(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0122566
(43) 공개일자 2022년09월02일

(51) 국제특허분류(Int. Cl.)
G06V 30/19 (2022.01) *G06N 3/08* (2006.01)
G06T 5/00 (2019.01)
(52) CPC특허분류
G06V 30/19147 (2022.01)
G06N 3/08 (2013.01)
(21) 출원번호 10-2022-0101802
(22) 출원일자 2022년08월16일
심사청구일자 2022년08월16일
(30) 우선권주장
202210279539.X 2022년03월22일 중국(CN)

(71) 출원인
베이징 바이두 넷컴 사이언스 테크놀로지 컴퍼니 리미티드
중국 베이징 하이디안 디스트릭트 샹디 10번가 넘버 10, 바이두 캠퍼스 2층
(72) 발명자
장 청콴
중국 베이징 100085 하이디안 디스트릭트 샹디 10번가 넘버 10 바이두 캠퍼스 2층
위 위에천
중국 베이징 100085 하이디안 디스트릭트 샹디 10번가 넘버 10 바이두 캠퍼스 2층
(뒷면에 계속)
(74) 대리인
특허법인태평양

전체 청구항 수 : 총 23 항

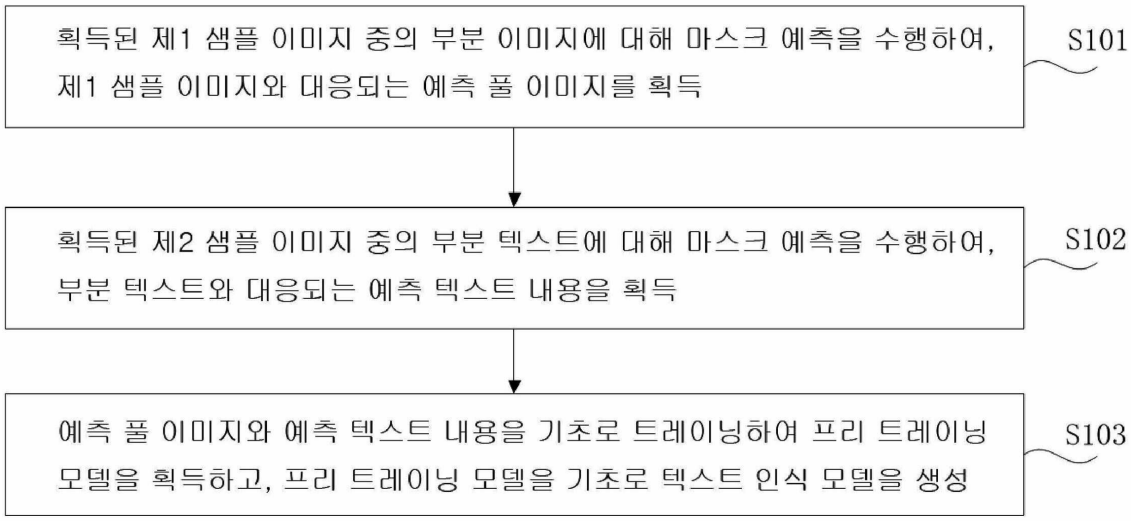
(54) 발명의 명칭 **텍스트 인식 모델의 트레이닝 방법, 텍스트 인식 방법 및 장치**

(57) 요 약

본 출원은 텍스트 인식 모델의 트레이닝 방법, 텍스트 인식 방법 및 장치를 제공하며, 인공지능 기술분야에 관한 것으로서, 구체적으로 딥러닝, 컴퓨터 비전 기술분야에 관한 것이며, 광학 문자 인식 등의 시나리오에 적용될 수 있다. 방안에 따르면, 획득된 제1 샘플 이미지 중의 부분 이미지에 대해 마스크 예측을 수행하여, 제1 샘플 이미

(뒷면에 계속)

대 표 도 - 도1



지와 대응되는 예측 풀 이미지를 획득하고, 획득된 제2 샘플 이미지 중의 부분 텍스트에 대해 마스크 예측을 수행하여, 부분 텍스트와 대응되는 예측 텍스트 내용을 획득하고, 예측 풀 이미지와 예측 텍스트 내용을 기초로 트레이닝하여 프리 트레이닝 모델을 획득하고, 프리 트레이닝 모델을 기초로 텍스트 인식 모델을 생성하고, 텍스트 인식 모델은 인식 대상 이미지에 대해 텍스트 인식을 수행하기 위한 것이며, 프리 트레이닝 모델이 보다 강한 이미지 비전 추리 능력과 텍스트 의미 추리 능력을 학습하도록 하고, 이에 따라 프리 트레이닝 모델을 기반으로 생성된 텍스트 인식 모델이 텍스트 인식을 수행할 때, 텍스트 인식의 정확성과 신뢰성을 향상시킨다.

(52) CPC특허분류
G06T 5/004 (2013.01)
G06T 2207/20081 (2013.01)

(72) 발명자
리 위린
중국 베이징 100085 하이디안 디스트릭트 샹디 10 번가 넘버 10 바이두 캠퍼스 2층
차오 지안지안
중국 베이징 100085 하이디안 디스트릭트 샹디 10 번가 넘버 10 바이두 캠퍼스 2층
친 샤멍
중국 베이징 100085 하이디안 디스트릭트 샹디 10 번가 넘버 10 바이두 캠퍼스 2층
야오 쿤
중국 베이징 100085 하이디안 디스트릭트 샹디 10 번가 넘버 10 바이두 캠퍼스 2층
한 준위
중국 베이징 100085 하이디안 디스트릭트 샹디 10 번가 넘버 10 바이두 캠퍼스 2층
리우 징투오
중국 베이징 100085 하이디안 디스트릭트 샹디 10 번가 넘버 10 바이두 캠퍼스 2층
딩 얼루이
중국 베이징 100085 하이디안 디스트릭트 샹디 10 번가 넘버 10 바이두 캠퍼스 2층
왕 징동
중국 베이징 100085 하이디안 디스트릭트 샹디 10 번가 넘버 10 바이두 캠퍼스 2층

명 세 서

청구범위

청구항 1

텍스트 인식 모델의 트레이닝 방법에 있어서,

획득된 제1 샘플 이미지 중의 부분 이미지에 대해 마스크 예측을 수행하여, 상기 제1 샘플 이미지와 대응되는 예측 풀 이미지를 획득하는 단계;

획득된 제2 샘플 이미지 중의 부분 텍스트에 대해 상기 마스크 예측을 수행하여, 상기 부분 텍스트와 대응되는 예측 텍스트 내용을 획득하는 단계;

상기 예측 풀 이미지와 상기 예측 텍스트 내용을 기초로 트레이닝하여 프리 트레이닝 모델을 획득하고, 상기 프리 트레이닝 모델을 기초로 텍스트 인식 모델을 생성하며, 여기서, 상기 텍스트 인식 모델은 인식 대상 이미지에 대해 텍스트 인식을 수행하기 위한 것인 단계;를 포함하는 것을 특징으로 하는 텍스트 인식 모델의 트레이닝 방법.

청구항 2

제1항에 있어서, 상기 마스크 예측은,

타겟 대상 중의 부분 대상을 랜덤으로 가리우는 단계;

상기 타겟 대상 중 가리워지지 않은 대상을 기초로, 상기 타겟 대상 중 가리워진 부분 대상에 대해 예측하여, 예측 결과를 획득하는 단계;를 포함하며,

여기서, 상기 타겟 대상이 제1 샘플 이미지이면, 상기 타겟 대상 중의 부분 대상은 부분 이미지이고, 상기 예측 결과는 상기 예측 풀 이미지이며; 상기 타겟 대상이 제2 샘플 이미지이면, 상기 타겟 대상 중의 부분 대상은 부분 텍스트이고, 상기 예측 결과는 상기 예측 텍스트 내용인 단계;를 포함하는 텍스트 인식 모델의 트레이닝 방법.

청구항 3

제2항에 있어서,

상기 타겟 대상 중 가리워지지 않은 대상을 기초로, 상기 타겟 대상 중 가리워진 부분 대상에 대해 예측하여, 예측 결과를 획득하는 단계는,

상기 타겟 대상 중 가리워지지 않은 대상에 대응되는 대상 특징을 추출하여, 제1 대상 특징을 획득하는 단계;

상기 제1 대상 특징을 기초로, 상기 타겟 대상 중 가리워진 부분 대상에 대해 예측하여, 상기 예측 결과를 획득하는 단계;를 포함하며,

여기서, 상기 타겟 대상이 제1 샘플 이미지이면, 상기 제1 대상 특징은 제1 비전 특징이고; 상기 타겟 대상이 제2 샘플 이미지이면, 상기 제1 대상 특징은 제1 의미 특징인 텍스트 인식 모델의 트레이닝 방법.

청구항 4

제3항에 있어서,

상기 타겟 대상은 제1 샘플 이미지이고, 상기 제1 대상 특징은 제1 비전 특징이며; 상기 제1 대상 특징을 기초로, 상기 타겟 대상 중 가리워진 부분 대상에 대해 예측하여, 상기 예측 결과를 획득하는 단계는,

상기 제1 비전 특징을 기초로, 상기 제1 샘플 이미지 중 가리워진 부분 이미지에 대응되는 비전 특징을 예측하여, 제2 비전 특징을 획득하는 단계;

상기 제2 비전 특징을 기초로, 상기 제1 샘플 이미지 중 가리워진 부분 이미지를 결정하는 단계;

상기 제1 샘플 이미지 중 가리워지지 않은 이미지, 및 결정된 상기 제1 샘플 이미지 중 가리워진 부분 이미지를

기초로, 상기 예측 풀 이미지를 생성하는 단계;를 포함하는 텍스트 인식 모델의 트레이닝 방법.

청구항 5

제3항 또는 제4항에 있어서,

상기 타겟 대상은 제2 샘플 이미지이고, 상기 제1 대상 특징은 제1 의미 특징이며; 상기 제1 대상 특징을 기초로, 상기 타겟 대상 중 가리워진 부분 대상에 대해 예측하여, 상기 예측 결과를 획득하는 단계는,

상기 제1 의미 특징을 기초로, 상기 제2 샘플 이미지 중 가리워진 부분 텍스트에 대응되는 의미 특징을 예측하여, 제2 의미 특징을 획득하는 단계;

상기 제2 의미 특징을 기초로, 상기 예측 텍스트 내용을 생성하는 단계;를 포함하는 텍스트 인식 모델의 트레이닝 방법.

청구항 6

제1항 내지 제4항 중 어느 한 항에 있어서,

상기 프리 트레이닝 모델을 기초로 텍스트 인식 모델을 생성하는 단계는,

인식 대상 태스크와 트레이닝 이미지를 획득하며, 여기서, 상기 트레이닝 이미지는 텍스트를 포함하는 단계;

상기 인식 대상 태스크와 상기 트레이닝 이미지를 기초로, 상기 프리 트레이닝 모델에 대해 트레이닝하여 상기 텍스트 인식 모델을 획득하는 단계;를 포함하는 텍스트 인식 모델의 트레이닝 방법.

청구항 7

제6항에 있어서,

상기 인식 대상 태스크와 상기 트레이닝 이미지를 기초로, 상기 프리 트레이닝 모델에 대해 트레이닝하여 상기 텍스트 인식 모델을 획득하는 단계는,

상기 트레이닝 이미지를 상기 프리 트레이닝 모델로 입력하여, 상기 트레이닝 이미지에 대응되는 멀티모달 특징맵을 획득하는 단계;

상기 인식 대상 태스크와 상기 멀티모달 특징맵을 기초로, 상기 텍스트 인식 모델을 생성하는 단계;를 포함하는 텍스트 인식 모델의 트레이닝 방법.

청구항 8

제7항에 있어서,

상기 인식 대상 태스크와 상기 멀티모달 특징맵을 기초로, 상기 텍스트 인식 모델을 생성하는 단계는,

상기 멀티모달 특징맵을 기초로, 상기 트레이닝 이미지의 상기 인식 대상 태스크에 따른 예측 인식 결과를 예측하는 단계;

상기 트레이닝 이미지의 기설정된 진실한 인식 결과, 및 상기 예측 인식 결과를 기초로, 상기 텍스트 인식 모델을 구축하는 단계;를 포함하는 텍스트 인식 모델의 트레이닝 방법.

청구항 9

텍스트 인식 방법에 있어서,

인식 대상 이미지를 획득하며, 여기서, 상기 인식 대상 이미지는 텍스트를 포함하는 단계;

사전에 트레이닝된 텍스트 인식 모델을 기반으로 상기 인식 대상 이미지에 대해 텍스트 인식을 수행하여, 상기 인식 대상 이미지 중의 텍스트 내용을 획득하는 단계;를 포함하고,

여기서, 상기 텍스트 인식 모델은 제1항 내지 제4항 중 어느 한 항에 따른 텍스트 인식 모델의 트레이닝 방법을 기반으로 획득된 것인 것을 특징으로 하는 텍스트 인식 방법.

청구항 10

제9항에 있어서,

사전에 트레이닝된 텍스트 인식 모델을 기반으로 상기 인식 대상 이미지에 대해 텍스트 인식을 수행하여, 상기 인식 대상 이미지 중의 텍스트 내용을 획득하는 단계는,

상기 텍스트 인식 모델을 기초로 상기 인식 대상 이미지의 멀티모달 특징맵을 결정하고, 상기 멀티모달 특징맵을 기초로 상기 인식 대상 이미지 중의 텍스트 내용을 결정하는 단계를 포함하고,

여기서, 상기 인식 대상 이미지의 멀티모달 특징맵은 상기 인식 대상 이미지의 비전 특징과 의미 특징을 나타내기 위한 것인 텍스트 인식 방법.

