(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0080156
(43) 공개일자 2025년06월05일

(51) 국제특허분류(Int. Cl.)
G06V 10/80 (2022.01) *G06N 3/0455* (2023.01)
G06V 10/25 (2022.01) *G06V 10/26* (2022.01)
G06V 10/46 (2022.01)
(52) CPC특허분류
G06V 10/806 (2023.08)
G06N 3/0455 (2023.01)
(21) 출원번호 10-2023-0167591
(22) 출원일자 2023년11월28일
심사청구일자 2023년11월28일

(71) 출원인
서강대학교산학협력단
서울특별시 마포구 백범로 35 (신수동, 서강대학교)
(72) 발명자
강석주
서울특별시 양천구 목동서로 70, 221동 1304호 (목동, 목동신시가지아파트2단지)
조유빈
서울특별시 노원구 공릉로 430, 102동 1002호 (중계동, 중앙하이츠아쿠아)
유현우
서울특별시 마포구 백범로 35, 1108호(신수동, 곤자가 국제학사)
(74) 대리인
이준영

전체 청구항 수 : 총 14 항

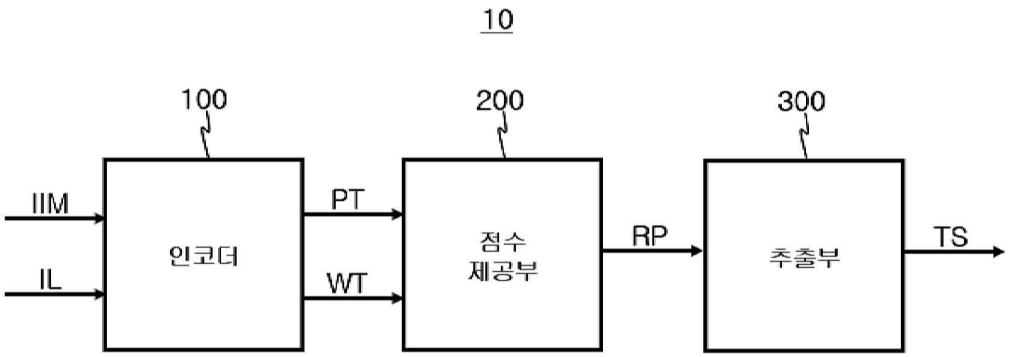
(54) 발명의 명칭 **타겟 추출장치**

(57) 요 약

본 발명의 실시예에 따른 타겟 추출장치는 인코더, 점수 제공부 및 추출부를 포함할 수 있다. 인코더는 입력 이미지 및 입력 이미지에 포함되는 타겟 세그먼트에 상응하는 언어 입력에 기초하여 복수의 픽셀 토큰들 및 복수의 단어 토큰들을 제공할 수 있다. 점수 제공부는 복수의 픽셀 토큰들 및 복수의 단어 토큰들에 따라 결정되는 결과점수를 제공할 수 있다. 추출부는 결과점수에 따라 결정되는 비쥬얼 익스프레션 가이드에 기초하여 타겟 세그먼트를 추출할 수 있다.

본 발명에 따른 타겟 추출장치는 입력 이미지에 따라 생성되는 픽셀 토큰들 및 입력 언어에 따라 생성되는 단어 토큰들에 기초하여 결정되는 비쥬얼 익스프레션 가이드에 따라 입력 이미지에서 타겟 세그먼트를 제공함으로써 타겟영역에 포함되는 객체를 보다 효과적으로 추출할 수 있다.

대 표 도 - 도1



(52) CPC특허분류
G06V 10/25 (2023.08)
G06V 10/26 (2023.08)
G06V 10/467 (2023.08)
G06V 10/803 (2023.08)
G06V 2201/07 (2022.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711198586
과제번호 00260091
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 대학ICT연구센터육성지원사업
연구과제명 초거대 AI를 위한 대규모데이터센터용 인공지능 시스템반도체 인력양성 및 핵심기술 개발
기 여 율 1/3
과제수행기관명 서강대학교산학협력단
연구기간 2023.07.01 ~ 2030.12.31
이 발명을 지원한 국가연구개발사업
과제고유번호 1711181038
과제번호 2021M3H2A1038042
부처명 과학기술정보통신부
과제관리(전문)기관명 (재)한국연구재단
연구사업명 원천기술개발사업 / 시스템반도체융합전문인력육성사업
연구과제명 지능형 사물 에너지 (iEoT) 반도체 시스템 융합 다빈치형 인력양성 센터
기 여 율 1/3
과제수행기관명 중앙대학교
연구기간 2021.04.09 ~ 2026.12.31
이 발명을 지원한 국가연구개발사업
과제고유번호 1415183166
과제번호 P0020535
부처명 산업통상자원부
과제관리(전문)기관명 한국산업기술진흥원
연구사업명 산업혁신인재성장지원
연구과제명 미래형자동차 핵심기술 R&D 전문인력양성
기 여 율 1/3
과제수행기관명 연세대학교
연구기간 2022.03.01 ~ 2027.02.28

명 세 서

청구범위

청구항 1

입력 이미지 및 상기 입력 이미지에 포함되는 타겟 세그먼트에 상응하는 언어 입력에 기초하여 복수의 픽셀 토큰들 및 복수의 단어 토큰들을 제공하는 인코더;

상기 복수의 픽셀 토큰들 및 상기 복수의 단어 토큰들에 따라 결정되는 결과점수를 제공하는 점수 제공부; 및

상기 결과점수에 따라 결정되는 비쥬얼 익스프레션 가이드에 기초하여 상기 타겟 세그먼트를 추출하는 추출부를 포함하는 타겟 추출장치.

청구항 2

제1항에 있어서,

상기 인코더는,

상기 입력 이미지에 포함되는 복수의 픽셀들을 기준으로 상기 복수의 픽셀들 각각의 정보를 포함하는 상기 복수의 픽셀 토큰들을 제공하는 비전 인코더; 및

상기 언어 입력에 포함되는 복수의 단어들을 기준으로 상기 복수의 단어들 각각의 정보를 포함하는 상기 복수의 단어 토큰들을 제공하는 언어 인코더를 포함하는 것을 특징으로 하는 타겟 추출장치.

청구항 3

제2항에 있어서,

상기 점수 제공부는,

상기 복수의 픽셀 토큰들의 각각을 제1 방향을 따라서 배치하고, 상기 복수의 단어 토큰들의 각각을 제1 방향과 수직인 제2 방향으로 배치하여 점수 매트릭스를 제공하는 매트릭스부; 및

상기 복수의 픽셀 토큰들 및 상기 복수의 단어 토큰들에 따라 점수 매트릭스에 포함되는 점수픽셀들 각각의 결과점수를 결정하는 점수 결정부를 포함하는 것을 특징으로 하는 타겟 추출장치.

