 손실 및 상기 제2 손실에 대해 가중 오버랩을 진행하여, 상기 뉴럴 네트워크의 전체 손실을 획득하는 단계를 포함하는 이미지의 디스크립션 스테이트먼트 포지셔닝 방법.

청구항 15

이미지의 디스크립션 스테이트먼트 포지셔닝 장치로서,

분석될 디스크립션 스테이트먼트 및 분석될 이미지에 대해 각각 분석 처리하여, 상기 분석될 디스크립션 스테이트먼트의 복수의 스테이트먼트 어텐션 가중치 및 상기 분석될 이미지의 복수의 이미지 어텐션 가중치를 획득하도록 구성되는 제1 가중치 획득 모듈;

상기 복수의 스테이트먼트 어텐션 가중치 및 분석될 이미지의 주제 특징, 위치 특징 및 관계 특징에 따라, 복수의 제1 매칭 스코어를 획득하도록 구성되는 제1 스코어 획득 모듈 - 상기 분석될 이미지는 복수의 객체를 포함하고, 주제 객체는 상기 복수의 객체 중 어텐션 가중치가 가장 높은 객체이며, 상기 주제 특징은 상기 주제 객체의 특징이고, 상기 위치 특징은 상기 복수의 객체의 위치 특징이며, 상기 관계 특징은 상기 복수의 객체 사이의 관계 특징임 - ;

상기 복수의 제1 매칭 스코어 및 상기 복수의 이미지 어텐션 가중치에 따라, 상기 분석될 디스크립션 스테이트먼트와 상기 분석될 이미지 사이의 제2 매칭 스코어를 획득하도록 구성되는 제2 스코어 획득 모듈; 및

상기 제2 매칭 스코어에 따라, 상기 분석될 이미지에서의 상기 분석될 디스크립션 스테이트먼트의 포지셔닝 결과를 결정하도록 구성되는 결과 결정 모듈을 포함하는 이미지의 디스크립션 스테이트먼트 포지셔닝 장치.

청구항 16

전자 기기로서,

프로세서; 및

프로세서 실행 가능 명령어가 저장된 메모리를 포함하고,

상기 프로세서는 청구항 1 또는 청구항 2에 따른 이미지의 디스크립션 스테이트먼트 포지셔닝 방법을 수행하도록 구성되는 전자 기기.

청구항 17

컴퓨터 프로그램 명령어가 저장된 컴퓨터 판독 가능 저장 매체로서,

상기 컴퓨터 프로그램 명령어가 프로세서에 의해 실행될 경우 청구항 1 또는 청구항 2에 따른 이미지의 디스크립션 스테이트먼트 포지셔닝 방법을 구현하는 컴퓨터 판독 가능 저장 매체.

청구항 18

삭제

청구항 19

삭제

청구항 20

삭제

청구항 21

삭제

청구항 22

삭제

청구항 23

삭제

청구항 24

삭제

청구항 25

삭제

청구항 26

삭제

청구항 27

삭제

청구항 28

삭제

청구항 29

삭제

청구항 30

삭제

발명의 설명

기술 분야

- [0001] 관련 출원의 상호 참조
- [0002] 본원 발명은 출원번호가 201811459428.7이고 출원일이 2018년 11월 30일인 중국 특허 출원에 기반하여 제출하였고 상기 중국 특허 출원의 우선권을 주장하는 바, 상기 중국 특허 출원의 모든 내용은 본원 발명에 원용된다.
- [0003] 본원 발명은 컴퓨터 시각 기술분야에 관한 것이지만 시각 기술분야에 한하지 않고, 특히는 이미지의 디스크립션 스테이트먼트 포지셔닝 방법 및 장치, 전자 기기 및 저장 매체에 관한 것이다.

배경 기술

- [0004] 문구 포지셔닝은 컴퓨터 시각 및 자연 언어 처리 교차 분야의 중요한 과제로서, 예를 들어 제시된 한단락의 말(스테이트먼트)에 따라 기계가 이미지에서 상기 스테이트먼트가 디스크립션하고자 하는 객체(사람 또는 물체 등)를 포지셔닝할 것을 요구한다. 관련 기술에서, 객체 및 그 관계를 식별하기 위해, 포지셔닝 모듈 및 관계 모듈 등으로 구성된 조합 모듈화 네트워크를 제출하였지만, 이러한 모듈은 특정 단어 또는 시각 개념에 지나치게 의존하고 자주 관찰되는 증거에 치중되어 스테이트먼트와 이미지의 대응 효과가 비교적 떨어지는 것을 초래할 수 있다.

발명의 내용

- [0005] 본원 발명은 이미지의 디스크립션 스테이트먼트 포지셔닝의 기술적 해결수단을 제공하였다.
- [0006] 본원 발명의 일 측면에 따라, 이미지의 디스크립션 스테이트먼트 포지셔닝 방법을 제공하며, 상기 방법은, 분석될 디스크립션 스테이트먼트 및 분석될 이미지에 대해 분석 처리하여, 상기 분석될 디스크립션 스테이트먼트의 복수의 스테이트먼트 어텐션 가중치 및 상기 분석될 이미지의 복수의 이미지 어텐션 가중치를 획득하는 단계, 상기 복수의 스테이트먼트 어텐션 가중치 및 분석될 이미지의 주체 특징, 위치 특징 및 관계 특징에 따라, 복수의 제1 매칭 스코어를 획득하는 단계 - 상기 분석될 이미지는 복수의 객체를 포함하고, 주체 객체는 상기 복수의 객체 중 어텐션 가중치가 가장 높은 객체이며, 상기 주체 특징은 상기 주체 객체의 특징이고, 상기 위치 특징은 상기 복수의 객체의 위치 특징이며, 상기 관계 특징은 상기 복수의 객체 사이의 관계 특징임 - ; 상기 복수의 제1 매칭 스코어 및 상기 복수의 이미지 어텐션 가중치에 따라, 상기 분석될 디스크립션 스테이트먼트와 상기 분석될 이미지 사이의 제2 매칭 스코어를 획득하는 단계 및 상기 제2 매칭 스코어에 따라, 상기 분석될 이미지에서의 상기 분석될 디스크립션 스테이트먼트의 포지셔닝 결과를 결정하는 단계를 포함한다.
- [0007] 본원 발명의 일 측면에 따라, 이미지의 디스크립션 스테이트먼트 포지셔닝 장치를 제공하며, 상기 장치는 분석될 디스크립션 스테이트먼트 및 분석될 이미지에 대해 분석 처리하여, 상기 분석될 디스크립션 스테이트먼트의 복수의 스테이트먼트 어텐션 가중치 및 상기 분석될 이미지의 복수의 이미지 어텐션 가중치를 획득하도록 구성되는 제1 가중치 획득 모듈, 상기 복수의 스테이트먼트 어텐션 가중치 및 분석될 이미지의 주체 특징, 위치 특징 및 관계 특징에 따라, 복수의 제1 매칭 스코어를 획득하도록 구성되는 제1 스코어 획득 모듈 - 상기 분석될 이미지는 복수의 객체를 포함하고, 주체 객체는 상기 복수의 객체 중 어텐션 가중치가 가장 높은 객체이며, 상기 주체 특징은 상기 주체 객체의 특징이고, 상기 위치 특징은 상기 복수의 객체의 위치 특징이며, 상기 관계 특징은 상기 복수의 객체 사이의 관계 특징임 - ; 상기 복수의 제1 매칭 스코어 및 상기 복수의 이미지 어텐션 가중치에 따라, 상기 분석될 디스크립션 스테이트먼트와 상기 분석될 이미지 사이의 제2 매칭 스코어를 획득하도록 구성되는 제2 스코어 획득 모듈 및 상기 제2 매칭 스코어에 따라, 상기 분석될 이미지에서의 상기 분석될 디스크립션 스테이트먼트의 포지셔닝 결과를 결정하도록 구성되는 결과 결정 모듈을 포함한다.
- [0008] 본원 발명의 일 측면에 따라, 전자 기기를 제공하였는 바, 프로세서 및 프로세서 실행 가능 명령어가 저장된 메모리를 포함하며, 상기 프로세서는 상기 이미지의 디스크립션 스테이트먼트 포지셔닝 방법을 수행하도록 구성된다.
- [0009] 본원 발명의 일 측면에 따라, 컴퓨터 프로그램 명령어가 저장되는 컴퓨터 판독 가능 저장 매체를 제공하였는 바, 상기 컴퓨터 프로그램 명령어가 프로세서에 의해 실행될 경우 상기 이미지의 디스크립션 스테이트먼트 포지셔닝 방법을 구현한다.

[0010] [발명의 효과]

[0011] 본원 발명의 실시예에서, 분석될 디스크립션 스테이트먼트의 스테이트먼트 어텐션 가중치 및 분석될 이미지의 이미지 어텐션 가중치를 획득할 수 있고, 스테이트먼트 어텐션 가중치 및 이미지의 주제 특징, 위치 특징 및 관계 특징에 따라 복수의 제1 매칭 스코어를 획득하고, 제1 매칭 스코어 및 이미지 어텐션 가중치에 따라 제2 매칭 스코어를 획득하며, 제2 매칭 스코어에 따라 포지셔닝 결과를 결정함으로써, 텍스트와 시각적 어의 사이의 대응 관계를 충분히 발견하여, 이미지에서의 디스크립션 스테이트먼트의 포지셔닝 정확도를 향상시킨다.

[0012] 상기의 일반적인 디스크립션과 후행문의 디테일 디스크립션은 단지 예시적 및 해석적인 것으로서, 본원 발명을 한정하지 않는다. 하기의 도면에 따라 예시적인 실시예를 상세하게 설명함으로써, 본원 발명의 기타 특징 및 방면은 더욱 분명해질 것이다.

도면의 간단한 설명

[0013] 여기서의 도면은 명세서에 병합되어 본명세서의 일부분을 구성하고, 본원 발명에 부합되는 실시예를 나타내며, 명세서와 함께 본 발명의 원리를 해석하기 위한 것이다.

도 1은 본원 발명의 실시예에 따른 이미지의 디스크립션 스테이트먼트 포지셔닝 방법의 흐름도이다.

도 2는 본원 발명의 실시예에 따른 뉴럴 네트워크의 모식도이다.

도 3은 본원 발명의 실시예에 따른 제2 샘플 디스크립션 스테이트먼트를 획득하는 단계의 모식도이다.

도 4는 본원 발명의 실시예에 따른 제2 샘플 이미지를 획득하는 단계의 모식도이다.

도 5는 본원 발명의 실시예에 따른 이미지의 디스크립션 스테이트먼트 포지셔닝 장치의 블록도이다.

도 6은 본원 발명의 실시예에 따른 전자 기기의 블록도이다.

도 7은 본원 발명의 실시예에 따른 전자 기기의 블록도이다.

발명을 실시하기 위한 구체적인 내용

[0014] 이하 도면을 참조하여 본원 발명의 각 예시적인 실시예, 특징 및 양태를 설명하도록 한다. 도면에서 동일한 도면 부호는 기능이 동일하거나 유사한 소자를 표시한다. 비록 도면에서 실시예의 각양태를 도시하였지만 별도의 지적이 없는 한비율에 따라 도면을 제작할 필요가 없다.

[0015] 여기서 전용되는 단어 "예시적"은 "예, 실시예 또는 설명형"적인 것을 의미한다. 여기서 "예시적"으로 설명되는 모든 실시예는 기타 실시예보다 반드시 더바람직하거나 좋다는 것으로 해석하지 아니한다.

[0016] 본문의 용어 "및/또는"은 단지 디스크립션 관련 객체의 관련 관계를 의미하며, 3가지 관계가 존재함을 의미한다. 예를 들면 A 및 /또는 B는, A만 존재, A와B가 동시에 존재, B만 존재하는 3가지 경우를 의미한다. 이 밖에, 본문에서 "적어도 한가지"는 여러가지 중의 임의의 한가지 또는 여러가지 중의 적어도 두가지의 임의의 조합을 의미한다. 예를 들면 A, B, C 중의 적어도 한가지를 포함하는 것은 A, B 및 C로 구성된 집합 중 선택되는 임의의 하나 또는 복수의 원소를 포함하는 것을 의미할 수 있다.

[0017] 이 밖에, 본원 발명을 더욱 잘 설명하기 위해, 이하의 구체적인 실시 형태에서 수많은 구체적인 세부 사항을 시사하였다. 본 기술분야의 통상의 기술자는 어떠한 구체적인 세부 사항이없어도 본원 발명을 실시할 수 있음을 반드시 이해해야 한다. 본원 발명의 취지를 강조하기 위해, 일부 구현예에서, 본 기술분야의 통상의 기술자에게 있어서 자명한 방법, 수단, 소자 및 회로에 대해 상세하게 설명하지 않는다.

[0018] 본 발명의 각실시예에서, 상기 이미지의 디스크립션 스테이트먼트 포지셔닝 방법은 단말 기기 또는 서버 등전자 기기로 수행될 수 있는 바, 단말 기기는 사용자 기기(User Equipment, UE), 모바일 기기, 사용자 단말기, 단말 장치, 셀룰러 폰, 무선 전화, 개인 휴대 정보 단말기(Personal Digital Assistant, PDA), 휴대용 기기, 컴퓨팅 기기, 차량 탑재 기기, 웨어러블 기기 동일 수 있으며, 상기 방법은 프로세서를 통해 메모리에 저장된 컴퓨터 관독 가능 명령어를 스케줄링하는 방식으로 구현될 수 있다. 또는, 서버를 통해 상기 방법을 수행할 수 있다.

[0019] 도 1은 본원 발명의 실시예에 따른 이미지의 디스크립션 스테이트먼트 포지셔닝 방법의 흐름도이다. 상기 방법은 하기의 단계를 포함한다.

[0020] 단계 S11에서, 분석될 디스크립션 스테이트먼트 및 분석될 이미지에 대해 분석 처리하여, 상기 분석될 디스크립

선 스테이트먼트의 복수의 스테이트먼트 어텐션 가중치 및 상기 분석될 이미지의 복수의 이미지 어텐션 가중치를 획득한다.