청구항 11

텍스트 인식 모델의 트레이닝 장치에 있어서,

예측 유닛, 트레이닝 유닛, 생성 유닛을 포함하고,

상기 예측 유닛은 획득된 제1 샘플 이미지 중의 부분 이미지에 대해 마스크 예측을 수행하여, 상기 제1 샘플 이미지와 대응되는 예측 풀 이미지를 획득하고;

상기 예측 유닛은 또한, 획득된 제2 샘플 이미지 중의 부분 텍스트에 대해 상기 마스크 예측을 수행하여, 상기 부분 텍스트와 대응되는 예측 텍스트 내용을 획득하고;

상기 트레이닝 유닛은 상기 예측 풀 이미지와 상기 예측 텍스트 내용을 기초로 트레이닝하여 프리 트레이닝 모델을 획득하고;

상기 생성 유닛은 상기 프리 트레이닝 모델을 기초로 텍스트 인식 모델을 생성하며, 여기서, 상기 텍스트 인식 모델은 인식 대상 이미지에 대해 텍스트 인식을 수행하기 위한 것을 특징으로 하는 텍스트 인식 모델의 트레이닝 장치.

청구항 12

제11항에 있어서, 상기 예측 유닛은,

타겟 대상 중의 부분 대상을 랜덤으로 가리우는 가림 서브 유닛;

상기 타겟 대상 중 가리워지지 않은 대상을 기초로, 상기 타겟 대상 중 가리워진 부분 대상에 대해 예측하여, 예측 결과를 획득하는 예측 서브 유닛;을 포함하고,

여기서, 상기 타겟 대상이 제1 샘플 이미지이면, 상기 타겟 대상 중의 부분 대상은 부분 이미지이고, 상기 예측 결과는 상기 예측 풀 이미지이며; 상기 타겟 대상이 제2 샘플 이미지이면, 상기 타겟 대상 중의 부분 대상은 부분 텍스트이고, 상기 예측 결과는 상기 예측 텍스트 내용인 텍스트 인식 모델의 트레이닝 장치.

청구항 13

제12항에 있어서, 상기 예측 서브 유닛은,

상기 타겟 대상 중 가리워지지 않은 대상에 대응되는 대상 특징을 추출하여, 제1 대상 특징을 획득하는 추출 모듈;

상기 제1 대상 특징을 기초로, 상기 타겟 대상 중 가리워진 부분 대상에 대해 예측하여, 상기 예측 결과를 획득하는 예측 모듈;을 포함하고,

여기서, 상기 타겟 대상이 제1 샘플 이미지이면, 상기 제1 대상 특징은 제1 비전 특징이고; 상기 타겟 대상이 제2 샘플 이미지이면, 상기 제1 대상 특징은 제1 의미 특징인 텍스트 인식 모델의 트레이닝 장치.

청구항 14

제13항에 있어서,

상기 타겟 대상은 제1 샘플 이미지이고, 상기 제1 대상 특징은 제1 비전 특징이며; 상기 예측 모듈은,

상기 제1 비전 특징을 기초로, 상기 제1 샘플 이미지 중 가리워진 부분 이미지에 대응되는 비전 특징을 예측하

여 제2 비전 특징을 획득하는 제1 예측 서브 모듈;

상기 제2 비전 특징을 기초로, 상기 제1 샘플 이미지 중 가리워진 부분 이미지를 결정하는 제1 결정 서브 모듈;

상기 제1 샘플 이미지 중 가리워지지 않은 이미지, 및 결정된 상기 제1 샘플 이미지 중 가리워진 부분 이미지를 기초로, 상기 예측 풀 이미지를 생성하는 제1 생성 서브 모듈;을 포함하는 텍스트 인식 모델의 트레이닝 장치.

청구항 15

제13항 또는 제14항에 있어서,

상기 타겟 대상은 제2 샘플 이미지이고, 상기 제1 대상 특징은 제1 의미 특징이며; 상기 예측 모듈은,

상기 제1 의미 특징을 기초로, 상기 제2 샘플 이미지 중 가리워진 부분 텍스트에 대응되는 의미 특징을 예측하여, 제2 의미 특징을 획득하는 제2 예측 서브 모듈;

상기 제2 의미 특징을 기초로, 상기 예측 텍스트 내용을 생성하는 제2 생성 서브 모듈;을 포함하는 텍스트 인식 모델의 트레이닝 장치.

청구항 16

제11항 내지 제14항 중 어느 한 항에 있어서, 상기 생성 유닛은,

인식 대상 태스크와 트레이닝 이미지를 획득하며, 여기서, 상기 트레이닝 이미지는 텍스트를 포함하는 획득 서브 유닛;

상기 인식 대상 태스크와 상기 트레이닝 이미지를 기초로, 상기 프리 트레이닝 모델에 대해 트레이닝하여 상기 텍스트 인식 모델을 획득하는 트레이닝 서브 유닛;을 포함하는 텍스트 인식 모델의 트레이닝 장치.

청구항 17

제16항에 있어서, 상기 트레이닝 서브 유닛은,

상기 트레이닝 이미지를 상기 프리 트레이닝 모델로 입력하여, 상기 트레이닝 이미지에 대응되는 멀티모달 특징 맵을 획득하는 입력 모듈;

상기 인식 대상 태스크와 상기 멀티모달 특징맵을 기초로, 상기 텍스트 인식 모델을 생성하는 생성 모듈;을 포함하는 텍스트 인식 모델의 트레이닝 장치.

청구항 18

제17항에 있어서, 상기 생성 모듈은,

상기 멀티모달 특징맵을 기초로, 상기 트레이닝 이미지의 상기 인식 대상 태스크에 따른 예측 인식 결과를 예측하는 제3 예측 서브 모듈;

상기 트레이닝 이미지의 기설정된 진실한 인식 결과, 및 상기 예측 인식 결과를 기초로, 상기 텍스트 인식 모델을 구축하는 구축 서브 모듈;을 포함하는 텍스트 인식 모델의 트레이닝 장치.

청구항 19

텍스트 인식 장치에 있어서,

인식 대상 이미지를 획득하며, 여기서, 상기 인식 대상 이미지는 텍스트를 포함하는 획득 유닛;

사전에 트레이닝된 텍스트 인식 모델을 기반으로 상기 인식 대상 이미지에 대해 텍스트 인식을 수행하여, 상기 인식 대상 이미지 중의 텍스트 내용을 획득하는 인식 유닛;을 포함하고,

여기서, 상기 텍스트 인식 모델은 제1항 내지 제4항 중 어느 한 항에 따른 텍스트 인식 모델의 트레이닝 방법을 기반으로 획득된 것인 것을 특징으로 하는 텍스트 인식 장치.

청구항 20

제19항에 있어서, 상기 인식 유닛은,

상기 텍스트 인식 모델을 기초로 상기 인식 대상 이미지의 멀티모달 특징맵을 결정하는 제1 결정 유닛;

상기 멀티모달 특징맵을 기초로 상기 인식 대상 이미지 중의 텍스트 내용을 결정하는 제2 결정 유닛;을 포함하고,

여기서, 상기 인식 대상 이미지의 멀티모달 특징맵은 상기 인식 대상 이미지의 비전 특징과 의미 특징을 나타내기 위한 것인 텍스트 인식 장치.

청구항 21

전자기기에 있어서,

적어도 하나의 프로세서; 및

상기 적어도 하나의 프로세서와 통신 연결되는 메모리;를 포함하며,

상기 메모리에 상기 적어도 하나의 프로세서에 의해 실행 가능한 명령이 저장되어 있고, 상기 명령은 상기 적어도 하나의 프로세서에 의해 실행되어, 상기 적어도 하나의 프로세서가 제1항 내지 제4항 중 어느 한 항에 따른 텍스트 인식 모델의 트레이닝 방법을 수행할 수 있도록 하거나; 또는, 상기 적어도 하나의 프로세서가 제9항에 따른 텍스트 인식 방법을 수행할 수 있도록 하는 것을 특징으로 하는 전자기기.

청구항 22

컴퓨터 명령이 저장되어 있는 비일시적 컴퓨터 판독 가능 저장매체에 있어서, 상기 컴퓨터 명령은 컴퓨터로 하여금 제1항 내지 제4항 중 어느 한 항에 따른 텍스트 인식 모델의 트레이닝 방법을 수행하도록 하거나; 또는, 상기 컴퓨터 명령은 상기 컴퓨터로 하여금 제9항에 따른 텍스트 인식 방법을 수행하도록 하는 것을 특징으로 하는 저장매체.

청구항 23

컴퓨터 판독 가능 저장매체에 저장되는 컴퓨터 프로그램에 있어서,

상기 컴퓨터 프로그램이 프로세서에 의해 실행될 때 제1항 내지 제4항 중 어느 한 항에 따른 텍스트 인식 모델의 트레이닝 방법을 구현하거나; 또는, 상기 컴퓨터 명령이 프로세서에 의해 실행될 때 제9항에 따른 텍스트 인식 방법을 구현하는 컴퓨터 프로그램.

발명의 설명

기 술 분 야

[0001] 본 출원은 인공지능 (Artificial Intelligence, AI) 기술분야에 관한 것으로서, 구체적으로 딥러닝, 컴퓨터 비전 기술분야에 관한 것이며, 광학 문자 인식(Optical Character Recognition, OCR) 등의 시나리오에 적용될 수 있으며, 특히 텍스트 인식 모델의 트레이닝 방법, 텍스트 인식 방법 및 장치에 관한 것이다.

배 경 기 술

[0002] OCR 기술은 교육, 금융, 의료, 교통 및 보험과 같은 다양한 산업 분야에서 모두 광범위한 관심을 받으며 적용되고 있다.

[0003] 종래기술에서, OCR 기술과 딥러닝 기술을 결합하여 텍스트 인식 모델을 구축하여, 텍스트 인식 모델을 기반으로 이미지에 대해 텍스트 인식을 수행할 수 있다.

[0004] 그러나 텍스트 인식 모델은 일반적으로 비전 정보에 의존하며, 비전 정보를 기반으로 이미지 중의 텍스트 내용을 분별하므로, 인식 정확도가 보다 낮은 단점이 존재한다.

발명의 내용

해결하려는 과제

[0005] 본 출원은 텍스트 인식의 신뢰성을 향상시키기 위한 텍스트 인식 모델의 트레이닝 방법, 텍스트 인식 방법 및

장치를 제공한다.

과제의 해결 수단

[0006] 본 출원의 제1 측면에 따르면, 텍스트 인식 모델의 트레이닝 방법을 제공하며, 상기 방법은,

[0007] 획득된 제1 샘플 이미지 중의 부분 이미지에 대해 마스크 예측을 수행하여, 상기 제1 샘플 이미지와 대응되는 예측 풀 이미지를 획득하는 단계;

[0008] 획득된 제2 샘플 이미지 중의 부분 텍스트에 대해 상기 마스크 예측을 수행하여, 상기 부분 텍스트와 대응되는 예측 텍스트 내용을 획득하는 단계;

[0009] 상기 예측 풀 이미지와 상기 예측 텍스트 내용을 기초로 트레이닝하여 프리 트레이닝 모델을 획득하고, 상기 프리 트레이닝 모델을 기초로 텍스트 인식 모델을 생성하며, 여기서, 상기 텍스트 인식 모델은 인식 대상 이미지에 대해 텍스트 인식을 수행하기 위한 것인 단계;를 포함한다.

[0010] 본 출원의 제2 측면에 따르면, 텍스트 인식 방법을 제공하며, 상기 방법은,

[0011] 인식 대상 이미지를 획득하며, 여기서, 상기 인식 대상 이미지는 텍스트를 포함하는 단계;

[0012] 사전에 트레이닝된 텍스트 인식 모델을 기반으로 상기 인식 대상 이미지에 대해 텍스트 인식을 수행하여, 상기 인식 대상 이미지 중의 텍스트 내용을 획득하는 단계;를 포함하고,

[0013] 여기서, 상기 텍스트 인식 모델은 제1 측면에 따른 방법을 기반으로 획득된 것이다.

[0014] 본 출원의 제3 측면에 따르면, 텍스트 인식 모델의 트레이닝 장치를 제공하며, 상기 장치는 예측 유닛, 트레이닝 유닛, 생성 유닛을 포함하고,

[0015] 상기 예측 유닛은 획득된 제1 샘플 이미지 중의 부분 이미지에 대해 마스크 예측을 수행하여, 상기 제1 샘플 이미지와 대응되는 예측 풀 이미지를 획득하고;

[0016] 상기 예측 유닛은 또한 획득된 제2 샘플 이미지 중의 부분 텍스트에 대해 상기 마스크 예측을 수행하여, 상기 부분 텍스트와 대응되는 예측 텍스트 내용을 획득하고;

[0017] 상기 트레이닝 유닛은 상기 예측 풀 이미지와 상기 예측 텍스트 내용을 기초로 트레이닝하여 프리 트레이닝 모델을 획득하고;

[0018] 상기 생성 유닛은 상기 프리 트레이닝 모델을 기초로 텍스트 인식 모델을 생성하며, 여기서, 상기 텍스트 인식 모델은 인식 대상 이미지에 대해 텍스트 인식을 수행하기 위한 것이다.

[0019] 본 출원의 제4 측면에 따르면, 텍스트 인식 장치를 제공하며, 상기 장치는,

[0020] 인식 대상 이미지를 획득하며, 여기서, 상기 인식 대상 이미지는 텍스트를 포함하는 획득 유닛;

[0021] 사전에 트레이닝된 텍스트 인식 모델을 기반으로 상기 인식 대상 이미지에 대해 텍스트 인식을 수행하여, 상기 인식 대상 이미지 중의 텍스트 내용을 획득하는 인식 유닛;을 포함하고,

[0022] 여기서, 상기 텍스트 인식 모델은 제1 측면에 따른 방법으로 획득된 것이다.

