(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년04월27일
(11) 등록번호 10-2245774
(24) 등록일자 2021년04월22일

(51) 국제특허분류(Int. Cl.)
G06F 16/583 (2019.01) *G06F 16/33* (2019.01)
G06F 16/55 (2019.01) *G06N 3/08* (2006.01)
(52) CPC특허분류
G06F 16/583 (2019.01)
G06F 16/3347 (2019.01)
(21) 출원번호 10-2019-0140763
(22) 출원일자 2019년11월06일
심사청구일자 2019년11월06일
(56) 선행기술조사문헌
KR1020170043582 A*
Nam, Hyeonseob 외2인. "Dual attention networks for multimodal reasoning and matching." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2017.07.26.공개)*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
변혜란
서울특별시 서대문구 연세로 50, 제4공학관 810호(신촌동, 연세대학교)
박성호
서울특별시 서대문구 연세로 50, 제4공학관 810호(신촌동, 연세대학교)
(74) 대리인
특허법인우인

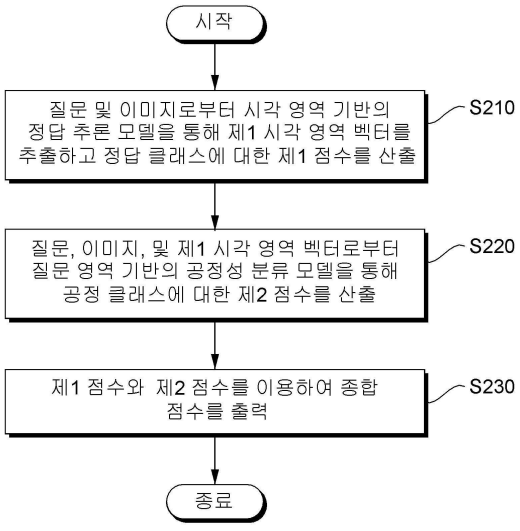
전체 청구항 수 : 총 10 항 심사관 : 이현중

(54) 발명의 명칭 **공정성 분류 네트워크를 이용한 시각 질의 응답 방법 및 장치**

(57) 요 약

본 실시예들은 공정성 분류 네트워크를 통해 질문에 맞는 공정 클래스를 분류하고 분류된 공정 클래스 점수를 정답 추론 네트워크를 통해 추론한 정답 점수에 적용하여, 공정성이 필요한 질문에 공정한 답변을 출력하는 시각 질의 응답 장치 및 방법을 제공한다.

대 표 도 - 도2



(52) CPC특허분류
G06F 16/55 (2019.01)
G06N 3/08 (2013.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 2019-11-0549
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 정부-과학기술정보통신부-정보통신기획평가원(한국연구재단부설)-정보통신방송연구개발사업-혁신성장동력프로젝트(인공지능)
연구과제명 [이지바로][주관/한국과학기술원] 인공지능 모델과 학습데이터의 편향성 분석-탐지-완화 제거 지원 프레임워크 개발 (1/4)
기 여 율 1/1
과제수행기관명 KAIST
연구기간 2019.04.01 ~ 2019.12.31

명 세 서

청구범위

청구항 1

컴퓨팅 디바이스에 의한 시각 질의 응답 방법에 있어서,

질문 및 이미지로부터 시각 영역 기반의 정답 추론 모델을 통해 제1 시각 영역 벡터를 추출하고 정답 클래스에 대한 제1 점수를 산출하는 단계;

상기 질문, 상기 이미지, 및 상기 제1 시각 영역 벡터로부터 질문 영역 기반의 공정성 분류 모델을 통해 공정 클래스에 대한 제2 점수를 산출하는 단계;

상기 제1 점수와 상기 제2 점수를 이용하여 종합 점수를 출력하는 단계를 포함하며,

상기 종합 점수를 출력하는 단계는,

상기 정답 클래스가 배경 클래스가 아니면 상기 제2 점수에 반영비율을 부여한 후 상기 제1 점수와 합산하여 상기 종합 점수를 출력하고,

상기 정답 클래스가 상기 배경 클래스이면 상기 제2 점수를 반영하지 않고 상기 제1 점수를 상기 종합 점수로 출력하는 것을 특징으로 하는 시각 질의 응답 방법.

청구항 2

제1항에 있어서,

상기 정답 추론 모델은 특징 추출 모델, 임베딩 모델, 시각 어텐션 모델, 정답 분류 모델을 포함하며,

상기 제1 점수를 산출하는 단계는,

상기 이미지로부터 상기 특징 추출 모델을 통해 시각 특징 벡터를 추출하고,

상기 질문으로부터 상기 임베딩 모델을 통해 질문 벡터를 추출하고,

상기 질문 벡터 및 상기 시각 특징 벡터에 대해 상기 시각 어텐션 모델을 통해 제1 시각 영역 가중치를 부여하여 상기 제1 시각 영역 벡터를 생성하고,

상기 질문 벡터와 상기 제1 시각 영역 벡터를 결합한 특징 정보를 상기 정답 분류 모델을 통해 상기 정답 클래스에 맞게 분류하는 것을 특징으로 하는 시각 질의 응답 방법.

청구항 3

제1항에 있어서,

상기 공정성 분류 모델은 특징 추출 모델, 임베딩 모델, 질문 어텐션 모델, 시각 어텐션 모델, 정답 분류 모델을 포함하며,

상기 제2 점수를 산출하는 단계는,

상기 이미지로부터 상기 특징 추출 모델을 통해 시각 특징 벡터를 추출하고, 상기 시각 특징 벡터와 상기 제1 시각 영역 벡터를 결합한 제2 시각 영역 벡터를 생성하고,

상기 질문으로부터 상기 임베딩 모델을 통해 질문 벡터를 추출하고,

상기 질문 벡터에 대해 상기 질문 어텐션 모델을 통해 질문 영역 가중치를 부여하여 질문 영역 벡터를 생성하고, 상기 질문 영역 벡터에 대해 상기 시각 어텐션 모델을 통해 제2 시각 영역 가중치를 산출하고,

상기 제2 시각 영역 벡터에 상기 제2 시각 영역 가중치를 부여하여 제3 시각 영역 벡터를 생성하고,

상기 제3 시각 영역 벡터를 상기 정답 분류 모델을 통해 상기 공정 클래스에 맞게 분류하는 것을 특징으로 하는

시각 질의 응답 방법.

청구항 4

제1항에 있어서,

상기 정답 추론 모델과 상기 공정성 분류 모델은 상기 제1 시각 영역 벡터를 공유하여 두 모델의 최적화를 위해 함께 학습되는 것을 특징으로 하는 시각 질의 응답 방법.

