



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0029622
(43) 공개일자 2025년03월05일

- | | |
|--|---|
| <p>(51) 국제특허분류(Int. Cl.)
 <i>G06N 3/0455</i> (2023.01) <i>G06N 3/0475</i> (2023.01)
 <i>G06N 3/09</i> (2023.01) <i>G06N 3/094</i> (2023.01)
 <i>G06V 10/82</i> (2022.01) <i>G06V 20/40</i> (2022.01)</p> <p>(52) CPC특허분류
 <i>G06N 3/0455</i> (2023.01)
 <i>G06N 3/0475</i> (2023.01)</p> <p>(21) 출원번호 10-2023-0110754
 (22) 출원일자 2023년08월23일
 심사청구일자 2023년08월23일</p> | <p>(71) 출원인
 경희대학교 산학협력단
 경기도 용인시 기흥구 덕영대로 1732 (서천동, 경희대학교 국제캠퍼스내)</p> <p>(72) 발명자
 김성태
 경기도 화성시 동탄공원로 21-39, 963동 701호 (능동, 동탄푸른마을신일해피트리)</p> <p>김정욱
 경기도 안양시 동안구 귀인로 258, 106동 802호 (평촌동, 꿈마을라이프아파트
 (뒷면에 계속))</p> <p>(74) 대리인
 두호특허법인</p> |
|--|---|

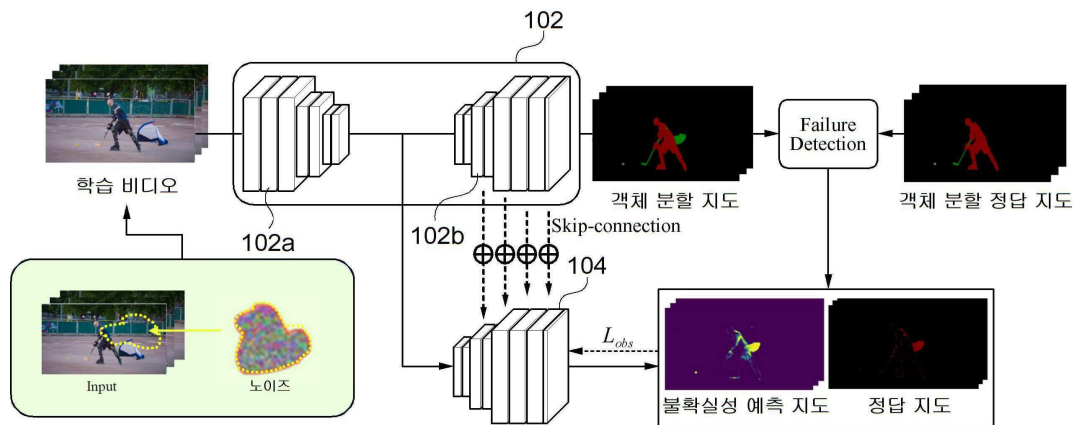
전체 청구항 수 : 총 19 항

(54) 발명의 명칭 **심층 신경망의 불확실성 예측 장치 및 방법**

(57) 요약

심층 신경망의 불확실성 예측 장치 및 방법이 개시된다. 개시되는 일 실시예에 따른 심층 신경망의 불확실성 예측 장치는, 입력 데이터를 기반으로 기 설정된 태스크를 수행하도록 학습된 제1 기계 학습 모델 및 제1 기계 학습 모델의 태스크 수행 결과에 대한 불확실성을 예측하도록 학습되는 제2 기계 학습 모델을 포함한다.

대표도



(52) CPC특허분류

G06N 3/09 (2023.01)
G06N 3/094 (2023.01)
G06V 10/82 (2022.01)
G06V 20/41 (2022.01)

조은기

경기도 안양시 동안구 동편로 135, 408동 (관양동, 동편마을)

(72) 발명자

김민국

부산광역시 북구 만덕대로155번길 15, 덕천동 904호 (덕천동, 삼성그린코아)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711179316
과제번호	00155911
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	인공지능융합혁신인재양성
연구과제명	인공지능융합혁신인재양성(경희대학교)
기 여 율	1/1
과제수행기관명	경희대학교산학협력단
연구기간	2023.01.01 ~ 2023.12.31

명세서

청구범위

청구항 1

입력 데이터를 기반으로 기 설정된 태스크를 수행하도록 학습된 제1 기계 학습 모델; 및

상기 제1 기계 학습 모델의 태스크 수행 결과에 대한 불확실성을 예측하도록 학습되는 제2 기계 학습 모델을 포함하는, 심층 신경망의 불확실성 예측 장치.

