

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)(11) 공개번호 10-2025-0056439
(43) 공개일자 2025년04월28일

(51) 국제특허분류(Int. Cl.)

G06V 10/424 (2022.01) G06F 40/205 (2020.01)

G06F 40/268 (2020.01) G06F 40/279 (2020.01)

G06N 3/08 (2023.01) G06T 11/60 (2006.01)

G06V 10/72 (2022.01) G06V 10/774 (2022.01)

(52) CPC특허분류

G06V 10/424 (2023.08)

G06F 40/205 (2020.01)

(21) 출원번호 10-2023-0140167

(22) 출원일자 2023년10월19일

심사청구일자 2023년10월19일

(71) 출원인

한국기술교육대학교 산학협력단

충청남도 천안시 동남구 병천면 충절로 1600 (한국기술교육대학교내)

(72) 발명자

오홍선

충청남도 천안시 동남구 병천면 충절로 1600, 434호(한국기술교육대학교 공학2관)

손성빈

경상북도 구미시 봉곡로24길 33-11, 107동 2001호(봉곡동, 봉곡e편한세상)

김경민

충청남도 천안시 서북구 불당1로 82, 610동 604호(불당동, 대원칸타빌아파트)

(74) 대리인

특허법인오암

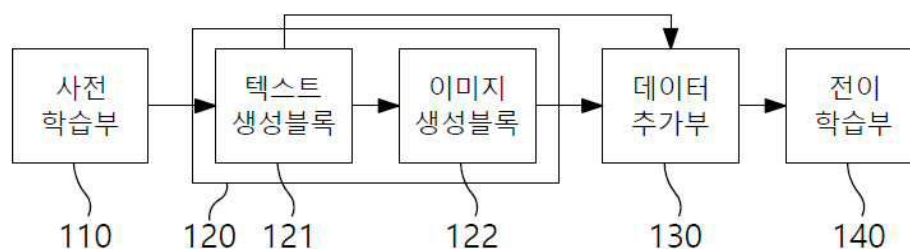
전체 청구항 수 : 총 8 항

(54) 발명의 명칭 이미지 캡서닝 장치 및 그 방법

(57) 요약

본 발명은 이미지 캡서닝 장치 및 그 방법에 관한 것으로, 기준데이터셋을 이용한 사전 학습을 통해 인공지능모델을 사전 학습시키는 사전학습부; 다운스트림 데이터셋이 입력될 경우 상기 다운스트림 데이터셋에 대응하는 추가텍스트와 추가이미지를 포함하는 데이터쌍을 생성하는 데이터쌍생성부; 생성된 상기 데이터쌍을 상기 다운스트림 데이터셋에 추가하여 보정데이터셋을 입력시키는 데이터추가부; 및 상기 보정데이터셋을 사전 학습된 상기 인공지능모델에 입력시켜 전이학습을 수행하는 전이학습부;를 포함함으로써, 적은 크기의 데이터셋의 입력으로도 이미지 캡서닝 성능을 향상시킬 수 있어 인공지능모델의 강건성을 효과적으로 확보할 수 있다.

대표도 - 도2



(52) CPC특허분류

G06F 40/268 (2020.01)

G06F 40/279 (2020.01)

G06N 3/08 (2023.01)

G06T 11/60 (2013.01)

G06V 10/72 (2023.08)

G06V 10/774 (2023.08)

이 발명을 지원한 국가연구개발사업

과제고유번호	1345370812
과제번호	2021RIS-004
부처명	교육부
과제관리(전문)기관명	한국연구재단
연구사업명	지자체-대학 협력기반지역혁신사업
연구과제명	(대전세종충남지역혁신플랫폼)충남대학교
기 여 율	1/1
과제수행기관명	(대전세종충남지역혁신플랫폼)충남대학교
연구기간	2023.04.01 ~ 2024.02.29

명세서

청구범위

청구항 1

기준데이터셋을 이용한 사전 학습을 통해 인공지능모델을 사전 학습시키는 사전학습부;

다운스트림 데이터셋이 입력될 경우 상기 다운스트림 데이터셋에 대응하는 추가텍스트와 추가이미지를 포함하는 데이터쌍을 생성하는 데이터쌍생성부;

생성된 상기 데이터쌍을 상기 다운스트림 데이터셋에 추가하여 보정데이터셋을 입력시키는 데이터추가부; 및

상기 보정데이터셋을 사전 학습된 상기 인공지능모델에 입력시켜 전이학습을 수행하는 전이학습부;

를 포함하는 이미지 캡셔닝 장치.

청구항 2

청구항 1에 있어서,

상기 데이터쌍생성부는,

상기 다운스트림 데이터셋의 각 문장(sentence)에 대응하는 형태를 갖는 상기 추가텍스트를 생성하는 텍스트생성블록; 및

상기 추가텍스트에 대응하여 상기 추가이미지를 복수개 생성하는 이미지생성블록;

을 포함하는 이미지 캡셔닝 장치.

청구항 3

청구항 2에 있어서,

상기 텍스트생성블록은,

상기 문장에서 명사 및 형용사를 추출하고, 상기 추출된 명사 및 형용사에서 객체(object) 및 상태(state)를 선택하는 방식으로 샘플링한 후에, 상기 객체 및 상태를 이용하여 상기 다운스트림 데이터셋의 각 문장에 대응하는 형태를 갖는 신규텍스트를 생성하고, 상기 신규텍스트의 스타일전이를 수행하는 방식으로 상기 추가텍스트를 생성하는

이미지 캡셔닝 장치.

청구항 4

청구항 3에 있어서,

상기 텍스트생성블록은,

상기 다운스트림 데이터셋으로부터 랜덤문장(random sentence)을 복수개 추출하고, 상기 랜덤문장과 상기 신규텍스트에 대해 이미지뷰(image view)와 관련된 단어구를 추가하는 방식으로 상기 스타일전이를 수행하는

이미지 캡셔닝 장치.

청구항 5

사전학습부에서 기준데이터셋을 이용한 사전 학습을 통해 인공지능모델을 사전 학습시키는 단계;

데이터쌍생성부에서 다운스트림 데이터셋이 입력될 경우 상기 다운스트림 데이터셋에 대응하는 추가텍스트와 추가이미지를 포함하는 데이터쌍을 생성하는 단계;

데이터추가부에서 생성된 상기 데이터쌍을 상기 다운스트림 데이터셋에 추가하여 보정데이터셋을 입력시키는 단계; 및

전이학습부에서 상기 보정데이터셋을 사전 학습된 상기 인공지능모델에 입력시켜 전이학습을 수행하는 단계;
를 포함하는 이미지 캡서닝 방법.