청구항 5

제1항에 있어서,

상기 종합 점수를 출력하는 단계는,

상기 정답 클래스의 개수가 상기 공정 클래스의 개수보다 많고, 상기 공정 클래스가 상기 정답 클래스에 대응되는지 여부에 따라 상기 공정 클래스의 레이블을 상기 정답 클래스의 레이블에 매핑하여 상기 제2 점수를 변환하고, 상기 제2 점수를 반영하여 상기 종합 점수를 출력하는 것을 특징으로 하는 시각 질의 응답 방법.

청구항 6

삭제

청구항 7

하나 이상의 프로세서 및 상기 하나 이상의 프로세서에 의해 실행되는 하나 이상의 프로그램을 저장하는 메모리를 포함하는 시각 질의 응답 장치에 있어서,

상기 프로세서는 질문 및 이미지로부터 시각 영역 기반의 정답 추론 모델을 통해 제1 시각 영역 벡터를 추출하고 정답 클래스에 대한 제1 점수를 산출하고,

상기 프로세서는 상기 질문, 상기 이미지, 및 상기 제1 시각 영역 벡터로부터 질문 영역 기반의 공정성 분류 모델을 통해 공정 클래스에 대한 제2 점수를 산출하고,

상기 프로세서는 상기 제1 점수와 상기 제2 점수를 이용하여 종합 점수를 출력하며,

상기 프로세서는,

상기 정답 클래스가 배경 클래스가 아니면 상기 제2 점수에 반영비율을 부여한 후 상기 제1 점수와 합산하여 상기 종합 점수를 출력하고,

상기 정답 클래스가 상기 배경 클래스이면 상기 제2 점수를 반영하지 않고 상기 제1 점수를 상기 종합 점수로 출력하는 것을 특징으로 하는 시각 질의 응답 장치.

청구항 8

제7항에 있어서,

상기 정답 추론 모델은 특징 추출 모델, 임베딩 모델, 시각 어텐션 모델, 정답 분류 모델을 포함하며,

상기 프로세서는,

상기 이미지로부터 상기 특징 추출 모델을 통해 시각 특징 벡터를 추출하고,

상기 질문으로부터 상기 임베딩 모델을 통해 질문 벡터를 추출하고,

상기 질문 벡터 및 상기 시각 특징 벡터에 대해 상기 시각 어텐션 모델을 통해 제1 시각 영역 가중치를 부여하여 상기 제1 시각 영역 벡터를 생성하고,

상기 질문 벡터와 상기 제1 시각 영역 벡터를 결합한 특징 정보를 상기 정답 분류 모델을 통해 상기 정답 클래스에 맞게 분류하는 것을 특징으로 하는 시각 질의 응답 장치.

청구항 9

제7항에 있어서,

상기 공정성 분류 모델은 특징 추출 모델, 임베딩 모델, 질문 어텐션 모델, 시각 어텐션 모델, 정답 분류 모델을 포함하며,

상기 프로세서는,

상기 이미지로부터 상기 특징 추출 모델을 통해 시각 특징 벡터를 추출하고, 상기 시각 특징 벡터와 상기 제1 시각 영역 벡터를 결합한 제2 시각 영역 벡터를 생성하고,

상기 질문으로부터 상기 임베딩 모델을 통해 질문 벡터를 추출하고,

상기 질문 벡터에 대해 상기 질문 어텐션 모델을 통해 질문 영역 가중치를 부여하여 질문 영역 벡터를 생성하고, 상기 질문 영역 벡터에 대해 상기 시각 어텐션 모델을 통해 제2 시각 영역 가중치를 산출하고,

상기 제2 시각 영역 벡터에 상기 제2 시각 영역 가중치를 부여하여 제3 시각 영역 벡터를 생성하고,

상기 제3 시각 영역 벡터를 상기 정답 분류 모델을 통해 상기 공정 클래스에 맞게 분류하는 것을 특징으로 하는 시각 질의 응답 장치.

청구항 10

제7항에 있어서,

상기 정답 추론 모델과 상기 공정성 분류 모델은 상기 제1 시각 영역 벡터를 공유하여 두 모델의 최적화를 위해 함께 학습되는 것을 특징으로 하는 시각 질의 응답 장치.

청구항 11

제7항에 있어서,

상기 프로세서는,

상기 정답 클래스의 개수가 상기 공정 클래스의 개수보다 많고, 상기 공정 클래스가 상기 정답 클래스에 대응되는지 여부에 따라 상기 공정 클래스의 레이블을 상기 정답 클래스의 레이블에 매핑하여 상기 제2 점수를 변환하고, 상기 제2 점수를 반영하여 상기 종합 점수를 출력하는 것을 특징으로 하는 시각 질의 응답 장치.

청구항 12

삭제

발명의 설명

기 술 분 야

[0001] 본 발명이 속하는 기술 분야는 시각 질의 응답 장치 및 방법에 관한 것이다.

배 경 기 술

[0002] 이 부분에 기술된 내용은 단순히 본 실시예에 대한 배경 정보를 제공할 뿐 종래기술을 구성하는 것은 아니다.

[0003] 시각적 질문 응답(Visual Question Answering, VQA)은 이미지에 관한 다양한 시각적인 의미론적 수준의 질문으로부터 단어를 찾고 이미지의 중요한 영역을 찾아서 정답을 추론하는 기술이다. 시각적 질문 응답은 동일한 이미지에 대하여 다양한 시각적인 의미론적 수준의 질문을 요구한다. 예컨대, 질문은 이미지에 대하여 "물체가 무엇인지", "물체의 색상이 무엇인지", "물체의 개수" 등을 요구할 수 있다.

[0004] 기존 VQA 모델은 성별 또는 인종 등의 공정성이 필요한 질문에 불공정한 답변을 출력하는 문제가 있다.

선행기술문헌

특허문헌

[0005] (특허문헌 0001) 한국등록특허공보 제10-1725885호 (2017.04.05)

발명의 내용

해결하려는 과제

[0006] 본 발명의 실시예들은 공정성 분류 네트워크를 통해 질문에 맞는 공정 클래스를 분류하고 분류된 공정 클래스 점수를 정답 추론 네트워크를 통해 추론한 정답 점수에 적용하여, 공정성이 필요한 질문에 공정한 답변을 출력하는데 주된 목적이 있다.

[0007] 본 발명의 명시되지 않은 또 다른 목적들은 하기의 상세한 설명 및 그 효과로부터 용이하게 추론할 수 있는 범위 내에서 추가적으로 고려될 수 있다.