청구항 4

제3항에 있어서,

상기 추출부는,

상기 점수픽셀들 각각의 상기 결과점수를 기준으로 미리 정해진 일정한 간격에 배치되는 상기 점수픽셀들을 추출하여 추출픽셀들을 제공하는 픽셀 추출부를 더 포함하는 것을 특징으로 하는 타겟 추출장치.

청구항 5

제4항에 있어서,

상기 추출부는,

상기 추출픽셀들에 상응하는 상기 결과점수를 상기 제1 방향으로 합산하여 합산점수를 제공하는 합산 점수부를 더 포함하는 것을 특징으로 하는 타겟 추출장치.

청구항 6

제5항에 있어서,

상기 추출부는,

상기 합산점수 중 가장 높은 점수에 해당하는 최고점수에 상응하는 상기 입력 이미지에 포함되는 상기 픽셀들의 정보를 상기 비쥬얼 익스프레션 가이드로 제공하는 가이드 제공부를 더 포함하는 것을 특징으로 하는 타겟 추출 장치.

청구항 7

제6항에 있어서,

상기 추출부는,

상기 비쥬얼 익스프레션 가이드 및 상기 입력 이미징에 기초하여 상기 타겟 세그먼트를 제공하는 타겟 제공부를 더 포함하는 것을 특징으로 하는 타겟 추출장치.

청구항 8

제7항에 있어서,

상기 픽셀 추출부는,

상기 점수픽셀들 각각의 상기 결과점수 및 미리 정해진 기준점수를 비교하여 상기 후보픽셀들을 제공하는 후보 추출부를 더 포함하는 것을 특징으로 하는 타겟 추출장치.

청구항 9

제8항에 있어서,

상기 픽셀 추출부는,

상기 후보픽셀들이 상기 점수 매트릭스 상에 복수의 영역들에 분산되어 배치되는 경우, 상기 복수의 영역들 각각에 배치되는 상기 후보픽셀들의 수에 따라 상기 추출픽셀들을 결정하는 결정부를 더 포함하는 것을 특징으로 하는 타겟 추출장치.

청구항 10

제9항에 있어서,

상기 결정부는,

상기 복수의 영역들 중 상기 후보픽셀들의 수가 가장 많이 배치되는 영역의 상기 후보픽셀들을 상기 추출픽셀들로 결정하는 것을 특징으로 하는 타겟 추출장치.

청구항 11

입력 이미지 및 상기 입력 이미지에 포함되는 타겟 세그먼트에 상응하는 언어 입력에 기초하여 복수의 픽셀 토큰들 및 복수의 단어 토큰들을 제공하는 인코더;

상기 복수의 픽셀 토큰들 및 상기 복수의 단어 토큰들에 따라 결정되는 결과점수를 제공하는 점수 제공부;

상기 결과점수에 따라 결정되는 비쥬얼 익스프레션 가이드에 따라 상기 타겟 세그먼트를 추출하는 추출부; 및

상기 입력 이미지에 상기 타겟 세그먼트를 표시하는 표시부를 포함하는 타겟 추출 시스템.

청구항 12

제11항에 있어서,

상기 인코더는,

상기 입력 이미지에 포함되는 복수의 픽셀들을 기준으로 상기 복수의 픽셀들 각각의 정보를 포함하는 상기 복수의 픽셀 토큰들을 제공하는 비전 인코더; 및

상기 언어 입력에 포함되는 복수의 단어들을 기준으로 상기 복수의 단어들 각각의 정보를 포함하는 상기 복수의 단어 토큰들을 제공하는 언어 인코더를 포함하는 것을 특징으로 하는 타겟 추출 시스템.

청구항 13

인코더가 입력 이미지 및 상기 입력 이미지에 포함되는 타겟 세그먼트에 상응하는 언어 입력에 기초하여 복수의 픽셀 토큰들 및 복수의 단어 토큰들을 제공하는 단계;

점수 제공부가 상기 복수의 픽셀 토큰들 및 상기 복수의 단어 토큰들에 따라 결정되는 결과점수를 제공하는 단계; 및

추출부가 상기 결과점수에 따라 결정되는 비쥬얼 익스프레션 가이드에 따라 상기 타겟 세그먼트를 추출하는 단계를 포함하는 타겟 추출장치의 동작방법.

청구항 14

인코더가 입력 이미지 및 상기 입력 이미지에 포함되는 타겟 세그먼트에 상응하는 언어 입력에 기초하여 복수의 픽셀 토큰들 및 복수의 단어 토큰들을 제공하는 단계;

점수 제공부가 상기 복수의 픽셀 토큰들 및 상기 복수의 단어 토큰들에 따라 결정되는 결과점수를 제공하는 단계;

추출부가 상기 결과점수에 따라 결정되는 비쥬얼 익스프레션 가이드에 따라 상기 타겟 세그먼트를 추출하는 단계; 및

표시부가 상기 입력 이미지에 상기 타겟 세그먼트를 표시하는 단계를 포함하는 타겟 추출 시스템의 동작방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 타겟 추출장치에 관한 것이다.

배 경 기 술

[0002] 입력 이미지에는 타겟 객체 영역과 아닌 영역의 정보가 모두 포함되어 있기 때문에 타겟 객체와 관련된 적절한 정보를 꼼꼼하게 식별하여 선택적으로 추출하는 방법론 개발이 필요할 수 있다. 최근, 이와 관련하여 다양한 연구들이 진행되고 있다.

선행기술문헌

특허문헌

[0003] (특허문헌 0001) (한국등록특허) 제10-2557912호 (등록일자, 2023.7.17)

발명의 내용

해결하려는 과제

[0004] 본 발명이 이루고자 하는 기술적 과제는 입력 이미지에 따라 생성되는 픽셀 토큰들 및 입력 언어에 따라 생성되는 단어 토큰들에 기초하여 결정되는 비쥬얼 익스프레션 가이드에 따라 입력 이미지에서 타겟 세그먼트를 용이하게 추출할 수 있는 타겟 추출장치를 제공하는 것이다.

