(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0088406
(43) 공개일자 2025년06월17일

(51) 국제특허분류(Int. Cl.)
G06T 11/60 (2006.01) *G06F 40/263* (2020.01)
G06F 40/279 (2020.01) *G06N 20/00* (2019.01)
G06N 3/0475 (2023.01) *G06T 5/70* (2024.01)
G06T 7/50 (2017.01)
(52) CPC특허분류
G06T 11/60 (2013.01)
G06F 40/263 (2020.01)
(21) 출원번호 10-2024-0180997
(22) 출원일자 2024년12월06일
심사청구일자 2024년12월06일
(30) 우선권주장
1020230178117 2023년12월08일 대한민국(KR)

(71) 출원인
조선대학교산학협력단
광주광역시 동구 조선대길 126(서석동, 산학협력단)
(72) 발명자
김판구
광주광역시 북구 설죽로 600
유경호
광주광역시 서구 풍암2로 62 (풍암동)
김형주
광주광역시 동구 학소로 109, 108동 104호(소태동, 무등산골드클래스)
(74) 대리인
특허법인지원

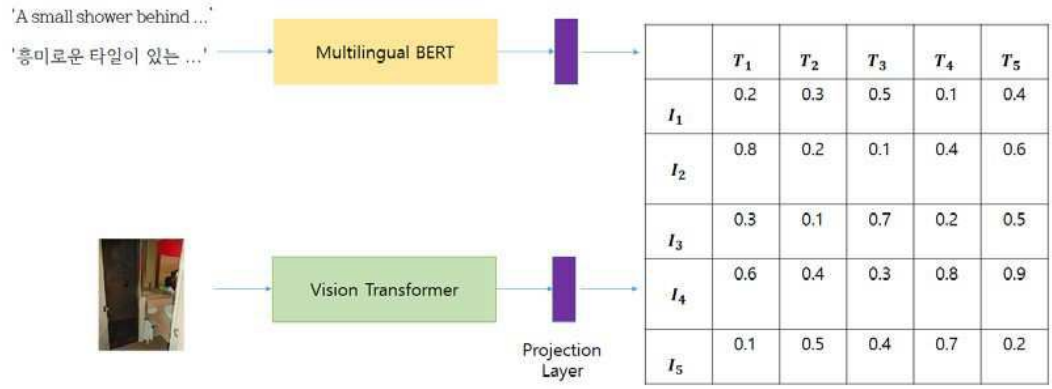
전체 청구항 수 : 총 14 항

(54) 발명의 명칭 **다중 모델을 기초로 웹툰 이미지 정보를 생성하기 위한 전자 장치 및 그 구동 방법**

(57) 요 약

다양한 실시예들에 따르면, 다중 모델을 기초로 웹툰 이미지 정보를 생성하기 위한 전자 장치에 있어서, 통신 인터페이스; 프로세서; 를 포함하고, 상기 프로세서는, 상기 통신 인터페이스를 통하여, 사용자의 외부 전자 장치로부터 적어도 하나의 텍스트 정보 및 적어도 하나의 이미지 정보를 수신하고, 상기 적어도 하나의 텍스트 정보를 제1 딥러닝 모델에 입력하여 적어도 하나의 키워드 추출 정보를 추출하고, 상기 적어도 하나의 이미지 정보를 제2 딥러닝 모델에 입력하여 적어도 하나의 변환 이미지 정보를 추출하고, 및 상기 적어도 하나의 키워드 추출 정보 및 상기 적어도 하나의 변환 이미지 정보를 멀티 모델에 입력하여 적어도 하나의 웹툰 이미지 정보를 출력하도록 설정되고, 상기 멀티 모델은, 복수의 키워드 추출 정보 및 복수의 변환 이미지 정보와, 복수의 웹툰 이미지 정보와, 복수의 키워드-이미지 변환 정보에 기반하여 학습된다.

대 표 도 - 도3



(52) CPC특허분류
G06F 40/279 (2020.01)
G06N 20/00 (2021.08)
G06N 3/0475 (2023.01)
G06T 5/70 (2024.01)
G06T 7/50 (2017.01)
G06T 2207/20081 (2013.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711177250
과제번호 2022-DD-RD-0065-02-101
부처명 과학기술정보통신부
과제관리(전문)기관명 연구개발특구진흥재단
연구사업명 기술사업화 협업 플랫폼 구축
연구과제명 인공지능 산업 육성 및 기술사업화를 위한 지능형 디지털 콘텐츠 제작 기술 개발 및 플랫폼 구축 사업
과제수행기관명 한국전자기술연구원
연구기간 2022.04.01 ~ 2024.12.31

명 세 서

청구범위

청구항 1

다중 모델을 기초로 웹툰 이미지 정보를 생성하기 위한 전자 장치에 있어서,

통신 인터페이스;

프로세서; 를 포함하고,

상기 프로세서는,

상기 통신 인터페이스를 통하여, 사용자의 외부 전자 장치로부터 적어도 하나의 텍스트 정보 및 적어도 하나의 이미지 정보를 수신하고,

상기 적어도 하나의 텍스트 정보를 제1 딥러닝 모델에 입력하여 적어도 하나의 키워드 추출 정보를 추출하고,

상기 적어도 하나의 이미지 정보를 제2 딥러닝 모델에 입력하여 적어도 하나의 변환 이미지 정보를 추출하고, 및

상기 적어도 하나의 키워드 추출 정보 및 상기 적어도 하나의 변환 이미지 정보를 멀티 모델에 입력하여 적어도 하나의 웹툰 이미지 정보를 출력하도록 설정되고,

상기 멀티 모델은,

복수의 키워드 추출 정보 및 복수의 변환 이미지 정보와, 복수의 웹툰 이미지 정보와, 복수의 키워드-이미지 변환 정보에 기반하여 학습되는,

전자 장치.

청구항 2

제1 항에 있어서,

상기 적어도 하나의 텍스트 정보는,

세계 전체에서 현재 사용되거나 과거에 사용되었던 모든 글자에 대한 적어도 하나의 글자 정보와, 세계 전체에서 현재 사용되거나 과거에 사용되었던 모든 언어에 대한 적어도 하나의 언어 정보를 포함하는,

전자 장치.

청구항 3

제2 항에 있어서,

상기 프로세서는,

상기 적어도 하나의 글자 정보를 기초로 적어도 하나의 중요 단어로 구성된 적어도 하나의 제1 키워드 정보를 추출하고,

상기 적어도 하나의 언어 정보를 기초로 적어도 하나의 중요 문장으로 구성된 적어도 하나의 제2 키워드 정보를 추출하고, 및

상기 적어도 하나의 제1 키워드 정보 및 상기 적어도 하나의 제2 키워드 정보를 상기 멀티 모델에 입력하여 상기 적어도 하나의 키워드 추출 정보를 출력하도록 설정되고,

상기 제1 딥러닝 모델은,

복수의 제1 키워드 정보 및 복수의 제2 키워드 정보, 복수의 텍스트 정보와, 복수의 키워드 추출 정보를 학습하고, 멀티모달 버트 모델(multilingual BERT model)로 구성되는,

전자 장치.

청구항 4

제1 항에 있어서,

상기 적어도 하나의 이미지 정보는,

상기 적어도 하나의 텍스트 정보에 입력된 텍스트 정보와 매칭되는 이미지 정보와, 다양한 표현을 나타내는 이미지 정보와, 각 언어별로 매칭된 이미지 정보를 포함하는,

전자 장치.

청구항 5

제4 항에 있어서,

상기 프로세서는,

상기 적어도 하나의 이미지 정보에서, 상기 언어별로 매칭된 이미지 정보를 상기 제2 딥러닝 모델에 입력하여 입력된 깊이 정보(depth image data)를 적어도 하나의 초기 이미지 정보로 출력하고, 및

상기 적어도 하나의 초기 이미지 정보를 상기 제2 딥러닝 모델에 입력하여 역방향 확산 (reverse diffusion)을 수행하여 노이즈를 제거한 적어도 하나의 초기 웹툰 이미지 정보로 구성된 상기 적어도 하나의 변환 이미지 정보를 출력하도록 설정되고,

상기 제2 딥러닝 모델은,

복수의 초기 이미지 정보, 복수의 초기 웹툰 이미지 정보와, 복수의 변환 이미지 정보에 기반하여 학습되고, 이미지-깊이 모델(depth-to-image model)로 구성되는,

전자 장치.

청구항 6

제1 항에 있어서,

상기 프로세서는,

상기 적어도 하나의 키워드 추출 정보를 키워드 벡터 정보로 변환하고, 상기 적어도 하나의 변환 이미지 정보를 이미지 벡터 정보로 변환하고,

상기 이미지 벡터 정보를 상기 멀티 모델에 입력하여 상기 키워드 벡터 정보와 동일한 크기, 동일한 거리, 및 동일한 차원으로 축소시키고,

상기 이미지 벡터 정보와 상기 키워드 벡터 정보 간의 내적을 하기의 [수학식 1]에 기초하여 산출한 유사도를 정량화하도록 설정된,

전자 장치.