- [0021] 한 가지 가능한 구현 방식에서, 분석될 이미지는 예를 들어 말을 타고 있는 복수의 사람과 같은 복수의 객체(사람, 동물 물체 등)를 포함할 수 있다. 분석될 디스크립션 스테이트먼트는 분석될 이미지 중의 어느 한 객체에 대한 디스크립션일 수 있는 바, 예를 들면 "여자 아이가 타고 있는 가운데의 갈색 말"이다. 분석될 이미지와 분석될 디스크립션 스테이트먼트 사이는 대응될 수도 안될 수도 있다. 본원 발명의 실시예에 따른 방법으로 스테이트먼트와 이미지 사이의 연관을 결정할 수 있다.
- [0022] 한 가지 가능한 구현 방식에서, 분석될 디스크립션 스테이트먼트의 복수의 스테이트먼트 어텐션 가중치는 스테이트먼트 주어 가중치, 스테이트먼트 위치 가중치 및 스테이트먼트 관계 가중치를 포함할 수 있고, 분석될 디스크립션 스테이트먼트의 상이한 타입의 워드에 대응되는 어텐션 가중치를 각각 표시하기 위한 것이다.
- [0023] 한 가지 가능한 구현 방식에서, 분석될 이미지의 복수의 이미지 어텐션 가중치는 주체 객체 가중치, 객체 위치 가중치 및 객체 관계 가중치를 포함할 수 있고, 분석될 이미지의 상이한 타입의 이미지 영역에 대응되는 어텐션 가중치를 각각 표시하기 위한 것이다.
- [0024] 단계 S12에서, 상기 복수의 스테이트먼트 어텐션 가중치 및 분석될 이미지의 주체 특징, 위치 특징 및 관계 특징에 따라, 복수의 제1 매칭 스코어를 획득하며, 상기 분석될 이미지는 복수의 객체를 포함하고, 주체 객체는 상기 복수의 객체 중 어텐션 가중치가 가장 높은 객체이며, 상기 주체 특징은 상기 주체 객체의 특징이고, 상기 위치 특징은 상기 복수의 객체의 위치 특징이며, 상기 관계 특징은 상기 복수의 객체 사이의 관계 특징이다.
- [0025] 한 가지 가능한 구현 방식에서, 상기 분석될 이미지는 복수의 객체(사람, 동물, 물체 등)를 포함하고, 주체 객체는 상기 복수의 객체 중 어텐션 가중치가 가장 높은 객체이다. 상기 주체 특징은 상기 주체 객체 자체의 이미지 특징이고, 상기 위치 특징은 상기 복수의 객체 사이의 상대적인 위치를 체현하는 위치 특징이며, 상기 관계 특징은 상기 복수의 객체 사이의 상대적인 관계를 체현하는 관계 특징이다.
- [0026] 한 가지 가능한 구현 방식에서, 복수의 제1 매칭 스코어는 주어 매칭 스코어, 위치 매칭 스코어 및 관계 매칭 스코어를 포함할 수 있다. 주어 매칭 스코어는 분석될 이미지 중의 주체 객체와 분석될 디스크립션 스테이트먼트의 객체 디스크립션 사이의 매칭 정도를 평가하기 위한 것이고, 위치 매칭 스코어는 분석될 이미지 중 복수의 객체의 상대적인 위치와 분석될 디스크립션 스테이트먼트의 위치 디스크립션 사이의 매칭 정도를 평가하며, 관계 매칭 스코어는 분석될 이미지 중 복수의 객체의 연관성과 분석될 디스크립션 스테이트먼트의 연관 디스크립션 사이의 매칭 정도를 평가하기 위한 것이다.
- [0027] 단계 S13에서, 상기 복수의 제1 매칭 스코어 및 상기 복수의 이미지 어텐션 가중치에 따라, 상기 분석될 디스크립션 스테이트먼트와 상기 분석될 이미지 사이의 제2 매칭 스코어를 획득한다.
- [0028] 한 가지 가능한 구현 방식에서, 주어 매칭 스코어, 위치 매칭 스코어 및 관계 매칭 스코어 및 주체 객체 가중치, 객체 위치 가중치 및 객체 관계 가중치에 따라, 상기 분석될 디스크립션 스테이트먼트와 상기 분석될 이미지 사이의 제2 매칭 스코어를 획득할 수 있다. 제2 매칭 스코어는 분석될 이미지와 분석될 디스크립션 스테이트먼트 사이의 전체 매칭 정도를 평가하기 위한 것이다.
- [0029] 단계 S14에서, 상기 제2 매칭 스코어에 따라, 상기 분석될 이미지에서의 상기 분석될 디스크립션 스테이트먼트의 포지셔닝 결과를 결정한다.
- [0030] 한 가지 가능한 구현 방식에서, 제2 매칭 스코어를 획득한 후, 분석될 이미지에서의 분석될 디스크립션 스테이트먼트의 포지셔닝 위치를 더욱 결정할 수 있고, 이로써 이미지에서의 디스크립션 스테이트먼트의 포지셔닝을 실현한다.
- [0031] 본원 발명의 실시예에 따르면, 분석될 디스크립션 스테이트먼트의 스테이트먼트 어텐션 가중치 및 분석될 이미지의 이미지 어텐션 가중치를 획득할 수 있고, 스테이트먼트 어텐션 가중치 및 이미지의 주체 특징, 위치 특징 및 관계 특징에 따라 복수의 제1 매칭 스코어를 획득하고, 제1 매칭 스코어 및 이미지 어텐션 가중치에 따라 제2 매칭 스코어를 획득하며, 제2 매칭 스코어에 따라 포지셔닝 결과를 결정함으로써, 텍스트와 시각적 어의 사이의 대응 관계를 충분히 발견하여, 이미지에서의 디스크립션 스테이트먼트의 포지셔닝 정확도를 향상시킨다.
- [0032] 한 가지 가능한 구현 방식에서, 단계 S11에서 분석될 디스크립션 스테이트먼트 및 분석될 이미지에 대해 분석 처리하여, 상기 분석될 디스크립션 스테이트먼트의 복수의 스테이트먼트 어텐션 가중치 및 상기 분석될 이미지

의 복수의 이미지 어텐션 가중치를 획득할 수 있다. 여기서, 단계 S11은,

- [0033] 상기 분석될 이미지에 대해 특징 추출을 진행하여, 상기 분석될 이미지의 이미지 특징 벡터를 획득하는 단계
- [0034] 상기 분석될 디스크립션 스테이트먼트에 대해 특징 추출을 진행하여, 상기 분석될 디스크립션 스테이트먼트의 복수의 워드의 워드 임베딩 벡터를 획득하는 단계 및
- [0035] 상기 이미지 특징 벡터 및 상기 복수의 워드의 워드 임베딩 벡터에 따라, 상기 분석될 디스크립션 스테이트먼트의 복수의 스테이트먼트 어텐션 가중치 및 상기 분석될 이미지의 복수의 이미지 어텐션 가중치를 획득하는 단계를 포함할 수 있다.
- [0036] 예를 들어 설명하면, 분석될 이미지 및 분석될 디스크립션 스테이트먼트에 대해 각각 특징 추출을 진행할 수 있다. 분석될 이미지에 대해, 분석될 이미지의 모든 픽셀 포인트에 대해 특징 추출을 진행할 수 있으며, 분석될 이미지의 이미지 특징 벡터(e_0)를 획득한다. 본원 발명은 분석될 이미지의 특징 추출 방식에 대해서 한정하지 않는다.
- [0037] 한 가지 가능한 구현 방식에서, 분석될 디스크립션 스테이트먼트에 대해, 워드를 처리하여 분석될 디스크립션 스테이트먼트의 복수의 워드를 결정할 수 있고, 각 워드에 대해 특징 추출을 진행하여, 복수의 워드의 워드 임베딩 벡터(word embedding) $\{e_t\}_{t=1}^T$ 를 획득한다. 여기서, T는 워드의 개수를 표시하고(T는 1보다 큰 정수임), e_t 는 제 t 번째의 워드 임베딩 벡터를 표시하며, $1 \leq t \leq T$ 이다. 본원 발명은 분석될 디스크립션 스테이트먼트의 구체적인 워드 방식 및 각 워드의 특징 추출의 구체적인 방식에 대해서 한정하지 않는다.
- [0038] 한 가지 가능한 구현 방식에서, 결정된 이미지 특징 벡터 및 상기 복수의 워드의 워드 임베딩 벡터에 따라, 상기 분석될 디스크립션 스테이트먼트의 복수의 스테이트먼트 어텐션 가중치 및 상기 분석될 이미지의 복수의 이미지 어텐션 가중치를 결정할 수 있다.
- [0039] 한 가지 가능한 구현 방식에서, 상기 방법은, 뉴럴 네트워크를 통해 상기 분석될 디스크립션 스테이트먼트의 복수의 스테이트먼트 어텐션 가중치 및 상기 분석될 이미지의 복수의 이미지 어텐션 가중치를 획득하는 단계를 더 포함할 수 있다. 상기 뉴럴 네트워크는 언어 어텐션 네트워크를 포함할 수 있고, 언어 어텐션 네트워크는 순환 뉴럴 네트워크(RNN), 장기 및 단기 기억 네트워크(LSTM) 등 네트워크를 통해 실현될 수 있다. 상기 분석될 이미지와 상기 분석될 디스크립션 스테이트먼트를 언어 어텐션 네트워크에 입력하고 처리를 거쳐, 상기 복수의 스테이트먼트 어텐션 가중치 및 상기 복수의 이미지 어텐션 가중치를 획득할 수 있다.
- [0040] 한 가지 가능한 구현 방식에서, 상기 분석될 이미지와 상기 분석될 디스크립션 스테이트먼트를 언어 어텐션 네트워크에 입력하고 처리를 거쳐, 상기 복수의 스테이트먼트 어텐션 가중치 및 상기 복수의 이미지 어텐션 가중치를 획득할 수 있다.
- [0041] 예를 들어 설명하면, 언어 어텐션 네트워크의 특징 추출 서브 네트워크를 통해 특징을 추출할 수 있고, 각각 이미지 특징 벡터 e_0 및 워드 임베딩 벡터 $\{e_t\}_{t=1}^T$ 를 획득할 수 있다. 상기 특징 추출 서브 네트워크는 콘볼루션 뉴럴 네트워크(CNN)(예를 들면 Faster CNN)일 수 있다.
- [0042] 한 가지 가능한 구현 방식에서, 언어 어텐션 네트워크는 어텐션 메커니즘에 기반한 LSTM 네트워크를 구비할 수 있다. 이미지 특징 벡터 e_0 를 LSTM 네트워크의 제1 레벨 입력으로 할 수 있고, 워드 임베딩 벡터 $\{e_t\}_{t=1}^T$ 를 LSTM 네트워크의 각레벨 순환의 입력으로 하여, LSTM 네트워크의 복수의 숨겨진 층(hidden layer)의 출력 상태 h_t 를 획득할 수 있다.
- [0043] 한 가지 가능한 구현 방식에서, 복수의 상태 h_t 에 따라, 이미지 어텐션 가중치 및 각 워드의 어텐션 가중치를 산출할 수 있고, 복수의 워드의 어텐션 가중치는 복수의 워드 임베딩 벡터 $\{e_t\}_{t=1}^T$ 에 대해 가중합(weighted sum)을 진행하여, 스테이트먼트 어텐션 가중치를 획득할 수 있다.
- [0044] 한 가지 가능한 구현 방식에서, 분석될 디스크립션 스테이트먼트의 복수의 스테이트먼트 어텐션 가중치는 워드 레벨 어텐션 가중치(word-level attention weights)이고, 스테이트먼트 주어 가중치(q^{subj}), 스테이트먼트 위치

가중치(q^{loc}) 및 스테이트먼트 관계 가중치(q^{rel})를 포함할 수 있으며, 분석될 디스크립션 스테이트먼트의 상이한 타입의 워드에 대응되는 어텐션 가중치를 각각 표시하기 위한 것이다.

[0045] 여기서, 스테이트먼트 주어 가중치는 스테이트먼트 중의 주어 워드를 주의할 경우 어텐션 가중치를 표시하기 위한 것이고, 예를 들어 스테이트먼트 "여자 아이가 타고 있는 가운데의 갈색 말"에서 주어의 워드는 "갈색 말" 또는 "말"의 어텐션 가중치이고, 스테이트먼트 위치 가중치는 스테이트먼트 중위치를 표시하는 워드를 주의할 경우 어텐션 가중치를 표시하기 위한 것이며, 예를 들어 상기 스테이트먼트 중위치를 표시하는 워드 "가운데"의 어텐션 가중치 스테이트먼트 관계 가중치는 스테이트먼트 중객체 사이의 관계를 표시하는 워드를 주의할 경우 어텐션 가중치를 표시하고, 예를 들어 상기 스테이트먼트 중객체 사이의 관계를 표시하는 워드 "여자 아이가 타고 있는"의 어텐션 가중치이다.

[0046] 한 가지 가능한 구현 방식에서, 분석될 이미지의 복수의 이미지 어텐션 가중치는 모듈 레벨 어텐션 가중치(module-level attention weights)이고, 주체 객체 가중치(ω^{subj}), 객체 위치 가중치(ω^{loc}) 및 객체 관계 가중치(ω^{rel})를 포함할 수 있고, 분석될 이미지의 상이한 타입의 이미지 영역에 대응되는 어텐션 가중치를 각각 표시하기 위한 것이다.

[0047] 여기서, 주체 객체 가중치는 이미지 중복수의 객체(사람, 동물, 물체 등)에서 가장 중요한 객체(주체 객체)를 주의할 경우의 어텐션 가중치를 표시할 수 있으며, 예를 들어 이미지 가운데의 인물이고, 객체 위치 가중치는 이미지 중복수의 객체의 상대적인 위치를 주의할 경우의 어텐션 가중치를 표시할 수 있고, 예를 들어 이미지의 중간, 좌측 및 우측 위치이며, 객체 관계 가중치는 이미지 중복수의 객체 사이의 연관성을 주의할 경우의 어텐션 가중치를 표시할 수 있으며, 예를 들어 이미지의 중간, 좌측 및 우측은 모두 사람들이 말을 타고 있다.