청구항 2

청구항 1에 있어서,

상기 제1 기계 학습 모델은,

입력되는 비디오에서 특징을 추출하도록 마련되는 인코더 및 상기 추출된 특징을 기초로 상기 비디오 내 각 객체들을 분할한 객체 분할 지도를 생성하도록 마련되는 디코더를 포함하며,

상기 제2 기계 학습 모델은, 상기 디코더와 동일한 심층 신경망 구조로 이루어지는, 심층 신경망의 불확실성 예측 장치.

청구항 3

청구항 2에 있어서,

상기 불확실성 예측 장치는,

상기 제1 기계 학습 모델의 학습이 완료된 상태에서 상기 제2 기계 학습 모델을 학습하되,

학습 비디오를 생성한 후, 생성한 상기 학습 비디오를 상기 제1 기계 학습 모델의 인코더로 입력하여 특징을 추출하도록 하고, 추출된 상기 특징을 각각 상기 제1 기계 학습 모델의 디코더 및 상기 제2 기계 학습 모델로 입력하는, 심층 신경망의 불확실성 예측 장치.

청구항 4

청구항 3에 있어서,

상기 학습 비디오는,

원래의 비디오에 노이즈가 추가되어 상기 기 학습된 제1 기계 학습 모델에서 객체 분할 예측이 틀리도록 하는, 심층 신경망의 불확실성 예측 장치.

청구항 5

청구항 4에 있어서,

상기 불확실성 예측 장치는,

상기 원래의 비디오에서 객체 영역 부근에 노이즈에 해당하는 공격 조각을 추가하여 상기 학습 비디오를 생성하는, 심층 신경망의 불확실성 예측 장치.

청구항 6

청구항 4에 있어서,

상기 디코더는, 상기 추출된 특징에 기초하여 상기 학습 비디오에서 각 객체들을 분할한 객체 분할 지도를 생성하고,

상기 제2 기계 학습 모델은, 상기 추출된 특징에 기초하여 상기 객체 분할 지도에서 틀린 픽셀들이 어느 부분인지를 나타내는 불확실성 예측 지도를 생성하는, 심층 신경망의 불확실성 예측 장치.

청구항 7

청구항 6에 있어서,

상기 불확실성 예측 장치는,

상기 디코더에서 생성한 객체 분할 지도와 해당 학습 비디오에 대한 객체 분할 정답 지도를 비교하여 상기 디코더의 예측이 틀린 픽셀들로 구성된 객체 분할 지도를 상기 제2 기계 학습 모델의 정답 지도로 설정하는, 심층 신경망의 불확실성 예측 장치.

청구항 8

청구항 7에 있어서,

상기 불확실성 예측 장치는,

상기 제2 기계 학습 모델에서 생성하는 불확실성 예측 지도가 상기 정답 지도를 닮아가도록 상기 제2 기계 학습 모델을 학습하는, 심층 신경망의 불확실성 예측 장치.

청구항 9

청구항 8에 있어서,

상기 제2 기계 학습 모델의 손실 함수(L_{obs})는 아래 수학식으로 표현되는, 심층 신경망의 불확실성 예측 장치.

(수학식)

$$L_{obs} = (y_{obs} - 1) \log(1 - Observer(x, Seg(x)))$$

y_{obs} : 제2 기계 학습 모델의 정답 지도

x : 학습 비디오

$Seg(x)$: 학습 비디오가 입력되는 경우 제1 기계 학습 모델에서 예측한 객체 분할 지도

$Observer$: 제2 기계 학습 모델에서 생성되는 불확실성 예측 지도

청구항 10

하나 이상의 프로세서들, 및

상기 하나 이상의 프로세서들에 의해 실행되는 하나 이상의 프로그램들을 저장하는 메모리를 구비한 컴퓨팅 장치에서 수행되는 방법으로서,

입력 데이터를 기반으로 기 설정된 태스크를 수행하도록 제1 기계 학습 모델을 학습시키는 단계; 및

상기 제1 기계 학습 모델의 태스크 수행 결과에 대한 불확실성을 예측하도록 제2 기계 학습 모델을 학습시키는 단계를 포함하는, 심층 신경망의 불확실성 예측 방법.