청구항 6

청구항 5에 있어서,

상기 데이터쌍을 생성하는 단계는,

상기 데이터쌍생성부에 구비되는 텍스트생성블록에서 상기 다운스트림 데이터셋의 각 문장에 대응하는 형태를 갖는 상기 추가텍스트를 생성하는 단계; 및

상기 데이터쌍생성부에 구비되는 이미지생성블록에서 상기 추가텍스트에 대응하여 상기 추가이미지를 복수개 생성하는 단계;

를 포함하는 이미지 캡서닝 방법.

청구항 7

청구항 6에 있어서,

상기 추가텍스트를 생성하는 단계는,

상기 텍스트생성블록에서 상기 문장에서 명사 및 형용사를 추출하고, 상기 추출된 명사 및 형용사에서 객체(object) 및 상태(state)를 선택하는 방식으로 샘플링한 후에, 상기 객체 및 상태를 이용하여 상기 다운스트림 데이터셋의 각 문장에 대응하는 형태를 갖는 신규텍스트를 생성하고, 상기 신규텍스트의 스타일전이를 수행하는 방식으로 상기 추가텍스트를 생성하는

이미지 캡서닝 방법.

청구항 8

청구항 7에 있어서,

상기 추가텍스트를 생성하는 단계는,

상기 텍스트생성블록에서 상기 다운스트림 데이터셋으로부터 랜덤문장(random sentence)을 복수개 추출하고, 상기 랜덤문장과 상기 신규텍스트에 대해 이미지뷰(image view)와 관련된 단어구를 추가하는 방식으로 상기 스타일전이를 수행하는

이미지 캡서닝 방법.

발명의 설명

기술 분야

본 발명은 기준데이터셋을 이용한 사전 학습을 통해 인공지능모델을 사전 학습시킨 후에, 다운스트림 데이터셋이 입력될 경우 다운스트림 데이터셋에 대응하는 추가텍스트와 추가이미지를 포함하는 데이터쌍을 생성하고, 이 데이터쌍을 다운스트림 데이터셋에 추가하여 보정데이터셋을 입력시키며, 보정데이터셋을 사전 학습된 인공지능

[0001]

모델에 입력시켜 전이학습을 수행함으로써, 적은 크기의 데이터셋의 입력으로도 이미지 캡셔닝 성능을 향상시킬 수 있어 인공지능모델의 강건성을 효과적으로 확보할 수 있는 이미지 캡셔닝 장치 및 그 방법에 관한 것이다.

배경 기술

- [0002] 잘 알려진 바와 같이, 최근 고성능 GPU(Graphic Processing Unit)의 사용으로 처리 가능한 연산량이 대폭 증가함에 따라, 패턴을 인식하는데 연산량이 많이 필요한 딥러닝기술이 계속 연구되고 있다.
- [0003] 예를 들면, 합성곱신경망(CNN : Convolution neural network) 등과 같은 신경망의 발달과 함께 객체인식, 이미지분류 등과 같은 이미지프로세싱 연구가 상당히 빠른 속도로 진행되고 있는데, 헬스케어 분야의 딥러닝기술 적용으로 인하여, 사람이 포함된 영상이해, 상황인식 등과 같은 연구가 심도있게 진행되면서, 딥러닝기반의 이미지 캡셔닝의 중요도가 부각되고 있다.
- [0004] 여기에서, 이미지 캡셔닝(Image captioning)은 입력된 이미지를 인공지능모델(예를 들면, 합성곱신경망(CNN) 등)을 통하여 특징을 추출하고, 학습된 단어특징공간에 매핑함으로써, 입력된 이미지의 설명문을 생산하는 기법으로서, 영상이해 및 상황인식에 가장 근접한 연구 중 하나이다.
- [0005] 예를 들면, 이미지 캡셔닝은 도 1에 도시한 바와 같이 이미지의 콘텐츠를 단어(word)와 문장(sentence)을 포함하는 언어(language)로 설명하는 기법으로, 대량의 데이터셋(large-scale dataset)을 이용하여 인공지능모델을 사전 학습(pre-training)시킨 후에, 상대적으로 적은 데이터(downstream)를 사전 학습된 인공지능모델에 입력시켜 파인튜닝(fin-tuning)하는 방식으로 수행될 수 있는데, 일반적으로 큰 모델(large model), 큰 데이터셋(large dataset)의 경우 다운스트림 입력에 대응하는 출력 성능이 높게 나타날 수 있다.
- [0006] 하지만, 큰 데이터셋(large data)을 사용한 사전학습모델들과는 달리, 다운스트림데이터는 상대적으로 적은 데이터이기 때문에, 인공지능모델의 강건성(robustness)을 달성하기 어려운 문제점이 있고, 현실세계(real world)의 분포가 다운스트림 데이터셋 분포보다 상대적으로 크기 때문에, 다운스트림 데이터셋은 실용응용분야(practical application)에 맞지 않아 제로샷러닝(Zero-shot learning), 퓨샷러닝(Few-shot learning)에 적용하기 어려운 문제점이 있다.
- [0007] 예를 들면, 다운스트림 태스크(task)마다 분포가 존재하기 때문에, 동일 인공지능모델을 사용할 경우 다운스트림 데이터셋에서의 성능이 일관되지 못하게 되는데, 개 검출(Dog Detection)을 기반으로 고양이 검출(Cat Detection)이 불가능하여 서로 다른 데이터셋들에 대해서는 인퍼런스(inference)성능을 보장하기 어려운 문제점이 있다.
- [0008] 또한, 이미지 캡셔닝의 성능이 데이터셋에 의존적이기 때문에, 다운스트림 테스트를 진행하기 위해서는 각 데이터셋으로 매번 학습을 진행해야 하는 문제점이 있다.