[0048] 일부 실시예에서, 이미지 어텐션 가중치는 이미지에서의 상기 객체의 각종 이미지 파라미터에 따라 결정될 수 있고, 상기 이미지 파라미터는 이미지에서의 상기 객체의 분포 위치, 상기 객체가 이미지에서 차지하는 면적, 이미지에서의 상기 객체의 주체색을 포함하지만 이에 한하지 않는다. 예를 들어, 분포 위치에 따라, 이미지 중간에 있는 객체는 이미지 가장자리에 있는 객체에 비해 비교적 높은 이미지 어텐션 가중치를 획득할 수 있다. 다시 예를 들어, 이미지에서 차지하는 면적이 비교적 큰 객체는 차지하는 면적이 비교적 작은 객체에 비해 비교적 높은 객체를 구비한다. 다시 예를 들어, 객체의 주체색이 추적되고 있는 객체의 색상이면, 기타 색상에 비해 비교적 높은 이미지 어텐션 가중치를 구비할 수 있다.

[0049] 또 일부 실시예에서, 객체가 부각시킨 각종 표현 상태에 따라 이미지 어텐션 객체를 결정하고, 예를 들어, 도로 모니터링 영상의 이미지 프레임으로 분석하며, 추적되는 객체가 차량이면, 규정 위반 현상이 있는 객체는 비교적 높은 이미지 어텐션을 구비하는 바, 예를 들어 이미지에 포함된 차량이 차선을 침범한 행위 특징을 포함하면, 상기 객체는 더욱 높은 이미지 어텐션 가중치를 구성할 수 있다.

[0050] 물론, 상기 내용은 단지 이미지 어텐션 가중치에 대해 예를 들어 설명한 것으로, 구체적으로 이미지 어텐션 가중치는 이미지 처리 수요에 따라 구성될 수 있으며, 이에 한하지 않는다.

[0051] 이러한 방식을 통해, 언어 어텐션 네트워크를 통해 시각(이미지) 및 텍스트(스테이트먼트) 중 상이한 타입의 정보를 포착하여 이미지 및 스테이트먼트가 각 양태에서의 대응 관계를 발견하여 처리 정밀도를 향상시켰다.

[0052] 한 가지 가능한 구현 방식에서, 단계 S12 이전에, 상기 이미지의 디스크립션 스테이트먼트 포지셔닝 방법은, 상기 분석될 이미지를 특징 추출 네트워크에 입력하고 처리를 거쳐, 상기 분석될 이미지의 주체 특징, 위치 특징 및 관계 특징을 획득하는 단계를 더 포함한다.

[0053] 예를 들어 설명하면, 특징 추출 네트워크는 미리 설정한 하나 또는 복수의 콘볼루션 뉴럴 네트워크(CNN)(예를 들어 Faster R-CNN)일 수 있고, 분석될 이미지의 주체 특징, 위치 특징 및 관계 특징을 추출하기 위한 것이다. 분석될 이미지의 모든 픽셀 포인트를 특징 추출 네트워크에 입력하고, ROI 풀링 전의 특징맵을 분석될 이미지의 전체 이미지 특징으로 한다.

[0054] 한 가지 가능한 구현 방식에서, 주체 특징에 관해, 분석될 이미지 중의 복수의 객체를 인식할 수 있고, 복수의 영역 중 어텐션 가중치가 가장 높은 객체를 주체 객체로 추출하며, 주체 객체의 영역의 특징맵을 주체 특징으로 결정한다. 예를 들어, 7x7의 특징맵을 추출하여 주체 특징으로 한다.

[0055] 한 가지 가능한 구현 방식에서, 위치 특징에 관해, 분석될 이미지 중의 복수의 객체가 위치하는 이미지 영역 사이의 상대적인 위치 편이량 및 상대 영역, 및 객체 자체의 위치 및 상대 영역에 따라, 위치 특징을 획득할 수

있다.

- [0056] 한 가지 가능한 구현 방식에서, 관계 특징에 관해, 영역 제안 중 평균값 풀링의 시각 특징, 상대적인 위치 편이량 및 상대 영역 사이의 연결에 따라, 상, 하문의 객체(복수의 객체) 사이의 관계 특징을 결정할 수 있다.
- [0057] 반드시 이해해야 하는 바, 본원 발명은 분석될 이미지의 주체 특징, 위치 특징 및 관계 특징을 추출하는 구체적인 방식에 대하여 한정하지 않는다.
- [0058] 한 가지 가능한 구현 방식에서, 단계 S12에서 상기 복수의 스테이트먼트 어텐션 가중치 및 분석될 이미지의 주체 특징, 위치 특징 및 관계 특징에 따라, 복수의 제1 매칭 스코어를 획득할 수 있다.
- [0059] 예를 들어 설명하면, 뉴럴 네트워크를 통해 복수의 제1 매칭 스코어를 획득할 수 있다. 상기 뉴럴 네트워크는 이미지 어텐션 네트워크를 포함할 수 있고, 상기 이미지 어텐션 네트워크는 주체 네트워크, 위치 네트워크 및 관계 네트워크를 포함한다. 여기서, 주체 네트워크, 위치 네트워크 및 관계 네트워크는 미리 구축되는 콘볼루션 뉴럴 네트워크(CNN)로 분류될 수 있다.
- [0060] 여기서, 주체 네트워크는 분석될 이미지 중 복수의 객체(사람, 동물, 물체 등)에서 가장 중요한 객체(주체 객체)와 분석될 디스크립션 스테이트먼트의 객체 디스크립션 사이의 매칭 정도를 평가하기 위한 것이고, 위치 네트워크는 분석될 이미지 중 복수의 객체의 상대적인 위치와 분석될 디스크립션 스테이트먼트의 위치 디스크립션 사이의 매칭 정도를 평가하기 위한 것이며, 관계 네트워크는 분석될 이미지 중 복수의 객체의 연관성과 분석될 디스크립션 스테이트먼트의 관련 디스크립션 사이의 매칭 정도를 평가하기 위한 것이다.
- [0061] 한 가지 가능한 구현 방식에서, 상기 복수의 스테이트먼트 어텐션 가중치 및 분석될 이미지의 주체 특징, 위치 특징 및 관계 특징을, 주체 네트워크, 위치 네트워크 및 관계 네트워크에 각각 입력하고 처리를 거쳐, 이미지 및 스테이트먼트 각양태의 매칭 정도를 평가할 수 있다.
- [0062] 여기서, 주체 객체는 분석될 이미지의 복수의 객체 중 어텐션 가중치가 가장 높은 객체이며, 주체 특징은 상기 주체 객체의 특징이고, 상기 위치 특징은 상기 복수의 객체의 위치 특징이며, 상기 관계 특징은 상기 복수의 객체 사이의 관계 특징이다.
- [0063] 한 가지 가능한 구현 방식에서, 단계 S12에서 획득된 복수의 제1 매칭 스코어는 주어 매칭 스코어, 위치 매칭 스코어 및 관계 매칭 스코어를 포함할 수 있다.
- [0064] 한 가지 가능한 구현 방식에서, 단계 S12는, 스테이트먼트 주어 가중치 및 주체 특징을 상기 주체 네트워크에 입력하고 처리를 거쳐, 주어 매칭 스코어를 획득하는 단계 스테이트먼트 위치 가중치 및 위치 특징을 상기 위치 네트워크에 입력하고 처리를 거쳐, 위치 매칭 스코어를 획득하는 단계 및 스테이트먼트 관계 가중치 및 관계 특징을 상기 관계 네트워크에 입력하고 처리를 거쳐, 관계 매칭 스코어를 획득하는 단계를 포함할 수 있다.
- [0065] 본 실시예에서, 스테이트먼트 주어 가중치 및 주체 특징을 주체 네트워크에 입력하고, 분석될 디스크립션 스테이트먼트의 주어와 분석될 이미지의 주체 객체 사이의 매칭 정도를 분석하여, 주어 매칭 스코어를 획득할 수 있으며 스테이트먼트 위치 가중치 및 위치 특징을 위치 네트워크에 입력하고, 분석될 디스크립션 스테이트먼트의 위치 워드와 분석될 이미지의 복수의 객체의 상대적인 위치 사이의 매칭 정도를 분석하여, 위치 매칭 스코어를 획득할 수 있고 스테이트먼트 관계 가중치 및 관계 특징을 관계 네트워크에 입력하고, 분석될 디스크립션 스테이트먼트의 관계 워드와 분석될 이미지의 복수의 객체의 연관성 사이의 매칭 정도를 분석하여, 관계 매칭 스코어를 획득할 수 있다.
- [0066] 예를 들어 설명하면, 복수의 스테이트먼트 어텐션 가중치(스테이트먼트 주어 가중치(q^{subj}), 스테이트먼트 위치 가중치(q^{loc}) 및 스테이트먼트 관계 가중치(q^{rel})), 및 복수의 객체 특징(주체 특징, 위치 특징, 관계 특징)을, 각각 주체 네트워크, 위치 네트워크 및 관계 네트워크에 입력하여 처리할 수 있다.
- [0067] 이러한 방식을 통해, 이미지와 디스크립션 스테이트먼트 각 양태의 매칭 정도를 결정할 수 있고, 매칭 판단의 정확성을 향상시켰다.
- [0068] 한 가지 가능한 구현 방식에서, 단계 S13에서 상기 복수의 제1 매칭 스코어 및 상기 복수의 이미지 어텐션 가중치에 따라, 상기 분석될 디스크립션 스테이트먼트와 상기 분석될 이미지 사이의 제2 매칭 스코어를 획득할 수 있다. 즉, 주어 매칭 스코어, 위치 매칭 스코어 및 관계 매칭 스코어 및 주체 객체 가중치(ω^{subj}), 객체 위치 가중치(ω^{loc}) 및 객체 관계 가중치(ω^{rel})에 따라, 상기 분석될 디스크립션 스테이트먼트와 상기 분석될 이미지

사이의 제2 매칭 스코어를 획득한다.

- [0069] 여기서, 단계 S13은,
- [0070] 상기 주체 객체 가중치, 상기 객체 위치 가중치 및 상기 객체 관계 가중치에 따라, 상기 주어 매칭 스코어, 상기 위치 매칭 스코어 및 상기 관계 매칭 스코어에 대해 가중 평균을 진행하여, 상기 제2 매칭 스코어를 결정하는 단계를 포함할 수 있다.
- [0071] 예를 들어 설명하면, 주어 매칭 스코어, 위치 매칭 스코어 및 관계 매칭 스코어를 획득한 후, 주체 객체 가중치(ω^{subj}), 객체 위치 가중치(ω^{loc}) 및 객체 관계 가중치(ω^{rel})에 따라, 주어 매칭 스코어, 위치 매칭 스코어 및 관계 매칭 스코어에 대해 각각 가중치를 진행하고, 가중치를 거친 후의 스코어 합계를 구한 후 평균값을 구한다. 상기 평균값을 분석될 디스크립션 스테이트먼트와 분석될 이미지 사이의 제1 매칭 스코어로 결정할 수 있다.
- [0072] 이러한 방식을 통해, 분석될 디스크립션 스테이트먼트와 분석될 이미지 사이의 정확한 매칭 스코어를 획득할 수 있다.
- [0073] 한 가지 가능한 구현 방식에서, 단계 S14에서 상기 제2 매칭 스코어에 따라 상기 분석될 이미지에서의 상기 분석될 디스크립션 스테이트먼트의 포지셔닝 결과를 결정할 수 있다. 즉, 제2 매칭 스코어를 획득한 후, 분석될 이미지에서의 분석될 디스크립션 스테이트먼트의 포지셔닝 결과를 부가적으로 결정할 수 있다. 여기서, 단계 S14는,
- [0074] 상기 제2 매칭 스코어가 기설정 임계값보다 크거나 같은 경우, 상기 주체 객체의 이미지 영역을 상기 분석될 디스크립션 스테이트먼트의 포지셔닝 위치로 결정하는 단계를 포함할 수 있다.
- [0075] 예를 들어 설명하면, 매칭 스코어의 임계값(예를 들어 기설정 임계값은 70점)을 미리 설정할 수 있는 바, 제2 매칭 스코어가 상기 기설정 임계값보다 크거나 같으면, 분석될 디스크립션 스테이트먼트는 분석될 이미지 중 주체 객체에 대한 디스크립션인 것으로 이해할 수 있고, 주체 객체가 위치한 이미지 영역을 분석될 디스크립션 스테이트먼트의 포지셔닝 위치로 결정할 수 있다. 반대로, 제2 매칭 스코어가 상기 기설정 임계값보다 작으면, 분석될 디스크립션 스테이트먼트가 분석될 이미지 중 주체 객체에 대한 디스크립션이 아닌 것으로 이해할 수 있으며, 포지셔닝 결과를 대응 불가로 결정할 수 있다. 본 기술분야의 통상의 기술자는 실제 상황에 따라 상기 기설정 임계값을 설정할 수 있고, 본원 발명은 기설정 임계값의 구체적인 값에 대해서 한정하지 않음을 반드시 이해해야 한다.
- [0076] 한 가지 가능한 구현 방식에서, 분석될 이미지에서 복수의 주체 객체를 각각 설정할 수 있고, 각각의 주체 객체의 주체 특징을 이미지 어텐션 네트워크에 각각 입력하고 처리를 거쳐, 각각의 주체 객체의 제2 매칭 스코어를 결정하며, 복수의 제2 매칭 스코어 중의 최고 스코어를 결정할 수 있다. 상기 상황에서, 분석될 디스크립션 스테이트먼트는 최고 스코어에 대응되는 주체 객체의 디스크립션인 것으로 이해할 수 있고, 상기 주체 객체가 위치한 이미지 영역을 분석될 디스크립션 스테이트먼트의 포지셔닝 위치로 결정할 수 있다.
- [0077] 이러한 방식을 통해, 분석될 이미지에서의 분석될 디스크립션 스테이트먼트의 정확한 포지셔닝을 실현할 수 있다.
- [0078] 도 2는 본원 발명의 실시예에 따른 뉴럴 네트워크의 모식도이다. 도 2에도시된 바와 같이, 상기 뉴럴 네트워크는 언어 어텐션 네트워크(21) 및 이미지 어텐션 네트워크를 포함할 수 있고, 이미지 어텐션 네트워크는 주체 네트워크(22), 위치 네트워크(23) 및 관계 네트워크(24)를 포함한다.
- [0079] 상기 구현예에서, 분석될 디스크립션 스테이트먼트 "여자 아이가 타고 있는 가운테의 갈색 말"(201)과 분석될 이미지(202)를 언어 어텐션 네트워크 21에 입력하고 처리를 거쳐, 세계의 이미지 어텐션 가중치(주체 객체 가중치(ω^{subj}), 객체 위치 가중치(ω^{loc}) 및 객체 관계 가중치(ω^{rel}))를 출력할 수 있고, 아울러 세계의 스테이트먼트 어텐션 가중치(스테이트먼트 주어 가중치(q^{subj}), 스테이트먼트 위치 가중치(q^{loc}) 및 스테이트먼트 관계 가중치(q^{rel}))를 출력한다.
- [0080] 상기 구현예에서, 특징 추출 네트워크(미도시)를 통해 분석될 이미지의 주체 특징(203), 위치 특징(204) 및 관계 특징(205)을 획득할 수 있다.
- [0081] 상기 구현예에서, 스테이트먼트 주어 가중치(q^{subj}) 및 주체 특징(203)을 주체 네트워크(22)에 입력하고 처리를