[0023] 본 출원의 제5 측면에 따르면, 전자기기를 제공하며, 상기 전자기기는,

[0024] 적어도 하나의 프로세서; 및

[0025] 상기 적어도 하나의 프로세서와 통신 연결되는 메모리;를 포함하며,

[0026] 상기 메모리에 상기 적어도 하나의 프로세서에 의해 실행 가능한 명령이 저장되어 있고, 상기 명령은 상기 적어도 하나의 프로세서에 의해 실행되어, 상기 적어도 하나의 프로세서가 제1 측면 또는 제2 측면에 따른 방법을 수행할 수 있도록 한다.

[0027] 본 출원의 제6 측면에 따르면, 컴퓨터 명령이 저장되어 있는 비일시적 컴퓨터 판독 가능 저장매체를 제공하며, 여기서, 상기 컴퓨터 명령은 컴퓨터로 하여금 제1 측면 또는 제2 측면에 따른 방법을 수행하도록 한다.

[0028] 본 출원의 제7 측면에 따르면, 컴퓨터 프로그램을 제공하며, 상기 컴퓨터 프로그램은 판독 가능 저장매체에 저장되고, 전자기기의 적어도 하나의 프로세서는 상기 판독 가능 저장매체로부터 상기 컴퓨터 프로그램을 판독할

수 있으며, 상기 적어도 하나의 프로세서는 상기 컴퓨터 프로그램을 실행하여 전자기기가 제1 측면 또는 제2 측면에 따른 방법을 수행하도록 한다.

발명의 효과

[0029] 본 출원의 마스크 예측을 기반으로 제1 샘플 이미지에 대응되는 예측 풀 이미지를 획득하고, 마스크 예측을 기반으로 제2 샘플 이미지 중의 부분 텍스트의 예측 텍스트 내용을 획득하고, 예측 풀 이미지와 예측 텍스트 내용을 결합하여 프리 트레이닝 모델을 생성하고, 프리 트레이닝 모델을 기반으로 텍스트 인식 모델을 생성하는 기술방안을 기초로, 프리 트레이닝 모델이 보다 강한 이미지 비전 추리 능력과 텍스트 의미 추리 능력을 학습하도록 하고, 이에 따라 프리 트레이닝 모델을 기반으로 생성된 텍스트 인식 모델이 텍스트 인식을 수행할 때, 텍스트 인식의 정확성과 신뢰성을 향상시킨다.

[0030] 본 부분에 기재되는 내용은 본 출원의 실시예의 핵심 또는 중요 특징을 특정하려는 목적이 아니며, 본 출원의 범위를 한정하는 것도 아니라는 점을 이해하여야 한다. 본 출원의 기타 특징은 아래의 명세서로부터 쉽게 이해할 수 있다.

도면의 간단한 설명

[0031] 첨부된 도면은 본 방안을 더 충분히 이해하도록 제공되는 것으로서, 본 출원에 대한 한정은 아니다. 여기서,

도 1은 본 출원의 제1 실시예에 따른 도면이다.

도 2는 본 출원의 제2 실시예에 따른 도면이다.

도 3은 본 출원의 제3 실시예에 따른 도면이다.

도 4는 본 출원의 제4 실시예에 따른 도면이다.

도 5는 본 출원의 제5 실시예에 따른 도면이다.

도 6은 본 출원의 제6 실시예에 따른 도면이다.

도 7은 본 출원의 제7 실시예에 따른 도면이다.

도 8은 본 출원의 제8 실시예에 따른 도면이다.

도 9는 본 출원의 제9 실시예에 따른 도면이다.

도 10은 본 출원의 실시예의 텍스트 인식 모델의 트레이닝 방법, 텍스트 인식 방법을 구현하기 위한 전자기기의 블록도이다.

발명을 실시하기 위한 구체적인 내용

[0032] 아래에서는 첨부 도면을 결합하여 본 출원의 예시적인 실시예에 대해 설명하며, 이해를 돕기 위하여 본 출원의 실시예의 다양한 세부 사항을 포함하며, 이들은 단지 예시적인 것으로만 간주하여야 한다. 따라서, 본 분야의 통상적인 지식을 가진 자라면, 여기에 기재되는 실시예에 대해 다양한 변경과 수정을 가할 수 있으며, 이는 본 출원의 범위와 정신을 벗어나지 않는다는 점을 이해하여야 한다. 마찬가지로, 명확성과 간결성을 위하여, 아래의 기재에서 공지 기능과 구조에 대한 설명을 생략한다.

[0033] OCR 기술과 딥러닝을 결합하여 텍스트 인식 모델을 구축할 때, "모듈 분리"의 방식을 사용하여 구현할 수 있고, "엔드 대 엔드 모델"의 방식을 사용할 수도 있다.

[0034] 예시적으로, "모듈 분리"의 방식이란, 텍스트 검출 모듈, 정보 추출 모듈, 텍스트 인식 모듈을 구축하고, 이 세 모듈을 결합하여 텍스트 인식 모델을 구축하는 것을 가리킨다.

[0035] "모듈 분리"의 방식을 사용할 경우, 사전에 각 모듈을 구축하고, 각 모듈을 결합하여야 하며, 과정이 상대적으로 번잡하고, 효율이 상대적으로 낮으며, 정확성이 누적 겹치므로, 상기 방식을 기반으로 구축된 텍스트 인식 모델의 인식 정확성이 보다 낮은 단점을 초래한다.

[0036] 예시적으로, "엔드 대 엔드 모델"의 방식이란, 입력단으로부터 출력단까지 하나의 예측 결과를 획득하며, 예를 들어 입력단에 이미지를 입력하고, 출력단에서 이미지에 대한 예측 텍스트 내용을 획득한다.

[0037] 하지만, "엔드 대 엔드 모델"의 방식을 사용하면 데이터 라벨링을 수행하여야 하며, 예를 들어 이미지의 진실한 텍스트 내용에 대해 라벨링하고, 트레이닝을 위해 제공되는 데이터가 비교적 유효해야 하므로, 트레이닝된 텍스트 인식 모델의 신뢰성이 보다 낮은 단점을 초래한다.

[0038] 상술한 어느 하나의 방법을 기반으로 트레이닝하여 획득된 텍스트 인식 모델은 일반적으로 두 가지 유형의 판단만 수행하며, 서로 다른 수직 유형마다 다른 클래스 필드 수요가 있을 때, 텍스트 인식 모델, 특히 분류된 채널 수량을 재설계하여야 하고, 텍스트 인식 모델도 다시 트레이닝하여야 하며, 멀티플렉싱될 수 없다.

[0039] 예를 들어, OCR 기술 중의 이미지 문자 검출 모델(EAST), 분할된 문자 검출 모델(DB), 및 텍스트 검출기(LOMO) 등은 일반적으로 텍스트(text) 유형과 비텍스트 유형(non-text)과 같은 두 유형의 판단에만 사용될 수 있다. 만약 특정 구체적인 수직 유형에서의 사용자가 관심하는 필드 인식 수요를 해결하여야 할 경우, 분류 클래스 수량을 증가시켜야 한다.

[0040] 일부 실시예에서, 클래스 검출 확장 방식을 통해, 트레이닝하여 새로운 텍스트 인식 모델을 획득할 수 있으며, 예를 들어 기존 텍스트 인식 모델의 기초 상에서, 별도의 언어 모델을 추가하여 필드 분류할 수 있다.

[0041] 예를 들어, 만약 텍스트 인식 모델이 OCR 기술 중의 엔드 대 엔드 텍스트 검출과 인식(FOTS) 및 텍스트 검출과 인식 모델(Mask Text Spotter)이면, 예컨대 양방향 인코더 표시(Bidirectional Encoder Representation from Transformers, BERT)와 같은 별도의 언어 모델을 추가하여, 새로운 텍스트 인식 모델을 획득하여야 하며, 별도의 언어 모델을 추가하므로, 별도의 트레이닝을 추가하여야 하며, 이에 따라 트레이닝 코스트가 보다 높고, 효율이 보다 낮은 단점을 초래한다.

[0042] 상술한 기술 문제점 중 적어도 하나를 방지하기 위하여, 본 출원의 발명자는 창조적 노동을 거쳐, 본 출원의 발명 사상에 이르게 되었다. 구체적으로, "엔드 대 엔드 모델"의 방식을 사용하여 트레이닝하여 프리 트레이닝 모델을 획득하고, 즉 모델 베이스에 대해 엔드 대 엔드의 프리 트레이닝을 수행하며, 비전 차원과 의미 차원을 결합하여 프리 트레이닝을 수행하고, 프리 트레이닝된 베이스를 기반으로 텍스트 인식 모델을 생성한다.

[0043] 상술한 발명 사상을 기반으로, 본 출원은 텍스트 인식 모델의 트레이닝 방법, 텍스트 인식 방법 및 장치를 제공하며, 인공지능 기술분야에 관한 것으로서, 구체적으로 딥러닝, 컴퓨터 비전 기술분야에 관한 것이며, OCR 등의 시나리오에 적용되어, 텍스트 인식 모델의 텍스트 인식에 대한 신뢰성을 향상시킬 수 있다.

[0044] 도 1은 본 출원의 제1 실시예에 따른 도면이다. 도 1에 도시된 바와 같이, 본 실시예에서 제공하는 텍스트 인식 모델의 트레이닝 방법은 아래의 단계(S101 - S1013)를 포함한다.

[0045] S101: 획득된 제1 샘플 이미지 중의 부분 이미지에 대해 마스크 예측을 수행하여, 제1 샘플 이미지와 대응되는 예측 풀 이미지를 획득한다.

[0046] 예시적으로, 본 실시예의 수행 주체는 텍스트 인식 모델의 트레이닝 장치(이하, '트레이닝 장치'로 약칭)일 수 있고, 트레이닝 장치는 서버(예컨대, 클라우드 서버, 또는 로컬 서버, 또는 서버 클러스터)일 수 있고, 단말기일 수도 있고, 컴퓨터일 수도 있고, 프로세서일 수도 있고, 칩 등일 수도 있으며, 본 실시예에서는 한정하지 않는다.

[0047] 여기서, 마스크 예측이란, 부분 이미지 또는 텍스트 등에 대해 마스크(mask) 처리(또는, '가림 처리'라고도 함)를 수행하고, mask 처리 전, 즉 가림 처리 전의 이미지 또는 텍스트 등의 완전한 이미지 또는 텍스트 등으로 복원하는 것을 가리킨다.

[0048] 상응하게, 상기 단계에 대해서는, 텍스트를 포함하는 제1 샘플 이미지를 획득하고, 제1 샘플 이미지의 부분 이미지에 대해 mask 처리를 수행하고, mask 처리 후의 이미지를 기반으로 완전한 제1 샘플 이미지(즉, 예측 풀 이미지)를 예측하는 것으로 이해할 수 있다.

[0049] 다시 말하면, 상기 단계에 대해서는, 이미지 재구성 태스크(mask image modelling)로서, 마스크 예측의 방식을 결합하여 제1 샘플 이미지에 대해 이미지 재구성을 수행하는 것으로 이해할 수 있다.

[0050] S102: 획득된 제2 샘플 이미지 중의 부분 텍스트에 대해 마스크 예측을 수행하여, 부분 텍스트와 대응되는 예측 텍스트 내용을 획득한다.

[0051] 상술한 분석을 결합하면, 상기 단계에 대해서는, 텍스트를 포함하는 제2 샘플 이미지를 획득하고, 제2 샘플 이미지 중의 부분 텍스트에 대해 mask 처리를 수행하고, mask 처리 후의 텍스트를 기반으로 mask 처리된 부분 텍

스트의 텍스트 내용(즉, 예측 텍스트 내용)을 예측하는 것으로 이해할 수 있다.

[0052] 다시 말하면, 상기 단계에 대해서는 텍스트 재구성 태스크(mask OCR modelling)로서, 마스크 예측의 방식을 결합하여 제2 샘플 이미지에 대해 텍스트 재구성을 수행하며, 구체적으로 제2 샘플 이미지 중의 부분 텍스트에 대해 재구성하는 것으로 이해할 수 있다.

[0053] 특별히 설명하여야 할 점은, 제1 샘플 이미지와 제2 샘플 이미지는 동일한 이미지일 수 있고, 상이한 이미지일 수도 있으며, 본 실시예에서는 한정하지 않는다.

[0054] S103: 예측 풀 이미지와 예측 텍스트 내용을 기초로 트레이닝하여 프리 트레이닝 모델을 획득하고, 프리 트레이닝 모델을 기초로 텍스트 인식 모델을 생성한다.

[0055] 여기서, 텍스트 인식 모델은 인식 대상 이미지에 대해 텍스트 인식을 수행하기 위한 것이다.

[0056] 프리 트레이닝 모델에 대해서는, 텍스트 인식 모델의 베이스로 이해하거나, 또는, 텍스트 인식 모델의 은닉층으로 이해할 수 있다.

[0057] 상술한 분석을 결합하여 알 수 있는 바와 같이, 프리 트레이닝 모델은 이미지 재구성과 텍스트 재구성을 기반으로 트레이닝하여 획득된 것으로서, 프리 트레이닝 모델이 보다 강한 이미지 비전 추리 능력과 텍스트 의미 추리 능력을 학습하도록 하고, 프리 트레이닝 모델을 기반으로 생성된 텍스트 인식 모델이 보다 강한 정확성과 신뢰성을 갖도록 한다.