선행기술문헌

특허문헌

- [0009] (특허문헌 0001) 1. 한국등록특허 제10-2293950호(2021.08.20.등록)

발명의 내용

해결하려는 과제

- [0010] 본 발명은 기준데이터셋을 이용한 사전 학습을 통해 인공지능모델을 사전 학습시킨 후에, 다운스트림 데이터셋이 입력될 경우 다운스트림 데이터셋에 대응하는 추가텍스트와 추가이미지를 포함하는 데이터쌍을 생성하고, 이 데이터쌍을 다운스트림 데이터셋에 추가하여 보정데이터셋을 입력시키며, 보정데이터셋을 사전 학습된 인공지능모델에 입력시켜 전이학습을 수행함으로써, 적은 크기의 데이터셋의 입력으로도 이미지 캡셔닝 성능을 향상시킬 수 있어 인공지능모델의 강건성을 효과적으로 확보할 수 있는 이미지 캡셔닝 장치 및 그 방법을 제공하고자 한다.
- [0011] 본 발명의 실시예들의 목적은 이상에서 언급한 목적으로 제한되지 않으며, 언급되지 않은 또 다른 목적들은 아래의 기재로부터 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

- [0012] 본 발명의 일 측면에 따르면, 기준데이터셋을 이용한 사전 학습을 통해 인공지능모델을 사전 학습시키는 사전 학습부; 다운스트림 데이터셋이 입력될 경우 상기 다운스트림 데이터셋에 대응하는 추가텍스트와 추가이미지를 포함하는 데이터쌍을 생성하는 데이터쌍생성부; 생성된 상기 데이터쌍을 상기 다운스트림 데이터셋에 추가하여 보정데이터셋을 입력시키는 데이터추가부; 및 상기 보정데이터셋을 사전 학습된 상기 인공지능모델에 입력시켜 전이학습을 수행하는 전이학습부;를 포함하는 이미지 캡서닝 장치가 제공될 수 있다.
- [0013] 또한, 본 발명의 일 측면에 따르면, 상기 데이터쌍생성부는, 상기 다운스트림 데이터셋의 각 문장에 대응하는 형태를 갖는 상기 추가텍스트를 생성하는 텍스트생성블록; 및 상기 추가텍스트에 대응하여 상기 추가이미지를 복수개 생성하는 이미지생성블록;을 포함하는 이미지 캡서닝 장치가 제공될 수 있다.
- [0014] 또한, 본 발명의 일 측면에 따르면, 상기 텍스트생성블록은, 상기 문장에서 명사 및 형용사를 추출하고, 상기 추출된 명사 및 형용사에서 객체(object) 및 상태(state)를 선택하는 방식으로 샘플링한 후에, 상기 객체 및 상태를 이용하여 상기 다운스트림 데이터셋의 각 문장에 대응하는 형태를 갖는 신규텍스트를 생성하고, 상기 신규텍스트의 스타일전이를 수행하는 방식으로 상기 추가텍스트를 생성하는 이미지 캡서닝 장치가 제공될 수 있다.
- [0015] 또한, 본 발명의 일 측면에 따르면, 상기 텍스트생성블록은, 상기 다운스트림 데이터셋으로부터 랜덤문장(random sentence)을 복수개 추출하고, 상기 랜덤문장과 상기 신규텍스트에 대해 이미지뷰(image view)와 관련된 단어구를 추가하는 방식으로 상기 스타일전이를 수행하는 이미지 캡서닝 장치가 제공될 수 있다.
- [0016] 본 발명의 다른 측면에 따르면, 사전학습부에서 기준데이터셋을 이용한 사전 학습을 통해 인공지능모델을 사전 학습시키는 단계; 데이터쌍생성부에서 다운스트림 데이터셋이 입력될 경우 상기 다운스트림 데이터셋에 대응하는 추가텍스트와 추가이미지를 포함하는 데이터쌍을 생성하는 단계; 데이터추가부에서 생성된 상기 데이터쌍을 상기 다운스트림 데이터셋에 추가하여 보정데이터셋을 입력시키는 단계; 및 전이학습부에서 상기 보정데이터셋을 사전 학습된 상기 인공지능모델에 입력시켜 전이학습을 수행하는 단계;를 포함하는 이미지 캡서닝 방법이 제공될 수 있다.
- [0017] 또한, 본 발명의 다른 측면에 따르면, 상기 데이터쌍을 생성하는 단계는, 상기 데이터쌍생성부에 구비되는 텍스트생성블록에서 상기 다운스트림 데이터셋의 각 문장(sentence)에 대응하는 형태를 갖는 상기 추가텍스트를 생성하는 단계; 및 상기 데이터쌍생성부에 구비되는 이미지생성블록에서 상기 추가텍스트에 대응하여 상기 추가이미지를 복수개 생성하는 단계;를 포함하는 이미지 캡서닝 방법이 제공될 수 있다.
- [0018] 또한, 본 발명의 다른 측면에 따르면, 상기 추가텍스트를 생성하는 단계는, 상기 텍스트생성블록에서 상기 문장에서 명사 및 형용사를 추출하고, 상기 추출된 명사 및 형용사에서 객체(object) 및 상태(state)를 선택하는 방식으로 샘플링한 후에, 상기 객체 및 상태를 이용하여 상기 다운스트림 데이터셋의 각 문장에 대응하는 형태를 갖는 신규텍스트를 생성하고, 상기 신규텍스트의 스타일전이를 수행하는 방식으로 상기 추가텍스트를 생성하는 이미지 캡서닝 방법이 제공될 수 있다.
- [0019] 또한, 본 발명의 다른 측면에 따르면, 상기 추가텍스트를 생성하는 단계는, 상기 텍스트생성블록에서 상기 다운스트림 데이터셋으로부터 랜덤문장(random sentence)을 복수개 추출하고, 상기 랜덤문장과 상기 신규텍스트에 대해 이미지뷰(image view)와 관련된 단어구를 추가하는 방식으로 상기 스타일전이를 수행하는 이미지 캡서닝 방법이 제공될 수 있다.

발명의 효과

- [0020] 본 발명은 기준데이터셋을 이용한 사전 학습을 통해 인공지능모델을 사전 학습시킨후에, 다운스트림 데이터셋이 입력될 경우 다운스트림 데이터셋에 대응하는 추가텍스트와 추가이미지를 포함하는 데이터쌍을 생성하고, 이 데이터쌍을 다운스트림 데이터셋에 추가하여 보정데이터셋을 입력시키며, 보정데이터셋을 사전 학습된 인공지능모델에 입력시켜 전이학습을 수행함으로써, 적은 크기의 데이터셋의 입력으로도 이미지 캡서닝 성능을 향상시킬 수 있어 인공지능모델의 강건성을 효과적으로 확보할 수 있다.

도면의 간단한 설명

- [0021] 도 1은 종래의 이미지 캡서닝을 설명하기 위한 도면이며,
도 2는 본 발명의 일 실시예에 따른 이미지 캡서닝 장치의 블록구성도이고,

도 3 내지 도 8은 본 발명의 일 실시예에 따른 이미지 캡서닝 장치의 상세구성을 설명하기 위한 도면이며, 도 9 및 도 10은 본 발명의 다른 실시예에 따라 이미지 캡서닝을 수행하는 과정을 나타낸 플로우차트이다.