거쳐, 주어 매칭 스코어를 획득할 수 있고, 스테이트먼트 위치 가중치(q^{loc}) 및 위치 특징(204)을 위치 네트워크(23)에 입력하고 처리를 거쳐, 위치 매칭 스코어를 획득할 수 있으며, 스테이트먼트 관계 가중치(q^{rel}) 및 관계 특징(205)을 관계 네트워크(24)에 입력하고 처리를 거쳐, 관계 매칭 스코어를 획득할 수 있다.

- [0082] 상기 구현예에서, 주체 객체 가중치(ω^{subj}), 객체 위치 가중치(ω^{loc}) 및 객체 관계 가중치(ω^{rel})에 따라, 주어 매칭 스코어, 위치 매칭 스코어 및 관계 매칭 스코어에 대해 가중치를 진행할 수 있고, 가중치를 거친 후의 스코어의 합계를 구한 후평균하여, 제2 매칭 스코어(206)를 획득하며, 제2 매칭 스코어(206)에 따라 분석될 이미지에서의 분석될 디스크립션 스테이트먼트의 포지셔닝 결과를 결정하여, 단계 S11 내지 S14의 전반적인 실현 과정을 완성한다.
- [0083] 상기 내용은 단지 본원 발명의 방법의 뉴럴 네트워크의 한구현예로서, 본원 발명은 뉴럴 네트워크의 구체적인 타입을 한정하지 않는다.
- [0084] 한 가지 가능한 구현 방식에서, 단계 S11 이전에, 상기 이미지의 디스크립션 스테이트먼트 포지셔닝 방법은, 샘플 세트를 이용하여 상기 뉴럴 네트워크를 트레이닝하는 단계를 더 포함하고, 상기 샘플 세트는 복수의 포지티브 샘플 페어 및 복수의 네거티브 샘플 페어를 포함한다.
- [0085] 여기서, 각각의 포지티브 샘플 페어는 제1 샘플 이미지 및 그의 제1 샘플 디스크립션 스테이트먼트를 포함하고,
- [0086] 각각의 네거티브 샘플 페어는 제1 샘플 이미지 및 상기 제1 샘플 디스크립션 스테이트먼트에서 워드를 제거한 후의 제2 샘플 디스크립션 스테이트먼트, 또는 제1 샘플 디스크립션 스테이트먼트 및 상기 제1 샘플 이미지에서 영역을 제거한 후의 제2 샘플 이미지를 포함한다.
- [0087] 한 가지 가능한 구현 방식에서, 어텐션 가이드에 기반한 크로스-모듈 제거 방법을 통해, 하이 어텐션 가중치의 시각 또는 텍스트 정보를 제거하여, 제거한 후의 트레이닝 샘플(제2 샘플 디스크립션 스테이트먼트 및 제2 샘플 이미지)을 획득하여, 트레이닝 정밀도를 향상시킨다.
- [0088] 예를 들어 설명하면, 복수의 트레이닝 샘플을 포함하는 샘플 세트를 미리 설정하여, 뉴럴 네트워크에 대해 트레이닝을 진행할 수 있다. 샘플 세트는 복수의 포지티브 샘플 페어를 포함하고, 각각의 포지티브 샘플 페어는 제1 샘플 이미지(0) 및 그 제1 샘플 디스크립션 스테이트먼트(Q)를 포함한다. 제1 샘플 이미지 중의 객체에 대해 디스크립션을 진행하는 스테이트먼트를 동일한 포지티브 샘플 페어 중의 제1 샘플 디스크립션 스테이트먼트로 할 수 있다. 샘플 세트는 복수의 네거티브 샘플 페어를 더 포함할 수 있고, 각각의 네거티브 샘플 페어는 제1 샘플 이미지 및 제1 샘플 디스크립션 스테이트먼트에서 워드를 제거한 후의 제2 샘플 디스크립션 스테이트먼트, 또는 제1 샘플 디스크립션 스테이트먼트 및 상기 제1 샘플 이미지에서 영역을 제거한 후의 제2 샘플 이미지를 포함한다. 본원 발명은 샘플 세트의 구체적인 구축 방식을 한정하지 않으며, 본원 발명은 각각의 샘플 페어 중의 샘플 이미지와 샘플 디스크립션 스테이트먼트 사이의 전후 순서에 대해 한정하지 않는다.
- [0089] 한 가지 가능한 구현 방식에서, 상기 방법은,
- [0090] 상기 포지티브 샘플 페어의 제1 샘플 디스크립션 스테이트먼트 및 제1 샘플 이미지를 상기 언어 어텐션 네트워크에 입력하여, 상기 제1 샘플 디스크립션 스테이트먼트의 복수의 워드의 어텐션 가중치를 획득하는 단계
- [0091] 사전 결정된 식별자를 사용하여 상기 제1 샘플 디스크립션 스테이트먼트 중 어텐션 가중치가 가장 높은 워드를 교체하여, 제2 샘플 디스크립션 스테이트먼트를 획득하는 단계 및
- [0092] 상기 제1 샘플 이미지 및 상기 제2 샘플 디스크립션 스테이트먼트를 네거티브 샘플 페어로 하는 단계를 더 포함할 수 있다.
- [0093] 한 가지 가능한 구현 방식에서, 언어 어텐션 네트워크를 통해 공간 어텐션 가이드를 진행하여, 가장 중요한 텍스트 정보를 제거하여 비교적 어려운 텍스트 트레이닝 샘플을 획득하여, 뉴럴 네트워크가 지나치게 특정 텍스트 정보(워드)에 의존하는 것을 방지하여, 트레이닝을 거쳐 획득된 뉴럴 네트워크의 정밀도를 향상한다.
- [0094] 도 3은 본원 발명의 실시예에 따른 제2 샘플 디스크립션 스테이트먼트를 획득하는 단계의 모식도이다. 예를 들어 설명하면, 도3에 도시된 바와 같이, 포지티브 샘플 페어의 제1 샘플 디스크립션 스테이트먼트(예를 들면 "여자 아이가 타고 있는 가운데의 갈색 말") 및 제1 샘플 이미지(예를 들어 말을 타고 있는 복수의 사람의 사진을 포함함)를 언어 어텐션 네트워크에 입력하여, 제1 샘플 디스크립션 스테이트먼트의 복수의 워드의 어텐션 가중치를 획득할 수 있다. 각 워드의 어텐션 가중치에 따라, 어텐션 가중치가 가장 높은 워드(예를 들어 "가운데")

를 결정할 수 있다. 상기 워드 "가운데"를 직접 제거하면 문법에 착오가 생겨 인식할 수 없기에, 따라서, 하나의 미지의 식별자로 상기 워드 "가운데"를 교체하여, 제2 샘플 디스크립션 스테이트먼트 Q*(여자 아이가 타고 있는 "미지"의 갈색 말)를 획득함으로써, 상기 제1 샘플 이미지 및 상기 제2 샘플 디스크립션 스테이트먼트를 네거티브 샘플 페어로 할 수 있다.

- [0095] 한 가지 가능한 구현 방식에서, 상기 방법은,
- [0096] 상기 포지티브 샘플 페어의 제1 샘플 디스크립션 스테이트먼트 및 제1 샘플 이미지를 상기 이미지 어텐션 네트워크에 입력하여, 상기 제1 샘플 이미지의 어텐션 가중치를 획득하는 단계
- [0097] 상기 제1 샘플 이미지 중 어텐션 가중치가 가장 높은 이미지 영역을 제거하여, 제2 샘플 이미지를 획득하는 단계 및
- [0098] 상기 제2 샘플 이미지 및 상기 제1 샘플 디스크립션 스테이트먼트를 네거티브 샘플 페어로 하는 단계를 더 포함할 수 있다.
- [0099] 한 가지 가능한 구현 방식에서, 이미지 어텐션 네트워크를 통해 가장 중요한 시각 정보를 인식 및 제거할 수 있고, 비교적 어려운 이미지 트레이닝 샘플을 획득하여, 뉴럴 네트워크가 특정 시각 정보에 지나치게 의존하는 것을 방지하여, 트레이닝하여 획득되는 뉴럴 네트워크의 정밀도를 향상시킨다.
- [0100] 도 4는 본원 발명의 실시예에 따른 제2 샘플 이미지를 획득하는 단계의 모식도이다. 예를 들어 설명하면, 도 4에 도시된 바와 같이, 포지티브 샘플 페어의 제1 샘플 이미지(예를 들어 말을 타고 있는 복수의 사람의 사진을 포함함) 및 제1 샘플 디스크립션 스테이트먼트(예를 들면 "여자 아이가 타고 있는 가운데의 갈색 말")를 이미지 어텐션 네트워크에 입력하여 처리할 수 있다. 여기서, 이미지 어텐션 네트워크의 주체 네트워크를 사용할 수 있고, 위치 네트워크 또는 관계 네트워크를 사용할 수도 있으며, 본원 발명은 이에 한하지 않는다.
- [0101] 한 가지 가능한 구현 방식에서, 제1 샘플 이미지 및 제1 샘플 디스크립션 스테이트먼트를 주체 네트워크에 입력하여, 제1 샘플 이미지 각영역의 어텐션 가중치를 획득할 수 있다. 각 영역의 어텐션 가중치에 따라, 어텐션 가중치가 가장 높은 객체가 가장 높은 객체 영역(예를 들어 가운데의 여자 아이가 위치한 이미지의 영역)을 결정할 수 있다. 제1 샘플 이미지에서 상기 객체 영역을 제거하여, 제2 샘플 이미지(0*)(도 4에 도시됨)를 획득함으로써, 상기 제2 샘플 이미지 및 상기 제1 샘플 디스크립션 스테이트먼트를 네거티브 샘플 페어로 할 수 있다.
- [0102] 한 가지 가능한 구현 방식에서, 샘플 세트를 이용하여 상기 뉴럴 네트워크를 트레이닝하는 단계는 상기 뉴럴 네트워크의 제1 손실 및 제2 손실에 따라, 상기 뉴럴 네트워크의 전체 손실을 결정하는 단계를 포함할 수 있다.
- [0103] 한 가지 가능한 구현 방식에서, 포지티브 샘플 페어(제1 샘플 이미지 및 그 제1 샘플 디스크립션 스테이트먼트)의 네트워크 손실을 획득하여 제1 손실로 할 수 있다. 제거한 후의 네거티브 샘플 페어(제2 샘플 이미지 및 제1 샘플 디스크립션 스테이트먼트, 또는 제1 샘플 이미지 및 제2 샘플 디스크립션 스테이트먼트)의 네트워크 손실을 획득한다.
- [0104] 한 가지 가능한 구현 방식에서, 샘플 세트를 이용하여 상기 뉴럴 네트워크를 트레이닝하는 단계는, 상기 전체 손실에 따라, 상기 뉴럴 네트워크를 트레이닝하는 단계를 더 포함할 수 있다.
- [0105] 한 가지 가능한 구현 방식에서, 전체 네트워크 손실(L)을 획득한 후, 전체 네트워크 손실(L)에 따라 상기 뉴럴 네트워크를 트레이닝할 수 있다. 이로써 트레이닝한 후의 뉴럴 네트워크를 결정한다. 본원 발명은 뉴럴 네트워크의 구체적인 트레이닝 방식에 대해 한정하지 않는다.
- [0106] 한 가지 가능한 구현 방식에서, 상기 뉴럴 네트워크의 제1 손실 및 제2 손실에 따라, 상기 뉴럴 네트워크의 전체 손실을 결정하는 단계 이전에, 상기 제1 손실을 획득하는 단계를 더 포함한다.
- [0107] 상기 제1 손실을 획득하는 단계는,
- [0108] 동일한 포지티브 샘플 페어의 제1 샘플 이미지 및 제1 샘플 디스크립션 스테이트먼트를 상기 뉴럴 네트워크에 입력하고 처리를 거쳐, 제1 트레이닝 스코어를 획득하는 단계 상이한 포지티브 샘플 페어의 제1 샘플 이미지 및 제1 샘플 디스크립션 스테이트먼트를 상기 뉴럴 네트워크에 입력하고 처리를 거쳐, 제2 트레이닝 스코어를 획득하는 단계 및 복수의 제1 트레이닝 스코어 및 복수의 제2 트레이닝 스코어에 따라, 제1 손실을 획득하는 단계를 포함한다.
- [0109] 예를 들어 설명하면, 포지티브 샘플 페어(제1 샘플 이미지 및 그 제1 샘플 디스크립션 스테이트먼트)의 네트워

크 손실을 획득할 수 있다. 트레이닝 세트 중 임의의 하나의 포지티브 샘플 페어에 대해, 동일한 포지티브 샘플 페어(O_i, Q_i)의 제1 샘플 이미지(O_i) 및 제1 샘플 디스크립션 스테이트먼트(Q_i)를 도2에 도시된 뉴럴 네트워크에 입력하고 처리를 거쳐, 제1 트레이닝 스코어($s(O_i, Q_i)$)를 획득한다. 여기서, i 는 샘플 번호이고, $1 \leq i \leq N$, N 은 샘플 세트 중의 포지티브 샘플 페어의 개수이다.