[0058] 본 실시예에서, 엔드 대 엔드의 모델 트레이닝을 구현할 수 있으며, 즉 바로 제1 샘플 이미지와 제2 샘플 이미지를 기반으로 각각에 대응되는 예측 결과를 출력할 수 있으며, 예를 들어 제1 샘플 이미지에 대응되는 예측 결과는 예측 풀 이미지이고, 제2 샘플 이미지에 대응되는 예측 결과는 예측 텍스트 내용이며, 예컨대 인공 또는 OCR 기술을 기반으로 제2 샘플 이미지에 대해 텍스트 검출을 수행하여 텍스트를 얻는 단계와 같은 기타 단계를 추가할 필요가 없으므로, 트레이닝 효율을 향상시키고, 트레이닝 리소스와 코스트를 절약한다.

[0059] 상술한 분석을 기반으로 알 수 있는 바와 같이, 본 출원의 실시예는 텍스트 인식 모델의 트레이닝 방법을 제공하며, 획득된 제1 샘플 이미지 중의 부분 이미지에 대해 마스크 예측을 수행하여, 제1 샘플 이미지와 대응되는 예측 풀 이미지를 획득하고, 획득된 제2 샘플 이미지 중의 부분 텍스트에 대해 마스크 예측을 수행하여, 부분 텍스트와 대응되는 예측 텍스트 내용을 획득하고, 예측 풀 이미지와 예측 텍스트 내용을 기초로 트레이닝하여 프리 트레이닝 모델을 획득하고, 프리 트레이닝 모델을 기초로 텍스트 인식 모델을 생성하며, 여기서, 텍스트 인식 모델은 인식 대상 이미지에 대해 텍스트 인식을 수행하기 위한 것인 것을 포함하며, 본 실시예에서, 마스크 예측을 기반으로 제1 샘플 이미지에 대응되는 예측 풀 이미지를 획득하고, 마스크 예측을 기반으로 제2 샘플 이미지 중의 부분 텍스트의 예측 텍스트 내용을 획득하고, 예측 풀 이미지와 예측 텍스트 내용을 결합하여 프리 트레이닝 모델을 생성하고, 프리 트레이닝 모델을 기반으로 텍스트 인식 모델을 생성하는 기술특징을 통해, 프리 트레이닝 모델이 보다 강한 이미지 비전 추리 능력과 텍스트 의미 추리 능력을 학습하도록 하고, 이에 따라 프리 트레이닝 모델을 기반으로 생성된 텍스트 인식 모델이 텍스트 인식을 수행할 때, 텍스트 인식의 정확성과 신뢰성을 향상시킨다.

[0060] 도 2는 본 출원의 제2 실시예에 따른 도면이다. 도 2에 도시된 바와 같이, 본 실시예에서 제공하는 텍스트 인식 모델의 트레이닝 방법은 단계(S201 - S203)를 포함한다.

[0061] S201: 타겟 대상을 획득한다.

[0062] 여기서, 타겟 대상은 제1 샘플 이미지와 제2 샘플 이미지를 포함한다.

[0063] 이해하여야 할 점은, 번잡한 설명을 피하기 위하여, 본 실시예에서는 본 실시예 중 상술한 실시예와 동일한 기술특징에 대한 반복되는 설명을 생략한다.

[0064] S202: 타겟 대상 중의 부분 대상을 랜덤으로 가리우고, 타겟 대상 중 가리워지지 않은 대상을 기초로, 타겟 대상 중 가리워진 부분 대상에 대해 예측하여, 예측 결과를 획득한다.

[0065] 여기서, 만약 타겟 대상이 제1 샘플 이미지이면, 타겟 대상 중의 부분 대상은 부분 이미지이고, 예측 결과는 예측 풀 이미지이다.

[0066] 만약 타겟 대상이 제2 샘플 이미지이면, 타겟 대상 중의 부분 대상은 부분 텍스트이고, 예측 결과는 예측 텍스트 내용이다.

[0067] 일부 실시예에서, 타겟 대상 중 가리워지지 않은 대상을 기초로, 타겟 대상 중 가리워진 부분 대상에 대해 예측하여, 예측 결과를 획득하는 것은 아래의 단계를 포함한다.

[0068] 제1 단계: 타겟 대상 중 가리워지지 않은 대상에 대응되는 대상 특징을 추출하여, 제1 대상 특징을 획득한다.

[0069] 제2 단계: 제1 대상 특징을 기초로, 타겟 대상 중 가리워진 부분 대상에 대해 예측하여, 예측 결과를 획득한다.

[0070] 여기서, 만약 타겟 대상이 제1 샘플 이미지이면, 제1 대상 특징은 제1 비전 특징이다. 만약 타겟 대상이 제2 샘플 이미지이면, 제1 대상 특징은 제1 의미 특징이다.

[0071] S203: 예측 풀 이미지와 예측 텍스트 내용을 기초로 트레이닝하여 프리 트레이닝 모델을 획득하고, 프리 트레이닝 모델을 기초로 텍스트 인식 모델을 생성한다.

[0072] 여기서, 텍스트 인식 모델은 인식 대상 이미지에 대해 텍스트 인식하기 위한 것이다.

[0073] 읽는 자가 본 출원의 구현 원리를 더욱 충분히 이해하도록, 이하 도 3을 참조하여 상술한 실시예(도 1과 도 2에 도시된 실시예)에 대해 상세하게 설명한다.

[0074] 도 3은 본 출원의 제3 실시예에 따른 도면이다. 도 3에 도시된 바와 같이, 본 실시예에서 제공하는 텍스트 인식 모델의 트레이닝 방법은 아래의 단계(S301 - S307)를 포함한다.

[0075] S301: 제1 샘플 이미지를 획득한다.

[0076] 마찬가지로, 번잡한 기재를 피하기 위하여, 본 실시예에서는 본 실시예 중 상술한 실시예와 동일한 기술특징에 대한 반복되는 설명을 생략한다.

[0077] S302: 제1 샘플 이미지 중의 부분 이미지를 랜덤으로 가리운다.

[0078] 이해하여야 할 점은, 네트워크 모델의 트레이닝은 일반적으로 반복 트레이닝하는 과정이며, 본 실시예에서, 매번의 반복 트레이닝은 모두 랜덤으로 제1 샘플 이미지의 부분 이미지를 가리우므로, 제1 샘플 이미지의 수량은 하나일 수 있고, 물론, 제1 샘플 이미지의 수량은 복수일 수도 있으며, 본 실시예에서는 한정하지 않는다.

[0079] S303: 제1 샘플 이미지 중 가리워지지 않은 이미지를 기초로, 제1 샘플 이미지 중 가리워진 부분 이미지에 대해 예측하여, 예측 풀 이미지를 획득한다.

[0080] 예시적으로, 제1 샘플 이미지에 대해 랜덤으로 가리운 후, 제1 샘플 이미지 중의 부분 이미지는 가리워지고, 다른 부분 이미지는 가리워지지 않으므로, 가리워지지 않은 이미지를 기반으로 완전한 제1 샘플 이미지(즉, 예측 풀 이미지)를 결정할 수 있다.

[0081] 본 실시예에서, "랜덤으로 가림 + 예측"의 방식을 결합하여, 예측 풀 이미지를 결정함으로써, 트레이닝 과정에서의 불가 결정성을 증가시킬 수 있으며, 이에 따라 트레이닝하여 획득된 프리 트레이닝 모델이 완전한 이미지를 복원하는 신뢰성을 향상시킨다.

[0082] 여기서, S302 - S303은 마스크 자동 인코더(MAE)를 기반으로 구현할 수 있다. 다시 말하면, 제1 샘플 이미지를 마스크 자동 인코더로 입력하여, 예측 풀 이미지를 출력할 수 있다.

[0083] 일부 실시예에서, S303은 아래의 단계들을 포함할 수 있다.

[0084] 제1 단계: 제1 샘플 이미지 중 가리워지지 않은 이미지에 대응되는 비전 특징을 추출하여, 제1 비전 특징을 획득한다.

[0085] 여기서, 비전 특징은 텍스처 특징, 윤곽 특징, 칼라 특징, 및 형상 특징과 같은 것들을 포함하며, 여기서는 일일이 나열하지 않는다.

[0086] 상응하게, 제1 비전 특징이란, 제1 샘플 이미지 중 가리워지지 않은 이미지에 대응되는 텍스처 특징, 윤곽 특징, 칼라 특징, 및 형상 특징과 같은 것들을 가리킨다.

[0087] 제2 단계: 제1 비전 특징을 기초로, 제1 샘플 이미지 중 가리워진 부분 이미지에 대해 예측하여, 예측 풀 이미지를 획득한다.

[0088] 본 실시예에서, 가리워지지 않은 이미지에 대응되는 텍스처 특징, 윤곽 특징, 칼라 특징, 및 형상 특징 등의 비전 특징을 결합하여, 예측 풀 이미지를 획득하는 것은, 비전 콘텍스트를 기반으로 예측 풀 이미지를 획득하여, 트레이닝하여 시각적 큐의 콘텍스트 지식 학습을 완성할 수 있는 프리 트레이닝 모델을 획득하는 것에

해당된다.

[0089] 일부 실시예에서, 제2 단계는 아래의 서브 단계들을 포함할 수 있다.

[0090] 제1 서브 단계: 제1 비전 특징을 기초로, 제1 샘플 이미지 중 가리워진 부분 이미지에 대응되는 비전 특징을 예측하여, 제2 비전 특징을 획득한다.

[0091] 예시적으로, 상술한 분석을 참조하면, 상기 서브 단계에 대해서는, 가리워지지 않은 이미지에 대응되는 예컨대 텍스처 특징, 윤곽 특징, 칼라 특징, 및 형상 특징 등의 비전 특징을 기초로, 예측하여 가리워진 부분 이미지에 대응되는 예컨대 텍스처 특징, 윤곽 특징, 칼라 특징, 및 형상 특징 등의 비전 특징을 획득하는 것으로 이해할 수 있다.

[0092] 제2 서브 단계: 제2 비전 특징을 기초로, 제1 샘플 이미지 중 가리워진 부분 이미지를 결정한다.

[0093] 예시적으로, 가리워진 부분 이미지에 대응되는 예컨대 텍스처 특징, 윤곽 특징, 칼라 특징, 및 형상 특징 등의 비전 특징을 획득한 후, 상기 비전 특징을 기반으로 가리워진 부분 이미지를 보충 및 복구할 수 있다.

[0094] 제3 서브 단계: 제1 샘플 이미지 중 가리워지지 않은 이미지, 및 결정된 제1 샘플 이미지 중 가리워진 부분 이미지를 기초로, 예측 풀 이미지를 생성한다.

[0095] 상술한 분석을 참조하면, 가리워진 부분 이미지에 대해 보충 및 복구한 후, 바로 가리워진 부분 이미지가 복원되고, 가리워지지 않은 부분 이미지와 복원된 가리워진 부분 이미지에 대해 스플라이싱하여, 예측 풀 이미지를 획득하고, 즉 제1 샘플 이미지를 복원하여, 예측 풀 이미지와 제1 샘플 이미지가 고도로 일치하도록 함으로써, 예측 풀 이미지의 정확성과 신뢰성을 향상시킨다.

[0096] S304: 제2 샘플 이미지를 획득한다.

[0097] 상술한 분석을 참조하여 알 수 있는 바와 같이, 제1 샘플 이미지와 제2 샘플 이미지는 동일한 이미지일 수 있고, 상응하게, 만약 제1 샘플 이미지와 제2 샘플 이미지가 동일한 이미지이면, 해당 단계를 생략할 수 있다.

[0098] S305: 제2 샘플 이미지 중의 부분 텍스트를 랜덤으로 가리운다.

[0099] 마찬가지로, 네트워크 모델의 트레이닝은 일반적으로 반복 트레이닝하는 과정이며, 본 실시예에서, 매번의 반복 트레이닝은 모두 제2 샘플 이미지의 부분 텍스트를 랜덤으로 가리우는 것이므로, 제2 샘플 이미지의 수량은 하나일 수 있고, 물론, 제2 샘플 이미지의 수량은 복수일 수도 있으며, 본 실시예에서는 한정하지 않는다.

[0100] 예를 들어, 제2 샘플 이미지 중의 부분 단어, 또는 부분 구절 등을 랜덤으로 가리울 수 있다.

[0101] S306: 제2 샘플 이미지 중 가리워지지 않은 텍스트를 기초로, 제2 샘플 이미지 중 가리워진 부분 텍스트에 대해 예측하여, 예측 텍스트 내용을 획득한다.

[0102] 예시적으로, 제2 샘플 이미지에 대해 랜덤으로 가리운 후, 제2 샘플 이미지 중의 부분 텍스트는 가리워지고, 다른 부분 텍스트는 가리워지지 않으며, 이때 가리워지지 않은 텍스트를 기반으로 가리워진 부분 텍스트의 텍스트 내용(즉, 예측 텍스트 내용)을 결정할 수 있다.

[0103] 본 실시예에서, "랜덤으로 가림 + 예측"의 방식을 결합하여, 텍스트 내용을 결정함으로써, 트레이닝 과정에서의 불가 결정성을 증가시킬 수 있으므로, 트레이닝하여 획득된 프리 트레이닝 모델이 완전한 이미지를 복원하는 신뢰성을 향상시킨다.

[0104] 여기서, S305 - S306은 마스크 언어 모델(Masked Language Model, MLM)을 기반으로 구현할 수 있다. 다시 말하면, 제2 샘플 이미지를 마스크 언어 모델로 입력하여, 예측 텍스트 내용을 출력할 수 있다.

[0105] 일부 실시예에서, S306는 아래의 단계들을 포함할 수 있다.

[0106] 제1 단계: 제2 샘플 이미지 중 가리워지지 않은 텍스트에 대응되는 의미 특징을 추출하여, 제1 의미 특징을 획득한다.