발명을 실시하기 위한 구체적인 내용

- [0022] 본 발명의 실시예들에 대한 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나 본 발명은 이하에서 개시되는 실시예들에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있으며, 단지 본 실시예들은 본 발명의 개시가 완전하도록 하고, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명은 청구항의 범주에 의해 정의될 뿐이다. 명세서 전체에 걸쳐 동일 참조 부호는 동일 구성 요소를 지칭한다.
- [0023] 본 발명의 실시예들을 설명함에 있어서 공지 기능 또는 구성에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략할 것이다. 그리고 후술되는 용어들은 본 발명의 실시예에서의 기능을 고려하여 정의된 용어들로서 이는 사용자, 운용자의 의도 또는 관례 등에 따라 달라질 수 있다. 그러므로 그 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다.
- [0024] 이하, 첨부된 도면을 참조하여 본 발명의 실시예를 상세히 설명하기로 한다.
- [0025] 도 2는 본 발명의 일 실시예에 따른 이미지 캡서닝 장치의 블록구성도이고, 도 3 내지 도 8은 본 발명의 일 실시예에 따른 이미지 캡서닝 장치의 상세구성을 설명하기 위한 도면이다.
- [0026] 도 2 내지 도 8을 참조하면, 본 발명의 일 실시예에 따른 이미지 캡서닝 장치는 사전학습부(110), 데이터쌍생성부(120), 데이터추가부(130), 전이학습부(140) 등을 포함할 수 있다.
- [0027] 사전학습부(110)는 기준데이터셋을 이용한 사전 학습을 통해 인공지능모델을 사전 학습시키는 구성부로, 기준데이터셋은 데이터베이스에 미리 수집된 대량의 데이터셋, 네트워크를 통해 크롤링 등의 방식으로 수집되는 대량의 데이터셋 등을 포함할 수 있으며, 이러한 데이터셋은 이미지(Image)와 캡션(Caption)의 데이터쌍으로 입력될 수 있다.
- [0028] 여기에서, 인공지능모델은 예를 들면, 합성곱신경망(CNN), 순환신경망(RNN : recurrent Neural Network), NIC(Neural Image Caption)모델, BUTD(Bottom-Up and Top-Down Attention)모델, BERT(Bidirectional Encoder Representations from Transformers)모델, CNN 및 RNN을 결합한 하이브리드모델 등을 포함할 수 있다.
- [0029] 데이터쌍생성부(120)는 다운스트림 데이터셋이 입력될 경우 다운스트림 데이터셋에 대응하는 추가텍스트와 추가 이미지를 포함하는 데이터쌍을 생성하는 구성부로, 텍스트생성블록(121), 이미지생성블록(122) 등을 포함할 수 있다.
- [0030] 여기에서, 텍스트생성블록(121)은 다운스트림 데이터셋의 각 문장(sentence)에 대응하는 형태(또는 표현, 단어, 스타일)를 갖는 추가텍스트를 생성하는 것으로, 문장에서 명사 및 형용사를 추출하고, 추출된 명사 및 형용사에서 객체(object) 및 상태(state)를 선택하는 방식으로 샘플링한 후에, 객체 및 상태를 이용하여 다운스트림 데이터셋의 각 문장에 대응하는 형태(또는 표현, 단어, 스타일)를 갖는 신규텍스트를 생성하고, 신규텍스트의 스타일전이를 수행하는 방식으로 추가텍스트를 생성할 수 있다.
- [0031] 또한, 텍스트생성블록(121)은 다운스트림 데이터셋으로부터 랜덤문장(random sentence)을 복수개 추출하고, 랜덤문장과 신규텍스트에 대해 이미지뷰(image view)와 관련된 단어구를 추가하는 방식으로 스타일전이를 수행할 수 있다.
- [0032] 예를 들면, 텍스트생성블록(121)은 도 4에 도시한 바와 같이 텍스트제너레이터(Text generator)를 포함하여 다운스트림 데이터셋에 존재하는 문장과 유사한 문장을 캡션항목에 대응하여 생성할 수 있는데, NER추출기(Named Entity Recognition extractor)를 통해 다운스트림 데이터셋의 캡션항목에서 랜덤으로 복수개의 문장을 추출하여 샘플링하고, 샘플링된 복수의 문장에서 각각의 객체 및 상태를 추출하는 방식으로 객체에 대응하는 명사와 상태에 대응하는 형용사를 NER셋으로 하여 추출할 수 있다.
- [0033] 또한, 텍스트생성블록(121)은 도 4에 도시한 바와 같이 NER샘플러(Named Entity Recognition sampler)를 이용하여 NER추출기를 통해 추출된 NER셋에 존재하는 객체와 상태를 랜덤하게 샘플링하여 복수개의 객체와 복수개의 상태를 생성하는 방식으로 추가텍스트에 추가할 객체 및 상태를 선택할 수 있다. 여기에서, 객체는 예를 들면,

3-4개를 생성할 수 있고, 상태는 예를 들면, 1-2개를 생성할 수 있다.