[0110] 한 가지 가능한 구현 방식에서, 상이한 포지티브 샘플 페어의 제1 샘플 이미지 및 이와 대응되지 않는 제1 샘플 디스크립션 스테이트먼트(O_i, Q_j)에 대해, 이를 도 2에 도시된 뉴럴 네트워크에 입력하고 처리를 거쳐, 제2 트레이닝 스코어 $s(O_i, Q_j)$ 를 획득할 수 있다. 여기서, j 는 샘플 번호이고, $1 \leq j \leq N$ 이며, j 는 i 와 같지 않다. 마찬가지로, 상이한 포지티브 샘플 페어의 제1 샘플 이미지 및 제1 샘플 디스크립션 스테이트먼트(O_j, Q_i)를 뉴럴 네트워크에 입력하여, 다른 한 제2 트레이닝 스코어 $s(O_j, Q_i)$ 를 획득할 수 있다.

[0111] 한 가지 가능한 구현 방식에서, 트레이닝 세트 중의 포지티브 샘플 페어(제1 샘플 이미지 및 제1 샘플 디스크립션 스테이트먼트)에 대해 각각 처리하여, 복수의 제1 트레이닝 스코어 및 복수의 제2 트레이닝 스코어를 획득할 수 있으며, 원시 샘플의 제1 손실 L_{rank} 을 획득할 수 있는 바,

$$L_{rank} = \sum_{i,j}^N ([m - s(O_i, Q_i) + s(O_i, Q_j)]_+ + [m - s(O_i, Q_i) + s(O_j, Q_i)]_+) \quad (1)$$

[0112] 공식 (1)에서, 연산자 $[x]_+$ 는 x 와 0 사이의 가장 큰 값을 표시할 수 있고, x 가 0보다 클 경우 x 의 값을 취하며, x 가 0보다 작거나 같을 경우 0을 취하고 m 은 상수일 수 있으며, 네트워크 손실의 간격을 표시한다. 본 기술분야의 통상의 기술자는 실제 상황에 따라 m 의 값(예를 들어 0.1)을 설정할 수 있으며, 본원 발명은 m 의 구체적인 값에 대해 한정하지 않음을 반드시 이해해야 한다.

[0114] 한 가지 가능한 구현 방식에서, 상기 뉴럴 네트워크의 제1 손실 및 제2 손실에 따라, 상기 뉴럴 네트워크의 전체 손실을 결정하는 단계 이전에, 상기 제2 손실을 획득하는 단계를 더 포함하고

[0115] 상기 제2 손실을 획득하는 단계는,

[0116] 동일한 네거티브 샘플 페어의 제2 샘플 이미지 및 제1 샘플 디스크립션 스테이트먼트를 상기 뉴럴 네트워크에 입력하고 처리를 거쳐, 제3 트레이닝 스코어를 획득하는 단계 상이한 네거티브 샘플 페어의 제2 샘플 이미지 및 제1 샘플 디스크립션 스테이트먼트를 상기 뉴럴 네트워크에 입력하고 처리를 거쳐, 제4 트레이닝 스코어를 획득하는 단계 동일한 네거티브 샘플 페어의 제1 샘플 이미지 및 제2 샘플 디스크립션 스테이트먼트를 상기 뉴럴 네트워크에 입력하고 처리를 거쳐, 제5 트레이닝 스코어를 획득하는 단계 상이한 네거티브 샘플 페어의 제1 샘플 이미지 및 제2 샘플 디스크립션 스테이트먼트를 상기 뉴럴 네트워크에 입력하고 처리를 거쳐, 제6 트레이닝 스코어를 획득하는 단계 및 복수의 제3 트레이닝 스코어, 복수의 제4 트레이닝 스코어, 복수의 제5 트레이닝 스코어 및 복수의 제6 트레이닝 스코어에 따라, 제2 손실을 획득하는 단계를 포함한다.

[0117] 예를 들어 설명하면, 제거한 후의 네거티브 샘플(제2 샘플 이미지 및 제2 샘플 디스크립션 스테이트먼트)의 네트워크 손실을 획득할 수 있다. 트레이닝 세트 중의 동일한 네거티브 샘플 페어 (O_i^*, Q_i) 에 대해, 제2 샘플 이미지 O_i^* 및 제1 샘플 디스크립션 스테이트먼트 Q_i 를 도 2에 도시된 뉴럴 네트워크에 입력하여 처리를 거쳐, 제3

트레이닝 스코어 $s(O_i^*, Q_i)$ 를 획득할 수 있다. 여기서, i 는 샘플 번호이고, $1 \leq i \leq N$, N 은 샘플 세트 중의 샘플 페어의 개수이다.

[0118] 한 가지 가능한 구현 방식에서, 트레이닝 세트 중 상이한 네거티브 샘플 페어(제2 샘플 이미지 O_i^* 및 대응되지 않는 제1 샘플 디스크립션 스테이트먼트 Q_i)에 대해, 이를 도 2에 도시된 뉴럴 네트워크에 입력하여 처리를 거쳐,

제4 트레이닝 스코어 $s(O_i^*, Q_j)$ 를 획득한다. 여기서, j 는 샘플 번호이고, $1 \leq j \leq N$, j 는 i 와 같지 않다.

[0119] 마찬가지로, 동일한 네거티브 샘플 페어 (O_i, Q_i^*) 의 제1 샘플 이미지 및 대응되는 제2 샘플 디스크립션 스테이트먼트를 뉴럴 네트워크에 입력하여, 제5 트레이닝 스코어 $s(O_i, Q_i^*)$ 를 획득할 수 있고 상이한 네거티브 샘플 페어 (O_j, Q_i^*) 의 제1 샘플 이미지 및 제2 샘플 디스크립션 스테이트먼트를 뉴럴 네트워크에 입력하여, 제6 트레이닝 스코어 $s(O_j, Q_i^*)$ 를 획득할 수 있다.

[0120] 한 가지 가능한 구현 방식에서, 트레이닝 세트 중의 복수의 포지티브 샘플 페어(제1 샘플 이미지 및 제1 샘플 디스크립션 스테이트먼트) 및 제거한 후의 네거티브 샘플 페어를 각각 처리하여, 복수의 제3 트레이닝 스코어, 복수의 제4 트레이닝 스코어, 복수의 제5 트레이닝 스코어 및 복수의 제6 트레이닝 스코어를 획득함으로써, 제거한 후의 샘플의 제2 손실 L_{erase} 을 획득할 수 있는 바,

$$L_{erase} = \sum_{i,j}^N ([m - s(O_i^*, Q_i) + s(O_i^*, Q_j)]_+ + [m - s(O_i, Q_i^*) + s(O_j, Q_i^*)]_+) \quad (2)$$

[0121] 공식 (2)에서, 연산자 $[x]_+$ 는 x 와 0 사이의 가장 큰 값을 표시할 수 있고, x 가 0보다 클 경우 x 의 값을 취하며, x 가 0보다 작거나 같을 경우 0을 취하고 m 은 상수일 수 있으며, 네트워크 손실의 간격을 표시한다. 본 기술분야의 통상의 기술자는 실제 상황에 따라 m 의 값(예를 들어 0.1)을 설정할 수 있으며, 본원 발명은 m 의 구체적인 값에 대해 한정하지 않음을 반드시 이해해야 한다.

[0122] 한 가지 가능한 구현 방식에서, 제1 손실 및 제2 손실을 결정한 후, 제1 손실 및 제2 손실에 따라 상기 뉴럴 네트워크의 전체 손실을 결정할 수 있고, 전체 손실에 따라 상기 뉴럴 네트워크를 트레이닝한다.

[0123] 여기서, 상기 뉴럴 네트워크의 제1 손실 및 제2 손실에 따라, 상기 뉴럴 네트워크의 전체 손실을 결정하는 단계는, 상기 제1 손실 및 상기 제2 손실에 대해 가중 오버랩을 진행하여, 상기 뉴럴 네트워크의 전체 손실을 획득하는 단계를 포함할 수 있다.

[0124] 예를 들어 설명하면, 하기의 공식으로 뉴럴 네트워크의 전체 네트워크 손실(L)을 산출할 수 있는 바,

$$L = \beta L_{erase} + \gamma L_{rank} \quad (3)$$

[0125] 공식 (3)에서, β 및 γ 는 각각 제1 손실 및 제2 손실의 가중치를 표시한다. 본 기술분야의 통상의 기술자는 실제 상황에 따라 β 및 γ 의 값을 설정할 수 있고, 본원 발명을 β 및 γ 의 구체적인 값에 대해 한정하지 않음을 반드시 이해해야 한다.

[0126] 한 가지 가능한 구현 방식에서, 전체 네트워크 손실(L)을 획득한 후, 전체 네트워크 손실(L)에 따라 상기 뉴럴 네트워크를 트레이닝할 수 있다. 예를 들어, 전체 네트워크 손실(L)에 따라, 역방향 경사법을 사용하여 상기 뉴럴 네트워크의 네트워크 파라미터 값을 조절할 수 있고 전체 네트워크 손실(L)을 제차 획득할 수 있다. 수차례 조절을 거쳐, 기설정된 트레이닝 조건(네트워크 손실(L)이 트레이닝 횟수에 수렴 또는 도달함)에 도달할 경우, 트레이닝한 후의 뉴럴 네트워크를 결정할 수 있다. 본원 발명은 뉴럴 네트워크의 구체적인 트레이닝 방식에 대해 한정하지 않는다.

[0127] 본원 발명의 실시예에 따른 이미지의 디스크립션 스테이트먼트 포지셔닝 방법은, 크로스-모드 제거 방식을 통해 하이 어텐션 가중치의 가장 주요한 시각 또는 텍스트 정보를 제거하여, 어려운 트레이닝 샘플을 생성함으로써, 뉴럴 네트워크 모형이 가장 주요한 증거 이외의 보충 증거를 찾도록 구동한다. 본원 발명의 실시예에 따르면, 원시 조회 스테이트먼트의 제거 이미지, 또는 원시 이미지의 제거 조회 스테이트먼트를 이용하여 비교적 어려운 트레이닝 샘플을 형성하여, 뉴럴 네트워크 모형이 트레이닝 데이터를 잠재된 문자-사진 대응관계에 더 잘이용하도록 하며, 또한 추리의 복잡성을 증가시키지 않는다.

[0128] 본원 발명의 실시예에 따르면, 로봇 또는 휴대폰 등 단말기에 응용될 수 있으며, 사람의 지도(문자 또는 음성)

에따라 이미지 중인물의 위치를 포지셔닝하여, 텍스트와 이미지 사이의 정확한 대응을 실현한다.