과제의 해결 수단

[0008] 본 실시예의 일 측면에 의하면, 컴퓨팅 디바이스에 의한 시각 질의 응답 방법에 있어서, 질문 및 이미지로부터 시각 영역 기반의 정답 추론 모델을 통해 제1 시각 영역 벡터를 추출하고 정답 클래스에 대한 제1 점수를 산출하는 단계, 상기 질문, 상기 이미지, 및 상기 제1 시각 영역 벡터로부터 질문 영역 기반의 공정성 분류 모델을 통해 공정 클래스에 대한 제2 점수를 산출하는 단계, 상기 제1 점수와 상기 제2 점수를 이용하여 종합 점수를 출력하는 단계를 포함하는 시각 질의 응답 방법을 제공한다.

[0009] 상기 정답 추론 모델은 특징 추출 모델, 임베딩 모델, 시각 어텐션 모델, 정답 분류 모델을 포함할 수 있다.

[0010] 상기 제1 점수를 산출하는 단계는, 상기 이미지로부터 상기 특징 추출 모델을 통해 시각 특징 벡터를 추출하고, 상기 질문으로부터 상기 임베딩 모델을 통해 질문 벡터를 추출하고, 상기 질문 벡터 및 상기 시각 특징 벡터에 대해 상기 시각 어텐션 모델을 통해 제1 시각 영역 가중치를 부여하여 상기 제1 시각 영역 벡터를 생성하고, 상기 질문 벡터와 상기 제1 시각 영역 벡터를 결합한 특징 정보를 상기 정답 분류 모델을 통해 상기 정답 클래스에 맞게 분류할 수 있다.

[0011] 상기 공정성 분류 모델은 특징 추출 모델, 임베딩 모델, 질문 어텐션 모델, 시각 어텐션 모델, 정답 분류 모델을 포함할 수 있다.

[0012] 상기 제2 점수를 산출하는 단계는, 상기 이미지로부터 상기 특징 추출 모델을 통해 시각 특징 벡터를 추출하고, 상기 시각 특징 벡터와 상기 제1 시각 영역 벡터를 결합한 제2 시각 영역 벡터를 생성하고, 상기 질문으로부터 상기 임베딩 모델을 통해 질문 벡터를 추출하고, 상기 질문 벡터에 대해 상기 질문 어텐션 모델을 통해 질문 영역 가중치를 부여하여 질문 영역 벡터를 생성하고, 상기 질문 영역 벡터에 대해 상기 시각 어텐션 모델을 통해 제2 시각 영역 가중치를 산출하고, 상기 제2 시각 영역 벡터에 상기 제2 시각 영역 가중치를 부여하여 상기 제3 시각 영역 벡터를 생성하고, 상기 제3 시각 영역 벡터를 상기 정답 분류 모델을 통해 상기 공정 클래스에 맞게 분류할 수 있다.

[0013] 상기 정답 추론 모델과 상기 공정성 분류 모델은 상기 제1 시각 영역 벡터를 공유하여 두 모델의 최적화를 위해 함께 학습될 수 있다.

[0014] 상기 종합 점수를 출력하는 단계는, 상기 정답 클래스의 개수가 상기 공정 클래스의 개수보다 많고, 상기 공정 클래스가 상기 정답 클래스에 대응되는지 여부에 따라 상기 공정 클래스의 레이블을 상기 정답 클래스의 레이블에 매핑하여 상기 제2 점수를 변환하고, 상기 제2 점수를 반영하여 상기 종합 점수를 출력할 수 있다.

[0015] 상기 종합 점수를 출력하는 단계는, 상기 정답 클래스가 배경 클래스가 아니면 상기 제2 점수에 가중치를 부여한 후 상기 제1 점수와 합산하여 상기 종합 점수를 출력하고, 상기 정답 클래스가 상기 배경 클래스이면 상기 제2 점수를 반영하지 않고 상기 제1 점수를 상기 종합 점수로 출력할 수 있다.

[0016] 본 실시예의 다른 측면에 의하면, 하나 이상의 프로세서 및 상기 하나 이상의 프로세서에 의해 실행되는 하나 이상의 프로그램을 저장하는 메모리를 포함하는 시각 질의 응답 장치에 있어서, 상기 프로세서는 질문 및 이미지로부터 시각 영역 기반의 정답 추론 모델을 통해 제1 시각 영역 벡터를 추출하고 정답 클래스에 대한 제1 점수를 산출하고, 상기 프로세서는 상기 질문, 상기 이미지, 및 상기 제1 시각 영역 벡터로부터 질문 영역 기반의 공정성 분류 모델을 통해 공정 클래스에 대한 제2 점수를 산출하고, 상기 프로세서는 상기 제1 점수와 상기 제2

점수를 이용하여 종합 점수를 출력하는 것을 특징으로 하는 시각 질의 응답 장치를 제공한다.

발명의 효과

[0017] 이상에서 설명한 바와 같이 본 발명의 실시예들에 의하면, 공정성 분류 네트워크를 통해 질문에 맞는 공정 클래스를 분류하고 분류된 공정 클래스 점수를 정답 추론 네트워크를 통해 추론한 정답 점수에 적용하여, 공정성이 필요한 질문에 공정한 답변을 출력할 수 있는 효과가 있다.

[0018] 여기에서 명시적으로 언급되지 않은 효과라 하더라도, 본 발명의 기술적 특징에 의해 기대되는 이하의 명세서에서 기재된 효과 및 그 잠정적인 효과는 본 발명의 명세서에 기재된 것과 같이 취급된다.

도면의 간단한 설명

[0019] 도 1은 본 발명의 일 실시예에 따른 시각 질의 응답 장치를 예시한 블록도이다.

도 2는 본 발명의 다른 실시예에 따른 시각 질의 응답 방법을 예시한 흐름도이다.

도 3은 본 발명의 실시예들에 따른 시각 질의 응답 장치의 정답 추론 모델 및 공정성 분류 모델을 예시한 도면이다.

도 4 및 도 5는 본 발명의 실시예들에 따라 수행된 모의실험 결과를 도시한 것이다.

발명을 실시하기 위한 구체적인 내용

[0020] 이하, 본 발명을 설명함에 있어서 관련된 공지기능에 대하여 이 분야의 기술자에게 자명한 사항으로서 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략하고, 본 발명의 일부 실시예들을 예시적인 도면을 통해 상세하게 설명한다.