[수학식 1]

$$similarity = \cos(\theta) = \frac{A \cdot B}{|A|\,|B|}$$

여기서, $\cos(\theta)$는 상기 유사도이고, A는 상기 이미지 벡터 정보이고, B는 상기 키워드 벡터 정보임.

청구항 7

제6 항에 있어서,

상기 프로세서는,

상기 유사도에 근거하여, 상기 이미지 벡터 정보와 상기 키워드 벡터 정보가 기 설정된 유사도를 초과하면, 상기 이미지 벡터 정보와 상기 키워드 벡터 정보를 근접시키고, 및

상기 유사도에 근거하여, 상기 이미지 벡터 정보와 상기 키워드 벡터 정보가 기 설정된 유사도를 초과하지 않으면, 상기 이미지 벡터 정보와 상기 키워드 벡터 정보의 거리를 이격시키도록 설정되고,

상기 멀티 모델은,

복수의 이미지 벡터 정보 및 복수의 키워드 벡터 정보와, 복수의 유사도 값과, 복수의 분류 정보 그룹을 추가로 학습하도록 설정된,

전자 장치.

청구항 8

다중 모델을 기초로 웹툰 이미지 정보를 생성하기 위한 전자 장치의 구동 방법에 있어서,

상기 방법은,

통신 인터페이스를 통하여, 사용자의 외부 방법으로부터 적어도 하나의 텍스트 정보 및 적어도 하나의 이미지 정보를 수신하고,

프로세서를 통하여, 상기 적어도 하나의 텍스트 정보를 제1 딥러닝 모델에 입력하여 적어도 하나의 키워드 추출 정보를 추출하고,

상기 프로세서를 통하여, 상기 적어도 하나의 이미지 정보를 제2 딥러닝 모델에 입력하여 적어도 하나의 변환 이미지 정보를 추출하고, 및

상기 프로세서를 통하여, 상기 적어도 하나의 키워드 추출 정보 및 상기 적어도 하나의 변환 이미지 정보를 멀티 모델에 입력하여 적어도 하나의 웹툰 이미지 정보를 출력하도록 설정되고,

상기 멀티 모델은,

복수의 키워드 추출 정보 및 복수의 변환 이미지 정보와, 복수의 웹툰 이미지 정보와, 복수의 키워드-이미지 변환 정보에 기반하여 학습되는,

방법.

청구항 9

제8 항에 있어서,

상기 적어도 하나의 텍스트 정보는,

세계 전체에서 현재 사용되거나 과거에 사용되었던 모든 글자에 대한 적어도 하나의 글자 정보와, 세계 전체에서 현재 사용되거나 과거에 사용되었던 모든 언어에 대한 적어도 하나의 언어 정보를 포함하는,

방법.

청구항 10

제9 항에 있어서,

상기 방법은,

상기 적어도 하나의 글자 정보를 기초로 적어도 하나의 중요 단어로 구성된 적어도 하나의 제1 키워드 정보를 추출하고,

상기 적어도 하나의 언어 정보를 기초로 적어도 하나의 중요 문장으로 구성된 적어도 하나의 제2 키워드 정보를 추출하고, 및

상기 적어도 하나의 제1 키워드 정보 및 상기 적어도 하나의 제2 키워드 정보를 상기 멀티 모델에 입력하여 상기 적어도 하나의 키워드 추출 정보를 출력하도록 설정되고,

상기 제1 딥러닝 모델은,

복수의 제1 키워드 정보 및 복수의 제2 키워드 정보, 복수의 텍스트 정보와, 복수의 키워드 추출 정보를 학습하고, 멀티모달 버트 모델(multilingual BERT model)로 구성되는,

방법.

청구항 11

제8 항에 있어서,

상기 적어도 하나의 이미지 정보는,

상기 적어도 하나의 텍스트 정보에 입력된 텍스트 정보와 매칭되는 이미지 정보와, 다양한 표현을 나타내는 이미지 정보와, 각 언어별로 매칭된 이미지 정보를 포함하는,

방법.

청구항 12

제11 항에 있어서,

상기 방법은,

상기 적어도 하나의 이미지 정보에서, 상기 언어별로 매칭된 이미지 정보를 상기 제2 딥러닝 모델에 입력하여 입력된 깊이 정보(depth image data)를 적어도 하나의 초기 이미지 정보로 출력하고, 및

상기 적어도 하나의 초기 이미지 정보를 상기 제2 딥러닝 모델에 입력하여 역방향 확산 (reverse diffusion)을 수행하여 노이즈를 제거한 적어도 하나의 초기 웹툰 이미지 정보로 구성된 상기 적어도 하나의 변환 이미지 정보를 출력하도록 설정되고,

상기 제2 딥러닝 모델은,

복수의 초기 이미지 정보, 복수의 초기 웹툰 이미지 정보와, 복수의 변환 이미지 정보에 기반하여 학습되고, 이미지-깊이 모델(depth-to-image model)로 구성되는,

방법.

청구항 13

제8 항에 있어서,

상기 방법은,

상기 적어도 하나의 키워드 추출 정보를 키워드 벡터 정보로 변환하고, 상기 적어도 하나의 변환 이미지 정보를

이미지 벡터 정보로 변환하고,

상기 이미지 벡터 정보를 상기 멀티 모델에 입력하여 상기 키워드 벡터 정보와 동일한 크기, 동일한 거리, 및 동일한 차원으로 축소시키고,

상기 이미지 벡터 정보와 상기 키워드 벡터 정보 간의 내적을 하기의 [수학식 1]에 기초하여 산출한 유사도를 정량화하도록 설정된,

방법.

[수학식 1]

$$similarity = \cos(\theta) = \frac{A \cdot B}{|A|\,|B|}$$

여기서, $\cos(\theta)$는 상기 유사도이고, A는 상기 이미지 벡터 정보이고, B는 상기 키워드 벡터 정보임.

청구항 14

제13 항에 있어서,

상기 방법은,

상기 유사도에 근거하여, 상기 이미지 벡터 정보와 상기 키워드 벡터 정보가 기 설정된 유사도를 초과하면, 상기 이미지 벡터 정보와 상기 키워드 벡터 정보를 근접시키고, 및

상기 유사도에 근거하여, 상기 이미지 벡터 정보와 상기 키워드 벡터 정보가 기 설정된 유사도를 초과하지 않으면, 상기 이미지 벡터 정보와 상기 키워드 벡터 정보의 거리를 이격시키도록 설정되고,

상기 멀티 모델은,

복수의 이미지 벡터 정보 및 복수의 키워드 벡터 정보와, 복수의 유사도 값과, 복수의 분류 정보 그룹을 추가로 학습하도록 설정된,

방법.

발명의 설명

기 술 분 야

[0001] 본 발명의 다양한 실시예는, 다중 모델을 기초로 웹툰 이미지 정보를 생성하기 위한 전자 장치 및 그 구동 방법에 관한 것이다.

배 경 기 술

[0003] 최근, 웹에서 만화에 관한 공급 및 수요가 많아지고 있어, 이에 따른 다양한 웹툰이 다양한 웹사이트에서 제공되고 있다.

[0004] 이러한, 기술에 대해 웹툰의 경우 실시간에 준하는 시간으로 웹툰을 생성하여 공급자들에게 제공해야 되는데, 이러한 웹툰을 빠르게 그릴 수 있는 방법이 추출되고 있다.

[0005] 그러나, 종래는, 영어에 편중되어 있어, 한국어와 같은 언어를 포함한 다국어 연구는 아직 미흡하고, 각 언어의 고유한 문법, 표현 방식, 어순 등을 반영하기 어려운 한계점이 있다. 또한, 종래의 웹툰 제작은 스토리 기획부터 배경 작업까지 모두 수작업으로 진행되며, 높은 비용과 시간적 자원이 필요이 필요한데, 컨텐츠 수요가 증가하는 상황에서 이러한 비효율성이 제작자의 부담을 가중시키는 문제점이 있었다.

[0006] 따라서, 이러한 상황을 극복하기 위해서, 제작자가 작성한 텍스트 및 이미지 만으로도 손쉽게 초안으로 웹툰 이

미지 정보를 생성할 수 있는 기술이 필요한 실정이다.

선행기술문헌

특허문헌

[0008] (특허문헌 0001) 한국등록특허공보 제10-2693273호 (2024.08.08)

(특허문헌 0002) 한국등록특허공보 제10-2690474호 (2024.07.26)

발명의 내용

해결하려는 과제

[0009] 따라서, 본 실시예는, 제작자의 전자 장치로부터 획득된 텍스트 정보 및 이미지 정보를 제1 딥러닝 모델, 제2 딥러닝 모델, 및 멀티 모델에 입력하여 출력된 웹툰 이미지 정보를 기초로 손쉽게 초안으로 구성된 웹툰 이미지 정보를 제작자의 전자 장치에 제공하기 위한 전자 장치 및 그 구동 방법에 관한 것이다.