- [0034] 그리고, 텍스트생성블록(121)은 도 4에 도시한 바와 같이 설명기반 텍스트제너레이터(Instruction-based Text Generator)를 통해 선택된 객체 및 상태를 포함하는 새로운 신규텍스트(generated text)를 생성할 수 있는데, 설명(Instruction)과 입력(input)을 통해 어떤 단어를 추가할 것인지에 대한 정보를 제너레이터에 전달할 수 있고, 입력은 선택된 객체 및 상태를 사용할 수 있다.
- [0035] 예를 들면, 설명(Instruction)은 "Make a description of specific picture, Use word in input. Do not write word like the picture shows"이고, 입력(Input)은 "wine, fishing, boat"인 포맷으로 제너레이터에 입력될 수 있다.
- [0036] 다음에, 텍스트제너레이터(Text generator)를 통해 획득된 문장(즉, 설명(Instruction)과 입력(Input))은 다운스트림 데이터셋의 스타일과 차이가 존재할 수 있기 때문에, 텍스트생성블록(121)은 도 5에 도시한 바와 같이 설명기반 스타일트랜스퍼(Instruction-based Style transfer)를 통해 문장의 스타일을 변환시킬 수 있다.
- [0037] 이러한 텍스트생성블록(121)은 다운스트림 데이터셋으로부터 복수개(N개)의 랜덤문장을 추출하여 추출된 문장을 설명(instruction)으로 구성하고, 생성된 신규텍스트(generated text)를 변환시킬 문장인 입력(input)으로 구성할 수 있다.
- [0038] 여기에서, 설명(Instruction)은 "Change text style of given input sentence into {sentence 1 },... {sentence N }"이고, 입력(Input)은 "generated text"인 포맷으로 구성할 수 있고, 예를 들면, "A bottle of wine and two glasses are placed outdoors, while a fishing boat sails in tanzania."를 "A bottle of wine and two glasses are placed outdoors, while a fishing boat sails in the Zanzibar Channel, Tanzania."로 하여 스타일을 변환시킬 수 있다.
- [0039] 또한, 텍스트생성블록(121)은 랜덤문장과 신규텍스트를 이용하여 스타일이 변환된 문장에 대해 이미지뷰와 관련된 단어구를 추가할 수 있는데, 도 5에 도시한 바와 같이 사진의 구도인 이미지뷰와 관련된 단어구를 뷰사전테이터베이스(View dictionary)에서 추출하여 변환된 문장에 추가할 수 있다.
- [0040] 이러한 텍스트생성블록(121)은 도 5에 도시한 바와 같이 뷰선택기(View selector)를 통해 사진의 구도와 관련된 단어구들(예를 들면, A wide shot of ~, A straight view ~ 등)을 변환된 문장에 추가한 후 설명(instruction)에 입력시킬 수 있고, 이 후 설명기반 모델을 이용하여 가장 자연스러운 구도를 선택할 수 있다.
- [0041] 예를 들면, 설명(Instruction)은 "Choose the best composition when you taking picture of {생성된 문장}"이고, 입력(Input)은 "무작위로 선정한 구도 N개"인 포맷으로 입력할 수 있고, 예를 들면, "A bottle of wine and two glasses are placed outdoors, while a fishing boat sails in the Zanzibar Channel, Tanzania."를 "A straight view of bottle of wine and two glasses are placed outdoors, while a fishing boat sails in the Zanzibar Channel, Tanzania."로 하여 스타일변환 텍스트로 생성할 수 있다.
- [0042] 이러한 스타일변환 텍스트는 추가텍스트로 하여 이미지생성블록(122) 및 데이터추가부(130)로 제공될 수 있다.
- [0043] 이미지생성블록(122)은 추가텍스트에 대응하여 추가이미지를 복수개 생성하는 것으로, 이미지제너레이터(Image generator)를 통해 스타일변환 텍스트(즉, 추가텍스트)에 대응하는 이미지를 생성할 수 있는데, 도 6에 도시한 바와 같이 다운스트림 데이터셋에 대응하여 유사하게 변환된 스타일변환 텍스트를 이용하여 신규이미지를 생성할 수 있다.
- [0044] 여기에서, 신규이미지는 하나의 스타일변환 텍스트에 대해 복수개(N개)의 이미지를 생성하여 스타일변환 텍스트에 대한 다양성을 보존할 수 있으며, 이미지제너레이터 모델은 디퓨전(diffusion)을 기반으로 하는 모델뿐만 아니라 종래에 제시된 다양한 모델을 사용할 수 있다.
- [0045] 예를 들면, 이미지생성블록(122)은 추가텍스트가 "A group of farmers can be seen discussing and examining the bright red tomatoes in the greenhouse."인 경우 및 "A young girl with freckles is seen in close up, smiling and waving at the camera on a sunny day."인 경우 도 7에 도시한 바와 같은 추가이미지를 각각 생성할 수 있다.
- [0046] 상술한 바와 같이 생성된 추가이미지는 데이터추가부(130)로 제공될 수 있다.
- [0047] 데이터추가부(130)는 생성된 데이터쌍을 다운스트림 데이터셋에 추가하여 보정데이터셋을 입력시키는 구성부로, 텍스트생성블록(121)을 통해 생성된 추가텍스트와 이미지생성블록(122)을 통해 생성된 추가이미지는 데이터쌍으

로 하여 데이터쌍생성부(120)로부터 제공될 수 있으며, 이러한 데이터쌍에서 추가텍스트는 도 8에 도시한 바와 같이 다운스트림 데이터셋의 캡션항목에 추가할 수 있고, 추가이미지는 다운스트림 데이터셋의 이미지항목에 추가하여 보정데이터셋을 입력시킬 수 있다.