[0021] 기존의 VQA 모델은 오로지 정답만을 출력할 수 있고, 성별 또는 인종 등의 공정성이 필요한 질문에 편향된 정답을 출력하는 경향이 있다.

[0022] 본 발명은 성별 또는 인종 등과 같은 공정한 답변을 출력하기 위해서 공정성 분류 네트워크를 통해 질문에 맞는 공정 클래스(예컨대, 성별 클래스)를 분류하고 분류된 공정 클래스 정보를 정답 추론 네트워크에 의해 추론된 시각 질의 정답에 매핑한다.

[0023] 도 1은 본 발명의 일 실시예에 따른 시각 질의 응답 장치를 예시한 블록도이다.

[0024] 시각 질의 응답 장치(110)는 적어도 하나의 프로세서(120), 컴퓨터 판독 가능한 저장매체(130) 및 통신 버스(170)를 포함한다.

[0025] 프로세서(120)는 시각 질의 응답 장치(110)로 동작하도록 제어할 수 있다. 예컨대, 프로세서(120)는 컴퓨터 판독 가능한 저장 매체(130)에 저장된 하나 이상의 프로그램들을 실행할 수 있다. 하나 이상의 프로그램들은 하나 이상의 컴퓨터 실행 가능 명령어를 포함할 수 있으며, 컴퓨터 실행 가능 명령어는 프로세서(120)에 의해 실행되는 경우 시각 질의 응답 장치(110)로 하여금 예시적인 실시예에 따른 동작들을 수행하도록 구성될 수 있다.

[0026] 컴퓨터 판독 가능한 저장 매체(130)는 컴퓨터 실행 가능 명령어 내지 프로그램 코드, 프로그램 데이터 및/또는 다른 적합한 형태의 정보를 저장하도록 구성된다. 컴퓨터 판독 가능한 저장 매체(130)에 저장된 프로그램(140)은 프로세서(120)에 의해 실행 가능한 명령어의 집합을 포함한다. 일 실시예에서, 컴퓨터 판독한 가능 저장 매체(130)는 메모리(랜덤 액세스 메모리와 같은 휘발성 메모리, 비휘발성 메모리, 또는 이들의 적절한 조합), 하나 이상의 자기 디스크 저장 디바이스들, 광학 디스크 저장 디바이스들, 플래시 메모리 디바이스들, 그 밖에 지식 그래프 완성 장치(110)에 의해 액세스되고 원하는 정보를 저장할 수 있는 다른 형태의 저장 매체, 또는 이들의 적합한 조합일 수 있다.

[0027] 통신 버스(170)는 프로세서(120), 컴퓨터 판독 가능한 저장 매체(140)를 포함하여 시각 질의 응답 장치(110)의 다른 다양한 컴포넌트들을 상호 연결한다.

[0028] 시각 질의 응답 장치(110)는 또한 하나 이상의 입출력 장치를 위한 인터페이스를 제공하는 하나 이상의 입출력 인터페이스(150) 및 하나 이상의 통신 인터페이스(160)를 포함할 수 있다. 입출력 인터페이스(150) 및 통신 인터페이스(160)는 통신 버스(170)에 연결된다. 입출력 장치는 입출력 인터페이스(150)를 통해 시각 질의 응답 장치(110)의 다른 컴포넌트들에 연결될 수 있다.

[0029] 시각 질의 응답 장치(110)는 공정성 분류 네트워크를 통해 질문에 맞는 공정 클래스를 분류하고 분류된 공정 클래스 점수를 정답 추론 네트워크를 통해 추론한 정답 점수에 적용하여, 공정성이 필요한 질문에 공정한 답변을 출력한다.

[0030] 도 2는 본 발명의 다른 실시예에 따른 시각 질의 응답 방법을 예시한 흐름도이다. 시각 질의 응답 방법은 시각 질의 응답 장치에 의해 수행될 수 있다.

[0031] 단계 S210에서 프로세서는 질문 및 이미지로부터 시각 영역 기반의 정답 추론 모델을 통해 제1 시각 영역 벡터를 추출하고 정답 클래스에 대한 제1 점수를 산출한다.

[0032] 제1 점수를 산출하는 단계(S210)는, 이미지로부터 특징 추출 모델을 통해 시각 특징 벡터를 추출한다. 질문으로부터 상기 임베딩 모델을 통해 질문 벡터를 추출한다. 질문 벡터 및 시각 특징 벡터에 대해 시각 어텐션 모델을 통해 제1 시각 영역 가중치를 부여하여 제1 시각 영역 벡터를 생성한다. 질문 벡터와 제1 시각 영역 벡터를 결합한 특징 정보를 정답 분류 모델을 통해 정답 클래스에 맞게 분류한다.

[0033] 단계 S220에서 프로세서는 질문, 이미지, 및 제1 시각 영역 벡터로부터 질문 영역 기반의 공정성 분류 모델을 통해 공정 클래스에 대한 제2 점수를 산출한다. 예컨대, 공정 클래스는 성별 클래스일 수 있다.

[0034] 제2 점수를 산출하는 단계(S220)는, 이미지로부터 특징 추출 모델을 통해 시각 특징 벡터를 추출하고, 시각 특징 벡터와 제1 시각 영역 벡터를 결합한 제2 시각 영역 벡터를 생성한다. 질문으로부터 상기 임베딩 모델을 통해 질문 벡터를 추출한다. 질문 벡터에 대해 상기 질문 어텐션 모델을 통해 질문 영역 가중치를 부여하여 질문 영역 벡터를 생성하고, 질문 영역 벡터에 대해 시각 어텐션 모델을 통해 제2 시각 영역 가중치를 산출한다. 제2 시각 영역 벡터에 제2 시각 영역 가중치를 부여하여 제3 시각 영역 벡터를 생성한다. 제3 시각 영역 벡터를 정답 분류 모델을 통해 공정 클래스에 맞게 분류한다.

[0035] 단계 S230에서 프로세서는 정답 클래스에 대한 제1 점수와 공정 클래스에 대한 제2 점수를 이용하여 종합 점수를 출력한다.

[0036] 종합 점수를 출력하는 단계(S230)는, 정답 클래스의 개수가 공정 클래스의 개수보다 많으므로, 공정 클래스가 정답 클래스에 대응되는지 여부에 따라 공정 클래스의 레이블을 정답 클래스의 레이블에 매핑하여 제2 점수를 변환하고, 제2 점수를 반영하여 종합 점수를 출력한다.

[0037] 종합 점수를 출력하는 단계(S230)는, 정답 클래스가 배경 클래스가 아니면 제2 점수에 가중치를 부여한 후 제1 점수와 합산하여 상기 종합 점수를 출력한다. 정답 클래스가 배경 클래스이면 제2 점수를 반영하지 않고 제1 점수를 종합 점수로 출력한다.