[0107] 여기서, 의미 특징이란 각 문자열 사이의 논리 관계의 특징을 가리킨다. 상응하게, 제1 의미 특징에 대해서는, 가리워지지 않은 텍스트에 포함된 각 문자열 사이의 논리 관계의 특징으로 이해할 수 있고, 가리워지지 않은 텍스트 중의 각 문자(글 및/또는 단어) 사이의 관련 관계의 특징으로 이해할 수도 있다.

[0108] 제2 단계: 제1 의미 특징을 기초로, 제2 샘플 이미지 중 가리워진 부분 텍스트에 대해 예측하여, 예측 텍스트

내용을 획득한다.

[0109] 본 실시예에서, 가리워지지 않은 텍스트에 대응되는 각 문자열 사이의 논리 관계 등의 의미 특징을 결합하여, 예측 텍스트 내용을 획득하는 것은, 의미 콘텍스트를 기반으로 예측 텍스트 내용을 획득하여, 트레이닝하여 의미 큐의 콘텍스트 지식 학습을 완성할 수 있는 프리 트레이닝 모델을 획득하는 것에 해당된다.

[0110] 일부 실시예에서, 제2 단계는 아래의 서브 단계들을 포함할 수 있다.

[0111] 제1 서브 단계: 제1 의미 특징을 기초로, 제2 샘플 이미지 중 가리워진 부분 텍스트에 대응되는 의미 특징을 예측하여, 제2 의미 특징을 획득한다.

[0112] 예시적으로, 상술한 분석을 참조하면, 상기 서브 단계에 대해서는, 가리워지지 않은 텍스트에 대응되는 예컨대 각 문자열 사이의 논리 관계의 특징 등의 의미 특징을 기초로, 예측하여 가리워진 부분 텍스트에 대응되는 예컨대 각 문자열 사이의 논리 관계의 특징 등의 의미 특징을 획득하는 것으로 이해할 수 있다.

[0113] 제2 서브 단계: 제2 의미 특징을 기초로, 예측 텍스트 내용을 생성한다.

[0114] 예시적으로, 가리워지지 않은 텍스트에 대응되는 예컨대 각 문자열 사이의 논리 관계의 특징 등의 의미 특징을 획득한 후, 상기 의미 특징을 기반으로 가리워진 부분 텍스트의 의미 특징을 보충 및 복구할 수 있다.

[0115] 상술한 분석을 참조하면, 가리워진 부분 텍스트의 의미 특징에 대해 보충 및 복구한 후, 바로 가리워진 부분 텍스트의 의미 특징이 복원되고, 상기 의미 특징에 대응되는 텍스트 내용(즉, 예측 텍스트 내용)을 결정하여, 예측 텍스트 내용과 가리워진 부분 텍스트의 텍스트 내용이 고도로 일치되도록 함으로써, 예측 텍스트 내용의 정확성과 신뢰성을 향상시킨다.

[0116] S307: 예측 풀 이미지와 예측 텍스트 내용을 기초로 트레이닝하여 프리 트레이닝 모델을 획득하고, 프리 트레이닝 모델을 기초로 텍스트 인식 모델을 생성한다.

[0117] 여기서, 텍스트 인식 모델은 인식 대상 이미지에 대해 텍스트 인식을 수행하기 위한 것이다.

[0118] 도 4는 본 출원의 제4 실시예에 따른 도면이다. 도 4에 도시된 바와 같이, 본 실시예에서 제공하는 텍스트 인식 모델의 트레이닝 방법은 아래의 단계(S401 - S405)를 포함한다.

[0119] S401: 획득된 제1 샘플 이미지 중의 부분 이미지에 대해 마스크 예측을 수행하여, 제1 샘플 이미지와 대응되는 예측 풀 이미지를 획득한다.

[0120] 마찬가지로, 번잡한 기재를 피하기 위하여, 본 실시예에서는 본 실시예 중 상술한 실시예와 동일한 기술특징에 대한 반복되는 설명을 생략한다.

[0121] S402: 획득된 제2 샘플 이미지 중의 부분 텍스트에 대해 마스크 예측을 수행하여, 부분 텍스트와 대응되는 예측 텍스트 내용을 획득한다.

[0122] S403: 예측 풀 이미지와 예측 텍스트 내용을 기초로 트레이닝하여 프리 트레이닝 모델을 획득한다.

[0123] 예시적으로, 예측 풀 이미지와 예측 텍스트 내용을 기반으로, 기초 네트워크 모델에 대해 트레이닝하여, 프리 트레이닝 모델을 획득할 수 있다.

[0124] 예를 들어, 예측 풀 이미지와 예측 텍스트 내용을 기반으로, 기초 네트워크 모델의 모델 파라미터에 대해 조정하여 프리 트레이닝 모델을 획득할 수 있다.

[0125] 여기서, 기초 네트워크 모델은 비전 변환기(Vision Transformer, ViT)일 수 있고, 뉴럴 네트워크 모델(Backbone), 예컨대 컨벌루션 뉴럴 네트워크 모델(CNN)일 수도 있고, 기타 네트워크 모델일 수도 있으며, 본 실시예에서는 한정하지 않는다.

[0126] S404: 인식 대상 태스크와 트레이닝 이미지를 획득한다.

[0127] 여기서, 트레이닝 이미지는 텍스트를 포함한다.

[0128] 여기서, 인식 대상 태스크는 텍스트 인식 모델의 인식 수요를 기반으로 결정된 것일 수 있고, 예컨대 인식 대상 태스크는 문자 검출 태스크일 수 있고, 텍스트 인식 태스크일 수도 있고, 필드 분류 태스크일 수도 있고, 기타 인식 태스크일 수도 있고, 여기서는 일일이 나열하지 않는다.

[0129] S405: 인식 대상 태스크와 트레이닝 이미지를 기초로, 프리 트레이닝 모델에 대해 트레이닝하여, 텍스트 인식

모델을 획득한다.

[0130] 여기서, 텍스트 인식 모델은 인식 대상 이미지에 대해 텍스트 인식을 수행하기 위한 것이다.

[0131] 상술한 분석을 참조하여 알 수 있는 바와 같이, 프리 트레이닝 모델은 비전 큐의 콘텍스트 지식 학습을 완성하기 위한 모델을 구비할 뿐만 아니라, 의미 큐 콘텍스트 지식 학습을 위한 모델도 구비하며, 즉 프리 트레이닝 모델은 멀티모달 특징 추출 베이스이므로, 프리 트레이닝 모델을 결합하여 트레이닝하여 획득된 텍스트 인식 모델은 비전 큐 기반 콘텍스트 지식 인식 능력을 구비할 뿐만 아니라, 의미 큐 기반 콘텍스트 지식 인식 능력도 구비한다.

[0132] 또한 인식 대상 태스크를 결합하여 프리 트레이닝 모델에 대해 트레이닝함으로써, 서로 다른 인식 수요에 기반하여 트레이닝하여 이에 대응되는 텍스트 인식 모델을 획득하여, 트레이닝하여 획득되는 텍스트 인식 모델의 유연성과 다양성을 향상시킬 수 있고, 다양한 인식 시나리오에 적용될 수 있으며, 상이한 인식 수요를 만족시킨다.

[0133] 일부 실시예에서, 프리 트레이닝 모델(즉 멀티모달 특징 추출 베이스)를 텍스트 검출 네트워크 모델(Efficient and Accuracy Scene Text, EAST), 분할 기반 문자 검출 네트워크(Differentiable Binarization, DB), 텍스트 검출 네트워크(Look More Than Once, LOMO) 등에 로딩하여, 텍스트 인식 모델의 문자 검출 태스크를 구현할 수 있고; 또 예를 들어, 프리 트레이닝 모델을 컨벌루션 순환 뉴럴 네트워크(Convolutional Recurrent Neural Network, CRNN)로 로딩할 수 있으며, 여기서, 컨벌루션 순환 뉴럴 네트워크는 연결성 시간 분류(Connectionist Temporal Classification, CTC) 복호화 방식을 사용할 수 있고, 주의 매커니즘(Attention) 복호화 방식을 사용할 수도 있고, 변환기(transformer) 복호화 방법 등을 사용할 수도 있으며, 이로부터 텍스트 인식 모델의 텍스트 인식 태스크를 구현하고; 또 예를 들어, 프리 트레이닝 모델을 풀 연결 네트워크 모델(Fully Connected, FC), 또는 컨벌루션 뉴럴 네트워크 모델(Convolutional Neural Networks, CNN)로 로딩하여, 텍스트 인식 모델의 필드 분류 태스크를 구현할 수 있다.

[0134] 일부 실시예에서, S405는 아래의 단계들을 포함할 수 있다.

[0135] 제1 단계: 트레이닝 이미지를 프리 트레이닝 모델로 입력하여, 트레이닝 이미지에 대응되는 멀티모달 특징맵(Multi-modal Feature Maps)을 획득한다.

[0136] 상술한 분석을 참조하면, 멀티모달 특징맵은 트레이닝 이미지의 복수의 차원의 특징, 예컨대 비전 차원의 특징과 의미 차원의 특징을 나타내기 위한 것이다. 예컨대 멀티모달 특징맵은 트레이닝 이미지에 대응되는 이미지 특징과 의미 특징을 나타낼 수 있다.

[0137] 일부 실시예에서, 멀티모달 특징맵은 (d*h*w)으로 나타낼 수 있고, 여기서, d는 특징 채널 수량을 나타내고, h와 w는 멀티모달 특징맵의 높이와 폭을 나타낸다.

[0138] 제2 단계: 인식 대상 태스크와 멀티모달 특징맵을 기초로, 텍스트 인식 모델을 생성한다.

[0139] 본 실시예에서, 멀티모달 특징맵은 복수의 차원으로부터 트레이닝 이미지의 특징에 대해 나타낼 수 있는 바, 트레이닝 이미지의 비전 특징을 나타낼 수 있을 뿐만 아니라, 트레이닝 이미지의 의미 특징을 나타낼 수도 있으며, 나타내는 비전 특징과 의미 특징은 보다 강한 신뢰성과 전면성을 가지므로, 멀티모달 특징맵을 결합하여 생성된 텍스트 인식 모델은 보다 강한 신뢰성과 정확성을 갖는다.

[0140] 일부 실시예에서, 제2 단계는 아래와 같은 서브 단계들을 포함할 수 있다.

[0141] 제1 서브 단계: 멀티모달 특징맵을 기초로, 트레이닝 이미지의 인식 대상 태스크에 따른 예측 인식 결과를 획득한다.

[0142] 예시적으로, 멀티모달 특징맵을 컨벌루션 순환 뉴럴 네트워크로 입력하여, 예측 인식 결과(예측 텍스트 결과)를 획득할 수 있다.

[0143] 제2 서브 단계: 트레이닝 이미지의 기설정된 진실한 인식 결과, 및 예측 인식 결과를 기초로, 텍스트 인식 모델을 구축한다.

[0144] 여기서, 진실한 인식 결과는 사전에 트레이닝 이미지에 대해 라벨링하여 얻어진 것일 수 있고, 본 실시예에서는 라벨링하는 방식에 대해 한정하지 않으며, 예컨대 인공 라벨링 방식일 수 있고, 자동 라벨링 방식일 수도 있다.

[0145] 예시적으로, 진실한 인식 결과와 예측 인식 결과 사이의 손실값을 연산할 수 있고, 만약 손실값이 기설정된 손

실 임계값보다 크면(또는 같으면), 트레이닝을 반복으로 수행하고, 반대로, 만약 손실값이 기설정된 손실 임계값보다 작으면, 텍스트 인식 모델의 구축을 완성하고, 또는, 만약 반복 횟수가 기설정된 반복 횟수에 도달하면, 텍스트 인식 모델의 구축을 완성한다.

[0146] 예를 들어, 만약 기차 승차권에 대해 텍스트 인식을 수행하기 위한 텍스트 인식 모델을 트레이닝해야 할 경우, 트레이닝 이미지는 기차 승차권 이미지이고, 기차표 이미지를 프리 트레이닝 모델로 입력하여, 기차 승차권 이미지의 멀티모달 특징맵을 출력하고, 멀티모달 특징맵을 예컨대 컨벌루션 순환 뉴럴 네트워크에 입력하고, 예컨대 기차 승차권 이미지 중의 "날짜, 기차 번호, 좌석 번호" 등의 예측 인식 결과를 출력하고, 상기 예측 인식 결과와 사전에 라벨링된 "날짜, 기차 번호, 좌석 번호"(즉, 진실한 인식 결과)를 비교하여, 트레이닝하여 텍스트 인식 모델을 획득하며, 트레이닝하여 획득된 텍스트 인식 모델은 인식 대상 승차권 이미지 중의 "날짜, 기차 번호, 좌석 번호" 텍스트 내용의 인식에 사용될 수 있다.

[0147] 도 5는 본 출원의 제5 실시예에 따른 도면이다. 도 5에 도시된 바와 같이, 본 실시예에서 제공하는 텍스트 인식 방법은 아래의 단계(S501 - S502)를 포함한다.

[0148] S501: 인식 대상 이미지를 획득한다.

[0149] 여기서, 인식 대상 이미지는 텍스트를 포함한다.

[0150] 예시적으로, 본 실시예의 수행 주체는 텍스트 인식 장치일 수 있고, 텍스트 인식 장치는 트레이닝 장치와 동일한 장치일 수 있고, 트레이닝 장치와 다른 장치일 수도 있으며, 본 실시예에서는 한정하지 않는다.

[0151] S502: 사전에 트레이닝된 텍스트 인식 모델을 기반으로 인식 대상 이미지에 대해 텍스트 인식을 수행하여, 인식 대상 이미지 중의 텍스트 내용을 획득한다.