과제의 해결 수단

[0011] 다양한 실시예들에 따르면, 다중 모델을 기초로 웹툰 이미지 정보를 생성하기 위한 전자 장치에 있어서, 통신 인터페이스; 프로세서; 를 포함하고, 상기 프로세서는, 상기 통신 인터페이스를 통하여, 사용자의 외부 전자 장치로부터 적어도 하나의 텍스트 정보 및 적어도 하나의 이미지 정보를 수신하고, 상기 적어도 하나의 텍스트 정보를 제1 딥러닝 모델에 입력하여 적어도 하나의 키워드 추출 정보를 추출하고, 상기 적어도 하나의 이미지 정보를 제2 딥러닝 모델에 입력하여 적어도 하나의 변환 이미지 정보를 추출하고, 및 상기 적어도 하나의 키워드 추출 정보 및 상기 적어도 하나의 변환 이미지 정보를 멀티 모델에 입력하여 적어도 하나의 웹툰 이미지 정보를 출력하도록 설정되고, 상기 멀티 모델은, 복수의 키워드 추출 정보 및 복수의 변환 이미지 정보와, 복수의 웹툰 이미지 정보와, 복수의 키워드-이미지 변환 정보에 기반하여 학습된다.

[0012] 또 다른 다양한 실시예들에 따르면, 다중 모델을 기초로 웹툰 이미지 정보를 생성하기 위한 전자 장치의 구동 방법에 있어서, 상기 방법은, 통신 인터페이스를 통하여, 사용자의 외부 방법으로부터 적어도 하나의 텍스트 정보 및 적어도 하나의 이미지 정보를 수신하고, 프로세서를 통하여, 상기 적어도 하나의 텍스트 정보를 제1 딥러닝 모델에 입력하여 적어도 하나의 키워드 추출 정보를 추출하고, 상기 프로세서를 통하여, 상기 적어도 하나의 이미지 정보를 제2 딥러닝 모델에 입력하여 적어도 하나의 변환 이미지 정보를 추출하고, 및 상기 프로세서를 통하여, 상기 적어도 하나의 키워드 추출 정보 및 상기 적어도 하나의 변환 이미지 정보를 멀티 모델에 입력하여 적어도 하나의 웹툰 이미지 정보를 출력하도록 설정되고, 상기 멀티 모델은, 복수의 키워드 추출 정보 및 복수의 변환 이미지 정보와, 복수의 웹툰 이미지 정보와, 복수의 키워드-이미지 변환 정보에 기반하여 학습된다.

발명의 효과

[0014] 본 실시예에 따른 전자 장치는, 사용자의 외부 전자 장치로부터 획득된 적어도 하나의 텍스트 정보 및 적어도 하나의 이미지 정보를 제1 딥러닝 모델, 제2 딥러닝 모델, 및 멀티 모델에 입력하여 적어도 하나의 웹툰 이미지 정보를 손쉽게 출력할 수 있고, 적어도 하나의 웹툰 이미지 정보를 사용자의 외부 전자 장치에 송신함으로써, 사용자가 텍스트 정보 및 이미지 정보만으로도 손쉽게 웹툰 이미지 정보를 획득하고 이를 활용할 수 있는 이점이 있다.

도면의 간단한 설명

[0016] 도 1은 본 발명의 다양한 실시예에 따른 전자 장치 및 네트워크의 블록도를 도시한다.

도 2는 다양한 실시예들에 따른, 전자 장치가 동작하는 방법을 설명하기 위한 흐름도이다.

도 3은 다양한 실시예들에 따른, 전자 장치에 사용되는 멀티 모델에 대한 구체적인 예시도이다.

도 4는 다양한 실시예들에 따른, 전자 장치에 사용되는 멀티 모델이 이미지 정보와 텍스트 정보의 유사도를 산출하는 구체적인 예시도이다.

도 5는 다양한 실시예들에 따른, 전자 장치에서 사용되는 멀티 모델이 입력된 텍스트 정보 및 이미지 정보를 기초로 생성한 다양한 이미지 정보를 출력하는 구체적인 예시도이다.

도 6 및 도 7은 다양한 실시예들에 따른, 전자 장치에서 사용되는 멀티 모델이 입력된 텍스트 정보를 기초로 실제 유사도 이미지 정보와 그린 이미지 정보를 출력한 구체적인 예시도이다.

발명을 실시하기 위한 구체적인 내용

[0017] 이하, 본 문서의 다양한 실시예들이 첨부된 도면을 참조하여 기재된다. 실시예 및 이에 사용된 용어들은 본 문서에 기재된 기술을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 해당 실시예의 다양한 변경, 균등물, 및/또는 대체물을 포함하는 것으로 이해되어야 한다. 도면의 설명과 관련하여, 유사한 구성요소에 대해서는 유사한 참조 부호가 사용될 수 있다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함할 수 있다. 본 문서에서, "A 또는 B" 또는 "A 및/또는 B 중 적어도 하나" 등의 표현은 함께 나열된 항목들의 모든 가능한 조합을 포함할 수 있다. "제 1," "제 2," "첫째," 또는 "둘째,"등의 표현들은 해당 구성요소들을, 순서 또는 중요도에 상관없이 수식할 수 있고, 한 구성요소를 다른 구성요소와 구분하기 위해 사용될 뿐 해당 구성요소들을 한정하지 않는다. 어떤(예: 제 1) 구성요소가 다른(예: 제 2) 구성요소에 "(기능적으로 또는 통신적으로) 연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 상기 어떤 구성요소가 상기 다른 구성요소에 직접적으로 연결되거나, 다른 구성요소(예: 제 3 구성요소)를 통하여 연결될 수 있다.

[0018] 본 문서에서, "~하도록 구성된(또는 설정된)(configured to)"은 상황에 따라, 예를 들면, 하드웨어적 또는 소프트웨어적으로 "~에 적합한," "~하는 능력을 가지는," "~하도록 변경된," "~하도록 만들어진," "~를 할 수 있는," 또는 "~하도록 설계된"과 상호 호환적으로(interchangeably) 사용될 수 있다. 어떤 상황에서는, "~하도록 구성된 장치"라는 표현은, 그 장치가 다른 장치 또는 부품들과 함께 "~할 수 있는" 것을 의미할 수 있다. 예를 들면, 문구 "A, B, 및 C를 수행하도록 구성된(또는 설정된) 프로세서"는 해당 동작을 수행하기 위한 전용 프로세서(예: 임베디드 프로세서), 또는 메모리 장치에 저장된 하나 이상의 소프트웨어 프로그램들을 실행함으로써, 해당 동작들을 수행할 수 있는 범용 프로세서(예: CPU 또는 application processor)를 의미할 수 있다.

[0019] 본 문서의 다양한 실시예들에 따른 전자 장치는, 예를 들면, 스마트폰, 태블릿 PC, 데스크탑 PC, 랩탑 PC, 넷북 컴퓨터, 워크스테이션, 서버 중 적어도 하나를 포함할 수 있다.

[0020] 도 1을 참조하여, 다양한 실시예에서의, 네트워크 환경(100) 내의 전자 장치(101)가 기재된다. 전자 장치(101)는 버스(110), 프로세서(120), 메모리(130), 입출력 인터페이스(140), 디스플레이(150), 및 통신 인터페이스(160)를 포함할 수 있다. 어떤 실시예에서는, 전자 장치(101)는, 구성요소들 중 적어도 하나를 생략하거나 다른 구성요소를 추가적으로 구비할 수 있다. 버스(110)는 구성요소들(110-170)을 서로 연결하고, 구성요소들 간의 통신(예: 제어 메시지 또는 데이터)을 전달하는 회로를 포함할 수 있다. 프로세서(120)는, 중앙처리장치, 어플리케이션 프로세서, 또는 커뮤니케이션 프로세서(communication processor(CP)) 중 하나 또는 그 이상을 포함할 수 있다. 프로세서(120)는, 예를 들면, 전자 장치(101)의 적어도 하나의 다른 구성요소들의 제어 및/또는 통신에 관한 연산이나 데이터 처리를 실행할 수 있다.

[0021] 메모리(130)는, 휘발성 및/또는 비휘발성 메모리를 포함할 수 있다. 메모리(130)는, 예를 들면, 전자 장치(101)의 적어도 하나의 다른 구성요소에 관계된 명령 또는 데이터를 저장할 수 있다. 한 실시예에 따르면, 메모리(130)는 소프트웨어 및/또는 프로그램(140)을 저장할 수 있다.

[0022] 입출력 인터페이스(140)는, 예를 들면, 사용자 또는 다른 외부 기기로부터 입력된 명령 또는 데이터를 전자 장치(101)의 다른 구성요소(들)에 전달하거나, 또는 전자 장치(101)의 다른 구성요소(들)로부터 수신된 명령 또는 데이터를 사용자 또는 다른 외부 기기로 출력할 수 있다.

[0023] 디스플레이(150)는, 예를 들면, 액정 디스플레이(LCD), 발광 다이오드(LED) 디스플레이, 유기 발광 다이오드(OLED) 디스플레이, 또는 마이크로 전자기계 시스템 (MEMS) 디스플레이, 또는 전자종이(electronic paper) 디스플레이를 포함할 수 있다. 디스플레이(150)는, 예를 들면, 사용자에게 각종 콘텐츠(예: 텍스트, 이미지,

비디오, 아이콘, 및/또는 심볼 등)을 표시할 수 있다. 디스플레이(150)는, 터치 스크린을 포함할 수 있으며, 예를 들면, 전자 펜 또는 사용자의 신체의 일부를 이용한 터치, 제스쳐, 근접, 또는 호버링 입력을 수신할 수 있다. 통신 인터페이스(160)는, 예를 들면, 전자 장치(101)와 외부 장치(예: 제 1 외부 전자 장치(102), 제 2 외부 전자 장치(104), 또는 서버(108)) 간의 통신을 설정할 수 있다. 예를 들면, 통신 인터페이스(160)는 무선 통신 또는 유선 통신을 통해서 네트워크(162)에 연결되어 외부 장치(예: 제 2 외부 전자 장치(104) 또는 서버(108))와 통신할 수 있다.