[0038] 도 3은 본 발명의 실시예들에 따른 시각 질의 응답 장치의 정답 추론 모델 및 공정성 분류 모델을 예시한 도면이다.

[0039] 공정성 분류 모델은 공정한 성별클래스를 예측한다. 예컨대, 레이블 은 여성(Female), 남성(Male), 배경(Background)으로 구분될 수 있다. 공정성 분류 모델은 특징 추출 모델, 임베딩 모델, 질문 어텐션 모델, 시각 어텐션 모델, 정답 분류 모델을 포함할 수 있다. 공정성 분류 모델은 멀티모달 분류 모델로 멀티모달은 다양한 자원으로부터 수집된 데이터가 하나의 정보를 표현한다. 멀티모달 분류 모델에 질문과 이미지가 입력될 수 있다.

[0040] 정답 추론 모델은 시각적 정보에 관한 질의에 따른 정답을 예측한다. 정답의 예시로는 바나나(Banana), 자동차(Car), 여자(Woman), 여성(Female), 남자(Man), 소년(Boy) 등이 있을 수 있다. 정답 추론 모델은 특징 추출 모델, 임베딩 모델, 시각 어텐션 모델, 정답 분류 모델을 포함할 수 있다.

[0041] 특징 추출 모델은 레이어가 연결된 네트워크이면 가중치 및 바이어스를 학습하는 모델이다. 특징 추출 모델은 CNN(Convolutional Neural Network) 등으로 구현될 수 있다.

[0042] 임베딩 모델은 단어를 벡터로 변경하는 모델이다. 임베딩 모델은 GRU(Gated Recurrent Unit) 등으로 구현될 수 있다.

[0043] 어텐션 모델은 예측 과정에서 특정 영역을 집중하여 관련된 영역에 어텐션 가중치를 부여하는 모델이다. 어텐션 메커니즘은 키-값 자료를 통해 매핑된 값을 추출할 수 있다. 주어진 쿼리에 대한 키의 유사도를 산출하고 키에 매핑된 값을 곱해 반환한다.

[0044] 정답 분류 모델은 데이터에 대한 클래스를 예측하고 해당하는 레이블을 부여한다. 신경 네트워크 등으로 구현된 다양한 분류 모델이 적용될 수 있다.

[0045] 시각 질의 응답 장치는 정답 추론 모델이 생성한 제1 시각 영역 벡터를 공정성 분류 모델에 적용하고, 공정성 분류 모델은 제1 시각 영역 벡터를 제2 시각 영역 벡터 및 제3 시각 영역 벡터로 변환하고, 분류된 성별 클래스를 시각 질의 정답에 매핑하는 과정을 수행한다. 예컨대, 성별 클래스의 점수가 남자(Man)가 0.8이고 여자(Woman)가 0.2이면, 정답의 점수에 남자(Man)는 0.8, 남성(Male)은 0.8, 소년(Boy)은 0.8이 반영될 수 있다.

[0046] 도 3에서 Main VQA Network를 참조하면, 프로세서는 이미지로부터 특징 추출 모델을 통해 시각 특징 벡터(v)를 추출한다. 질문으로부터 임베딩 모델을 통해 질문 벡터(q)를 추출한다. 질문 벡터(q) 및 시각 특징 벡터(v)에 대해 시각 어텐션 모델을 통해 제1 시각 영역 가중치(Att_v)를 부여하여 제1 시각 영역 벡터($\acute{v}$)를 생성한다. Att_v는 중요 영역을 선정하고 벡터에 어텐션 가중치를 부여한다. ◎는 요소별 곱을 의미한다. 질문 벡터(q)와 제1 시각 영역 벡터($\acute{v}$)를 결합한 특징 정보를 정답 분류 모델을 통해 정답 클래스에 맞게 분류한다.

[0047] 도 3에서 Multimodal Classifier를 참조하면, 프로세서는 이미지로부터 특징 추출 모델을 통해 시각 특징 벡터(v)를 추출하고, 시각 특징 벡터(v)와 제1 시각 영역 벡터($\acute{v}$)를 결합한 제2 시각 영역 벡터($\dot{v}$)를 생성한다. 질문으로부터 임베딩 모델을 통해 질문 벡터(q)를 추출하고, 질문 벡터(q)에 대해 질문 어텐션 모델을 통해 질문 영역 가중치(Att_q)를 부여하여 질문 영역 벡터를 생성한다. Att_q는 중요 단어를 선정하고 벡터에 어텐션 가중치를 부여한다. 질문 영역 벡터에 대해 시각 어텐션 모델을 통해 제2 시각 영역 가중치(Att_v)를 산출한다. 제2 시각 영역 벡터($\dot{v}$)에 제2 시각 영역 가중치를 부여하여 제3 시각 영역 벡터($\hat{v}$)를 생성한다. 제3 시각 영역 벡터를 정답 분류 모델을 통해 공정 클래스에 맞게 분류한다.

[0048] 정답 추론 모델과 공정성 분류 모델은 제1 시각 영역 벡터($\acute{v}$)를 공유하여 두 모델의 최적화를 위해 함께 학습된다.

[0049] 프로세서는 정답 클래스의 개수가 공정 클래스의 개수보다 많으므로, 공정 클래스가 정답 클래스에 대응되는지 여부에 따라 공정 클래스의 레이블을 정답 클래스의 레이블에 매핑하여 제2 점수를 변환한다. 변환된 제2 점수를 반영하여 종합 점수를 출력한다. 매핑 식은 수학식 1과 같이 표현된다.

수학식 1

[0050]
$$f_{l_n, s^* \rightarrow \{0,1\}} : f(l_n, s^*) = \begin{cases} 1, & l_n \in s^* \\ 0, & l_n \notin s^* \end{cases}$$

[0051] 수학식 1에서 s^*는 예측된 클래스이다. 정답 클래스가 예측된 클래스에 대응하면 1, 대응하지 않으면 0으로 매핑된다. 제2 점수는 매핑 레이블 $M_{label} = [m_1, m_2, \ldots, m_M]$, $m_n = f(ln, s*)$을 통해 산출된다. Pc는 소프트맥스 점수이다.

수학식 2

[0052]
$$MC_{score} = Max(P_c) * M_{label}$$

[0053] 프로세서는 정답 클래스가 배경 클래스가 아니면 제2 점수에 가중치를 부여한 후 제1 점수와 합산하여 종합 점수를 출력하고, 정답 클래스가 배경 클래스이면 제2 점수를 반영하지 않고 제1 점수를 종합 점수로 출력한다. 종합 점수는 수학식 3과 같이 표현된다.