- [0131] 본 기술분야의 통상의 기술자는 구체적인 실시 형태의 상기 방법에서, 각 단계의 기재 순서는 엄격한 수행 순서를 의미하지 않으며 실시 과정에 임의의 한정을 구성하지 않으며, 각 단계의 구체적인 수행 순서는 그기능 및 가능한 내재적인 논리에 의해 결정되어야 함을 이해할 수 있다.
- [0132] 본원 발명에 언급된 상기 각 방법 실시예는 원리 논리를 벗어나지 않는 상황하에서 모두 서로 결합하여 구현한 후의 실시예를 형성할 수 있으며, 편폭의 제한으로 인해 본원 발명은 더서술하지 않음을 이해해야 한다.
- [0133] 도 5는 본원 발명의 실시예에 따른 이미지의 디스크립션 스테이트먼트 포지셔닝 장치의 블록도이고, 도 5에도시된 바와 같이, 상기 이미지의 디스크립션 스테이트먼트 포지셔닝 장치는,
- [0134] 분석될 디스크립션 스테이트먼트와 분석될 이미지를 분석 처리하여, 상기 분석될 디스크립션 스테이트먼트의 복수의 스테이트먼트 어텐션 가중치 및 상기 분석될 이미지의 복수의 이미지 어텐션 가중치를 획득하도록 구성되는 제1 가중치 획득 모듈(51);
- [0135] 상기 복수의 스테이트먼트 어텐션 가중치 및 분석될 이미지의 주체 특징, 위치 특징 및 관계 특징에 따라, 복수의 제1 매칭 스코어를 획득하도록 구성되는 제1 스코어 획득 모듈(52) - 상기 분석될 이미지는 복수의 객체를 포함하고, 주체 객체는 상기 복수의 객체 중어텐션 가중치가 가장 높은 객체이며, 상기 주체 특징은 상기 주체 객체의 특징이고, 상기 위치 특징은 상기 복수의 객체의 위치 특징이며, 상기 관계 특징은 상기 복수의 객체 사이의 관계 특징임 - ;
- [0136] 상기 복수의 제1 매칭 스코어 및 상기 복수의 이미지 어텐션 가중치에 따라, 상기 분석될 디스크립션 스테이트먼트와 상기 분석될 이미지 사이의 제2 매칭 스코어를 획득하도록 구성되는 제2 스코어 획득 모듈(53); 및
- [0137] 상기 제2 매칭 스코어에 따라, 상기 분석될 이미지에서의 상기 분석될 디스크립션 스테이트먼트의 포지셔닝 결과를 결정하도록 구성되는 결과 결정 모듈(54)을포함한다.
- [0138] 한 가지 가능한 구현 방식에서, 상기 제1 가중치 획득 모듈은,
- [0139] 상기 분석될 이미지에 대해 특징 추출을 진행하여, 상기 분석될 이미지의 이미지 특징 벡터를 획득하도록 구성되는 이미지 특징 추출 서브 모듈
- [0140] 상기 분석될 디스크립션 스테이트먼트에 대해 특징 추출을 진행하여, 상기 분석될 디스크립션 스테이트먼트의 복수의 워드의 워드 임베딩 벡터를 획득하도록 구성되는 워드 특징 추출 서브 모듈 및
- [0141] 상기 이미지 특징 벡터 및 상기 복수의 워드의 워드 임베딩 벡터에 따라, 상기 분석될 디스크립션 스테이트먼트의 복수의 스테이트먼트 어텐션 가중치 및 상기 분석될 이미지의 복수의 이미지 어텐션 가중치를 획득하도록 구성되는 제1 가중치 획득 서브 모듈을 포함한다.
- [0142] 한 가지 가능한 구현 방식에서, 상기 이미지의 디스크립션 스테이트먼트 포지셔닝 장치는, 뉴럴 네트워크를 통해 상기 분석될 디스크립션 스테이트먼트의 복수의 스테이트먼트 어텐션 가중치 및 상기 분석될 이미지의 복수의 이미지 어텐션 가중치를 획득하도록 구성되는 제2 가중치 획득 모듈을 더 포함한다.
- [0143] 한 가지 가능한 구현 방식에서, 상기 복수의 스테이트먼트 어텐션 가중치는 스테이트먼트 주어 가중치, 스테이트먼트 위치 가중치 및 스테이트먼트 관계 가중치를 포함하고, 상기 뉴럴 네트워크는 이미지 어텐션 네트워크를 포함하며, 상기 이미지 어텐션 네트워크는 주체 네트워크, 위치 네트워크 및 관계 네트워크를 포함하고, 상기 복수의 제1 매칭 스코어는 주어 매칭 스코어, 위치 매칭 스코어 및 관계 매칭 스코어를 포함하며, 상기 제1 스코어 획득 모듈은,
- [0144] 상기 스테이트먼트 주어 가중치 및 주체 특징을 상기 주체 네트워크에 입력하고 처리를 거쳐, 상기 주어 매칭 스코어를 획득하도록 구성되는 제1 스코어 획득 서브 모듈
- [0145] 상기 스테이트먼트 위치 가중치 및 위치 특징을 상기 위치 네트워크에 입력하고 처리를 거쳐, 상기 위치 매칭 스코어를 획득하도록 구성되는 제2 스코어 획득 서브 모듈 및
- [0146] 상기 스테이트먼트 관계 가중치 및 관계 특징을 상기 관계 네트워크에 입력하고 처리를 거쳐, 상기 관계 매칭 스코어를 획득하도록 구성되는 제3 스코어 획득 서브 모듈을 포함한다.
- [0147] 한 가지 가능한 구현 방식에서, 상기 복수의 이미지 어텐션 가중치는 주체 객체 가중치, 객체 위치 가중치 및

객체 관계 가중치를 포함하고, 상기 제2 스코어 획득 모듈은,

- [0148] 상기 주체 객체 가중치, 상기 객체 위치 가중치 및 상기 객체 관계 가중치에 따라, 상기 주어 매칭 스코어, 상기 위치 매칭 스코어 및 상기 관계 매칭 스코어에 대해 가중 평균을 진행하여, 상기 제2 매칭 스코어를 결정하는 제4 스코어 획득 서브 모듈을 포함한다.
- [0149] 한 가지 가능한 구현 방식에서, 상기 이미지의 디스크립션 스테이트먼트 포지셔닝 장치는,
- [0150] 상기 분석될 이미지를 특징 추출 네트워크에 입력하고 처리를 거쳐, 상기 주체 특징, 상기 위치 특징 및 상기 관계 특징을 획득하는 제3 가중치 획득 모듈을 더 포함한다.
- [0151] 한 가지 가능한 구현 방식에서, 상기 결과 결정 모듈은,
- [0152] 상기 제2 매칭 스코어가 기설정 임계값보다 크거나 같은 경우, 상기 주체 객체의 이미지 영역을 상기 분석될 디스크립션 스테이트먼트의 포지셔닝 위치로 결정하는 위치 결정 서브 모듈을 포함한다.
- [0153] 한 가지 가능한 구현 방식에서, 상기 제2 가중치 획득 모듈 이전에, 샘플 세트를 이용하여 상기 뉴럴 네트워크를 트레이닝하는 트레이닝 모듈을 더 포함하고, 상기 샘플 세트는 복수의 포지티브 샘플 페어 및 복수의 네거티브 샘플 페어를 포함하고,
- [0154] 여기서, 각각의 포지티브 샘플 페어는 제1 샘플 이미지 및 그 제1 샘플 디스크립션 스테이트먼트를 포함하며,
- [0155] 각각의 네거티브 샘플 페어는 제1 샘플 이미지 및 상기 제1 샘플 디스크립션 스테이트먼트에서 워드를 제거한 후의 제2 샘플 디스크립션 스테이트먼트, 또는 제1 샘플 디스크립션 스테이트먼트 및 상기 제1 샘플 이미지에서 영역을 제거한 후의 제2 샘플 이미지를 포함한다.
- [0156] 한 가지 가능한 구현 방식에서, 상기 뉴럴 네트워크는 언어 어텐션 네트워크를 더 포함하고, 상기 이미지의 디스크립션 스테이트먼트 포지셔닝 장치는,
- [0157] 상기 포지티브 샘플 페어의 제1 샘플 디스크립션 스테이트먼트 및 제1 샘플 이미지를 상기 언어 어텐션 네트워크에 입력하여, 상기 제1 샘플 디스크립션 스테이트먼트의 복수의 워드의 어텐션 가중치를 획득하는 워드 가중치 결정 모듈
- [0158] 사전 결정된 식별자를 사용하여 상기 제1 샘플 디스크립션 스테이트먼트 중 어텐션 가중치가 가장 높은 워드를 교체하여, 제2 샘플 디스크립션 스테이트먼트를 획득하는 워드 교체 모듈 및
- [0159] 상기 제1 샘플 이미지 및 상기 제2 샘플 디스크립션 스테이트먼트를 네거티브 샘플 페어로 하는 제1 네거티브 샘플 페어 결정 모듈을 더 포함한다.
- [0160] 한 가지 가능한 구현 방식에서, 상기 이미지의 디스크립션 스테이트먼트 포지셔닝 장치는,
- [0161] 상기 포지티브 샘플 페어의 제1 샘플 디스크립션 스테이트먼트 및 제1 샘플 이미지를 상기 이미지 어텐션 네트워크에 입력하여, 상기 제1 샘플 이미지의 어텐션 가중치를 획득하는 이미지 가중치 결정 모듈
- [0162] 상기 제1 샘플 이미지 중 어텐션 가중치가 가장 높은 이미지 영역을 제거하여, 제2 샘플 이미지를 획득하는 영역 제거 모듈 및
- [0163] 상기 제2 샘플 이미지 및 상기 제1 샘플 디스크립션 스테이트먼트를 네거티브 샘플 페어로 하도록 구성되는 제2 네거티브 샘플 페어 결정 모듈을 더 포함한다.
- [0164] 한 가지 가능한 구현 방식에서, 상기 트레이닝 모듈은,
- [0165] 상기 뉴럴 네트워크의 제1 손실 및 제2 손실에 따라, 상기 뉴럴 네트워크의 전체 손실을 결정하도록 구성되는 전체 손실 결정 서브 모듈 및
- [0166] 상기 전체 손실에 따라, 상기 뉴럴 네트워크를 트레이닝하도록 구성되는 트레이닝 서브 모듈을 포함한다.
- [0167] 한 가지 가능한 구현 방식에서, 상기 이미지의 디스크립션 스테이트먼트 포지셔닝 장치는, 상기 전체 손실 결정 서브 모듈 이전에 상기 제1 손실을 획득하도록 구성되는 제1 손실 획득 서브 모듈을 더 포함하고 상기 제1 손실 획득 서브 모듈은,
- [0168] 동일한 포지티브 샘플 페어의 제1 샘플 이미지 및 제1 샘플 디스크립션 스테이트먼트를 상기 뉴럴 네트워크에 입력하고 처리를 거쳐, 제1 트레이닝 스코어를 획득하고

- [0169] 상이한 포지티브 샘플 페어의 제1 샘플 이미지 및 제1 샘플 디스크립션 스테이트먼트를 상기 뉴럴 네트워크에 입력하고 처리를 거쳐, 제2 트레이닝 스코어를 획득하며
- [0170] 복수의 제1 트레이닝 스코어 및 복수의 제2 트레이닝 스코어에 따라, 제1 손실을 획득하도록 구성된다.
- [0171] 한 가지 가능한 구현 방식에서, 상기 이미지의 디스크립션 스테이트먼트 포지셔닝 장치는, 상기 전체 손실 결정 서버 모듈 이전에, 상기 제2 손실을 획득하도록 구성되는 제2 손실 획득 서버 모듈을 더 포함하고 상기 제2 손실 획득 서버 모듈은,
- [0172] 동일한 네거티브 샘플 페어의 제2 샘플 이미지 및 제1 샘플 디스크립션 스테이트먼트를 상기 뉴럴 네트워크에 입력하고 처리를 거쳐, 제3 트레이닝 스코어를 획득하고
- [0173] 상이한 네거티브 샘플 페어의 제2 샘플 이미지 및 제1 샘플 디스크립션 스테이트먼트를 상기 뉴럴 네트워크에 입력하고 처리를 거쳐, 제4 트레이닝 스코어를 획득하며
- [0174] 동일한 네거티브 샘플 페어의 제1 샘플 이미지 및 제2 샘플 디스크립션 스테이트먼트를 상기 뉴럴 네트워크에 입력하고 처리를 거쳐, 제5 트레이닝 스코어를 획득하고
- [0175] 상이한 네거티브 샘플 페어의 제1 샘플 이미지 및 제2 샘플 디스크립션 스테이트먼트를 상기 뉴럴 네트워크에 입력하고 처리를 거쳐, 제6 트레이닝 스코어를 획득하며
- [0176] 복수의 제3 트레이닝 스코어, 복수의 제4 트레이닝 스코어, 복수의 제5 트레이닝 스코어 및 복수의 제6 트레이닝 스코어에 따라, 제2 손실을 획득하도록 구성된다.
- [0177] 한 가지 가능한 구현 방식에서, 상기 전체 손실 결정 서버 모듈은,
- [0178] 상기 제1 손실 및 상기 제2 손실에 대해 가중 오버랩을 진행하여, 상기 뉴럴 네트워크의 전체 손실을 획득하도록 구성된다.
- [0179] 일부 실시예에서, 본원 발명의 실시예에서 제공하는 장치가 구비하는 기능 또는 포함되는 모듈은 상문의 방법 실시예에 서술된 방법을 수행하기 위한 것으로서, 그구체적인 구현은 상문의 방법 실시예의 서술을 참조 바라며, 간결함을 위해 여기서 더서술하지 않는다.
- [0180] 본원 발명의 실시예는 컴퓨터 판독 가능 저장 매체를 더제공하는 바, 이에 컴퓨터 프로그램 명령어가 저장되어 있으며, 상기 컴퓨터 프로그램 명령어가 프로세서에 의해 실행될 경우 상기 방법을 구현한다. 컴퓨터 판독 가능 저장 매체는 비휘발성 컴퓨터 판독 가능 저장 매체일 수 있다.
- [0181] 본원 발명의 실시예는 전자 기기를 더제공하는 바, 프로세서 프로세서 실행 가능 명령어가 저장된 메모리를 포함하며 여기서, 상기 프로세서는 상기 방법을 수행하도록 구성된다.
- [0182] 전자 기기는 단말기, 서버 또는 기타 형태의 기기로 제공될 수 있다.
- [0183] 도 6은 본원 발명의 실시예에 따른 전자 기기(800)의블록도이다. 예를 들어, 전자 기기(800)는모바일 전화, 컴퓨터, 디지털 방송 단말기, 메시지 송수신기, 게임 콘솔, 태블릿 기기, 의료 기기, 헬스 기기, PDA 등 단말기일 수 있다.
- [0184] 도 6을 참조하면, 전자 기기(800)는프로세싱 컴포넌트(802), 메모리(804), 전원 컴포넌트(806), 멀티미디어 컴포넌트(808), 오디오 컴포넌트(810), 입력/출력(I/O) 인터페이스(812), 센서 컴포넌트(814) 및 통신 컴포넌트(816) 중 하나 또는 복수의 컴포넌트를 포함할 수 있다.
- [0185] 프로세싱 컴포넌트(802)는디스플레이, 전화 통화, 데이터 통신, 카메라 동작 및 기록 동작과 관련된 동작과 같은 전자 기기(800)의 전체 동작을 제어한다. 프로세싱 컴포넌트(802)는, 상기 방법의 전부 또는 일부 단계를 완료하도록 하나 또는 복수의 프로세서(820)을 포함하여 명령어를 실행한다. 이외에, 프로세싱 컴포넌트(802)는, 프로세싱 컴포넌트(802)와 다른 컴포넌트 사이의 인터랙션을 진행하도록 하나 또는 복수의 모듈을 포함할 수 있다. 예를 들어, 프로세싱 컴포넌트(802)는, 멀티미디어 모듈(808)과프로세싱 컴포넌트(802) 사이의 인터랙션을 편리하게 진행하도록 멀티미디어 모듈을 포함할 수 있다.
- [0186] 메모리(804)는 다양한 유형의 데이터를 저장하여 전자 기기(800)에서의 동작을 지원한다. 이러한 데이터의 예시는 전자 기기(800)에서 동작하는 임의의 응용 프로그램 또는 방법의 명령어, 연락처 데이터, 전화번호부 데이터, 메시지, 이미지, 영상 등을 포함한다. 메모리(804)는, 정적 랜덤 액세스 메모리(SRAM), 전기적으로 소

거 가능한 프로그램 가능 판독 전용 메모리(EEPROM), 소거 가능한 프로그램 가능 판독 전용 메모리(EPROM), 프로그램 가능 판독 전용 메모리(PROM), 판독 전용 메모리(ROM), 자기 메모리, 플래시 메모리, 디스크 또는 광 디스크와 같은 임의의 유형의 휘발성 또는 비휘발성 저장 기기 또는 이들의 조합에 의해 구현될 수 있다.