청구항 11

청구항 10에 있어서,

상기 제1 기계 학습 모델은,

입력되는 비디오에서 특징을 추출하도록 마련되는 인코더 및 상기 추출된 특징을 기초로 상기 비디오 내 각 객체들을 분할한 객체 분할 지도를 생성하도록 마련되는 디코더를 포함하며,

상기 제2 기계 학습 모델은, 상기 디코더와 동일한 심층 신경망 구조로 이루어지는, 심층 신경망의 불확실성 예측 방법.

청구항 12

청구항 11에 있어서,

상기 컴퓨팅 장치는, 상기 제1 기계 학습 모델의 학습이 완료된 상태에서 상기 제2 기계 학습 모델을 학습하며,

상기 제2 기계 학습 모델을 학습하는 단계는,

학습 비디오를 생성한 후, 생성한 상기 학습 비디오를 상기 제1 기계 학습 모델의 인코더로 입력하여 특징을 추출하도록 하고, 추출된 상기 특징을 각각 상기 제1 기계 학습 모델의 디코더 및 상기 제2 기계 학습 모델로 입력하는, 심층 신경망의 불확실성 예측 방법.

청구항 13

청구항 12에 있어서,

상기 학습 비디오는,

원래의 비디오에 노이즈가 추가되어 상기 기 학습된 제1 기계 학습 모델에서 객체 분할 예측이 틀리도록 하는, 심층 신경망의 불확실성 예측 방법.

청구항 14

청구항 13에 있어서,

상기 컴퓨팅 장치는,

상기 원래의 비디오에서 객체 영역 부근에 노이즈에 해당하는 공격 조각을 추가하여 상기 학습 비디오를 생성하는, 심층 신경망의 불확실성 예측 방법.

청구항 15

청구항 13에 있어서,

상기 디코더는, 상기 추출된 특징에 기초하여 상기 학습 비디오에서 각 객체들을 분할한 객체 분할 지도를 생성하고,

상기 제2 기계 학습 모델은, 상기 추출된 특징에 기초하여 상기 객체 분할 지도에서 틀린 픽셀들이 어느 부분인지를 나타내는 불확실성 예측 지도를 생성하는, 심층 신경망의 불확실성 예측 방법.

청구항 16

청구항 15에 있어서,

상기 컴퓨팅 장치는,

상기 디코더에서 생성한 객체 분할 지도와 해당 학습 비디오에 대한 객체 분할 정답 지도를 비교하여 상기 디코더의 예측이 틀린 픽셀들로 구성된 객체 분할 지도를 상기 제2 기계 학습 모델의 정답 지도로 설정하는, 심층 신경망의 불확실성 예측 방법.

청구항 17

청구항 16에 있어서,

상기 컴퓨팅 장치는,

상기 제2 기계 학습 모델에서 생성하는 불확실성 예측 지도가 상기 정답 지도를 닮아가도록 상기 제2 기계 학습 모델을 학습하는, 심층 신경망의 불확실성 예측 방법.

청구항 18

청구항 17에 있어서,

상기 제2 기계 학습 모델의 손실 함수(L_{obs})는 아래 수학적식으로 표현되는, 심층 신경망의 불확실성 예측 방법.