과제의 해결 수단

[0005] 이러한 과제를 해결하기 위하여 본 발명의 실시예에 따른 타겟 추출장치는 인코더, 점수 제공부 및 추출부를 포함할 수 있다. 인코더는 입력 이미지 및 상기 입력 이미지에 포함되는 타겟 세그먼트에 상응하는 언어 입력에 기초하여 복수의 픽셀 토큰들 및 복수의 단어 토큰들을 제공할 수 있다. 점수 제공부는 상기 복수의 픽셀 토큰들 및 상기 복수의 단어 토큰들에 따라 결정되는 결과점수를 제공할 수 있다. 추출부는 상기 결과점수에 따라 결정되는 비쥬얼 익스프레션 가이드에 기초하여 상기 타겟 세그먼트를 추출할 수 있다.

[0006] 일 실시예에 있어서, 상기 인코더는 비전 인코더 및 언어 인코더를 포함할 수 있다. 비전 인코더는 상기 입력 이미지에 포함되는 복수의 픽셀들을 기준으로 상기 복수의 픽셀들 각각의 정보를 포함하는 상기 복수의 픽셀 토큰들을 제공할 수 있다. 언어 인코더는 상기 언어 입력에 포함되는 복수의 단어들을 기준으로 상기 복수의 단어들 각각의 정보를 포함하는 상기 복수의 단어 토큰들을 제공할 수 있다.

[0007] 일 실시예에 있어서, 상기 점수 제공부는 매트릭스부 및 점수 결정부를 포함할 수 있다. 매트릭스부는 상기 복수의 픽셀 토큰들의 각각을 제1 방향을 따라서 배치하고, 상기 복수의 단어 토큰들의 각각을 제1 방향과 수직인 제2 방향으로 배치하여 점수 매트릭스를 제공할 수 있다. 점수 결정부는 상기 복수의 픽셀 토큰들 및 상기 복수의 단어 토큰들에 따라 점수 매트릭스에 포함되는 점수픽셀들 각각의 결과점수를 결정할 수 있다.

[0008] 일 실시예에 있어서, 상기 추출부는 픽셀 추출부를 더 포함할 수 있다. 픽셀 추출부는 상기 점수픽셀들 각각의 상기 결과점수를 기준으로 미리 정해진 일정한 간격에 배치되는 상기 점수픽셀들을 추출하여 추출픽셀들을 제공할 수 있다.

[0009] 일 실시예에 있어서, 상기 추출부는 합산 점수부를 더 포함할 수 있다. 합산 점수부는 상기 추출픽셀들에 상응하는 상기 결과점수를 상기 제1 방향으로 합산하여 합산점수를 제공할 수 있다.

[0010] 일 실시예에 있어서, 상기 추출부는 가이드 제공부를 더 포함할 수 있다. 가이드 제공부는 상기 합산점수 중 가장 높은 점수에 해당하는 최고점수에 상응하는 상기 입력 이미지에 포함되는 상기 픽셀들의 정보를 상기 비쥬얼 익스프레션 가이드로 제공할 수 있다.

[0011] 일 실시예에 있어서, 상기 추출부는 타겟 제공부를 더 포함할 수 있다. 타겟 제공부는 상기 비쥬얼 익스프레션 가이드 및 상기 입력 이미징에 기초하여 상기 타겟 세그먼트를 제공할 수 있다.

[0012] 일 실시예에 있어서, 상기 픽셀 추출부는 후보 추출부를 더 포함할 수 있다. 후보 추출부는 상기 점수픽셀들 각각의 상기 결과점수 및 미리 정해진 기준점수를 비교하여 상기 후보픽셀들을 제공할 수 있다.

[0013] 일 실시예에 있어서, 상기 픽셀 추출부는 결정부를 더 포함할 수 있다. 결정부는 상기 후보픽셀들이 상기 점수 매트릭스 상에 복수의 영역들에 분산되어 배치되는 경우, 상기 복수의 영역들 각각에 배치되는 상기 후보픽셀들의 수에 따라 상기 추출픽셀들을 결정할 수 있다.

[0014] 일 실시예에 있어서, 상기 결정부는 상기 복수의 영역들 중 상기 후보픽셀들의 수가 가장 많이 배치되는 영역의 상기 후보픽셀들을 상기 추출픽셀들로 결정할 수 있다.

[0015] 이러한 과제를 해결하기 위하여 본 발명의 실시예에 따른 타겟 추출 시스템은 인코더, 점수 제공부, 추출부 및 표시부를 포함할 수 있다. 인코더는 입력 이미지 및 상기 입력 이미지에 포함되는 타겟 세그먼트에 상응하는 언어 입력에 기초하여 복수의 픽셀 토큰들 및 복수의 단어 토큰들을 제공할 수 있다. 점수 제공부는 상기 복수의 픽셀 토큰들 및 상기 복수의 단어 토큰들에 따라 결정되는 결과점수를 제공할 수 있다. 추출부는 상기 결과점수에 따라 결정되는 비쥬얼 익스프레션 가이드에 따라 상기 타겟 세그먼트를 추출할 수 있다. 표시부는 상기 입력 이미지에 상기 타겟 세그먼트를 표시할 수 있다.

[0016] 일 실시예에 있어서, 상기 인코더는 비전 인코더 및 언어 인코더를 포함할 수 있다. 비전 인코더는 상기 입력 이미지에 포함되는 복수의 픽셀들을 기준으로 상기 복수의 픽셀들 각각의 정보를 포함하는 상기 복수의 픽셀 토큰들을 제공할 수 있다. 언어 인코더는 상기 언어 입력에 포함되는 복수의 단어들을 기준으로 상기 복수의 단어들 각각의 정보를 포함하는 상기 복수의 단어 토큰들을 제공할 수 있다.

[0017] 이러한 과제를 해결하기 위하여 본 발명의 실시예에 따른 타겟 추출장치의 동작방법에서는, 인코더가 입력 이미지 및 상기 입력 이미지에 포함되는 타겟 세그먼트에 상응하는 언어 입력에 기초하여 복수의 픽셀 토큰들 및 복수의 단어 토큰들을 제공할 수 있다. 점수 제공부가 상기 복수의 픽셀 토큰들 및 상기 복수의 단어 토큰들에 따라 결정되는 결과점수를 제공할 수 있다. 추출부가 상기 결과점수에 따라 결정되는 비쥬얼 익스프레션 가이드에 따라 상기 타겟 세그먼트를 추출할 수 있다.

[0018] 이러한 과제를 해결하기 위하여 본 발명의 실시예에 따른 타겟 추출 시스템의 동작방법에서는, 인코더가 입력 이미지 및 상기 입력 이미지에 포함되는 타겟 세그먼트에 상응하는 언어 입력에 기초하여 복수의 픽셀 토큰들 및 복수의 단어 토큰들을 제공할 수 있다. 점수 제공부가 상기 복수의 픽셀 토큰들 및 상기 복수의 단어 토큰들에 따라 결정되는 결과점수를 제공할 수 있다. 추출부가 상기 결과점수에 따라 결정되는 비쥬얼 익스프레션 가이드에 따라 상기 타겟 세그먼트를 추출할 수 있다. 표시부가 상기 입력 이미지에 상기 타겟 세그먼트를 표시할

수 있다.