- [0187] 전원 컴포넌트(806)는 전자 기기(800)의 다양한 컴포넌트에게 전력을 제공한다. 전원 컴포넌트(806)는 전원 관리 시스템, 하나 또는 복수의 전원, 및 전자 기기(800)의 생성, 관리, 및 전원 할당과 관련된 다른 컴포넌트를 포함할 수 있다.
- [0188] 멀티미디어 컴포넌트(808)는 상기 전자 기기(800)와 사용자 사이에 출력 인터페이스를 제공하는 스크린을 포함한다. 일부 실시예에서, 스크린은 액정 디스플레이(LCD) 및 터치 패널(TP)을 포함할 수 있다. 만약, 스크린이 터치 패널을 포함하면, 사용자로부터 입력 신호를 수신하도록 스크린은 터치 스크린으로 구현될 수 있다. 터치 패널은, 패널에서의 터치, 슬라이드, 터치 패널의 제스처를 감지하기 위한 하나 또는 복수의 터치 센서를 포함한다. 상기 터치 센서는 터치 또는 슬라이드 동작의 경계를 감지할 뿐만 아니라 상기 터치 또는 슬라이드에 관련된 지속 시간 및 압력을 감지할 수 있다. 일부 실시예에서, 멀티미디어 컴포넌트는 하나의 전방 카메라 및 /또는 후방 카메라를 포함한다. 전자 기기(800)가, 촬영 모드 또는 영상 모드일 경우, 전방 카메라 및 /또는 후방 카메라는 외부의 멀티미디어 데이터를 수신할 수 있다. 각각의 전방 카메라와 후방 카메라는 하나의 고정식 광학 렌즈 시스템 또는 초점 거리 및 광학 줌기능을 가질 수 있다.
- [0189] 오디오 컴포넌트(810)는 오디오 신호를 출력 및 /또는 입력한다. 예를 들어, 오디오 컴포넌트(810)는 하나의 마이크(MIC)를 포함하고 전자 기기(800)가 통화 모드, 녹음 모드 및 음성 인식 모드와 같은 동작 모드일 경우, 마이크는 외부 오디오 신호를 수신한다. 수신된 오디오 신호는 메모리(804)에 저장되거나 통신 컴포넌트(816)를 통해 발송될 수 있다. 일부 실시예에서, 오디오 컴포넌트(810)는 오디오 신호를 출력하기 위한 스피커를 포함한다.
- [0190] I/O 인터페이스(812)는 프로세싱 컴포넌트(802)와 주변 장치 인터페이스 모듈 사이의 인터페이스를 제공하며, 상기 주변 장치 인터페이스 모듈은 키보드, 클릭 휠(Click Wheel), 버튼 등일 수 있다. 이러한 버튼은 홈버튼, 볼륨 버튼, 시작 버튼 및 잠금 버튼을 포함할 수 있지만 이에 한정되지는 않는다.
- [0191] 센서 컴포넌트(814)는 다양한 양태의 상태 평가를 전자 기기(800)에 제공하기 위한 하나 또는 복수의 센서를 포함한다. 예를 들어, 센서 컴포넌트(814)는 전자 기기(800)의 온/오프 상태, 컴포넌트의 상대적 위치를 감지할 수 있고, 예를 들어, 상기 컴포넌트는 전자 기기(800)의 디스플레이 및 키패드이고 센서 컴포넌트(814)는 전자 기기(800) 또는 전자 기기(800)의 컴포넌트의 위치 변화, 사용자와 전자 기기(800) 사이의 접촉 여부, 전자 기기(800) 방위 또는 가속/감속 및 전자 기기(800)의 온도 변화를 감지할 수 있다. 센서 컴포넌트(814)는 물리적 접촉없이 주변 물체의 존재를 감지하는 근접 센서를 포함할 수 있다. 센서 컴포넌트(814)는 이미징 애플리케이션에 사용하기 위한 CMOS 또는 CCD 이미지 센서와 같은 광센서를 더 포함할 수 있다. 일부 실시예에서, 상기 센서 컴포넌트(814)는 가속도 센서, 자이로 센서, 자기 센서, 압력 센서 또는 온도 센서를 더 포함할 수 있다.
- [0192] 통신 컴포넌트(816)는 전자 기기(800)와 다른 기기 사이의 유선 또는 무선 방식의 통신이 용이하도록 구성된다. 전자 기기(800)는 WiFi, 2G 또는 3G 또는 이들의 조합과 같은 통신 표준에 기반한 무선 네트워크에 액세스할 수 있다. 예시적인 일 실시예에서, 통신 컴포넌트(816)는 방송 채널을 통해 외부 방송 관리 시스템으로부터 방송 신호 또는 방송 관련 정보를 수신한다. 예시적인 일 실시예에서, 상기 통신 컴포넌트(816)는 근거리 통신을 촉진하는 근거리 통신(NFC) 모듈을 더 포함한다. 예를 들어, NFC 모듈은 무선 주파수 식별(RFID) 기술, 적외선 통신 규격(IrDA) 기술, 초광대역(UWB) 기술, 블루투스(BT) 기술 및 다른 기술 기반으로 구현될 수 있다.
- [0193] 예시적인 실시예에서, 전자 기기(800)는 하나 또는 복수의 주문형 집적 회로(ASIC), 디지털 신호 프로세서(DSP), 디지털 신호 프로세서 기기(DSPD), 프로그램 가능 논리 소자(PLD), 필드 프로그램 가능 게이트 어레이(FPGA), 컨트롤러, 마이크로 컨트롤러, 마이크로 프로세서 또는 다른 전자 소자에 의해 상기 방법을 실행할 수 있다.
- [0194] 예시적인 실시예에서, 상기 방법을 수행하기 위해 전자 기기(800)의 프로세서(820)에 의해 실행 가능한 명령어를 포함하는 메모리(804)와 같은 명령어를 포함하는 비 일시적 컴퓨터 판독 가능한 기록매체를 더 제공한다.
- [0195] 도 7은 본원 발명의 실시예에 따른 전자 기기(1900)의 블록도이다. 예를 들어, 전자 기기(1900)는 하나의 서버로 제공될 수 있다. 도 7을 참조하면 전자 기기(1900)는 프로세싱 컴포넌트(1922)를 포함하고 이는 또한 하나 또는 다수의 프로세서 및 프로세싱 컴포넌트(1922)가 실행 가능한 명령어, 예를 들어 애플리케이션 프로그램을 저장하기 위한 메모리(1932)를 대표로 하는 메모리 리소스를 포함한다. 메모리(1932)에 저장된 애플리케이션 프로그램

램은 하나 또는 하나 이상의 각각의 명령어와 대응되는 모듈을 포함할 수 있다. 이 밖에 프로세싱 컴포넌트 (1922)는 상기 방법을 수행하도록 명령어를 실행한다.

- [0196] 전자 기기(1900)는 전자 기기(1900)의 전원관리를 실행하도록 구성된 하나의 전원 컴포넌트(1926), 전자 기기 (1900)를 네트워크에 연결하도록 구성된 하나의 유선 또는 무선 네트워크 인터페이스(1950), 및 하나의 입력/출력(I/O) 인터페이스(1958)를 더 포함할 수 있다. 전자 기기(1900)는 메모리(1932)에 저장된 운영체제, 예를 들어 Windows Server™, Mac OS X™, Unix™, Linux™, FreeBSD™ 또는 유사한 운영체제를 작동시킬 수 있다.
- [0197] 예시적인 실시예에서, 비밀시적인 컴퓨터 관독 가능 저장 매체를 더 제공하였는 바, 예를 들어 컴퓨터 프로그램 명령어를 포함하는 메모리(1932)이며, 상기 컴퓨터 프로그램 명령어는 전자 기기(1900)의 프로세싱 컴포넌트 (1922)에 의해 수행되어 상기 방법을 완성한다.
- [0198] 본원 발명은 시스템, 방법 및 /또는 컴퓨터 프로그램 제품일 수 있다. 컴퓨터 프로그램 제품은 컴퓨터 관독 가능 저장 매체를 포함할 수 있고, 프로세서가 본원 발명의 각 양태를 실현하도록 하기 위한 컴퓨터 관독 가능 프 로그램 명령어가 저장되어 있다.
- [0199] 컴퓨터 관독 가능 저장 매체는 명령어로 수행 기기가 사용하는 명령어를 유지 및 저장하는 유형의 기기일 수 있 다. 컴퓨터 관독 가능 저장 매체는 예를 들어, 전기 저장 장치, 자기 저장 장치, 광 저장 장치, 전자 자기 저장 장치, 반도체 저장 장치 또는 상기의 임의의 적합한 조합일 수 있지만 이에 한하지 않는다. 컴퓨터 관독 가능 저장 매체의 더욱 구체적인 예(철저히 않은 리스트)는, 휴대용 컴퓨터 디스크, 하드 디스크, 램(RAM), 롬 (ROM), 소거 가능 프로그래머블 롬(EPROM 또는 플래시 메모리), 정적 램(SRAM), CD-ROM, 디지털 비디오 디스크 (DVD), 메모리스틱, 플로피 디스켓, 기계적 코딩 장치, 예를 들어 명령어가 저장된 천공 카드 또는 요홈 내의 돌기 구조, 및 상기 임의의 적합한 조합을 포함한다. 여기서 사용되는 컴퓨터 관독 가능 저장 매체는 예컨대 무 선 전자파 또는 기타 자유 전파되는 전자파, 웨이브 가이드 또는 기타 전송 매체를 통해 전파되는 전자파(예를 들어, 광섬유 케이블을 통한 라이트 펄스), 또는 전선을 통해 전송되는 전자 신호 등 순간 신호 자체인 것으로 해석되지 않는다.
- [0200] 여기서 서술되는 컴퓨터 관독 가능 프로그램 명령어는 컴퓨터 관독 가능 저장 매체에서 각 컴퓨팅/프로세싱 기 기에 다운될 수 있거나, 또는 인터넷, 랜, 광역 통신망 및 /또는 무선망을 통해 외부 컴퓨터 또는 외부 저장 기 기에 다운될 수 있다. 네트워크는 구리 전송 케이블, 광섬유 전송, 무선 전송, 라우터, 방화벽, 교환기, 게이트 웨이 컴퓨터 및 /또는 가장자리 서버를 포함할 수 있다. 각각의 컴퓨팅/프로세싱 기기 중의 네트워크 어댑터 카 드 또는 네트워크 인터페이스가 네트워크에서 컴퓨터 관독 가능 프로그램 명령어를 수신하고, 상기 컴퓨터 관독 가능 프로그램 명령어를 전달하여, 각 컴퓨팅/프로세싱 기기에 저장된 컴퓨터 관독 가능 저장 매체에 제공한다.
- [0201] 본원 발명의 동작을 수행하기 위한 컴퓨터 프로그램 명령어는 어셈블리 명령, 명령어 집합 (ISA) 명령, 기계 명 령, 기계 관련 명령, 마이크로코드, 펌웨어 명령어, 상태 설정 데이터, 또는 하나 또는 복수의 프로그래밍 언어 의 임의의 조합으로 프로그래밍된 소스 코드 또는 객체 코드일 수 있고, 상기 프로그래밍 언어는 Smalltalk, C++를 비롯한 객체 지향 프로그래밍 언어와 "C" 언어 또는 유사한 프로그래밍 언어를 비롯한 프로그래밍 언어를 포함한다. 컴퓨터 관독 가능 프로그램 명령어는 완전히 사용자의 컴퓨터에서 실행되거나, 부분적으로 사용자의 컴퓨터에서 실행되거나, 하나의 독립형 소프트웨어 패키지로서 실행되거나, 일부는 사용자의 컴퓨터에서 실행되 고 일부는 원격 컴퓨터에서 실행되거나, 또는 완전히 원격 컴퓨터 또는 서버에서 실행될 수 있다. 원격 컴퓨터 의 경우, 원격 컴퓨터는 구내 정보 통신망 (LAN) 또는 광역 통신망(WAN)을 포함한 모든 종류의 네트워크를 통해 사용자의 컴퓨터에 연결되거나 외부 컴퓨터에 연결될 수 있다(예를 들어, 인터넷 서비스 제공 업체를 이용하여 인터넷을 통해 연결). 일부 실시예에서, 컴퓨터 관독 가능 프로그램 명령어의 상태 정보를 이용하여 예를 들어 프로그래머블 논리 회로, 필드 프로그래머블 어레이(FPGA) 또는 프로그래머블 논리 어레이(PLA)와 같은 전자 회로의 개인 맞춤을 실현하며, 상기 전자 회로는 컴퓨터 관독 가능 프로그램 명령어를 실행하여 본원 발명의 각 양태를 구현할 수 있다.
- [0202] 여기서 본원 발명의 실시예에 따른 방법, 장치(시스템) 및 컴퓨터 프로그램 제품의 흐름도 및 /또는 블록도를 참조하여 본원 발명의 각 양태를 서술하였다. 흐름도 및 /또는 블록도의 각각의 프레임 및 흐름도 및 /또는 블 록도 중 각 프레임의 조합은 모두 컴퓨터 관독 가능 프로그램 명령어로 구현될 수 있음을 반드시 이해해야 한다.
- [0203] 이러한 컴퓨터 관독 가능 프로그램 명령어는 범용 컴퓨터, 전용 컴퓨터 또는 기타 프로그래머블 데이터 처리 장 치의 프로세서에 제공되어 기계를 생산할 수 있으며 이로써 이러한 명령어가 컴퓨터 또는 기타 프로그래머블 데 이터 처리 장치의 프로세서를 통해 수행될 경우, 흐름도 및 /또는 블록도 중의 하나 또는 복수의 프레임에 규정

된 기능/동작을 구현하는 장치를 생산하였다. 또한 이러한 컴퓨터 판독 가능 프로그램 명령어를 컴퓨터 판독 가능 저장 매체에 저장할 수 있으며, 이러한 명령어는 컴퓨터, 프로그래머블 데이터 처리 장치 및 /또는 기타 기기가 특정 방식으로 작동되도록 하여 명령어가 저장된 컴퓨터 판독 가능 매체는 하나의 제조품을 포함하고, 이는 흐름도 및 /또는 블록도 중의 하나 또는 복수의 프레임에 규정된 기능/동작의 각 양태를 구현하는 명령어를 포함한다.