[0152] 여기서, 텍스트 인식 모델은 상술한 어느 하나의 실시예에 따른 텍스트 인식 모델의 트레이닝 방법을 기반으로 획득된 것이다.

[0153] 일부 실시예에서, S502는 아래의 단계를 포함할 수 있다.

[0154] 제1 단계: 텍스트 인식 모델을 기초로 인식 대상 이미지의 멀티모달 특징맵을 결정한다.

[0155] 제2 단계: 멀티모달 특징맵을 기초로 인식 대상 이미지 중의 텍스트 내용을 결정한다.

[0156] 여기서, 인식 대상 이미지의 멀티모달 특징맵은 인식 대상 이미지의 비전 특징과 의미 특징을 나타내기 위한 것이다.

[0157] 예시적으로, 상술한 분석을 참조하면, 텍스트 인식 모델은 프리 트레이닝 모델을 포함하고, 만약 텍스트 인식 모델이 프리 트레이닝 모델을 컨벌루션 순환 뉴럴 네트워크에 로딩하여 트레이닝하여 획득된 것이면, 즉 텍스트 인식 모델이 컨벌루션 순환 뉴럴 네트워크를 더 포함하며, 본 실시예는 아래와 같이 이해할 수 있다.

[0158] 인식 대상 이미지를 프리 트레이닝 모델로 입력하여, 멀티모달 특징맵을 출력하고, 멀티모달 특징맵을 컨벌루션 순환 뉴럴 네트워크로 입력하여, 인식 대상 이미지 중의 텍스트 내용을 출력한다.

[0159] 도 6은 본 출원의 제6 실시예에 따른 도면이다. 도 6에 도시된 바와 같이, 본 실시예에서 제공하는 텍스트 인식 모델의 트레이닝 장치(600)는, 예측 유닛(601), 트레이닝 유닛(602), 생성 유닛(603)을 포함하고,

[0160] 예측 유닛(601)은 획득된 제1 샘플 이미지 중의 부분 이미지에 대해 마스크 예측을 수행하여, 제1 샘플 이미지와 대응되는 예측 풀 이미지를 획득하고;

[0161] 예측 유닛(601)은 또한 획득된 제2 샘플 이미지 중의 부분 텍스트에 대해 마스크 예측을 수행하여, 부분 텍스트와 대응되는 예측 텍스트 내용을 획득하고;

[0162] 트레이닝 유닛(602)은 예측 풀 이미지와 예측 텍스트 내용을 기초로 트레이닝하여 프리 트레이닝 모델을 획득하고;

[0163] 생성 유닛(603)은 프리 트레이닝 모델을 기초로 텍스트 인식 모델을 생성하며, 여기서, 텍스트 인식 모델은 인식 대상 이미지에 대해 텍스트 인식을 수행하기 위한 것이다.

[0164] 도 7은 본 출원의 제7 실시예에 따른 도면이다. 도 7에 도시된 바와 같이, 본 실시예에서 제공하는 텍스트 인식 모델의 트레이닝 장치(700)는 예측 유닛(701)을 포함한다.

[0165] 예측 유닛(701)은 획득된 제1 샘플 이미지 중의 부분 이미지에 대해 마스크 예측을 수행하여, 제1 샘플 이미지와 대응되는 예측 풀 이미지를 획득하기 위한 것이다.

[0166] 예측 유닛(701)은 또한 획득된 제2 샘플 이미지 중의 부분 텍스트에 대해 마스크 예측을 수행하여, 부분 텍스트와 대응되는 예측 텍스트 내용을 획득하기 위한 것이다.

[0167] 도 7을 참조하면, 일부 실시예에서, 예측 유닛(701)은,

[0168] 타겟 대상 중의 부분 대상을 랜덤으로 가리우는 가림 서브 유닛(7011);

[0169] 타겟 대상 중 가리워지지 않은 대상을 기초로, 타겟 대상 중 가리워진 부분 대상에 대해 예측하여, 예측 결과를 획득하는 예측 서브 유닛(7012);을 포함한다.

[0170] 여기서, 만약 타겟 대상이 제1 샘플 이미지이면, 타겟 대상 중의 부분 대상은 부분 이미지이고, 예측 결과는 예측 풀 이미지이며; 만약 타겟 대상이 제2 샘플 이미지이면, 타겟 대상 중의 부분 대상은 부분 텍스트이고, 예측 결과는 예측 텍스트 내용이다.

[0171] 일부 실시예에서, 예측 서브 유닛(7012)은,

[0172] 타겟 대상 중 가리워지지 않은 대상에 대응되는 대상 특징을 추출하여, 제1 대상 특징을 획득하는 추출 모듈;

[0173] 제1 대상 특징을 기초로, 타겟 대상 중 가리워진 부분 대상에 대해 예측하여, 예측 결과를 획득하는 예측 모듈;을 포함한다.

[0174] 여기서, 만약 타겟 대상이 제1 샘플 이미지이면, 제1 대상 특징은 제1 비전 특징이고; 만약 타겟 대상이 제2 샘플 이미지이면, 제1 대상 특징은 제1 의미 특징이다.

[0175] 일부 실시예에서, 타겟 대상은 제1 샘플 이미지이고, 상기 제1 대상 특징은 제1 비전 특징이며; 예측 모듈은,

[0176] 제1 비전 특징을 기초로, 제1 샘플 이미지 중 가리워진 부분 이미지에 대응되는 비전 특징을 예측하여, 제2 비전 특징을 획득하는 제1 예측 서브 모듈;

[0177] 제2 비전 특징을 기초로, 제1 샘플 이미지 중 가리워진 부분 이미지를 결정하는 제1 결정 서브 모듈;

[0178] 제1 샘플 이미지 중 가리워지지 않은 이미지, 및 결정된 제1 샘플 이미지 중 가리워진 부분 이미지를 기초로, 예측 풀 이미지를 생성하는 제1 생성 서브 모듈;을 포함한다.

[0179] 일부 실시예에서, 타겟 대상은 제2 샘플 이미지이고, 상기 제1 대상 특징은 제1 의미 특징이며; 예측 모듈은,

[0180] 제1 의미 특징을 기초로, 제2 샘플 이미지 중 가리워진 부분 텍스트에 대응되는 의미 특징을 예측하여, 제2 의미 특징을 획득하는 제2 예측 서브 모듈;

[0181] 제2 의미 특징을 기초로, 예측 텍스트 내용을 생성하는 제2 생성 서브 모듈;을 포함한다.

[0182] 텍스트 인식 모델의 트레이닝 장치(700)는

[0183] 예측 풀 이미지와 예측 텍스트 내용을 기초로 트레이닝하여 프리 트레이닝 모델을 획득하기 위한 것인 트레이닝 유닛(702);

[0184] 프리 트레이닝 모델을 기초로 텍스트 인식 모델을 생성하며, 여기서, 텍스트 인식 모델은 인식 대상 이미지에 대해 텍스트 인식을 수행하기 위한 것인 생성 유닛(703)을 더 포함한다.

[0185] 도 7을 참조하면, 일부 실시예에서, 생성 유닛(703)은,

[0186] 인식 대상 태스크와 트레이닝 이미지를 획득하며, 여기서, 트레이닝 이미지는 텍스트를 포함하는 획득 서브 유닛(7031);

[0187] 인식 대상 태스크와 트레이닝 이미지를 기초로, 프리 트레이닝 모델에 대해 트레이닝하여, 텍스트 인식 모델을 획득하는 트레이닝 서브 유닛(7032);을 포함한다.

[0188] 일부 실시예에서, 트레이닝 서브 유닛(7032)은,

[0189] 트레이닝 이미지를 프리 트레이닝 모델로 입력하여, 트레이닝 이미지에 대응되는 멀티모달 특징맵을 획득하는 입력 모듈;

[0190] 인식 대상 태스크와 멀티모달 특징맵을 기초로, 텍스트 인식 모델을 생성하는 생성 모듈;을 포함한다.

[0191] 일부 실시예에서, 생성 모듈은,

[0192] 멀티모달 특징맵을 기초로, 트레이닝 이미지의 인식 대상 태스크에 따른 예측 인식 결과를 예측하는 제3 예측 서브 모듈;

[0193] 트레이닝 이미지의 기설정된 진실한 인식 결과, 및 예측 인식 결과를 기초로, 텍스트 인식 모델을 구축하는 구축 서브 모듈;을 포함한다.

[0194] 도 8은 본 출원의 제8 실시예에 따른 도면이다. 도 8에 도시된 바와 같이, 본 실시예에서 제공하는 텍스트 인식 장치(800)는,

[0195] 인식 대상 이미지를 획득하며, 여기서, 인식 대상 이미지는 텍스트를 포함하는 획득 유닛(801);

[0196] 사전에 트레이닝된 텍스트 인식 모델을 기반으로 인식 대상 이미지에 대해 텍스트 인식을 수행하여, 인식 대상 이미지 중의 텍스트 내용을 획득하는 인식 유닛(802);을 포함한다.

[0197] 여기서, 텍스트 인식 모델은 상술한 어느 하나의 실시예에 따른 텍스트 인식 모델의 트레이닝 방법을 기반으로 획득된 것이다.

[0198] 도 8을 참조하여 알 수 있는 바와 같이, 일부 실시예에서, 인식 유닛(802)은,

[0199] 텍스트 인식 모델을 기초로 인식 대상 이미지의 멀티모달 특징맵을 결정하는 제1 결정 유닛(8021);

[0200] 멀티모달 특징맵을 기초로 인식 대상 이미지 중의 텍스트 내용을 결정하는 제2 결정 유닛(8022);을 포함한다.

[0201] 여기서, 인식 대상 이미지의 멀티모달 특징맵은 인식 대상 이미지의 비전 특징과 의미 특징을 나타내기 위한 것이다.

[0202] 도 9는 본 출원의 제9 실시예에 따른 도면이다. 도 9에 도시된 바와 같이, 본 출원에 따른 전자기기(900)는 프로세서(901)와 메모리(902)를 포함할 수 있다.

[0203] 메모리(902)는 프로그램을 저장하기 위한 것이고; 메모리(902)는 휘발성 메모리(volatile memory)를 포함할 수 있고, 예를 들어 정적 랜덤 액세스 메모리(static random-access memory, SRAM), 더블 데이터 레이트 동기식 동적 랜덤 액세스 메모리(Double Data Rate Synchronous Dynamic Random Access Memory, DDR SDRAM) 등과 같은 랜덤 액세스 메모리(random-access memory, RAM)를 들 수 있고; 메모리는 비휘발성 메모리(non-volatile memory)를 포함할 수도 있고, 예를 들어 플래쉬 메모리(flash memory)를 들 수 있다. 메모리(902)는 컴퓨터 프로그램(상술한 방법을 구현하는 응용 프로그램, 기능 모듈 등), 컴퓨터 명령 등을 저장하기 위한 것이며, 상술한 컴퓨터 프로그램, 컴퓨터 명령 등은 섹션을 나누어 하나 또는 복수의 메모리(902)에 저장될 수도 있다. 상술한 컴퓨터 프로그램, 컴퓨터 명령, 데이터 등은 프로세서(901)에 의해 호출될 수 있다.

[0204] 상술한 컴퓨터 프로그램, 컴퓨터 명령 등은 섹션을 나누어 하나 또는 복수의 메모리(902)에 저장될 수 있다. 또한 상술한 컴퓨터 프로그램, 컴퓨터 명령 등은 프로세서(901)에 의해 호출될 수 있다.

[0205] 프로세서(901)는 메모리(902)에 저장된 컴퓨터 프로그램을 실행하여, 상술한 실시예에 따른 방법 중의 각각의 단계를 구현하기 위한 것이다.

[0206] 구체적으로 상술한 방법 실시예 중의 관련 기재를 참조할 수 있다.

[0207] 프로세서(901)와 메모리(902)는 별도의 구성일 수 있고, 일체로 집적된 구성일 수도 있다. 프로세서(901)와 메모리(902)가 별도의 구성일 때, 메모리(902), 프로세서(901)는 버스(903)를 통해 커플링 연결될 수 있다.

[0208] 본 실시예의 전자기기는 상술한 방법 중의 기술방안을 수행할 수 있으며, 그 구체적 구현 과정과 기술원리가 동일하므로, 여기서는 반복되는 설명을 생략한다.

[0209] 본 출원의 기술방안에서, 관련되는 사용자 개인 정보(예컨대, 얼굴 이미지 등)의 수집, 저장, 사용, 가공, 전송, 제공 및 공개 등의 처리는 모두 관련 법률 법규의 규정에 부합되며, 공서양속에 어긋나지 않는다.

[0210] 본 출원의 실시예에 따르면, 본 출원은 전자기기, 판독 가능 저장매체 및 컴퓨터 프로그램 제품을 더 제공한다.

[0211] 본 출원의 실시예에 따르면, 본 출원은 컴퓨터 프로그램을 더 제공하며, 컴퓨터 프로그램은 판독 가능 저장매체에 저장되고, 전자기기의 적어도 하나의 프로세서는 판독 가능 저장매체로부터 컴퓨터 프로그램을 판독할 수 있

으며, 적어도 하나의 프로세서는 컴퓨터 프로그램을 실행하여 전자기기가 상술한 어느 하나의 실시예에 따른 방안을 수행하도록 한다.