[0024] 무선 통신은, 예를 들면, LTE, LTE-A(LTE Advance), CDMA(code division multiple access), WCDMA(wideband CDMA), UMTS(universal mobile telecommunications system), WiBro(Wireless Broadband), 또는 GSM(Global System for Mobile Communications) 등 중 적어도 하나를 사용하는 셀룰러 통신을 포함할 수 있다. 한 실시예에 따르면, 무선 통신은, 예를 들면, WiFi(wireless fidelity), 블루투스, 블루투스 저전력(BLE), 지그비(Zigbee), NFC(near field communication), 자력 시큐어 트랜스미션(Magnetic Secure Transmission), 라디오 프리퀀시(RF), 또는 보디 에어리어 네트워크(BAN) 중 적어도 하나를 포함할 수 있다. 한실시예에 따르면, 무선 통신은 GNSS를 포함할 수 있다. GNSS는, 예를 들면, GPS(Global Positioning System), Glonass(Global Navigation Satellite System), Beidou Navigation Satellite System(이하 "Beidou") 또는 Galileo, the European global satellite-based navigation system일 수 있다. 이하, 본 문서에서는, "GPS"는 "GNSS"와 상호 호환적으로 사용될 수 있다. 유선 통신은, 예를 들면, USB(universal serial bus), HDMI(high definition multimedia interface), RS-232(recommended standard232), 전력선 통신, 또는 POTS(plain old telephone service) 등 중 적어도 하나를 포함할 수 있다. 네트워크(162)는 텔레커뮤니케이션 네트워크, 예를 들면, 컴퓨터 네트워크(예: LAN 또는 WAN), 인터넷, 또는 텔레폰 네트워크 중 적어도 하나를 포함할 수 있다.

[0025] 제 1 및 제 2 외부 전자 장치(102, 104, 106) 각각은 전자 장치(101)와 동일한 또는 다른 종류의 장치일 수 있다. 다양한 실시예에 따르면, 전자 장치(101)에서 실행되는 동작들의 전부 또는 일부는 다른 하나 또는 복수의 전자 장치(예: 전자 장치(102, 104, 106), 또는 서버(108)에서 실행될 수 있다. 한 실시예에 따르면, 전자 장치(101)가 어떤 기능이나 서비스를 자동으로 또는 요청에 의하여 수행해야 할 경우에, 전자 장치(101)는 기능 또는 서비스를 자체적으로 실행시키는 대신에 또는 추가적으로, 그와 연관된 적어도 일부 기능을 다른 장치(예: 전자 장치(102, 104, 106), 또는 서버(108))에게 요청할 수 있다. 다른 전자 장치(예: 전자 장치(102, 104, 106), 또는 서버(108))는 요청된 기능 또는 추가 기능을 실행하고, 그 결과를 전자 장치(101)로 전달할 수 있다. 전자 장치(101)는 수신된 결과를 그대로 또는 추가적으로 처리하여 요청된 기능이나 서비스를 제공할 수 있다. 이를 위하여, 예를 들면, 클라우드 컴퓨팅, 분산 컴퓨팅, 또는 클라이언트-서버 컴퓨팅 기술이 이용될 수 있다.

[0027] 도 2는 다양한 실시예들에 따른, 전자 장치가 동작하는 방법을 설명하기 위한 흐름도이다.

[0028] 도 3은 다양한 실시예들에 따른, 전자 장치에 사용되는 멀티 모델에 대한 구체적인 예시도이다.

[0029] 도 4는 다양한 실시예들에 따른, 전자 장치에 사용되는 멀티 모델이 이미지 정보와 텍스트 정보의 유사도를 산출하는 구체적인 예시도이다.

[0030] 도 5는 다양한 실시예들에 따른, 전자 장치에서 사용되는 멀티 모델이 입력된 텍스트 정보 및 이미지 정보를 기초로 생성한 다양한 이미지 정보를 출력하는 구체적인 예시도이다.

[0031] 도 6 및 도 7은 다양한 실시예들에 따른, 전자 장치에서 사용되는 멀티 모델이 입력된 텍스트 정보를 기초로 실제 유사도 이미지 정보와 그린 이미지 정보를 출력한 구체적인 예시도이다.

[0033] 201 동작에서, 전자 장치(101)(예: 도 1의 프로세서(120))는, 통신 인터페이스(예: 도 1의 통신 인터페이스(160))를 통하여, 사용자의 외부 전자 장치(102, 104, 106)로부터 적어도 하나의 텍스트 정보 및 적어도 하나의 이미지 정보를 수신할 수 있다. 일 실시예에 따르면, 적어도 하나의 텍스트 정보는, 세계 전체에서 현재 사용되거나 과거에 사용되었던 모든 글자에 대한 적어도 하나의 글자 정보와, 세계 전체에서 현재 사용되거나 과거에 사용되었던 모든 언어에 대한 적어도 하나의 언어 정보를 포함할 수 있다. 구체적으로, 적어도 하나의 글자 정보는, 한글, 알파벳, 한자구조와 같은 글자 구조를 포함할 수 있다. 또한, 적어도 하나의 언어 정보는, 한국어에 대한 단어 및 문장, 구절과, 일본어, 영어, 프랑스어 등과 같은 언어계열에서 현재 사용되는 모든 언어 및/또는 라틴어, 산스크리트어와 같은 과거의 언어에 대한 것까지 포함할 수 있다. 이후, 전자 장치(101)는, 적어

도 하나의 글자 정보 및 적어도 하나의 언어 정보를 제1 딥러닝 모델에 입력할 수 있다.

[0034] 또 다른 일 실시예에 따르면, 적어도 하나의 이미지 정보는, 적어도 하나의 텍스트 정보에 입력된 텍스트 정보와 매칭되는 이미지 정보와, 다양한 표현을 나타내는 이미지 정보와, 각 언어별로 매칭된 이미지 정보를 포함할 수 있다. 구체적으로, 적어도 하나의 이미지 정보는, 적어도 하나의 글자 정보와 매칭되는 이미지 정보로 구성되도록 사전에 사용자가 미리 학습하여 설정할 수 있거나, 추후 학습된 모델에서 출력할 때 매칭 될 수 있는 기본 이미지 정보로 구성될 수 있다. 또한, 각 언어별로 매칭된 이미지 정보는, 도 5에 도시된 바와 같이, 영어 및/또는 한국어와 매칭되는 이미지 정보로 구성될 수 있다. 이후, 전자 장치(101)는, 적어도 하나의 이미지 정보를 제2 딥러닝 모델에 입력할 수 있다.

[0035] 203 동작에서, 전자 장치(101)(예: 도 1의 프로세서(120))는, 적어도 하나의 텍스트 정보를 제1 딥러닝 모델에 입력하여 적어도 하나의 키워드 추출 정보를 추출할 수 있다.

[0036] 일 실시예에 따르면, 전자 장치(101)는, 적어도 하나의 글자 정보를 기초로 적어도 하나의 중요 단어로 구성된 적어도 하나의 제1 키워드 정보를 추출할 수 있다. 구체적으로, 전자 장치(101)는, 다중언어 모델로 구성된 멀티 버트 모델(multilingual BERT model)인 제1 딥러닝 모델에 적어도 하나의 글자 정보를 입력할 수 있다. 이는 제1 딥러닝 모델이 적어도 하나의 텍스트 정보에서 언어적 특징을 효과적으로 반영할 수 있다. 이후, 전자 장치(101)는, 제1 딥러닝 모델에서 출력된 적어도 하나의 키워드 정보를 획득할 수 있고, 메모리(예: 도 1의 메모리(130))에 저장시킬 수 있다.

[0037] 또 다른 일 실시예에 따르면, 전자 장치(101)는, 다수의 문장으로 구성된 적어도 하나의 언어 정보를 제1 딥러닝 모델에 입력할 수 있다. 구체적으로, 전자 장치(101)는, 다중언어 모델로 구성된 멀티 버트 모델인 제1 딥러닝 모델에 적어도 하나의 문자 정보를 입력할 수 있다. 이는, 제1 딥러닝 모델이 적어도 하나의 텍스트 정보에서 문맥적 특징을 효과적으로 반영할 수 있다. 이후, 전자 장치(101)는, 제1 딥러닝 모델에서 출력된 적어도 하나의 키워드 정보를 획득할 수 있고, 메모리(예: 도 1의 메모리(130))에 저장시킬 수 있으며, 적어도 하나의 언어 정보를 기초로 적어도 하나의 중요 문장으로 구성된 적어도 하나의 제2 키워드 정보를 추출할 수 있다. 이후, 전자 장치(101)는, 적어도 하나의 제1 키워드 정보 및 적어도 하나의 제2 키워드 정보를 제1 딥러닝 모델에 입력하여 적어도 하나의 키워드 추출 정보를 출력할 수 있다.