[0019] 위에서 언급된 본 발명의 기술적 과제 외에도, 본 발명의 다른 특징 및 이점들이 이하에서 기술되거나, 그러한 기술 및 설명으로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

발명의 효과

[0020] 이상과 같은 본 발명에 따르면 다음과 같은 효과가 있다.

[0021] 본 발명에 따른 타겟 추출장치는 입력 이미지에 따라 생성되는 픽셀 토큰들 및 입력 언어에 따라 생성되는 단어 토큰들에 기초하여 결정되는 비쥬얼 익스프레션 가이드에 따라 입력 이미지에서 타겟 세그먼트를 제공함으로써 타겟영역에 포함되는 객체를 보다 효과적으로 추출할 수 있다.

[0022] 이 밖에도, 본 발명의 실시 예들을 통해 본 발명의 또 다른 특징 및 이점들이 새롭게 파악될 수도 있을 것이다.

도면의 간단한 설명

[0023] 도 1은 본 발명의 실시예들에 따른 타겟 추출장치를 나타내는 도면이다.

도 2는 도 1의 타겟 추출장치에서 사용되는 입력 이미지의 일 예를 나타내는 도면이다.

도 3은 도 1의 타겟 추출장치에서 사용되는 언어 입력의 일 예를 나타내는 도면이다.

도 4는 도 1의 타겟 추출장치에 포함되는 인코더의 일 예를 나타내는 도면이다.

도 5는 도 1의 타겟 추출장치에 포함되는 점수 제공부의 일 예를 나타내는 도면이다.

도 6은 도 1의 타겟 추출장치에 포함되는 추출부의 일 예를 나타내는 도면이다.

도 7은 도 6의 추출부에 포함되는 픽셀 추출부 및 합산 점수부의 동작을 설명하기 위한 도면이다.

도 8은 도 1의 타겟 추출장치에 포함되는 추출부의 다른 예를 나타내는 도면이다.

도 9는 도 1의 타겟 추출장치에 포함되는 픽셀 추출부의 일 예를 나타내는 도면이다.

도 10은 도 9의 픽셀 추출부의 동작을 설명하기 위한 도면이다.

도 11은 본 발명의 실시예들에 따른 타겟 추출 시스템을 나타내는 도면이다.

도 12는 도 11의 타겟 추출 시스템에 포함되는 표시부를 설명하기 위한 도면이다.

도 13은 본 발명의 실시예들에 따른 타겟 추출장치의 동작방법을 나타내는 순서도이다.

도 14는 본 발명의 실시예들에 따른 타겟 추출 시스템의 동작방법을 나타내는 순서도이다.

발명을 실시하기 위한 구체적인 내용

[0024] 본 명세서에서 각 도면의 구성 요소들에 참조번호를 부가함에 있어서 동일한 구성 요소들에 한해서는 비록 다른 도면상에 표시되더라도 가능한한 동일한 번호를 가지도록 하고 있음에 유의하여야 한다.

[0025] 한편, 본 명세서에서 서술되는 용어의 의미는 다음과 같이 이해되어야 할 것이다.

[0026] 단수의 표현은 문맥상 명백하게 다르게 정의하지 않는 한, 복수의 표현을 포함하는 것으로 이해되어야 하는 것으로, 이들 용어들에 의해 권리범위가 한정되어서는 아니 된다.

[0027] "포함하다" 또는 "가지다" 등의 용어는 하나 또는 그 이상의 다른 특징이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0028] 이하, 첨부되는 도면을 참고하여 상기 문제점을 해결하기 위해 고안된 본 발명의 바람직한 실시예들에 대해 상세히 설명한다.

[0029] 도 1은 본 발명의 실시예들에 따른 타겟 추출장치를 나타내는 도면이고, 도 2는 도 1의 타겟 추출장치에서 사용되는 입력 이미지의 일 예를 나타내는 도면이고, 도 3은 도 1의 타겟 추출장치에서 사용되는 언어 입력의 일 예를 나타내는 도면이고, 도 4는 도 1의 타겟 추출장치에 포함되는 인코더의 일 예를 나타내는 도면이다.

[0030] 도 1 내지 4를 참조하면, 본 발명의 실시예에 따른 타겟 추출장치(10)는 인코더(100), 점수 제공부(200) 및 추출부(300)를 포함할 수 있다. 인코더(100)는 입력 이미지(IIM) 및 입력 이미지(IIM)에 포함되는 타겟 세그먼트(TS)에 상응하는 언어 입력(IL)에 기초하여 복수의 픽셀 토큰들(PT) 및 복수의 단어 토큰들(WT)을 제공할 수 있다.

[0031] 일 실시예에 있어서, 인코더(100)는 비전 인코더(110) 및 언어 인코더(120)를 포함할 수 있다. 비전 인코더(110)는 입력 이미지(IIM)에 포함되는 복수의 픽셀들(PX)을 기준으로 복수의 픽셀들(PX) 각각의 정보를 포함하는 복수의 픽셀 토큰들(PT)을 제공할 수 있다. 예를 들어, 입력 이미지(IIM)는 복수의 픽셀들(PX)을 포함할 수 있고, 복수의 픽셀들(PX)은 제1 픽셀(PX1) 내지 제N 픽셀(PXN)을 포함할 수 있다. 제1 픽셀(PX1)에 관한 정보는 입력 이미지(IIM)에서 제1 픽셀(PX1)이 배치되는 위치, 제1 픽셀(PX1)의 위치에서 입력 이미지(IIM)의 색상, 모양 등을 포함하여 입력 이미지(IIM)에 포함되는 제1 픽셀(PX1)과 관련된 다양한 정보를 포함할 수 있다.

[0032] 또한, 복수의 픽셀 토큰들(PT)은 제1 픽셀 토큰(PT1) 내지 제N 픽셀 토큰(PTN)을 포함할 수 있다. 제1 픽셀(PX1)에 대한 정보는 제1 픽셀 토큰(PT1)에 포함될 수 있다. 위에서 설명한 제1 픽셀(PX1)과 관련된 정보에 대한 내용은 제2 픽셀(PX2) 내지 제N 픽셀(PXN)에도 동일하게 적용될 수 있고, 제1 픽셀 토큰(PT1)에 대한 설명도 제2 픽셀 토큰(PT2) 내지 제N 픽셀 토큰(PTN)에 동일하게 적용될 수 있다.