수학식 3

[0054]
$$Total_{score} = \begin{cases} VQA_{score} + \alpha MC_{score}, & s^* \neq \text{`background'} class \\ VQA_{score}, & s^* = \text{`background'} class \end{cases}$$

[0055] 프로세서는 공정 분류 점수(멀티모달 분류 점수)의 반영비율을 N 배로 설정할 수 있다. 예컨대, N은 1, 2, 3 등의 자연수로 설정될 수 있다. 시뮬레이션 결과에 의하면 반영비율을 3으로 설정할 때, 성별 편향 오차가 제일 적게 나타났다.

[0056] 도 4 및 도 5는 본 발명의 실시예들에 따라 수행된 모의실험 결과를 도시한 것이다.

[0057] 도 4를 참조하면, 공정 분류 점수(멀티모달 분류 점수)를 반영하지 않으면 성별에 오차가 발생할 수 있다. 각각의 이미지에 대해 복수의 어텐션 가중치가 적용되고 밝은 영역은 어텐션 가중치가 높은 영역을 나타낸다. 어텐션 생성뿐만 아니라 성별에 관한 정확한 정답을 예측할 수 있음을 파악할 수 있다.

[0058] 도 5를 참조하면, 공정 분류 모델이 시각 질의 응답 정보를 사용하지 않으면 클래스 예측에 오차가 발생할 수 있고, 공정 분류 모델이 멀티모달 분류 모델로서 시각 질의 응답 정보를 사용할 때 정확한 정답을 예측할 수 있음을 파악할 수 있다.

[0059] 시각 질의 응답 장치는 하드웨어, 펌웨어, 소프트웨어 또는 이들의 조합에 의해 로직회로 내에서 구현될 수 있고, 범용 또는 특정 목적 컴퓨터를 이용하여 구현될 수도 있다. 장치는 고정배선형(Hardwired) 기기, 필드 프로그램 가능한 게이트 어레이(Field Programmable Gate Array, FPGA), 주문형 반도체(Application Specific Integrated Circuit, ASIC) 등을 이용하여 구현될 수 있다. 또한, 장치는 하나 이상의 프로세서 및 컨트롤러를 포함한 시스템온칩(System on Chip, SoC)으로 구현될 수 있다.

[0060] 시각 질의 응답 장치는 하드웨어적 요소가 마련된 컴퓨팅 디바이스 또는 서버에 소프트웨어, 하드웨어, 또는 이들의 조합하는 형태로 탑재될 수 있다. 컴퓨팅 디바이스 또는 서버는 각종 기기 또는 유무선 통신망과 통신을 수행하기 위한 통신 모뎀 등의 통신장치, 프로그램을 실행하기 위한 데이터를 저장하는 메모리, 프로그램을 실행하여 연산 및 명령하기 위한 마이크로프로세서 등을 전부 또는 일부 포함한 다양한 장치를 의미할 수 있다.

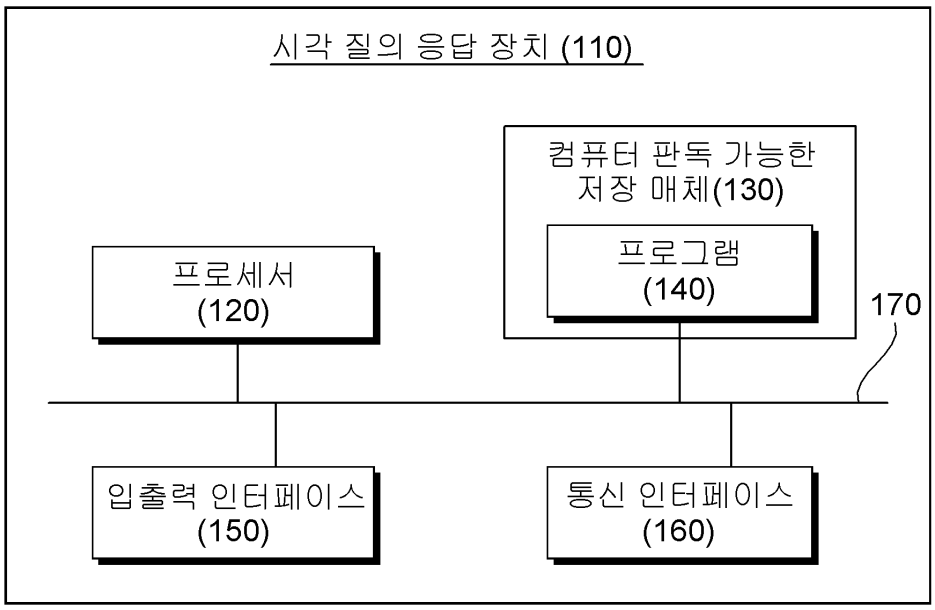
[0061] 도 2에서는 각각의 과정을 순차적으로 실행하는 것으로 기재하고 있으나 이는 예시적으로 설명한 것에 불과하고, 이 분야의 기술자라면 본 발명의 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 도 2에 기재된 순서를 변경하여 실행하거나 또는 하나 이상의 과정을 병렬적으로 실행하거나 다른 과정을 추가하는 것으로 다양하게 수정 및 변형하여 적용 가능할 것이다.

[0062] 본 실시예들에 따른 동작은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능한 매체에 기록될 수 있다. 컴퓨터 판독 가능한 매체는 실행을 위해 프로세서에 명령어를 제공하는데 참여한 임의의 매체를 나타낸다. 컴퓨터 판독 가능한 매체는 프로그램 명령, 데이터 파일, 데이터 구조 또는 이들의 조합을 포함할 수 있다. 예를 들면, 자기 매체, 광기록 매체, 메모리 등이 있을 수 있다. 컴퓨터 프로그램은 네트워크로 연결된 컴퓨터 시스템 상에 분산되어 분산 방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수도 있다. 본 실시예를 구현하기 위한 기능적인(Functional) 프로그램, 코드, 및 코드 세그먼트들은 본 실시예가 속하는 기술분야의 프로그래머들에 의해 용이하게 추론될 수 있을 것이다.