[0212] 도 10은 본 출원의 실시예를 수행할 수 있는 예시적인 전자기기(1000)를 나타내는 블록도이다. 전자기기는 랩톱 컴퓨터, 데스크톱 컴퓨터, 워크 스테이션, 개인 정보 단말, 서버, 블레이드 서버, 대형 컴퓨터, 및 기타 적합한 컴퓨터와 같은 다양한 형태의 디지털 컴퓨터를 의미한다. 전자기기는 개인 정보 단말, 셀룰러폰, 스마트 폰, 웨어러블 기기 및 기타 유사한 컴퓨팅 장치와 같은 다양한 형태의 모바일 장치를 의미할 수도 있다. 본문에 개시된 부재, 이들의 연결 및 관계, 및 이들의 기능은 단지 예시적인 것이며, 본문에 개시된 것 및/또는 요구하는 본 출원의 구현을 한정하려는 의도가 아니다.

[0213] 도 10에 도시된 바와 같이, 기기(1000)는, 읽기 전용 메모리(ROM, 1002)에 저장된 컴퓨터 프로그램 또는 저장 유닛(1008)으로부터 랜덤 액세스 메모리(RAM, 1003)에 로딩된 컴퓨터 프로그램을 기초로, 다양한 적합한 동작 및 처리를 수행할 수 있는 컴퓨팅 유닛(1001)을 포함한다. RAM(1003)에는, 기기(1000)의 동작에 필요한 다양한 프로그램과 데이터를 더 저장할 수 있다. 컴퓨팅 유닛(1001), ROM(1002) 및 RAM(1003)은 버스(1004)를 통해 서로 연결된다. 입력/출력(I/O) 인터페이스(1005)도 버스(1004)에 연결된다.

[0214] 기기(1000) 중의 복수의 부재는 I/O 인터페이스(1005)에 연결되고, 예를 들어 키보드, 마우스 등과 같은 입력 유닛(1006); 예를 들어 다양한 유형의 디스플레이, 스피커 등과 같은 출력 유닛(1007); 예를 들어 자기 디스크, 광 디스크 등과 같은 저장 유닛(1008); 및 예를 들어 네트워크 카드, 모뎀, 무선 통신 트랜시버 등과 같은 통신 유닛(1009)을 포함한다. 통신 유닛(1009)은 기기(1000)가 인터넷과 같은 컴퓨터 네트워크 및/또는 다양한 통신 네트워크를 통해 기타 기기와 정보/데이터를 교환하는 것을 허용한다.

[0215] 컴퓨팅 유닛(1001)은 처리 및 연산 능력을 갖춘 다양한 범용 및/또는 전용 처리 모듈일 수 있다. 컴퓨팅 유닛(1001)의 일부 예시로서 중앙 처리 유닛(CPU), 그래픽 처리 유닛(GPU), 다양한 전용 인공지능(AI) 연산 칩, 다양한 기계 학습 모델 알고리즘을 실행하는 컴퓨팅 유닛, 디지털 신호 프로세서(DSP), 및 임의의 적합한 프로세서, 컨트롤러, 마이크로 컨트롤러 등을 포함하지만 이에 제한되는 것은 아니다. 컴퓨팅 유닛(1001)은 상술한 각각의 방법과 처리, 예컨대 텍스트 인식 모델의 트레이닝 방법, 텍스트 인식 방법을 수행한다. 예를 들어, 일부 실시예에서, 텍스트 인식 모델의 트레이닝 방법, 텍스트 인식 방법은 컴퓨터 소프트웨어 프로그램으로 구현되어, 명시적으로 저장 유닛(1008)과 같은 기계 판독 가능 매체에 저장될 수 있다. 일부 실시예에서, 컴퓨터 프로그램의 부분 또는 전부는 ROM(1002) 및/또는 통신 유닛(1009)을 통해 기기(1000) 상에 로딩 및/또는 설치될 수 있다. 컴퓨터 프로그램이 RAM(1003)에 로딩되어 컴퓨팅 유닛(1001)에 의해 실행될 때, 상술한 텍스트 인식 모델의 트레이닝 방법, 텍스트 인식 방법의 하나 또는 복수의 단계를 수행할 수 있다. 대안으로서, 기타 실시예에서, 컴퓨팅 유닛(1001)은 기타 임의의 적합한 방식을 통해(예를 들어, 펌웨어를 통해) 텍스트 인식 모델의 트레이닝 방법, 텍스트 인식 방법을 수행하도록 구성될 수 있다.

[0216] 본 명세서에 기재되는 시스템 및 기술의 다양한 실시형태는 디지털 전자 회로 시스템, 집적 회로 시스템, 필드 프로그래머블 어레이(FPGA), 전용 집적 회로(ASIC), 전용 표준 제품(ASSP), 시스템 온 칩 시스템(SOC), 부하 프로그래머블 논리 장치, 컴퓨터 하드웨어, 펌웨어, 소프트웨어, 및/또는 이들의 조합에서 구현될 수 있다. 이러한 다양한 실시형태는 하나 또는 복수의 컴퓨터 프로그램에서 실시되는 것을 포함할 수 있고, 해당 하나 또는 복수의 컴퓨터 프로그램은 적어도 하나의 프로그래머블 프로세서를 포함하는 프로그래머블 시스템 상에서 실행 및/또는 해석될 수 있으며, 해당 프로그래머블 프로세서는 전용 또는 범용 프로그래머블 프로세서일 수 있고, 저장 시스템, 적어도 하나의 입력 장치, 및 적어도 하나의 출력 장치로부터 데이터와 명령을 수신하고, 데이터와 명령을 해당 저장 시스템, 해당 적어도 하나의 입력 장치, 및 해당 적어도 하나의 출력 장치로 전송할 수 있다.

[0217] 본 출원의 방법을 실시하기 위한 프로그램 코드는 하나 또는 복수의 프로그래밍 언어의 임의의 조합으로 작성될 수 있다. 이러한 프로그램 코드는 범용 컴퓨터, 전용 컴퓨터 또는 기타 프로그래머블 데이터 처리 장치의 프로세서 또는 컨트롤러에 제공되어, 프로그램 코드가 프로세서 또는 컨트롤러에 의해 실행될 때 흐름도 및/또는 블록도에서 규정하는 기능/조작이 실시되도록 할 수 있다. 프로그램 코드는 완전히 기계 상에서 실행되거나, 부분적으로 기계 상에서 실행될 수 있으며, 독립 소프트웨어 패키지로서 부분적으로 기계 상에서 실행되고 부분적으로 원격 기계 상에서 실행되거나 완전히 원격 기계 또는 서버 상에서 실행될 수도 있다.

[0218] 본 출원의 문맥에서, 기계 판독 가능 매체는 유형의 매체일 수 있고, 명령 실행 시스템, 장치 또는 기기에 의해 사용되거나 명령 실행 시스템, 장치 또는 기기와 결합되어 사용되는 프로그램을 포함하거나 저장할 수 있다. 기계 판독 가능 매체는 기계 판독 가능 신호 매체이거나 기계 판독 가능 저장 매체일 수 있다. 기계 판독 가능 매

체는 전자적, 자기적, 광학적, 전자기적, 적외선, 또는 반도체 시스템, 장치 또는 기기, 또는 상술한 내용의 임의의 적합한 조합을 포함할 수 있지만 이에 제한되는 것은 아니다. 기계 판독 가능 저장매체의 더 구체적인 예시로서 하나 또는 복수의 선을 기반으로 하는 전기적 연결, 휴대형 컴퓨터 디스크, 하드 디스크, 랜덤 액세스 메모리(RAM), 읽기 전용 메모리(ROM), 소거 가능 및 프로그래머블 읽기 전용 메모리(EPROM 또는 플래쉬 메모리), 광섬유, 휴대용 컴팩트 읽기 전용 메모리(CD-ROM), 광학 저장 장치, 자기 저장 장치, 또는 상술한 내용의 임의의 조합을 포함한다.

[0219] 사용자와의 인터랙션을 제공하기 위하여, 컴퓨터 상에서 본 명세서에 기재되는 시스템 및 기술을 실시할 수 있으며, 해당 컴퓨터는 사용자에게 정보를 디스플레이하기 위한 디스플레이 장치(예를 들어, CRT(캐소드레이 튜브) 또는 LCD(액정 디스플레이) 모니터); 및 키보드와 지향 장치(예를 들어, 마우스 또는 트랙볼)를 구비하고, 사용자는 해당 키보드와 해당 지향 장치를 통해 입력을 컴퓨터로 제공할 수 있다. 기타 종류의 장치는 사용자와의 인터랙션을 제공할 수도 있다. 예를 들어, 사용자에게 제공되는 피드백은 임의의 형태의 센싱 피드백(예를 들어, 시각적 피드백, 청각적 피드백, 또는 촉각적 피드백)일 수 있고; 임의의 형태(사운드 입력, 음성 입력 또는 촉각 입력)을 통해 사용자로부터의 입력을 수신할 수 있다.

[0220] 여기에 기재되는 시스템과 기술은 백그라운드 부재를 포함하는 컴퓨팅 시스템(예를 들어, 데이터 서버로서), 또는 중간부재를 포함하는 컴퓨팅 시스템(예를 들어, 응용 서버), 또는 프론트 엔드 부재를 포함하는 컴퓨팅 시스템(예를 들어, 그래픽 유저 인터페이스 또는 인터넷 브라우저를 구비하는 사용자 컴퓨터, 사용자는 해당 그래픽 유저 인터페이스 또는 해당 인터넷 브라우저를 통해 여기에 기재되는 시스템 및 기술의 실시형태와 인터랙션할 수 있다), 또는 이러한 백그라운드 부재, 중간 부재, 또는 프론트 엔드 부재를 포함하는 임의의 조합의 컴퓨팅 시스템에서 실시할 수 있다. 임의의 형태 또는 매체의 디지털 데이터 통신(예를 들어, 통신 네트워크)을 통해 시스템의 부재를 서로 연결시킬 수 있다. 통신 네트워크의 예시로서, 근거리 통신망(LAN), 광역 통신망(WAN) 및 인터넷을 포함한다.

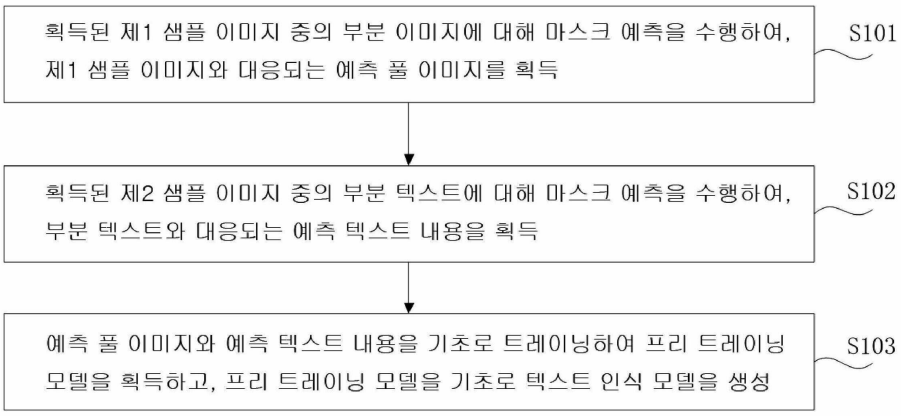
[0221] 컴퓨터 시스템은 클라이언트와 서버를 포함할 수 있다. 클라이언트와 서버는 일반적으로 서로 멀리 떨어져 있으며, 통상적으로 통신 네트워크를 통해 인터랙션한다. 상응한 컴퓨터 상에서 실행되며 서로 클라이언트 - 서버 관계를 가지는 컴퓨터 프로그램을 통해 클라이언트와 서버의 관계를 생성한다. 서버는 클라우드 서버일 수 있고, 클라우드 컴퓨팅 서버 또는 클라우드 호스트라고도 불리우며, 클라우드 컴퓨팅 서비스 시스템 중의 일 호스트 제품으로서, 기존의 물리 호스트와 가상 사설 서버("Virtual Private Server", 또는 "VPS"로 약칭)에 존재하는 관리 상의 어려움이 크고, 서비스 확장이 약한 흠결을 해결한다. 서버는 분포식 시스템의 서버, 또는 블록체인이 결합된 서버일 수도 있다.

[0222] 상술한 다양한 형태의 프로세스를 사용하여, 단계를 재배열, 추가 또는 삭제할 수 있다는 점을 이해하여야 한다. 예를 들어, 본 출원에 기재된 각 단계는 병열로 수행될 수 있고 순차적으로 수행될 수도 있고 서로 다른 순서로 수행될 수도 있으며, 본 출원에 개시된 기술방안이 원하는 결과를 얻을 수만 있다면, 본 명세서에서는 이에 대해 제한하지 않는다.

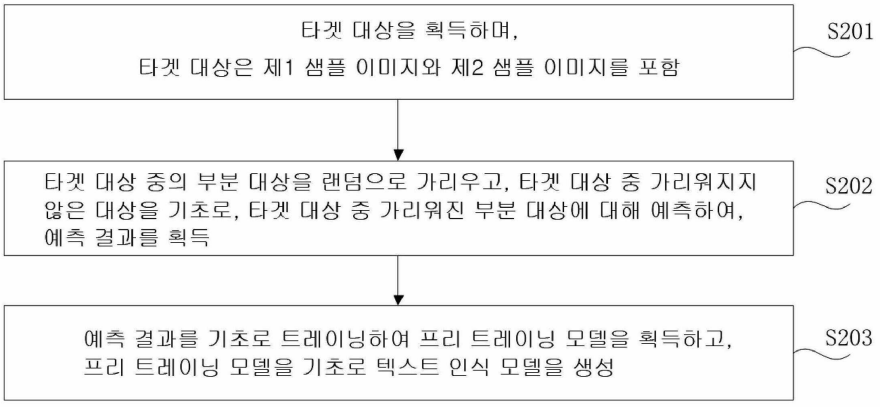
[0223] 상술한 구체적인 실시형태는 본 출원의 보호범위에 대한 한정이 아니다. 본 분야의 통상의 지식을 가진 자라면, 설계 요구와 기타 요소를 기초로, 다양한 수정, 조합, 서브 조합 및 대체를 수행할 수 있다는 점을 이해하여야 한다. 본 출원의 사상과 원칙 내에서 이루어진 모든 수정, 동등한 치환 및 개선 등은 모두 본 출원의 보호 범위 내에 포함되어야 한다.