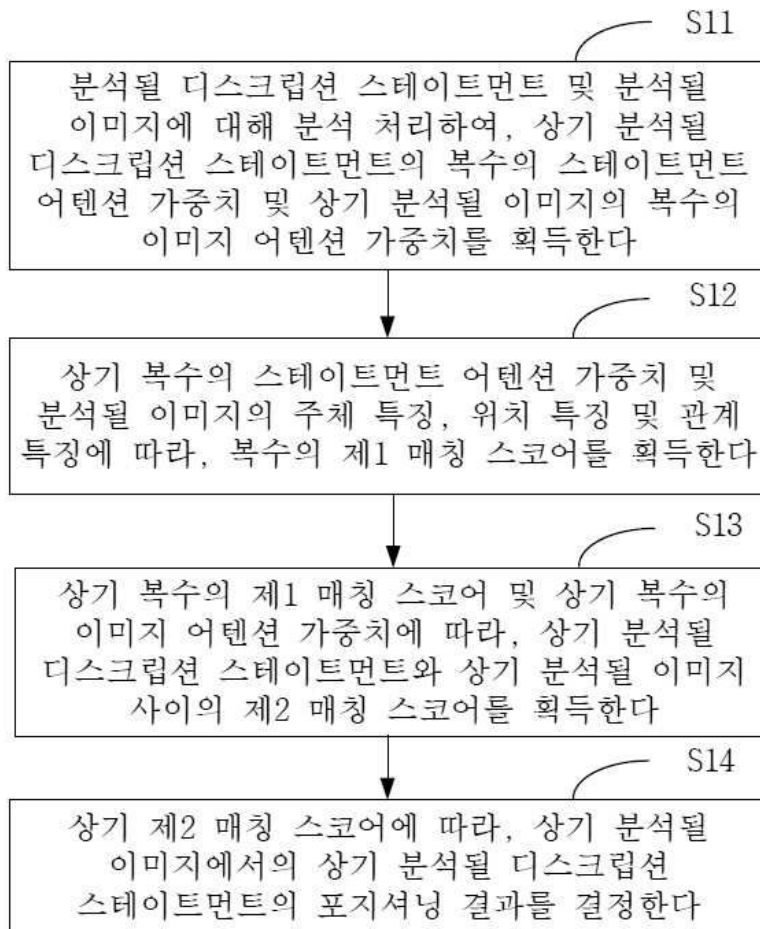
[0204] 컴퓨터 판독 가능 프로그램 명령어를 컴퓨터, 기타 프로그래머블 데이터 처리 장치, 또는 기타 기기에 로딩하여 컴퓨터, 기타 프로그래머블 데이터 처리 장치, 또는 기타 기기가 일련의 작동 단계를 수행 가능하도록 하여 컴퓨터 구현의 과정을 생성함으로써, 컴퓨터, 기타 프로그래머블 데이터 처리 장치, 또는 기타 기기에서 실행되는 명령어는 흐름도 및 /또는 블록도 중의 하나 또는 복수의 프레임에 규정된 기능/동작을 구현한다.

[0205] 도면 중의 흐름도 및 블록도는 본원 발명의 복수의 실시예에 따른 시스템, 방법 및 컴퓨터 프로그램 제품의 구현 가능한 시스템 아키텍처, 기능 및 동작을 도시하였다. 이 점으로부터, 흐름도 및 블록도 중의 각각의 프레임은 하나의 모듈, 프로그램 세그먼트 또는 명령어의 일부분을 대표할 수 있으며, 상기 모듈, 프로그램 세그먼트 또는 명령어의 일부분은 규정된 논리 기능을 구현하는 실행 가능한 하나 또는 복수의 명령어를 포함한다. 교체를 구현함에 있어서, 블록 중 마킹한 기능도 도면에 마킹된 순서와 상이하게 발생할 수 있다. 예를 들어, 두 개의 연속적인 블록은 실제상에서 기본적으로 병행하여 수행되며, 이들은 때로 반대되는 순서로 수행될 수도 있으며, 이는 관련 기능에 따라 결정된다. 블록도 및 /또는 흐름도 중의 각각의 블록, 블록도 및 /또는 흐름도 중의 블록의 조합은 규정된 기능 또는 동작을 수행하는 전용적인, 하드웨어에 기반한 시스템에 따라 구현될 수 있거나, 전용 하드웨어와 컴퓨터 명령어의 조합으로 구현될 수 있다.

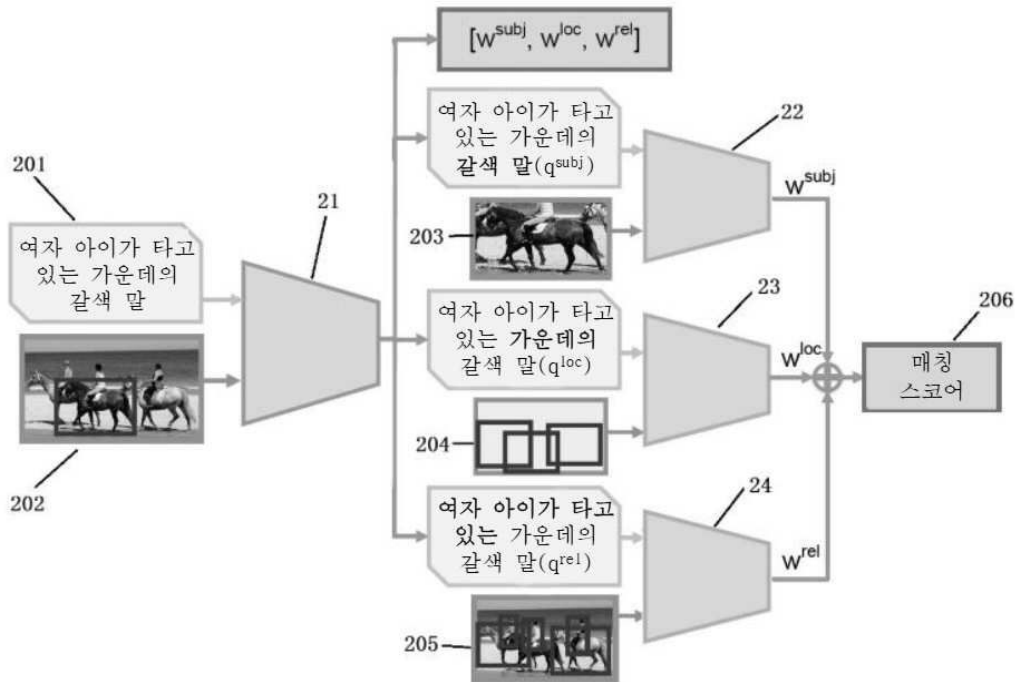
[0206] 상기와 같이 본원 발명의 각 실시예를 서술하였으며 상기 설명은 예시적인 것으로서 철저한 것이 아니고 또한 공개된 각 실시예에 한정되지 않는다. 서술되는 각 실시예의 범위 및 정신을 벗어나지 않는 상황하에, 본기술분야의 통상의 기술자에게 있어서 수많은 보정과 변경은 모두 자명한 것이다. 본문에서 사용되는 용어의 선택은 각 실시예의 원리, 실제 응용 또는 시장의 기술 개선을 가장 확실하게 해석하거나, 또는 본 기술 분야의 기타 통상의 기술자가 본문에 개시된 각 실시예를 더욱 잘 이해하도록 한다.

도면

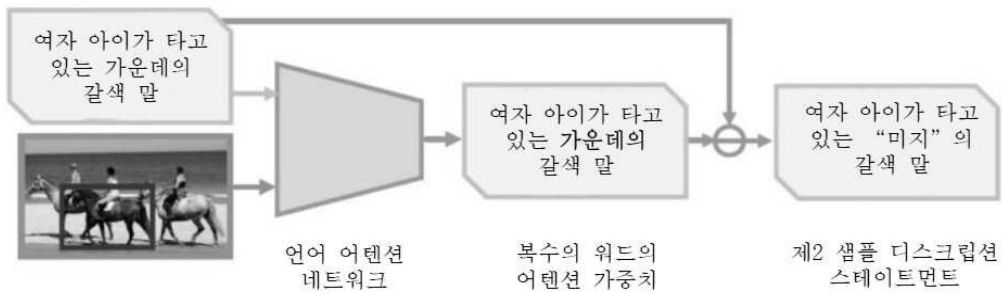
도면1



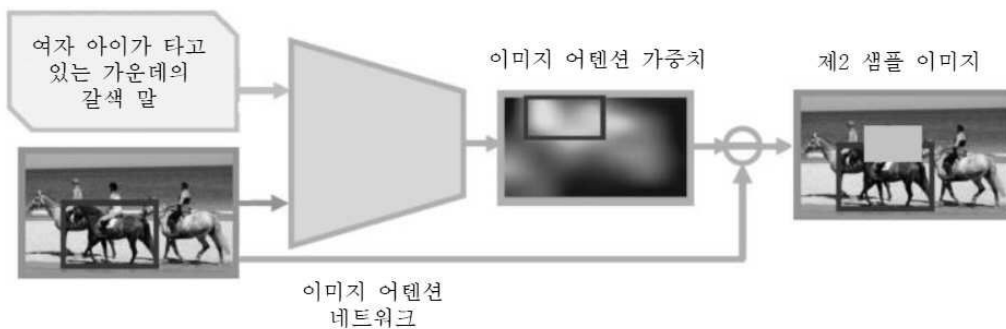
도면2



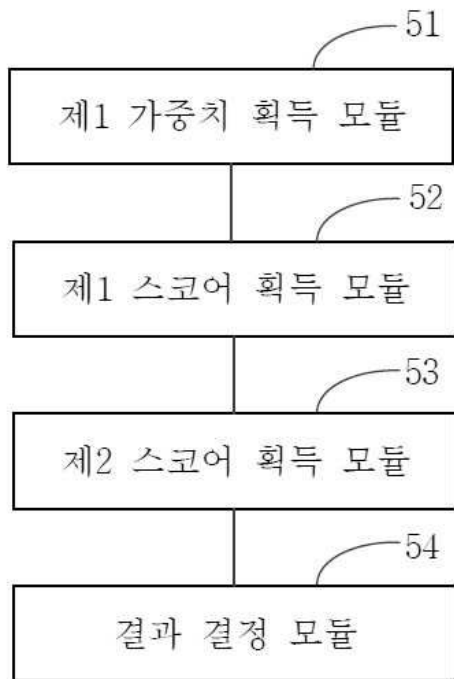
도면3



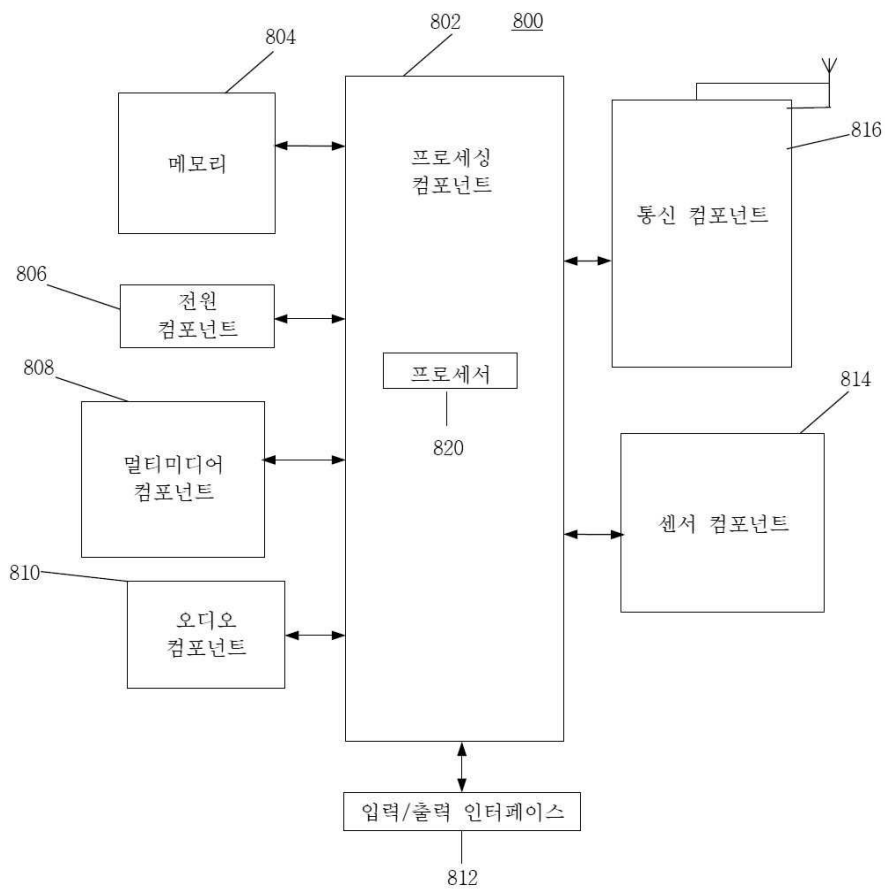
도면4



도면5



도면6



도면7