- [0048] 전이학습부(140)는 보정데이터셋을 사전 학습된 인공지능모델에 입력시켜 전이학습을 수행하는 구성부로, 데이터추가부(130)로부터 입력되는 보정데이터셋을 이용하여 사전학습부(110)를 통해 사전 학습된 인공지능모델에 해당 보정데이터셋을 입력하여 파인튜닝(fine-tuning)을 수행한 후 출력함으로써, 이미지캡서닝을 수행할 수 있다.
- [0049] 따라서, 본 발명의 일 실시예에 따르면, 기준데이터셋을 이용한 사전 학습을 통해 인공지능모델을 사전 학습시킨 후에, 다운스트림 데이터셋이 입력될 경우 다운스트림 데이터셋에 대응하는 추가텍스트와 추가이미지를 포함하는 데이터쌍을 생성하고, 이 데이터쌍을 다운스트림 데이터셋에 추가하여 보정데이터셋을 입력시키며, 보정데이터셋을 사전 학습된 인공지능모델에 입력시켜 전이학습을 수행함으로써, 적은 크기의 데이터셋의 입력으로도 이미지 캡서닝 성능을 향상시킬 수 있어 인공지능모델의 강건성을 효과적으로 확보할 수 있다.
- [0050] 도 9 및 도 10은 본 발명의 다른 실시예에 따라 이미지 캡서닝을 수행하는 과정을 나타낸 플로우차트이다.
- [0051] 도 9를 참조하면, 사전학습부(110)에서 기준데이터셋을 이용한 사전 학습을 통해 인공지능모델을 사전 학습시킬 수 있다(단계210).
- [0052] 여기에서, 기준데이터셋은 데이터베이스에 미리 수집된 대량의 데이터셋, 네트워크를 통해 크롤링 등의 방식으로 수집되는 대량의 데이터셋 등을 포함할 수 있으며, 이러한 데이터셋은 이미지(Image)와 캡션(Caption)의 데이터쌍으로 입력될 수 있다.
- [0053] 여기에서, 인공지능모델은 예를 들면, 합성곱신경망(CNN), 순환신경망(RNN), NIC모델, BUTD모델, BERT모델, CNN 및 RNN을 결합한 하이브리드모델 등을 포함할 수 있다.
- [0054] 그리고, 데이터쌍생성부(120)에서 다운스트림 데이터셋이 입력될 경우 다운스트림 데이터셋에 대응하는 추가텍스트와 추가이미지를 포함하는 데이터쌍을 생성할 수 있다(단계220).
- [0055] 상기 데이터쌍을 생성하는 단계(220)는, 데이터쌍생성부(120)에 구비되는 텍스트생성블록(121)에서 다운스트림 데이터셋의 각 문장(sentence)에 대응하는 형태(또는 표현, 단어, 스타일)를 갖는 추가텍스트를 생성하는 단계(221)와, 데이터쌍생성부(120)에 구비되는 이미지생성블록(122)에서 추가텍스트에 대응하여 추가이미지를 복수개 생성하는 단계(222)를 포함할 수 있다.
- [0056] 여기에서, 상기 추가텍스트를 생성하는 단계(221)는, 텍스트생성블록(221)에서 문장에서 명사 및 형용사를 추출하고, 추출된 명사 및 형용사에서 객체(object) 및 상태(state)를 선택하는 방식으로 샘플링한 후에, 객체 및 상태를 이용하여 상기 다운스트림 데이터셋의 각 문장에 대응하는 형태(또는 표현, 단어, 스타일)를 갖는 신규텍스트를 생성하고, 신규텍스트의 스타일전이를 수행하는 방식으로 추가텍스트를 생성할 수 있다.
- [0057] 또한, 상기 추가텍스트를 생성하는 단계(221)는, 텍스트생성블록(221)에서 다운스트림 데이터셋으로부터 랜덤문장(random sentence)을 복수개 추출하고, 랜덤문장과 신규텍스트에 대해 이미지뷰(image view)와 관련된 단어구를 추가하는 방식으로 스타일전이를 수행할 수 있다.
- [0058] 예를 들면, 추가텍스트를 생성하는 단계(221)에서는 텍스트생성블록(121)에서 텍스트제너레이터를 포함하여 다운스트림 데이터셋에 존재하는 문장과 유사한 문장을 캡션항목에 대응하여 생성할 수 있는데, NER추출기를 통해 다운스트림 데이터셋의 캡션항목에서 랜덤으로 복수개의 문장을 추출하여 샘플링하고, 샘플링된 복수의 문장에서 각각의 객체 및 상태를 추출하는 방식으로 객체에 대응하는 명사와 상태에 대응하는 형용사를 NER셋으로 하여 추출할 수 있다.
- [0059] 또한, 추가텍스트를 생성하는 단계(221)에서는 텍스트생성블록(121)에서 NER샘플러를 이용하여 NER추출기를 통해 추출된 NER셋에 존재하는 객체와 상태를 랜덤하게 샘플링하여 복수개의 객체와 복수개의 상태를 생성하는 방식으로 추가텍스트에 추가할 객체 및 상태를 선택할 수 있다. 여기에서, 객체는 예를 들면, 3-4개를 생성할 수 있고, 상태는 예를 들면, 1-2개를 생성할 수 있다.
- [0060] 그리고, 추가텍스트를 생성하는 단계(221)에서는 텍스트생성블록(121)에서 설명기반 텍스트제너레이터를 통해 선택된 객체 및 상태를 포함하는 새로운 신규텍스트를 생성할 수 있는데, 설명과 입력을 통해 어떤 단어를 추가할 것인지에 대한 정보를 제너레이터에 전달할 수 있고, 입력은 선택된 객체 및 상태를 사용할 수 있다.

- [0061] 다음에, 추가텍스트를 생성하는 단계(221)에서는 텍스트제너레이터를 통해 획득된 문장(즉, 설명과 입력)은 다운스트림 데이터셋의 스타일과 차이가 존재할 수 있기 때문에, 텍스트생성블록(121)에서 설명기반 스타일트랜스퍼를 통해 문장의 스타일을 변환시킬 수 있다.
- [0062] 이러한 텍스트생성블록(121)에서는 다운스트림 데이터셋으로부터 복수개(N개)의 랜덤문장을 추출하여 추출된 문장을 설명으로 구성하고, 생성된 신규텍스트를 변환시킬 문장인 입력으로 구성할 수 있다.
- [0063] 또한, 추가텍스트를 생성하는 단계(221)에서는 텍스트생성블록(121)에서 랜덤문장과 신규텍스트를 이용하여 스타일이 변환된 문장에 대해 이미지뷰와 관련된 단어구를 추가할 수 있는데, 사진의 구도인 이미지뷰와 관련된 단어구를 뷰사전데이터베이스에서 추출하여 변환된 문장에 추가할 수 있다.
- [0064] 이러한 텍스트생성블록(121)에서는 뷰선택기를 통해 사진의 구도와 관련된 단어구들(예를 들면, A wide shot of ~, A straight view ~ 등)을 변환된 문장에 추가한 후 설명(instruction)에 입력시킬 수 있고, 이 후 설명기반 모델을 이용하여 가장 자연스러운 구도를 선택할 수 있다.
- [0065] 이러한 스타일변환 텍스트는 추가텍스트로 하여 이미지생성블록(122) 및 데이터추가부(130)로 제공될 수 있다.
- [0066] 한편, 추가이미지를 복수개 생성하는 단계(222)에서는 이미지생성블록(122)에서 이미지제너레이터를 통해 스타일변환 텍스트(즉, 추가텍스트)에 대응하는 이미지를 생성할 수 있는데, 다운스트림 데이터셋에 대응하여 유사하게 변환된 스타일변환 텍스트를 이용하여 신규이미지를 생성할 수 있다.
- [0067] 여기에서, 신규이미지는 하나의 스타일변환 텍스트에 대해 복수개(N개)의 이미지를 생성하여 스타일변환 텍스트에 대한 다양성을 보존할 수 있으며, 이미지제너레이터 모델은 디퓨전(diffusion)을 기반으로 하는 모델뿐만 아니라 종래에 제시된 다양한 모델을 사용할 수 있다.
- [0068] 상술한 바와 같이 생성된 추가이미지는 데이터추가부(130)로 제공될 수 있다.
- [0069] 다음에, 데이터추가부(130)에서 생성된 데이터쌍을 다운스트림 데이터셋에 추가하여 보정데이터셋을 입력시킬 수 있다(단계230).
- [0070] 예를 들면, 텍스트생성블록(121)을 통해 생성된 추가텍스트와 이미지생성블록(122)을 통해 생성된 추가이미지는 데이터쌍으로 하여 데이터쌍생성부(120)로부터 데이터추가부(130)에 제공될 수 있으며, 데이터추가부(130)에서는 이 데이터쌍에서 추가텍스트는 다운스트림 데이터셋의 캡션항목에 추가할 수 있고, 추가이미지는 다운스트림 데이터셋의 이미지항목에 추가하여 보정데이터셋을 입력시킬 수 있다.
- [0071] 이어서, 전이학습부(140)에서 보정데이터셋을 사전 학습된 인공지능모델에 입력시켜 전이학습을 수행할 수 있다(단계240).
- [0072] 이러한 전이학습부(140)에서는 데이터추가부(130)로부터 입력되는 보정데이터셋을 이용하여 사전학습부(110)를 통해 사전 학습된 인공지능모델에 해당 보정데이터셋을 입력하여 파인튜닝(fine-tuning)을 수행한 후 출력함으로써, 이미지캡처능을 수행할 수 있다.
- [0073] 따라서, 본 발명의 다른 실시예에 따르면, 기준데이터셋을 이용한 사전 학습을 통해 인공지능모델을 사전 학습시킨 후에, 다운스트림 데이터셋이 입력될 경우 다운스트림 데이터셋에 대응하는 추가텍스트와 추가이미지를 포함하는 데이터쌍을 생성하고, 이 데이터쌍을 다운스트림 데이터셋에 추가하여 보정데이터셋을 입력시키며, 보정데이터셋을 사전 학습된 인공지능모델에 입력시켜 전이학습을 수행함으로써, 적은 크기의 데이터셋의 입력으로도 이미지 캡처 성능을 향상시킬 수 있어 인공지능모델의 강건성을 효과적으로 확보할 수 있다.
- [0074] 이상의 설명에서는 본 발명의 다양한 실시예들을 제시하여 설명하였으나 본 발명이 반드시 이에 한정되는 것은 아니며, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자라면 본 발명의 기술적 사상을 벗어나지 않는 범위 내에서 여러 가지 치환, 변형 및 변경이 가능함을 쉽게 알 수 있을 것이다.