[0038] 또 다른 일 실시예에 따르면, 제1 딥러닝 모델은, 복수의 제1 키워드 정보 및 복수의 제2 키워드 정보, 복수의 텍스트 정보와, 복수의 키워드 추출 정보를 학습하고, 멀티모달 버트 모델(multilingual BERT model)로 구성될 수 있다. 구체적으로, 제1 딥러닝 모델은, 100개 이상의 언어로 훈련된 모델로, 다국어 데이터를 이해하고 처리할 수 있고, 단일 모델로 여러 언어를 동시에 처리할 수 있어 다국어 작업에 적합할 수 있다. 더 구체적으로, 제1 딥러닝 모델은, 입력된 텍스트의 단어들을 문맥적으로 이해하고, 문장 내 의미를 반영한 벡터 표현(임베딩)을 생성할 수 있으며, 각 단어의 의미가 해당 문맥에 따라 동적으로 변화하는 과정으로 학습될 수 있다. 또한, 제1 딥러닝 모델은 대규모의 다국어 코퍼스를 기반으로 언어 모델링(마스크드 언어 모델링, Next Sentence Prediction) 방식으로 사전 훈련되어 있고, 추가적인 미세 조정을 통해 특정 작업에 맞춘 성능 개선 가능한 이점이 있다. 또한, 제1 딥러닝 모델은 하나의 모델로 여러 언어를 지원하며, 언어 간 공통된 표현을 학습해 작업 간의 전이 학습(Transfer Learning)이 가능하고 예시로 영어 데이터로 학습한 모델이 유사한 구조를 가진 독일어나 프랑스어 작업에서도 사용할 수 있다. 또한, 제1 딥러닝 모델은, 같은 단어라도 문맥에 따라 다른 벡터를 할당하여, 중의성을 줄이고 텍스트의 정확한 의미를 파악할 수 있고, Transformer 아키텍처를 기반으로 하며, 양방향(Bidirectional)으로 문장을 분석할 수 있으며, 입력된 텍스트의 양쪽 방향에서 정보를 수집하여 더 정밀한 문장 표현 제공할 수 있다. 또한, 제1 딥러닝 모델은, 번역, 분류, 감정 분석, 질의응답 등 다양한 자연어 처리(NLP) 작업에서 높은 성능을 보일 수 있다. 따라서, 제1 딥러닝 모델은, 다양한 언어 간의 공통 표현을 학습하여, 특정 언어에 제한되지 않는 강력한 일반화 성능 제공할 수 있고, 번역 데이터 없이도 저자원 언어(low-resource languages)에 대한 모델 성능을 높이는 데 유리하며, 다국어 NLP 작업에서 별도의 언어별 모델을 만들 필요 없이 단일 모델로 처리 가능, 리소스 절약을 수행할 수 있다. 또한, 제1 딥러닝 모델은, 대규모 다국어 코퍼스로 학습되어 추가적인 데이터 없이도 강력한 성능을 발휘할 수 있고, 특정 작업에 맞는 데이터셋으로 미세 조정(Fine-tuning)을 통해 다양한 NLP 작업에 쉽게 적용 가능할 수 있다. 따라서, 본 실시예의 전자 장치(101)는 적어도 하나의 제1 키워드 정보 및 적어도 하나의 제2 키워드 정보를 제1 딥러닝 모델에 출력하여 손쉽게 적어도 하나의 키워드 추출 정보를 출력할 수 있다.

[0039] 205 동작에서, 전자 장치(101)(예: 도 1의 프로세서(120))는, 적어도 하나의 이미지 정보를 제2 딥러닝 모델에 입력하여 적어도 하나의 변환 이미지 정보를 추출할 수 있다. 일 실시예에 따르면, 전자 장치(101)는, 적어도

하나의 이미지 정보에서, 언어별로 매칭된 이미지 정보를 제2 딥러닝 모델에 입력하여 입력된 깊이 정보(depth image data)를 적어도 하나의 초기 이미지 정보로 출력하고, 적어도 하나의 초기 이미지 정보를 제2 딥러닝 모델에 입력하여 역방향 확산 (reverse diffusion)을 수행하여 노이즈를 제거한 적어도 하나의 초기 웹툰 이미지 정보로 구성된 적어도 하나의 변환 이미지 정보를 출력하도록 설정될 수 있다. 구체적으로, 전자 장치(101)는, 언어별로 매칭된 적어도 하나의 텍스트 정보와 관련된 적어도 하나의 이미지 정보를 획득하면, 각각을 임베딩하여 가장 유사한 깊이 정보를 포함한 적어도 하나의 이미지 정보를 출력할 수 있도록 한다. 이후, 전자 장치(101)는 제2 딥러닝 모델을 통하여 선택된 깊이 정보를 이용하여 적어도 하나의 초기 이미지 정보를 출력하고, 출력된 적어도 하나의 초기 이미지 정보에 구조적 특징 및 벡터에 대한 근거의 정보를 생성할 수 있다. 이후, 전자 장치(101)는, 적어도 하나의 초기 이미지 정보를 역방향 확산 과정을 수행하여 노이즈 제거를 수행하면서 제2 딥러닝 모델에 입력하여, 적어도 하나의 초기 웹툰 이미지 정보로 구성된 적어도 하나의 변환 이미지 정보를 출력할 수 있다.

[0040] 또 다른 일 실시예에 따르면, 제2 딥러닝 모델은, 복수의 초기 이미지 정보, 복수의 초기 웹툰 이미지 정보와, 복수의 변환 이미지 정보에 기반하여 학습되고, 이미지-깊이 모델(depth-to-image model)로 구성될 수 있다. 구체적으로, 제2 딥러닝 모델은, ViT 모델(vision transformer model)로써, Transformer 아키텍처를 기반으로 이미지 정보를 처리하는 딥러닝 모델로 구성될 수 있다. 이에 대해서 제2 딥러닝 모델의 구체적인 기능으로는, 도 4에 도시된 바와 같이, 초기 이미지 정보를 작은 패치로 분할하고, 각 패치를 일종의 단어 및/또는 텍스트로 간주하여 transformer에 입력할 수 있다. 이후, 제2 딥러닝 모델에서는 각 패치에 대한 고차원 임베딩을 생성하고, 패치 정보를 Transformer 입력에 적합한 형태로 변환하며, 패치 간의 공간적 관계를 유지하기 위해 위치 정보를 임베딩에 추가할 수 있다. 이후, 제2 딥러닝 모델은, 셀프 어텐션(Self-Attention) 메커니즘을 사용하여 패치 간의 관계와 전역적 정보를 학습함으로써, 기존 CNN(Convolutional Neural Network)과 달리, 지역 정보만 아니라 이미지 전반의 맥락적 정보를 이해할 수 있고, Transformer의 출력에서 클래스 토큰(Class Token)을 활용하여 이미지 정보를 분류하거나, 세그멘테이션, 특징 추출 등 다양한 작업 수행할 수 있다. 이를 위해 제2 딥러닝 모델은, NLP(자연어 처리)에서 성공적으로 사용된 Transformer 아키텍처를 이미지 정보 처리에 적용하고, CNN 대신 어텐션 메커니즘을 통해 이미지 정보를 처리하며, Self-Attention 메커니즘으로 이미지 전체의 전역적 관계를 학습할 수 있다. 이는, 제2 딥러닝 모델이 Self-Attention을 통해 이미지를 전역적으로 학습하여, 전체 이미지 맥락을 더 잘 이해하여 CNN의 국소적 한계를 극복할 수 있고, 다양한 이미지 정보 작업(분류, 세그멘테이션, 객체 검출 등)에 적용 가능하며, 기존 CNN 모델(CNN + Transformer 하이브리드)이나 멀티모달 모델(텍스트-이미지 통합)과 쉽게 결합 가능하고, 대규모 데이터셋에서 학습할 경우, CNN 기반 모델보다 더 나은 성능을 보여줄 수 있는 이점이 있다.

[0041] 207 동작에서, 전자 장치(101)(예: 도 1의 프로세서(120))는, 적어도 하나의 키워드 추출 정보 및 적어도 하나의 변환 이미지 정보를 멀티 모델에 입력하여 적어도 하나의 웹툰 이미지 정보를 출력할 수 있다. 일 실시예에 따르면, 멀티 모델은, 복수의 키워드 추출 정보 및 복수의 변환 이미지 정보와, 복수의 웹툰 이미지 정보와, 복수의 키워드-이미지 변환 정보에 기반하여 학습될 수 있다. 또한, 멀티 모델은, 복수의 이미지 벡터 정보 및 복수의 키워드 벡터 정보와, 복수의 유사도 값과, 복수의 분류 정보 그룹을 추가로 학습할 수 있다. 구체적으로, 멀티 모델은, 하나의 정보를 여러 정보 형태(모달리티)로 표현하여 학습하는 접근 방식으로, 이미지 정보, 텍스트 정보, 및 오디오 정보를 결합하여 특징을 학습할 수 있다. 또한, 멀티 모델은, 도 5에 도시된 바와 같이, 이미지 정보와 텍스트 정보를 동시에 학습하여, 동일한 잠재 공간(Latent Space)에 매핑할 수 있고, 대조 학습(Contrastive Learning)을 통해 유사한 정보를 가까이하면서, 비유사한 정보를 최대한 멀리 배치하는 방식으로 정보간의 관계를 학습할 수 있다. 또한, 멀티 모델은, 도 3에 도시된 바와 같이, 텍스트 정보와 이미지 정보에 대해 각각의 인코더를 사용하여 최종 프로젝션 레이어를 통해 동일한 차원의 임베딩 벡터를 생성할 수 있다. 이는, 텍스트 정보와 이미지 정보와의 유사도를 측정하거나 관계를 이해하는데 매우 유용할 수 있고, 텍스트-이미지 생성 및 편집에 CLIP 임베딩을 활용하여 품질 향상할 수 있다. 텍스트-이미지 정보 쌍에서 유사한 쌍은 가까이, 비유사한 쌍은 멀리 배치할 수 있도록 한다.