[0033] 언어 인코더(120)는 언어 입력(IL)에 포함되는 복수의 단어들을 기준으로 복수의 단어들 각각의 정보를 포함하는 복수의 단어 토큰들(WT)을 제공할 수 있다. 예를 들어, 언어 입력(IL)은 "uniform number 10"일 수 있다. 언어 입력(IL)에 포함되는 복수의 단어들은 3개일 수 있고, 복수의 단어들 중 제1 단어는 "uniform"일 수 있고, 제2 단어는 "number"일 수 있다. 또한, 제3 단어는 "10"일 수 있다.

[0034] 또한, 복수의 단어 토큰들(WT)은 제1 단어 토큰(WT1) 내지 제3 단어 토큰(WT3)을 포함할 수 있다. 제1 단어 토큰(WT1)은 제1 단어에 해당하는"uniform"과 관련된 정보를 포함할 수 있고, 제2 단어 토큰(WT2)은 제2 단어에 해당하는"number"와 관련된 정보를 포함할 수 있다. 또한, 제3 단어 토큰(WT3)은 제3 단어에 해당하는"10"과 관련된 정보를 포함할 수 있다.

[0035] 도 5는 도 1의 타겟 추출장치에 포함되는 점수 제공부의 일 예를 나타내는 도면이고, 도 6은 도 1의 타겟 추출장치에 포함되는 추출부의 일 예를 나타내는 도면이고, 도 7은 도 6의 추출부에 포함되는 픽셀 추출부 및 합산 점수부의 동작을 설명하기 위한 도면이다.

[0036] 도 1 내지 7을 참조하면, 점수 제공부(200)는 복수의 픽셀 토큰들(PT) 및 복수의 단어 토큰들(WT)에 따라 결정되는 결과점수(RP)를 제공할 수 있다.

[0037] 일 실시예에 있어서, 점수 제공부(200)는 매트릭스부(210) 및 점수 결정부(220)를 포함할 수 있다. 매트릭스부(210)는 복수의 픽셀 토큰들(PT)의 각각을 제1 방향(D1)을 따라서 배치하고, 복수의 단어 토큰들(WT)의 각각을 제1 방향(D1)과 수직인 제2 방향(D2)으로 배치하여 점수 매트릭스(SM)를 제공할 수 있다. 예를 들어, 복수의 픽셀 토큰들(PT)은 제1 픽셀 토큰(PT1) 내지 제N 픽셀 토큰(PTN)을 포함할 수 있고, 복수의 단어 토큰들(WT)은 제1 단어 토큰(WT1) 내지 제K 단어 토큰(WTK)을 포함할 수 있다. 매트릭스부(210)는 제1 방향(D1)을 따라 제1 픽셀 토큰(PT1) 내지 제N 픽셀 토큰(PTN)을 순서대로 배치할 수 있고, 제1 방향(D1)과 수직인 제2 방향(D2)을 따라 제1 단어 토큰(WT1) 내지 제K 단어 토큰(WTK)을 순서대로 배치할 수 있다. 위와 같이 복수의 픽셀 토큰들(PT) 및 복수의 단어 토큰들(WT)을 배치하면, 미리 정해진 유사도 결정방식에 따라 산출되는 복수의 픽셀 토큰들(PT) 및 복수의 단어 토큰들(WT) 사이의 결과점수(RP)를 설정할 수 있는 점수 매트릭스(SM)를 구현할 수 있다.

[0038] 점수 결정부(220)는 복수의 픽셀 토큰들(PT) 및 복수의 단어 토큰들(WT)에 따라 점수 매트릭스(SM)에 포함되는 점수픽셀들(SP) 각각의 결과점수(RP)를 결정할 수 있다. 예를 들어, 점수 매트릭스(SM)는 복수의 점수픽셀들(SP)을 포함할 수 있다. 복수의 점수픽셀들(SP)은 제1 점수픽셀(SP1) 내지 제Q 점수픽셀(SPQ)을 포함할 수 있다. 제1 점수픽셀(SP1)에는 미리 정해진 유사도 결정방식에 따라 제1 픽셀 토큰(PT1) 및 제1 단어 토큰(WT1) 사이의 유사도를 점수화하여 제1 결과점수(RP1)가 저장될 수 있다. 이와 동일한 방식으로 제2 점수픽셀(SP2) 내지 제Q 점수픽셀(SPQ) 각각에 제2 결과점수(RP2) 내지 제Q 결과점수(RPQ)가 저장될 수 있다.

[0039] 추출부(300)는 결과점수(RP)에 따라 결정되는 비쥬얼 익스프레션 가이드(VEG)에 기초하여 타겟 세그먼트(TS)를 추출할 수 있다.

[0040] 일 실시예에 있어서, 추출부(300)는 픽셀 추출부(310)를 더 포함할 수 있다. 픽셀 추출부(310)는 점수픽셀들

(SP) 각각의 결과점수(RP)를 기준으로 미리 정해진 일정한 간격(K)에 배치되는 점수픽셀들(SP)을 추출하여 추출픽셀들(EP)을 제공할 수 있다. 예를 들어, 점수 매트릭스(SM)에 포함되는 결과점수(RP) 중 가장 높은 점수가 배치되는 영역은 제1 영역(RE1)일 수 있다. 이 경우, 픽셀 추출부(310)는 제1 영역(RE1)에 포함되는 4개의 점수픽셀들(SP)을 선택할 수 있다. 이 경우, 제1 영역(RE1)에 포함되는 4개의 점수픽셀들(SP)은 제1 추출픽셀(EP1) 내지 제4 추출픽셀(EP4)일 수 있다. 여기서, 미리 정해진 간격(K)은 2개의 점수픽셀로 설정하고 있으나, 본 발명은 이에 한정되지 않고, 다양한 방식으로 간격(K)을 설정할 수 있다.

[0041] 일 실시예에 있어서, 추출부(300)는 합산 점수부(320)를 더 포함할 수 있다. 합산 점수부(320)는 추출픽셀들(EP)에 상응하는 결과점수(RP)를 제1 방향(D1)으로 합산하여 합산점수(HP)를 제공할 수 있다. 예를 들어, 복수의 추출픽셀들(EP)은 제1 추출픽셀(EP1) 내지 제4 추출픽셀(EP4)을 포함할 수 있다. 이 경우, 합산 점수부(320)는 제1 추출픽셀(EP1)에 상응하는 결과점수(RP) 및 제2 추출픽셀(EP2)에 상응하는 결과점수(RP)를 제1 방향(D1)으로 합하여 제1 합산점수(HP1)를 제공할 수 있고, 또한 제3 추출픽셀(EP3)에 상응하는 결과점수(RP) 및 제4 추출픽셀(EP4)에 상응하는 결과점수(RP)를 제1 방향(D1)으로 합하여 제2 합산점수(HP2)를 제공할 수 있다.