부호의 설명

- [0075] 110 : 사전학습부
120 : 데이터쌍생성부
121 : 텍스트생성블록

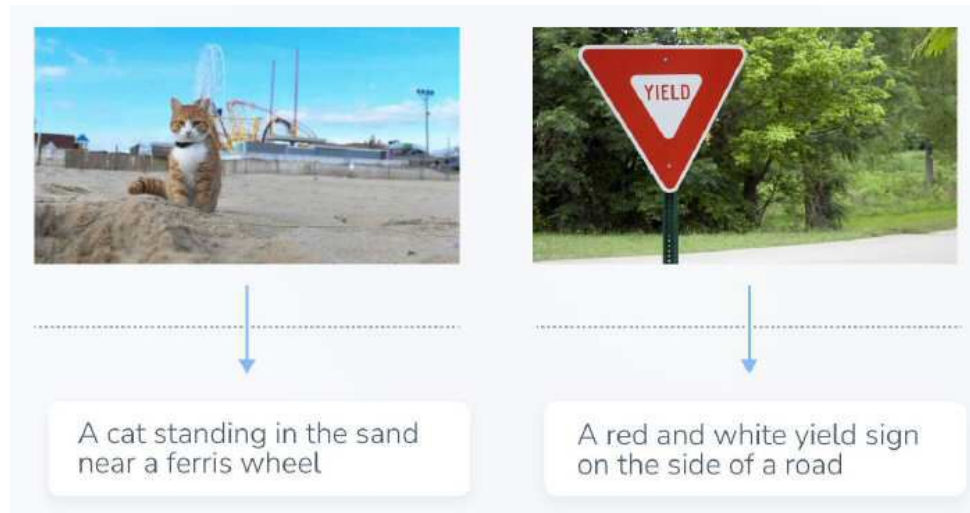
122 : 이미지생성블록

130 : 데이터추가부

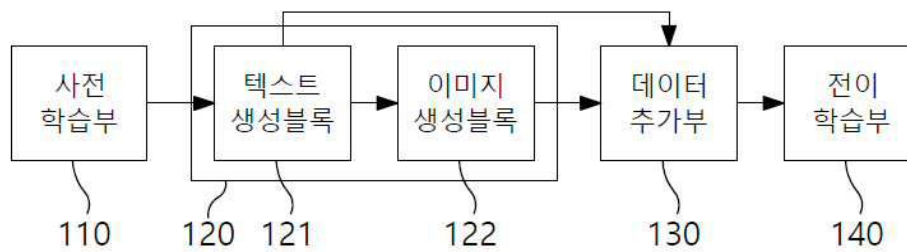
140 : 전이학습부

도면

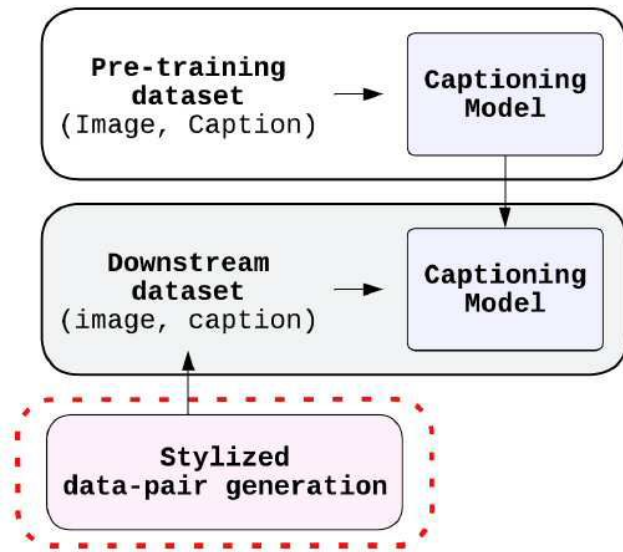
도면1



도면2

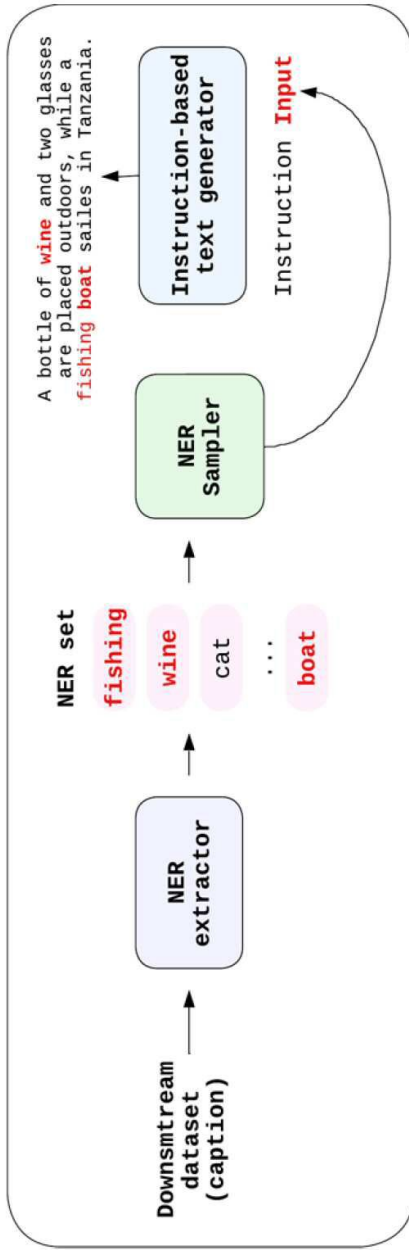


도면3

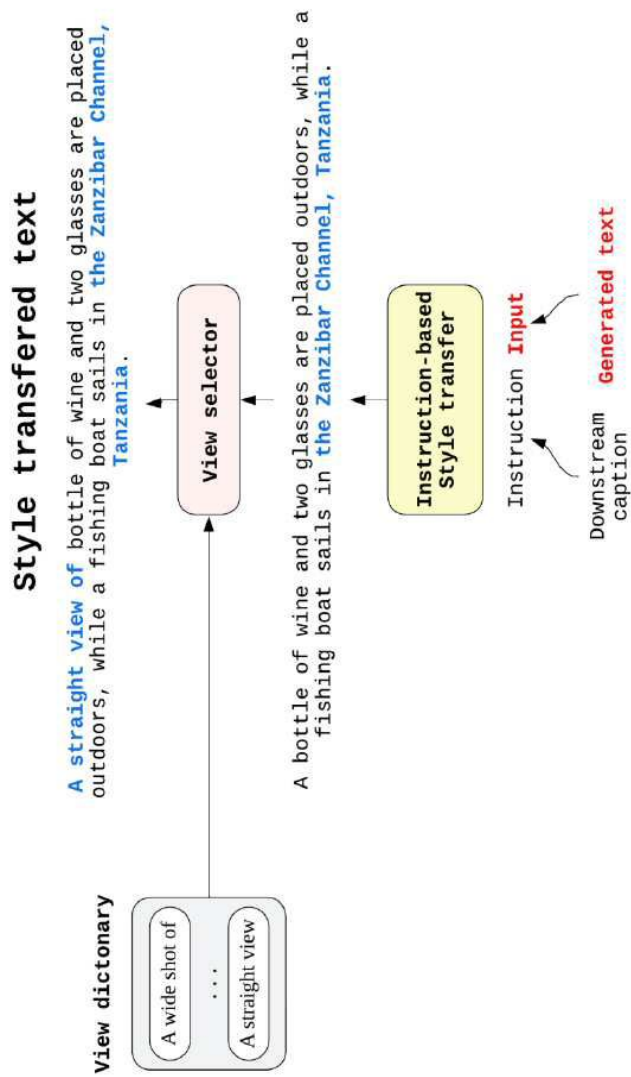


도면4

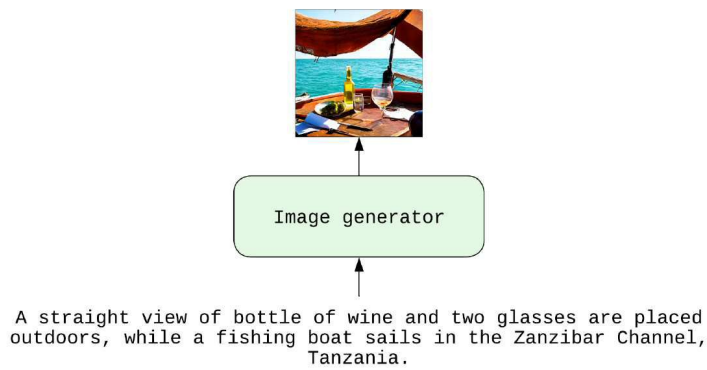
Text generator