[0042] 또 다른 일 실시예에 따르면, 전자 장치(101)는, 적어도 하나의 키워드 추출 정보를 키워드 벡터 정보로 변환하고, 적어도 하나의 변환 이미지 정보를 이미지 벡터 정보로 변환하고, 이미지 벡터 정보를 멀티 모델에 입력하여 키워드 벡터 정보와 동일한 크기, 동일한 거리, 및 동일한 차원으로 축소시키고, 이미지 벡터 정보와 키워드 벡터 정보 간의 내적을 하기의 [수학식 1]에 기초하여 산출한 유사도를 정량화할 수 있다.

[0043] [수학식 1]

[0044] $$similarity = \cos(\theta) = \frac{A \cdot B}{|A|\,|B|}$$

[0045] 여기서, $\cos(\theta)$는 유사도이고, A는 이미지 벡터 정보이고, B는 키워드 벡터 정보임.

[0046] 구체적으로, 멀티 모델은 훈련 과정에서 텍스트 정보 및 이미지 정보를 각각 텍스트 정보 벡터 및 이미지 정보 벡터로 변환할 수 있다. 각 인코더는 동일한 차원의 텍스트-이미지 정보 쌍에서 유사한 쌍을 가까이, 비유사한 쌍을 멀리 배치할 수 있다. 또한, 전자 장치(101)는, 텍스트 정보 벡터와 이미지 정보 벡터를 정규화한 후, 내적을 계산해 상기의 [수학식 1]에 기초하여 코사인 유사도를 산출할 수 있고, 해당 값은 -1~1 사이로, 높은 값일수록 텍스트와 이미지 간의 유사성이 높음을 의미할 수 있다. 또한, 전자 장치(101)는, 도 4에 도시된 바와 같이, 각 로짓의 최대값이 실제 레이블과 얼마나 일치하는지 확인하여 정확도를 계산할 수 있고, 더 구체적으로 예측 배열 [2, 0, 2, 3, 3]과 정답 레이블 배열 [0, 1, 2, 3, 4]을 비교하면 정확도는 0.4(40%)로 산출할 수 있어, 기 설정된 유사도와 비교하여 낮은 경우로 판단하고 각 벡터간의 거리를 최대한 이격시킬 수 있다. 특히, 전자 장치(101)는, 로짓을 사용해 텍스트 정보 벡터와 이미지 정보 벡터 간의 교차 엔트로피 손실을 행(row) 및 열(column) 방향으로 계산하고 평균을 산출하여 출력된 손실 값이 텍스트 정보 벡터와 이미지 정보 벡터 간의 관계를 학습하기 위해 인코더를 업데이트하는 데 사용될 수 있다.

[0047] 또한, 전자 장치(101)는, 유사도에 근거하여, 이미지 벡터 정보와 키워드 벡터 정보가 기 설정된 유사도를 초과하면, 이미지 벡터 정보와 키워드 벡터 정보를 근접시킬 수 있다. 구체적으로, 전자 장치(101)는, 이미지 벡터 정보와 키워드 벡터 정보에서 상기의 [수학식 1]에 출력된 값이 0.8이고, 기 설정된 유사도가 0.7이면, 유사도가 높은 것으로 판단하고 이미지 벡터 정보와 키워드 벡터 정보를 근접시킬 수 있다. 또한, 전자 장치(101)는, 유사도에 근거하여, 이미지 벡터 정보와 키워드 벡터 정보가 기 설정된 유사도를 초과하지 않으면, 이미지 벡터 정보와 키워드 벡터 정보의 거리를 이격시킬 수 있다. 구체적으로, 전자 장치(101)는, 이미지 벡터 정보와 키워드 벡터 정보에서 상기의 [수학식 1]에 출력된 값이 0.3이고, 기 설정된 유사도가 0.7이면, 유사도가 낮은 것으로 판단하고 이미지 벡터 정보와 키워드 벡터 정보를 최대한 멀리 이격시킬 수 있다.

[0048] 또 다른 일 실시예에 따르면, 전자 장치(101)는, 상단의 과정을 거쳐 멀티 모델에서 출력된 적어도 하나의 웹툰 이미지 정보에 대해서 MSCOCO 데이터를 Cartoon-GAN으로 변환하여 웹툰 스타일로 변경한 82만 개의 텍스트-이미지 쌍을 사용하고, 25 에폭, 학습률 0.00001, 배치 크기 128(4개의 GPU에서 32씩 분산)하는 훈련 환경을 생성할 수 있으며, 멀티링구얼 BERT와 Vision Transformer(ViT)를 텍스트와 이미지 인코더로 사용하고, 각각 512차원 프로젝션 레이어를 추가할 수 있다. 이후, 전자 장치(101)는, 멀티 모델을 이용하여 텍스트 정보 벡터와 이미지 정보 벡터의 코사인 유사도를 기반으로 교차 엔트로피 손실을 계산하여 학습할 수 있음으로써, 유사도가 높은 상위 N개의 이미지 정보를 반환하고, 도 6에 도시된 바와 같이, 한국어와 영어 쿼리를 모두 실험에 사용하여 의미적으로 적합한 이미지 정보를 검색할 수 있다. 또한, 전자 장치(101)는, 도 7에 도시된 바와 같이, 입력된 적어도 하나의 이미지 정보에 깊이 정보(Depth)를 초기값으로 설정할 수 있고, 역방향 확산(reverse diffusion) 과정을 통해 노이즈를 제거하며 RGB 이미지 정보를 생성하며, “webtoon” 키워드를 사용해 만화 스타일 이미지 정보를 생성할 수 있다. 즉, 전자 장치(101)는, 도 7에 도시된 바와 같이, 제1 딥러닝 모델 및 제2 딥러닝 모델, 및 멀티 모델을 기초로 적어도 하나의 텍스트 정보 및 적어도 하나의 이미지 정보를 기초로 적어도 하나의 웹툰 이미지 정보를 생성할 수 있으며, 수많은 검증을 통해 정확한 웹툰 이미지 정보를 생성할 수 있다.

[0049] 209 동작에서, 전자 장치(101)(예: 도 1의 프로세서(120))는, 통신 인터페이스(예: 도 1의 통신 인터페이스(160))를 통하여, 적어도 하나의 웹툰 이미지 정보, 적어도 하나의 키워드 추출 정보, 및 적어도 하나의 변환 이미지 정보를 외부 전자 장치(102, 104, 106)에 송신할 수 있다. 일 실시예에 따르면, 전자 장치(101)는 생성된 적어도 하나의 웹툰 이미지 정보와 적어도 하나의 웹툰 이미지 정보를 추출한 적어도 하나의 키워드 정보 및 적어도 하나의 변환 이미지 정보를 하나의 파일에 묶고, 사용자가 보유한 외부 전자 장치(102, 104, 106)에 송신할 수 있다. 이후, 사용자는 사용자의 외부 전자 장치(102, 104, 106)를 기초로 사용자가 입력한 텍스트 정보 및 이미지 정보로 웹툰 이미지 정보를 출력한 결과 값을 손쉽게 확인할 수 있다.

[0051] 본 실시예에 따른 전자 장치(101)는, 사용자의 외부 전자 장치(102, 104, 106)로부터 획득된 적어도 하나의 텍스트 정보 및 적어도 하나의 이미지 정보를 제1 딥러닝 모델, 제2 딥러닝 모델, 및 멀티 모델에 입력하여 적어도 하나의 웹툰 이미지 정보를 손쉽게 출력할 수 있고, 적어도 하나의 웹툰 이미지 정보를 사용자의 외부 전자 장치(102, 104, 106)에 송신함으로써, 사용자가 텍스트 정보 및 이미지 정보만으로도 손쉽게 웹툰 이미지 정보

를 획득하고 이를 활용할 수 있는 이점이 있다.