[0042] 일 실시예에 있어서, 추출부(300)는 가이드 제공부(330)를 더 포함할 수 있다. 가이드 제공부(330)는 합산점수(HP) 중 가장 높은 점수에 해당하는 최고점수에 상응하는 입력 이미지(IIM)에 포함되는 픽셀들(PX)의 정보를 비쥬얼 익스프레션 가이드(VEG)로 제공할 수 있다. 예를 들어, 가이드 제공부(330)는 합산점수(HP) 중 가장 높은 최고점수인 제1 영역(RE1)에 상응하는 픽셀들(PX)의 정보를 비쥬얼 익스프레션 가이드(VEG)로서 타겟 제공부(340)에 전달할 수 있다.

[0043] 일 실시예에 있어서, 추출부(300)는 타겟 제공부(340)를 더 포함할 수 있다. 타겟 제공부(340)는 비쥬얼 익스프레션 가이드(VEG) 및 입력 이미지(IIM)에 기초하여 타겟 세그먼트(TS)를 제공할 수 있다. 본 발명에 따른 타겟 제공부(340)는 빅데이터를 이용하여 학습한 상태로 입력 이미지(IIM) 및 비쥬얼 익스프레션 가이드(VEG)가 입력되는 경우, 입력 이미지(IIM)에서 타겟 세그먼트(TS)를 추출할 수 있다.

[0044] 본 발명에 따른 타겟 추출장치(10)는 입력 이미지(IIM)에 따라 생성되는 픽셀 토큰들(PT) 및 입력 언어에 따라 생성되는 단어 토큰들(WT)에 기초하여 결정되는 비쥬얼 익스프레션 가이드(VEG)에 따라 입력 이미지(IIM)에서 타겟 세그먼트(TS)를 제공함으로써 타겟영역에 포함되는 객체를 보다 효과적으로 추출할 수 있다.

[0045] 도 8은 도 1의 타겟 추출장치에 포함되는 추출부의 다른 예를 나타내는 도면이고, 도 9는 도 1의 타겟 추출장치에 포함되는 픽셀 추출부의 일 예를 나타내는 도면이고, 도 10은 도 9의 픽셀 추출부의 동작을 설명하기 위한 도면이다.

[0046] 도 1 내지 10을 참조하면, 일 실시예에 있어서, 픽셀 추출부(310)는 후보 추출부(311)를 더 포함할 수 있다. 후보 추출부(311)는 점수픽셀들(SP) 각각의 결과점수(RP) 및 미리 정해진 기준점수(REP)를 비교하여 후보픽셀들(HBP)을 제공할 수 있다. 예를 들어, 복수의 점수픽셀들(SP)은 제1 점수픽셀(SP1) 내지 제Q 점수픽셀(SPQ)들을 포함할 수 있다. 제1 점수픽셀(SP1) 내지 제Q 점수픽셀(SPQ)에 상응하는 결과점수(RP)는 제1 결과점수(RP1) 내지 제Q 결과점수(RPQ)일 수 있다. 또한, 미리 정해진 기준점수(REP)는 60점일 수 있다. 이 경우, 후보 추출부(311)는 제1 결과점수(RP1) 내지 제Q 결과점수(RPQ)를 기준점수(REP) 60점과 비교할 수 있고, 제1 영역(RE1) 및 제2 영역(RE2)에 배치되는 점수픽셀들(SP)의 결과점수(RP)가 기준점수(REP)보다 높을 수 있다. 이 경우, 후보 추출부(311)는 제1 영역(RE1)에 배치되는 점수픽셀들(SP)을 제1 후보픽셀(HBP1) 내지 제4 후보픽셀(HBP4)로 선택할 수 있고, 제2 영역(RE2)에 배치되는 점수픽셀을 제5 후보픽셀로 선택할 수 있다.

[0047] 일 실시예에 있어서, 픽셀 추출부(310)는 결정부(312)를 더 포함할 수 있다. 결정부(312)는 후보픽셀들(HBP)이 점수 매트릭스(SM) 상에 복수의 영역들에 분산되어 배치되는 경우, 복수의 영역들 각각에 배치되는 후보픽셀들(HBP)의 수에 따라 추출픽셀들(EP)을 결정할 수 있다.

[0048] 다른 일 실시예에 있어서, 결정부(312)는 복수의 영역들 중 후보픽셀들(HBP)의 수가 가장 많이 배치되는 영역의 후보픽셀들(HBP)을 추출픽셀들(EP)로 결정할 수 있다. 예를 들어, 복수의 영역들은 제1 영역(RE1) 및 제2 영역(RE2)을 포함할 수 있다. 제1 영역(RE1)에는 4개의 후보픽셀들(HBP)이 배치될 수 있고, 제2 영역(RE2)에는 1개의 후보픽셀이 배치될 수 있다. 이 경우, 결정부(312)는 제2 영역(RE2)보다 후보픽셀들(HBP)의 수가 많은 제1 영역(RE1)에 배치되는 제1 후보픽셀(HBP1) 내지 제4 후보픽셀(HBP4)을 추출픽셀들(EP)로 결정할 수 있다.

[0049] 도 11은 본 발명의 실시예들에 따른 타겟 추출 시스템을 나타내는 도면이고, 도 12는 도 11의 타겟 추출 시스템에 포함되는 표시부를 설명하기 위한 도면이다.

[0050] 도 1 내지 12를 참조하면, 본 발명의 실시예에 따른 타겟 추출 시스템은 인코더(100), 점수 제공부(200), 추출

부(300) 및 표시부(400)를 포함할 수 있다. 인코더(100)는 입력 이미지(IIM) 및 입력 이미지(IIM)에 포함되는 타겟 세그먼트(TS)에 상응하는 언어 입력(IL)에 기초하여 복수의 픽셀 토큰들(PT) 및 복수의 단어 토큰들(WT)을 제공할 수 있다. 점수 제공부(200)는 복수의 픽셀 토큰들(PT) 및 복수의 단어 토큰들(WT)에 따라 결정되는 결과점수(RP)를 제공할 수 있다. 추출부(300)는 결과점수(RP)에 따라 결정되는 비쥬얼 익스프레션 가이드(VEG)에 따라 타겟 세그먼트(TS)를 추출할 수 있다. 표시부(400)는 입력 이미지(IIM)에 타겟 세그먼트(TS)를 표시할 수 있다. 도 12에 도시되는 바와 같이, 표시부(400)는 입력 이미지(IIM)에 포함되는 타겟 세그먼트(TS)를 색상을 달리하여 표시할 수도 있다.