도면5



도면6



도면7



Input text

A group of farmers can be seen discussing and examining the bright red tomatoes in the greenhouse.

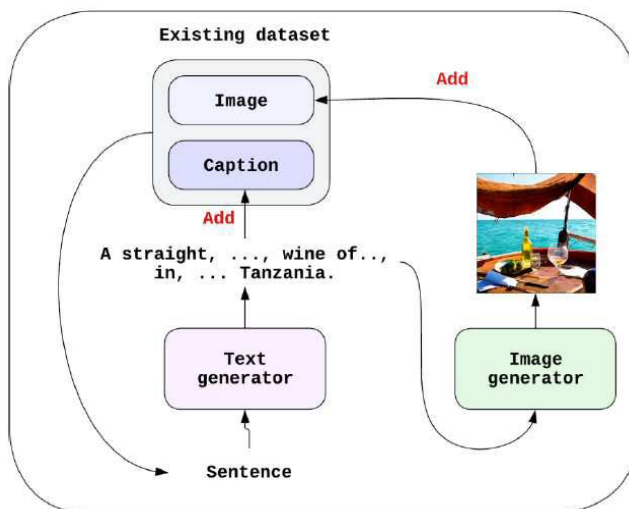


Input text

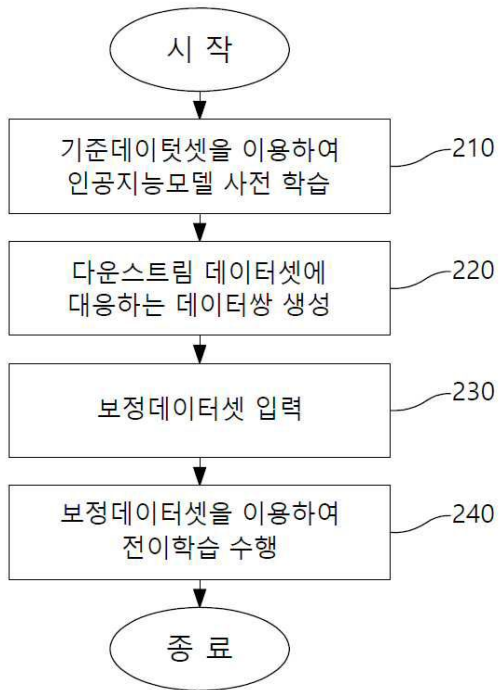
A young girl with freckles is seen in close up, smiling and waving at the camera on a sunny day.

도면8

Stylized data-pair generation



도면9



도면10