[0053] 다양한 실시예들에 따르면, 다중 모델을 기초로 웹툰 이미지 정보를 생성하기 위한 전자 장치에 있어서, 통신 인터페이스; 프로세서; 를 포함하고, 상기 프로세서는, 상기 통신 인터페이스를 통하여, 사용자의 외부 전자 장치로부터 적어도 하나의 텍스트 정보 및 적어도 하나의 이미지 정보를 수신하고, 상기 적어도 하나의 텍스트 정보를 제1 딥러닝 모델에 입력하여 적어도 하나의 키워드 추출 정보를 추출하고, 상기 적어도 하나의 이미지 정보를 제2 딥러닝 모델에 입력하여 적어도 하나의 변환 이미지 정보를 추출하고, 및 상기 적어도 하나의 키워드 추출 정보 및 상기 적어도 하나의 변환 이미지 정보를 멀티 모델에 입력하여 적어도 하나의 웹툰 이미지 정보를 출력하도록 설정되고, 상기 멀티 모델은, 복수의 키워드 추출 정보 및 복수의 변환 이미지 정보와, 복수의 웹툰 이미지 정보와, 복수의 키워드-이미지 변환 정보에 기반하여 학습된다.

[0054] 다양한 실시예들에 따르면, 상기 적어도 하나의 텍스트 정보는, 세계 전체에서 현재 사용되거나 과거에 사용되었던 모든 글자에 대한 적어도 하나의 글자 정보와, 세계 전체에서 현재 사용되거나 과거에 사용되었던 모든 언어에 대한 적어도 하나의 언어 정보를 포함한다.

[0055] 다양한 실시예들에 따르면, 상기 프로세서는, 상기 적어도 하나의 글자 정보를 기초로 적어도 하나의 중요 단어로 구성된 적어도 하나의 제1 키워드 정보를 추출하고, 상기 적어도 하나의 언어 정보를 기초로 적어도 하나의 중요 문장으로 구성된 적어도 하나의 제2 키워드 정보를 추출하고, 및 상기 적어도 하나의 제1 키워드 정보 및 상기 적어도 하나의 제2 키워드 정보를 상기 멀티 모델에 입력하여 상기 적어도 하나의 키워드 추출 정보를 출력하도록 설정되고, 상기 제1 딥러닝 모델은, 복수의 제1 키워드 정보 및 복수의 제2 키워드 정보, 복수의 텍스트 정보와, 복수의 키워드 추출 정보를 학습하고, 멀티모달 버트 모델(multilingual BERT model)로 구성된다.

[0056] 다양한 실시예들에 따르면, 상기 적어도 하나의 이미지 정보는, 상기 적어도 하나의 텍스트 정보에 입력된 텍스트 정보와 매칭되는 이미지 정보와, 다양한 표현을 나타내는 이미지 정보와, 각 언어별로 매칭된 이미지 정보를 포함한다.

[0057] 다양한 실시예들에 따르면, 상기 프로세서는, 상기 적어도 하나의 이미지 정보에서, 상기 언어별로 매칭된 이미지 정보를 상기 제2 딥러닝 모델에 입력하여 입력된 깊이 정보(depth image data)를 적어도 하나의 초기 이미지 정보로 출력하고, 및 상기 적어도 하나의 초기 이미지 정보를 상기 제2 딥러닝 모델에 입력하여 역방향 확산(reverse diffusion)을 수행하여 노이즈를 제거한 적어도 하나의 초기 웹툰 이미지 정보로 구성된 상기 적어도 하나의 변환 이미지 정보를 출력하도록 설정되고, 상기 제2 딥러닝 모델은, 복수의 초기 이미지 정보, 복수의 초기 웹툰 이미지 정보와, 복수의 변환 이미지 정보에 기반하여 학습되고, 이미지-깊이 모델(depth-to-image model)로 구성된다.

[0058] 다양한 실시예들에 따르면, 상기 프로세서는, 상기 적어도 하나의 키워드 추출 정보를 키워드 벡터 정보로 변환하고, 상기 적어도 하나의 변환 이미지 정보를 이미지 벡터 정보로 변환하고, 상기 이미지 벡터 정보를 상기 멀티 모델에 입력하여 상기 키워드 벡터 정보와 동일한 크기, 동일한 거리, 및 동일한 차원으로 축소시키고, 상기 이미지 벡터 정보와 상기 키워드 벡터 정보 간의 내적을 하기의 [수학식 1]에 기초하여 산출한 유사도를 정량화하도록 설정된다.

[0059] [수학식 1]

[0060] $$similarity = \cos(\theta) = \frac{A \cdot B}{|A|\,|B|}$$

[0061] 여기서, $\cos(\theta)$는 상기 유사도이고, A는 상기 이미지 벡터 정보이고, B는 상기 키워드 벡터 정보임.

[0062] 다양한 실시예들에 따르면, 상기 프로세서는, 상기 유사도에 근거하여, 상기 이미지 벡터 정보와 상기 키워드 벡터 정보가 기 설정된 유사도를 초과하면, 상기 이미지 벡터 정보와 상기 키워드 벡터 정보를 근접시키고, 및 상기 유사도에 근거하여, 상기 이미지 벡터 정보와 상기 키워드 벡터 정보가 기 설정된 유사도를 초과하지 않으면, 상기 이미지 벡터 정보와 상기 키워드 벡터 정보의 거리를 이격시키도록 설정되고, 상기 멀티 모델은, 복수의 이미지 벡터 정보 및 복수의 키워드 벡터 정보와, 복수의 유사도 값과, 복수의 분류 정보 그룹을 추가로 학습하도록 설정된다.

[0063] 또 다른 다양한 실시예들에 따르면, 다중 모델을 기초로 웹툰 이미지 정보를 생성하기 위한 전자 장치의 구동

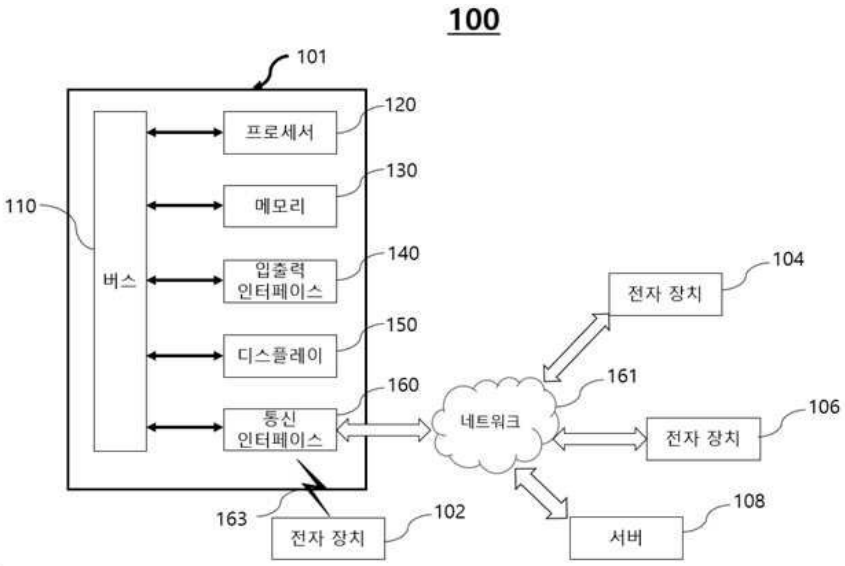
방법에 있어서, 상기 방법은, 통신 인터페이스를 통하여, 사용자의 외부 방법으로부터 적어도 하나의 텍스트 정보 및 적어도 하나의 이미지 정보를 수신하고, 프로세서를 통하여, 상기 적어도 하나의 텍스트 정보를 제1 딥러닝 모델에 입력하여 적어도 하나의 키워드 추출 정보를 추출하고, 상기 프로세서를 통하여, 상기 적어도 하나의 이미지 정보를 제2 딥러닝 모델에 입력하여 적어도 하나의 변환 이미지 정보를 추출하고, 및 상기 프로세서를 통하여, 상기 적어도 하나의 키워드 추출 정보 및 상기 적어도 하나의 변환 이미지 정보를 멀티 모델에 입력하여 적어도 하나의 웹툰 이미지 정보를 출력하도록 설정되고, 상기 멀티 모델은, 복수의 키워드 추출 정보 및 복수의 변환 이미지 정보와, 복수의 웹툰 이미지 정보와, 복수의 키워드-이미지 변환 정보에 기반하여 학습된다.