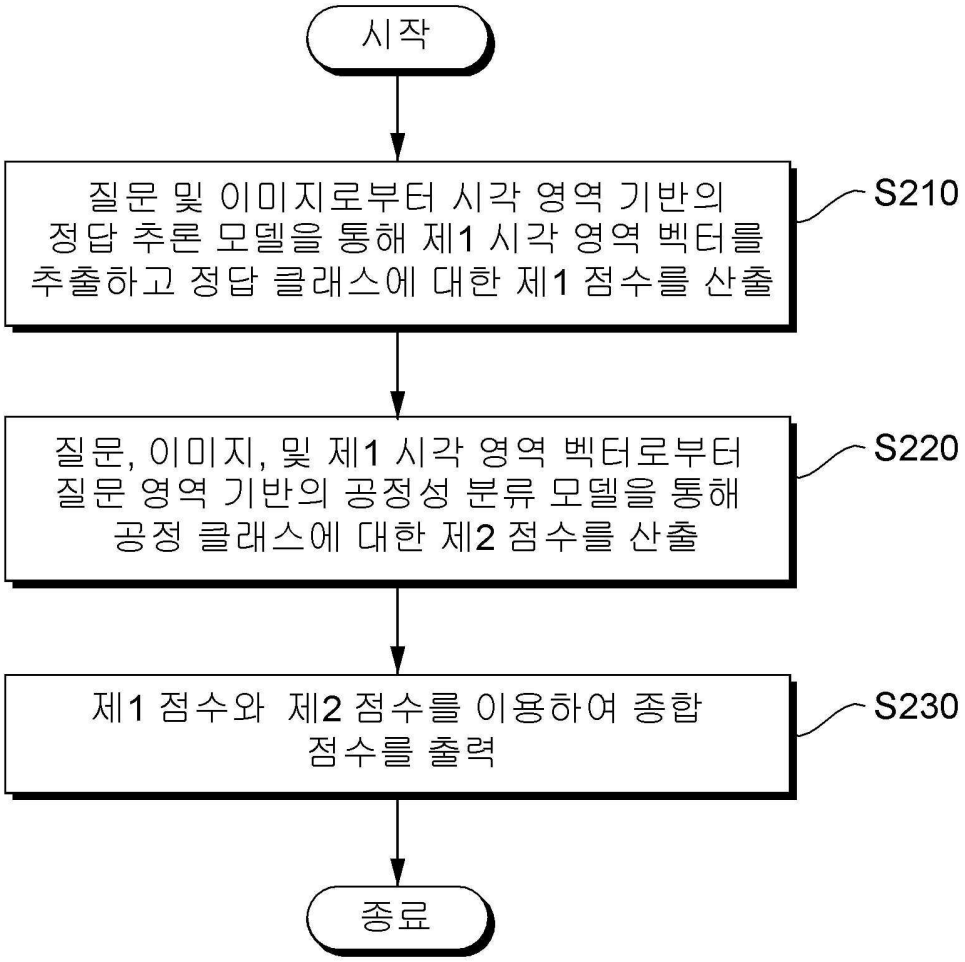
[0063] 본 실시예들은 본 실시예의 기술 사상을 설명하기 위한 것이고, 이러한 실시예에 의하여 본 실시예의 기술 사상의 범위가 한정되는 것은 아니다. 본 실시예의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 실시예의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

도면

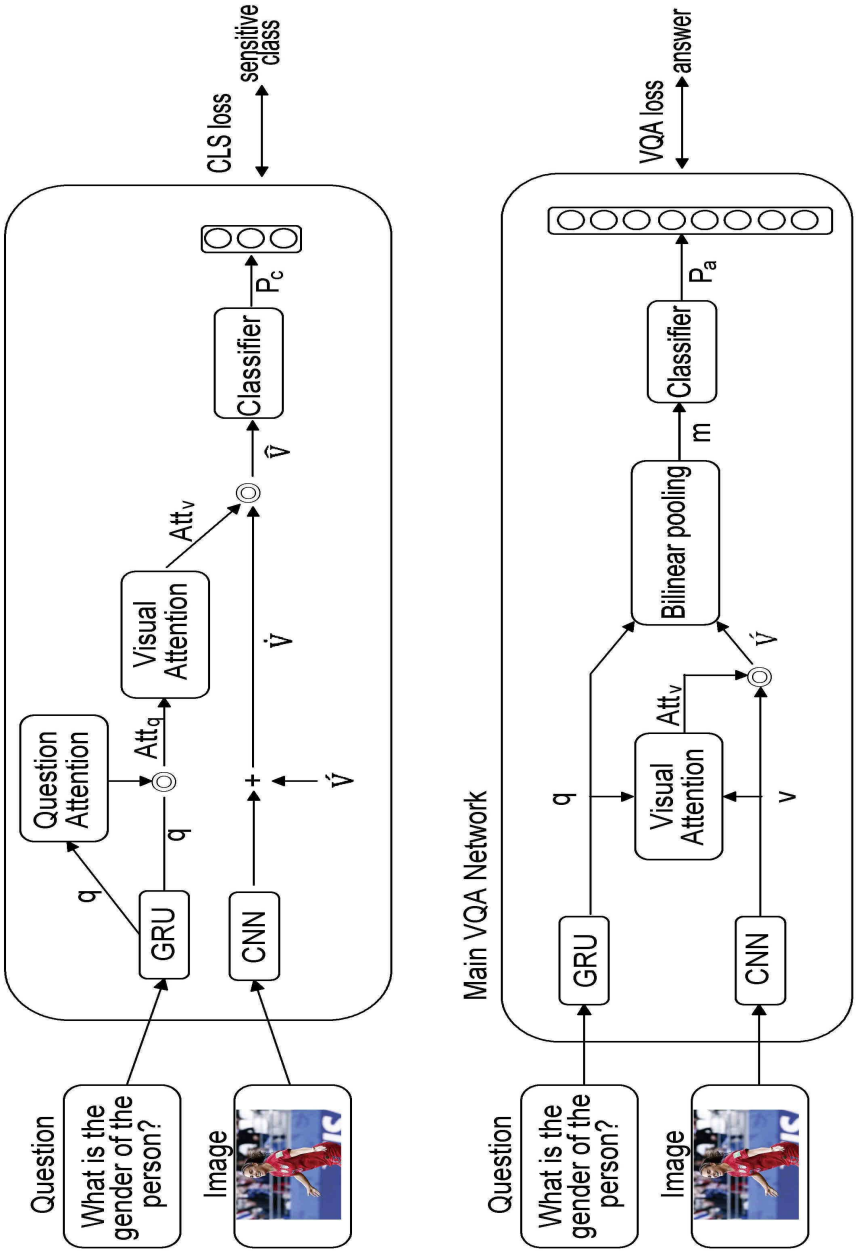
도면1


시각 질의 응답 장치 (110)
컴퓨터 판독 가능한
저장 매체(130)
프로그램
(140)
프로세서
(120)
170
입출력 인터페이스
(150)
통신 인터페이스
(160)