[0051] 일 실시예에 있어서, 인코더(100)는 비전 인코더(110) 및 언어 인코더(120)를 포함할 수 있다. 비전 인코더(110)는 입력 이미지(IIM)에 포함되는 복수의 픽셀들(PX)을 기준으로 복수의 픽셀들(PX) 각각의 정보를 포함하는 복수의 픽셀 토큰들(PT)을 제공할 수 있다. 언어 인코더(120)는 언어 입력(IL)에 포함되는 복수의 단어들을 기준으로 복수의 단어들 각각의 정보를 포함하는 복수의 단어 토큰들(WT)을 제공할 수 있다.

[0052] 도 13은 본 발명의 실시예들에 따른 타겟 추출장치의 동작방법을 나타내는 순서도이다.

[0053] 도 1 내지 13을 참조하면, 본 발명의 실시예에 따른 타겟 추출장치(10)의 동작방법에서는, 인코더(100)가 입력 이미지(IIM) 및 입력 이미지(IIM)에 포함되는 타겟 세그먼트(TS)에 상응하는 언어 입력(IL)에 기초하여 복수의 픽셀 토큰들(PT) 및 복수의 단어 토큰들(WT)을 제공할 수 있다(S100). 점수 제공부(200)가 복수의 픽셀 토큰들(PT) 및 복수의 단어 토큰들(WT)에 따라 결정되는 결과점수(RP)를 제공할 수 있다(S200). 추출부(300)가 결과점수(RP)에 따라 결정되는 비쥬얼 익스프레션 가이드(VEG)에 따라 타겟 세그먼트(TS)를 추출할 수 있다(S300).

[0054] 본 발명에 따른 타겟 추출장치(10)는 입력 이미지(IIM)에 따라 생성되는 픽셀 토큰들(PT) 및 입력 언어에 따라 생성되는 단어 토큰들(WT)에 기초하여 결정되는 비쥬얼 익스프레션 가이드(VEG)에 따라 입력 이미지(IIM)에서 타겟 세그먼트(TS)를 제공함으로써 타겟영역에 포함되는 객체를 보다 효과적으로 추출할 수 있다.

[0055] 도 14는 본 발명의 실시예들에 따른 타겟 추출 시스템의 동작방법을 나타내는 순서도이다.

[0056] 도 1 내지 14를 참조하면, 본 발명의 실시예에 따른 타겟 추출 시스템의 동작방법에서는, 인코더(100)가 입력 이미지(IIM) 및 입력 이미지(IIM)에 포함되는 타겟 세그먼트(TS)에 상응하는 언어 입력(IL)에 기초하여 복수의 픽셀 토큰들(PT) 및 복수의 단어 토큰들(WT)을 제공할 수 있다(S100). 점수 제공부(200)가 복수의 픽셀 토큰들(PT) 및 복수의 단어 토큰들(WT)에 따라 결정되는 결과점수(RP)를 제공할 수 있다(S200). 추출부(300)가 결과점수(RP)에 따라 결정되는 비쥬얼 익스프레션 가이드(VEG)에 따라 타겟 세그먼트(TS)를 추출할 수 있다(S300). 표시부(400)가 입력 이미지(IIM)에 타겟 세그먼트(TS)를 표시할 수 있다(S400).

[0057] 위에서 언급된 본 발명의 기술적 과제 외에도, 본 발명의 다른 특징 및 이점들이 이하에서 기술되거나, 그러한 기술 및 설명으로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

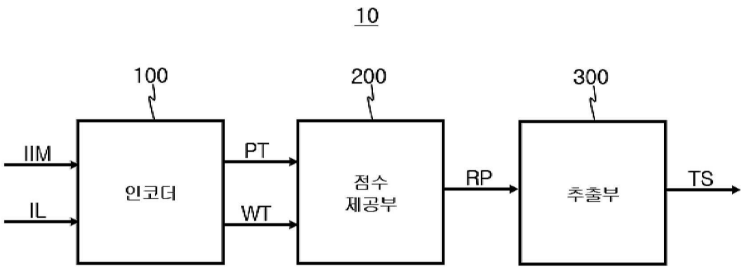
부호의 설명

[0058] 10: 타겟 추출장치 100: 인코더

200: 점수 제공부 300: 추출부

도면

도면1



도면2

입력 이미지 (IIM)

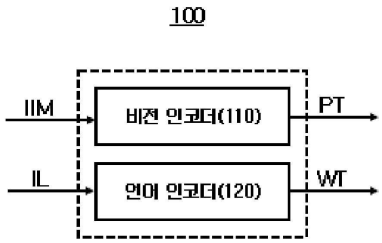


도면3

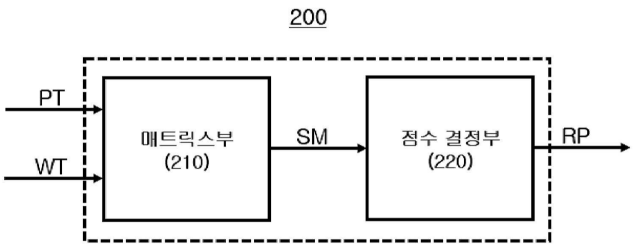
언어 입력 (IL)

“ uniform number 10”