도면

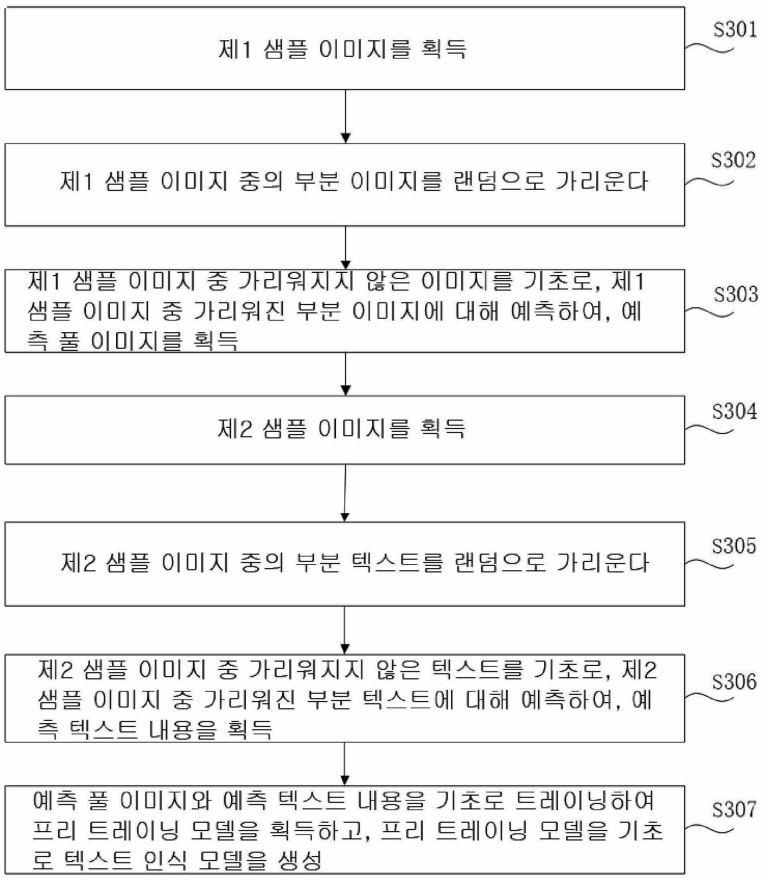
도면1


획득된 제1 샘플 이미지 중의 부분 이미지에 대해 마스크 예측을 수행하여, 제1 샘플 이미지와 대응되는 예측 풀 이미지를 획득
S101
획득된 제2 샘플 이미지 중의 부분 텍스트에 대해 마스크 예측을 수행하여, 부분 텍스트와 대응되는 예측 텍스트 내용을 획득
S102
예측 풀 이미지와 예측 텍스트 내용을 기초로 트레이닝하여 프리 트레이닝 모델을 획득하고, 프리 트레이닝 모델을 기초로 텍스트 인식 모델을 생성
S103

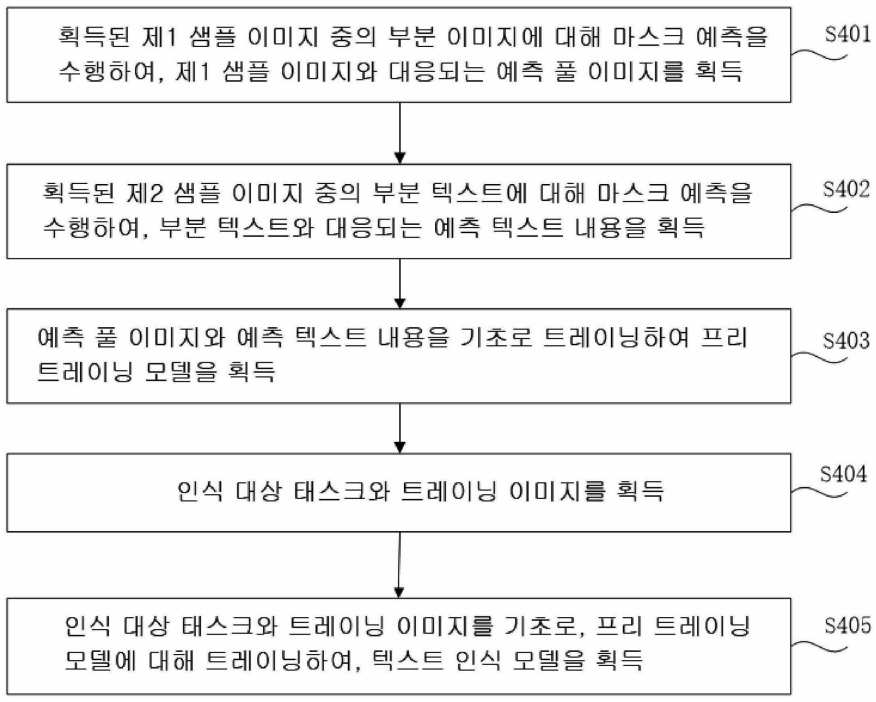
도면2


타겟 대상을 획득하며, 타겟 대상은 제1 샘플 이미지와 제2 샘플 이미지를 포함
S201
타겟 대상 중의 부분 대상을 랜덤으로 가리우고, 타겟 대상 중 가리워지지 않은 대상을 기초로, 타겟 대상 중 가리워진 부분 대상에 대해 예측하여, 예측 결과를 획득
S202
예측 결과를 기초로 트레이닝하여 프리 트레이닝 모델을 획득하고, 프리 트레이닝 모델을 기초로 텍스트 인식 모델을 생성
S203

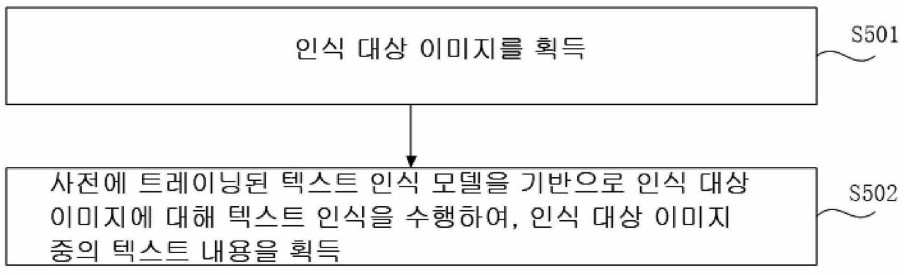
도면3


제1 샘플 이미지를 획득
S301
제1 샘플 이미지 중의 부분 이미지를 랜덤으로 가리운다
S302
제1 샘플 이미지 중 가리워지지 않은 이미지를 기초로, 제1 샘플 이미지 중 가리워진 부분 이미지에 대해 예측하여, 예측 풀 이미지를 획득
S303
제2 샘플 이미지를 획득
S304
제2 샘플 이미지 중의 부분 텍스트를 랜덤으로 가리운다
S305
제2 샘플 이미지 중 가리워지지 않은 텍스트를 기초로, 제2 샘플 이미지 중 가리워진 부분 텍스트에 대해 예측하여, 예측 텍스트 내용을 획득
S306
예측 풀 이미지와 예측 텍스트 내용을 기초로 트레이닝하여 프리 트레이닝 모델을 획득하고, 프리 트레이닝 모델을 기초로 텍스트 인식 모델을 생성
S307

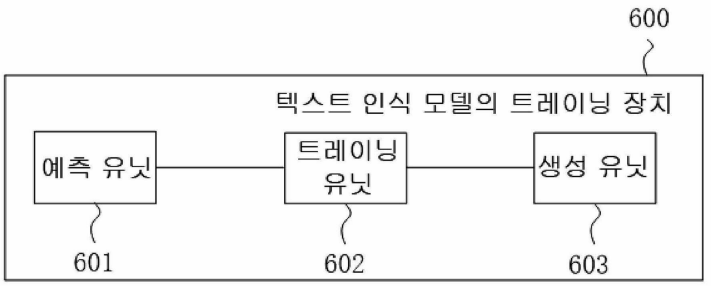
도면4


획득된 제1 샘플 이미지 중의 부분 이미지에 대해 마스크 예측을 수행하여, 제1 샘플 이미지와 대응되는 예측 풀 이미지를 획득
S401
획득된 제2 샘플 이미지 중의 부분 텍스트에 대해 마스크 예측을 수행하여, 부분 텍스트와 대응되는 예측 텍스트 내용을 획득
S402
예측 풀 이미지와 예측 텍스트 내용을 기초로 트레이닝하여 프리 트레이닝 모델을 획득
S403
인식 대상 태스크와 트레이닝 이미지를 획득
S404
인식 대상 태스크와 트레이닝 이미지를 기초로, 프리 트레이닝 모델에 대해 트레이닝하여, 텍스트 인식 모델을 획득
S405

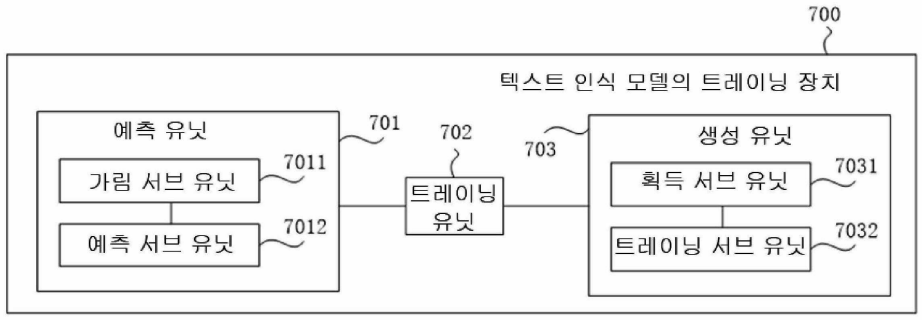
도면5



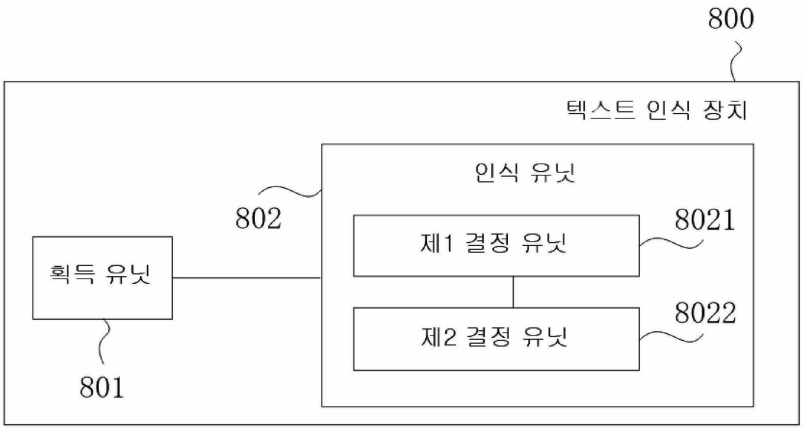
도면6



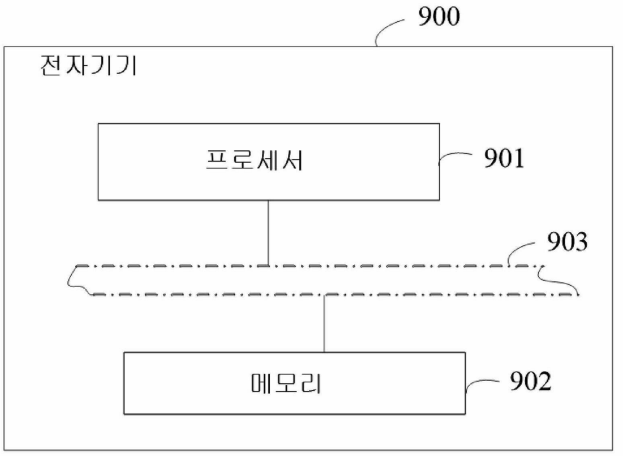
도면7



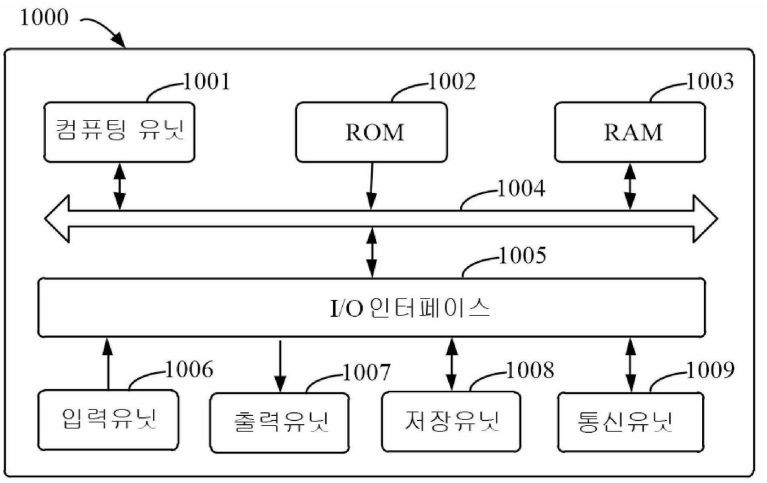
도면8



도면9


900
전자기기
프로세서
901
903
메모리
902

도면10


1000
1001
컴퓨팅 유닛
1002
ROM
1003
RAM
1004
1005
I/O 인터페이스
1006
입력유닛
1007
출력유닛
1008
저장유닛
1009
통신유닛