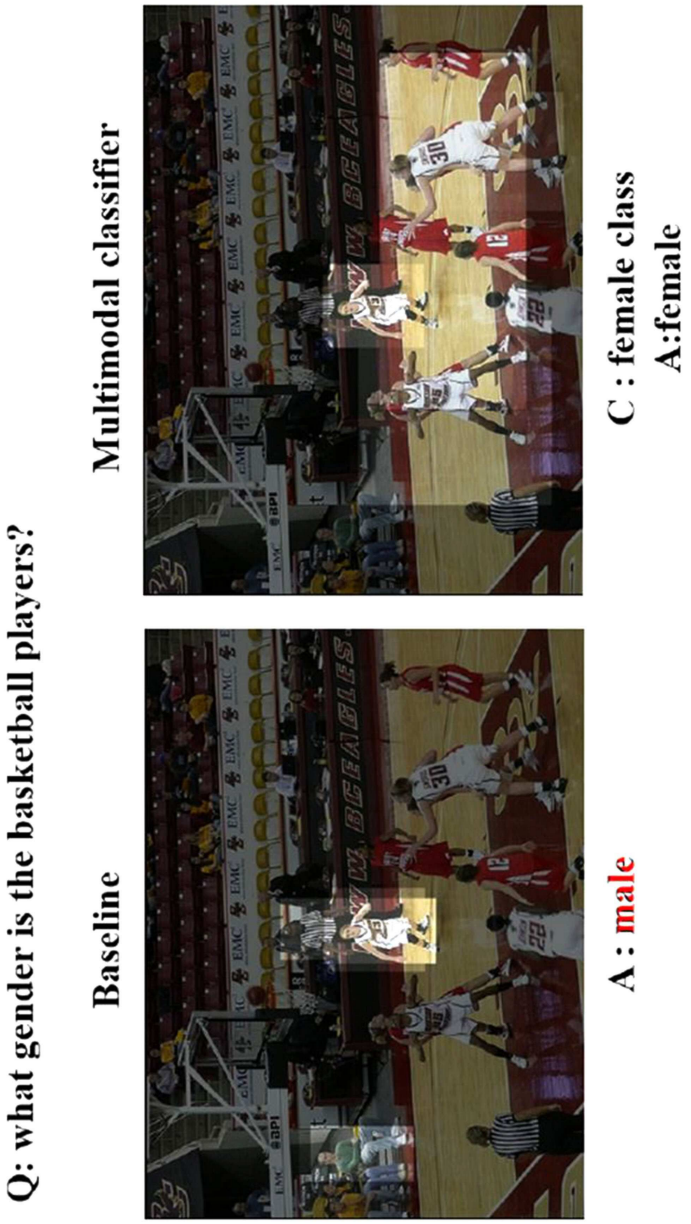
도면2


시작
질문 및 이미지로부터 시각 영역 기반의 정답 추론 모델을 통해 제1 시각 영역 벡터를 추출하고 정답 클래스에 대한 제1 점수를 산출
S210
질문, 이미지, 및 제1 시각 영역 벡터로부터 질문 영역 기반의 공정성 분류 모델을 통해 공정 클래스에 대한 제2 점수를 산출
S220
제1 점수와 제2 점수를 이용하여 종합 점수를 출력
S230
종료

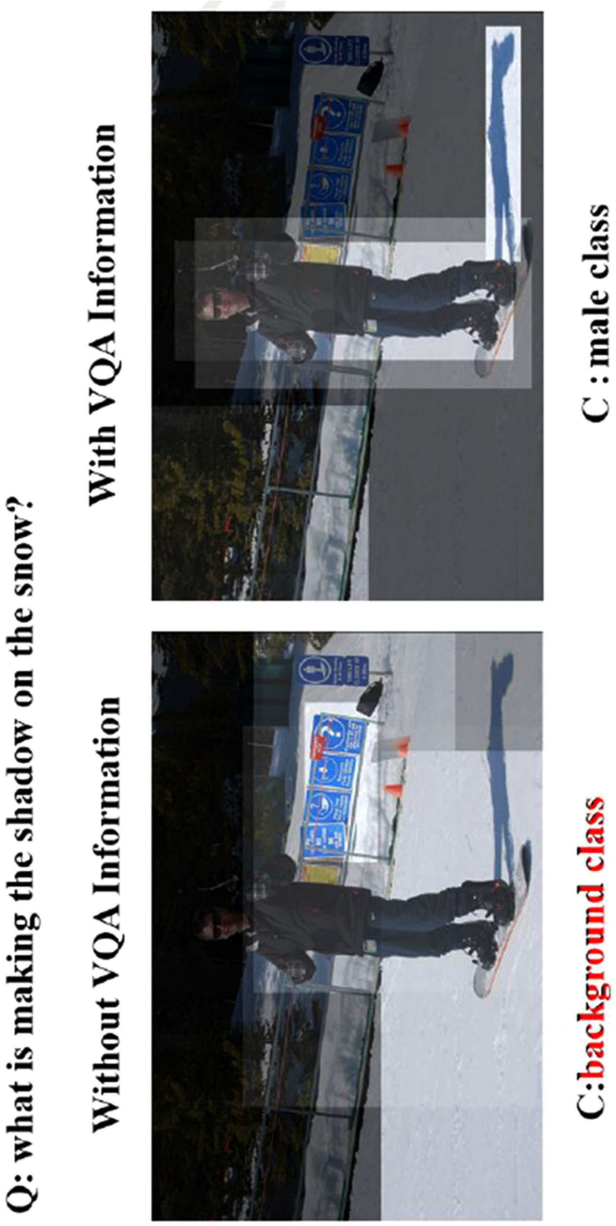
도면3


CLS loss
sensitive class
Classifier
P_c
$\tilde{V}$
Att_v
Visual Attention
Att_q
Question Attention
q
GRU
CNN
$\dot{V}$
Question
What is the gender of the person?
Image
Main VQA Network
VQA loss
answer
Classifier
P_a
m
Bilinear pooling
$\acute{V}$
Att_v
Visual Attention
q
v
GRU
CNN
Question
What is the gender of the person?
Image

도면4


Q: what gender is the basketball players?
Baseline
A : male
Multimodal classifier
C : female class
A:female

도면5


Q: what is making the shadow on the snow?
Without VQA Information
C:background class
With VQA Information
C : male class