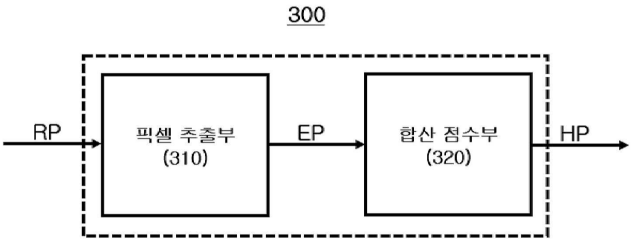
도면4



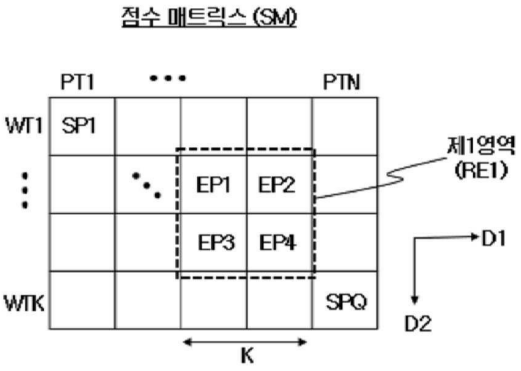
도면5



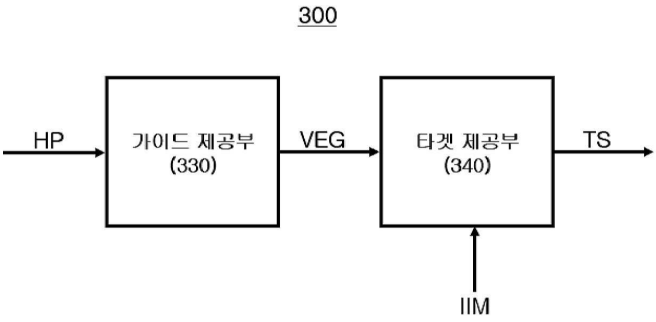
도면6



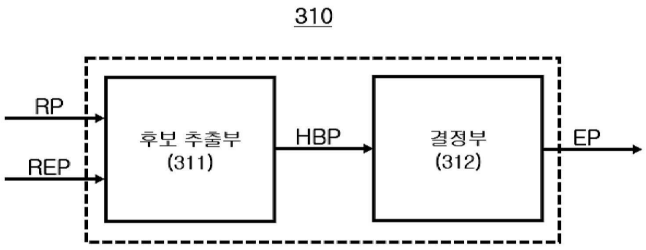
도면7



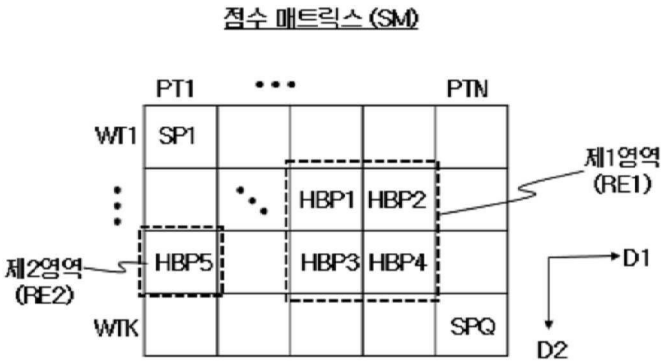
도면8



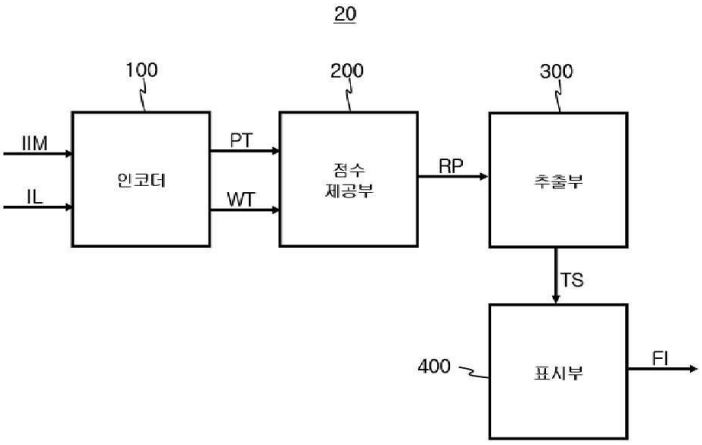
도면9



도면10



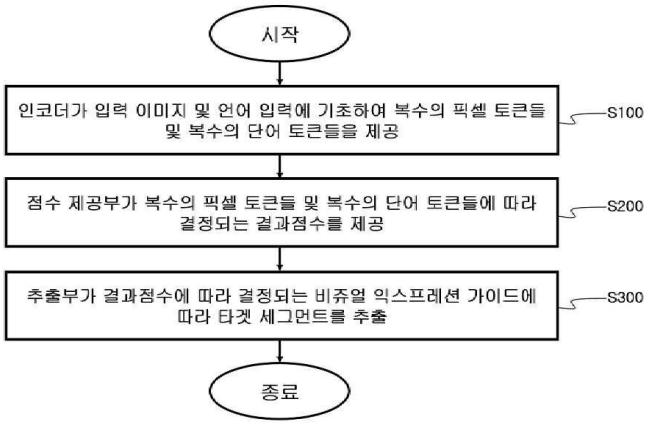
도면11



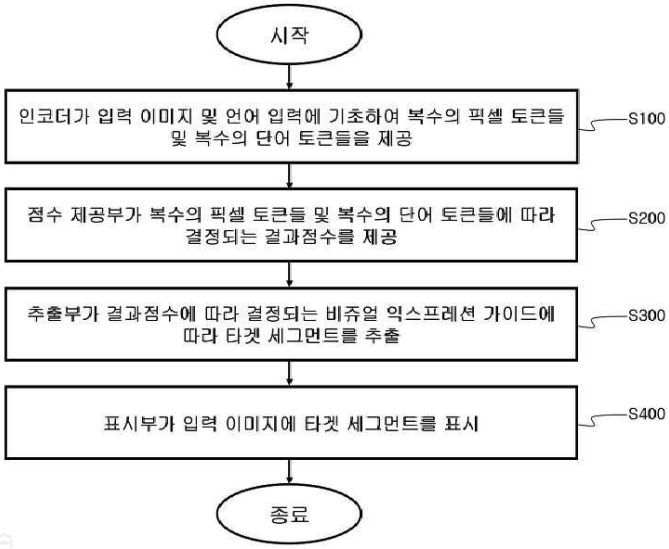
도면12



도면13


시작
인코더가 입력 이미지 및 언어 입력에 기초하여 복수의 픽셀 토큰들 및 복수의 단어 토큰들을 제공
S100
점수 제공부가 복수의 픽셀 토큰들 및 복수의 단어 토큰들에 따라 결정되는 결과점수를 제공
S200
추출부가 결과점수에 따라 결정되는 비쥬얼 익스프레션 가이드에 따라 타겟 세그먼트를 추출
S300
종료

도면14


시작
인코더가 입력 이미지 및 언어 입력에 기초하여 복수의 픽셀 토큰들 및 복수의 단어 토큰들을 제공
S100
점수 제공부가 복수의 픽셀 토큰들 및 복수의 단어 토큰들에 따라 결정되는 결과점수를 제공
S200
추출부가 결과점수에 따라 결정되는 비쥬얼 익스프레션 가이드에 따라 타겟 세그먼트를 추출
S300
표시부가 입력 이미지에 타겟 세그먼트를 표시
S400
종료