[0065] 본 문서에서 사용된 용어 "모듈" 또는 "~부"은 하드웨어, 소프트웨어 또는 펌웨어로 구성된 유닛을 포함하며, 예를 들면, 로직, 논리 블록, 부품, 또는 회로 등의 용어와 상호 호환적으로 사용될 수 있다. "모듈" 또는 "~부"는, 일체로 구성된 부품 또는 하나 또는 그 이상의 기능을 수행하는 최소 단위 또는 그 일부가 될 수 있다. "모듈" 또는 "~부"는 기계적으로 또는 전자적으로 구현될 수 있으며, 예를 들면, 어떤 동작들을 수행하는, 알려졌거나 앞으로 개발될, ASIC(application-specific integrated circuit) 칩, FPGAs(field-programmable gate arrays), 또는 프로그램 가능 논리 장치를 포함할 수 있고, 프로세서(120)에 의하여 실행될 수 있다. 다양한 실시예에 따른 장치(예: 모듈들 또는 그 기능들) 또는 방법(예: 동작들)의 적어도 일부는 프로그램 모듈의 형태로 컴퓨터로 판독 가능한 저장 매체(예: 메모리(130))에 저장된 명령어로 구현될 수 있다. 상기 명령어가 프로세서(예: 프로세서(120))에 의해 실행될 경우, 프로세서가 상기 명령어에 해당하는 기능을 수행할 수 있다. 컴퓨터로 판독 가능한 기록 매체는, 하드디스크, 플로피디스크, 마그네틱 매체(예: 자기테이프), 광기록 매체(예: CD-ROM, DVD, 자기-광 매체 (예: 플롭티컬 디스크), 내장 메모리 등을 포함할 수 있다. 명령어는 컴파일러에 의해 만들어지는 코드 또는 인터프리터에 의해 실행될 수 있는 코드를 포함할 수 있다. 다양한 실시예에 따른 모듈 또는 프로그램 모듈은 전술한 구성요소들 중 적어도 하나 이상을 포함하거나, 일부가 생략되거나, 또는 다른 구성요소를 더 포함할 수 있다. 다양한 실시예에 따른, 모듈, 프로그램 모듈 또는 다른 구성요소에 의해 수행되는 동작들은 순차적, 병렬적, 반복적 또는 휴리스틱하게 실행되거나, 적어도 일부 동작이 다른 순서로 실행되거나, 생략되거나, 또는 다른 동작이 추가될 수 있다.

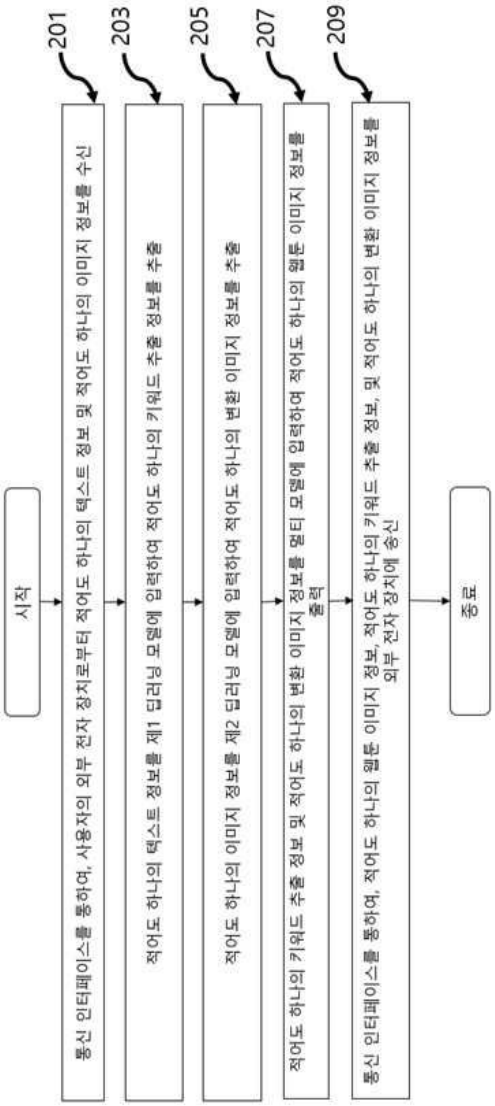
[0066] 그리고 본 문서에 개시된 실시예는 개시된, 기술 내용의 설명 및 이해를 위해 제시된 것이며, 본 개시의 범위를 한정하는 것은 아니다. 따라서, 본 개시의 범위는, 본 개시의 기술적 사상에 근거한 모든 변경 또는 다양한 다른 실시예를 포함하는 것으로 해석되어야 한다.

도면

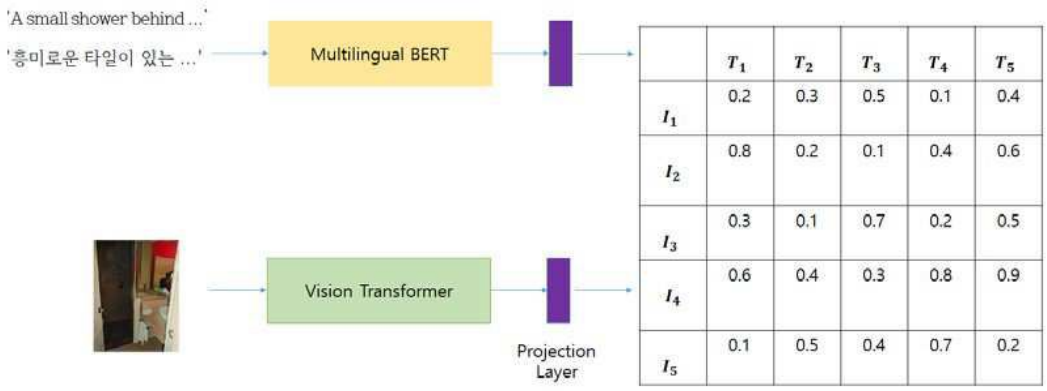
도면1



도면2



도면3



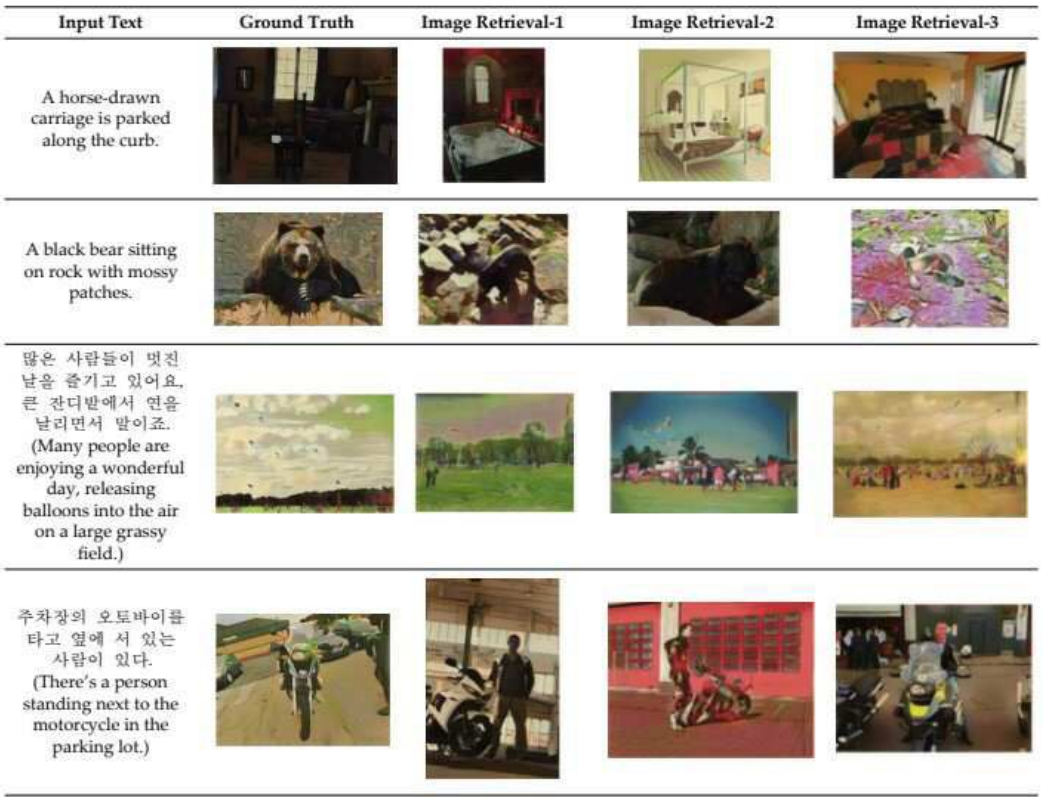
	T_1	T_2	T_3	T_4	T_5
I_1	0.2	0.3	0.5	0.1	0.4
I_2	0.8	0.2	0.1	0.4	0.6
I_3	0.3	0.1	0.7	0.2	0.5
I_4	0.6	0.4	0.3	0.8	0.9
I_5	0.1	0.5	0.4	0.7	0.2

도면4

	T_1	T_2	T_3	T_4	T_5
I_1	0.2	0.3	0.5	0.1	0.4
I_2	0.8	0.2	0.1	0.4	0.6
I_3	0.3	0.1	0.7	0.2	0.5
I_4	0.6	0.4	0.3	0.8	0.9
I_5	0.1	0.5	0.4	0.7	0.2

```
labels = np.arange(len(logits))
acc_i = ((np.argmax(logits, 1) == labels)).mean()
acc_t = ((np.argmax(logits, 0) == labels)).mean()
acc = (acc_i + acc_t) / 2
```

도면5



도면6

(a) CLIP Input Text	(b) Most Similar Image	(c) Generated Image 1	(d) Generated Image 2
Six teens playing frisbee in a field of grass			
A horse drawn carriage is parked along the curb.			
Electric appliances crammed on a counter in a kitchen.			
A person in a ski suit is skiing.			

도면7

(a) CLIP Input Text	(b) Most Similar Image	(c) Generated Image
Man riding a motor bike on a dirt road on the countryside.		
The wind whipped through the rider's hair as he navigated the twists and turns of the path.		
The sound of the engine echoed across the fields, disturbing the peaceful silence of the countryside.		
He couldn't help but feel a sense of freedom as he sped along on his trusty machine.		