(수학적식)

$$L_{obs} = (y_{obs} - 1) \log(1 - Observer(x, Seg(x)))$$

y_{obs} : 제2 기계 학습 모델의 정답 지도

x : 학습 비디오

$Seg(x)$: 학습 비디오가 입력되는 경우 제1 기계 학습 모델에서 예측한 객체 분할 지도

$Observer$: 제2 기계 학습 모델에서 생성되는 불확실성 예측 지도

청구항 19

비일시적 컴퓨터 판독 가능한 저장 매체(non-transitory computer readable storage medium)에 저장된 컴퓨터 프로그램으로서,

상기 컴퓨터 프로그램은 하나 이상의 명령어들을 포함하고, 상기 명령어들은 하나 이상의 프로세서들을 갖는 컴퓨팅 장치에 의해 실행될 때, 상기 컴퓨팅 장치로 하여금,

입력 데이터를 기반으로 기 설정된 태스크를 수행하도록 제1 기계 학습 모델을 학습시키는 단계; 및

상기 제1 기계 학습 모델의 태스크 수행 결과에 대한 불확실성을 예측하도록 제2 기계 학습 모델을 학습시키는 단계를 수행하도록 하는, 비일시적 컴퓨터 판독 가능한 저장 매체에 저장된 컴퓨터 프로그램.

발명의 설명

기술 분야

[0001] 본 발명의 실시예는 심층 신경망의 불확실성 예측 기술과 관련된다.

배경 기술

[0003] 뉴럴 네트워크의 불확실성 예측을 목적으로 하는 불확실성 예측(Uncertainty Estimation) 기술은 이미지 인식, 의미적 분할, 비디오 객체 분할 등 컴퓨터 비전 분야에서 활발히 연구되고 있다. 불확실성 예측 기술은 특히 뉴럴 네트워크의 결정이 매우 치명적인 영향을 끼칠 수 있는 분야(예를 들어, 의료 수술 비디오 및 자율주행 비디오 등)에서 중요성이 대두되고 있다.

[0004] 현재 뉴럴 네트워크의 발전으로, 뉴럴 네트워크의 예측 정확도가 매우 높은 수준으로 향상되었으나, 여전히 뉴럴 네트워크의 블랙박스(black-box) 특징으로 인해 뉴럴 네트워크의 의사 결정 과정이 완벽히 설명되지 않으며, 따라서 그 최종 의사 결정을 완벽히 신뢰할 수는 없다. 이러한 뉴럴 네트워크의 의사 결정의 불확실성을 탐지할 수 있다면, 뉴럴 네트워크의 예측이 불확실하거나 확신도가 낮은 경우 최종 사용자와의 상호 작용을 통해 상황을 개선시킬 수 있다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) 한국등록특허공보 제10-2200212호(2021.01.08)

발명의 내용

해결하려는 과제

[0007] 본 발명의 실시예는 심층 신경망의 불확실성 예측을 위한 새로운 기법을 제공하기 위한 것이다.

과제의 해결 수단

[0009] 개시되는 일 실시예에 따른 심층 신경망의 불확실성 예측 장치는, 입력 데이터를 기반으로 기 설정된 태스크를 수행하도록 학습된 제1 기계 학습 모델; 및 상기 제1 기계 학습 모델의 태스크 수행 결과에 대한 불확실성을 예측하도록 학습되는 제2 기계 학습 모델을 포함한다.

[0010] 상기 제1 기계 학습 모델은, 입력되는 비디오에서 특징을 추출하도록 마련되는 인코더 및 상기 추출된 특징을 기초로 상기 비디오 내 각 객체들을 분할한 객체 분할 지도를 생성하도록 마련되는 디코더를 포함하며, 상기 제2 기계 학습 모델은, 상기 디코더와 동일한 심층 신경망 구조로 이루어질 수 있다.

[0011] 상기 불확실성 예측 장치는, 상기 제1 기계 학습 모델의 학습이 완료된 상태에서 상기 제2 기계 학습 모델을 학습하되, 학습 비디오를 생성한 후, 생성한 상기 학습 비디오를 상기 제1 기계 학습 모델의 인코더로 입력하여 특징을 추출하도록 하고, 추출된 상기 특징을 각각 상기 제1 기계 학습 모델의 디코더 및 상기 제2 기계 학습 모델로 입력할 수 있다.

[0012] 상기 학습 비디오는, 원래의 비디오에 노이즈가 추가되어 상기 기 학습된 제1 기계 학습 모델에서 객체 분할 예측이 틀리도록 할 수 있다.

[0013] 상기 불확실성 예측 장치는, 상기 원래의 비디오에서 객체 영역 부근에 노이즈에 해당하는 공격 조각을 추가하여 상기 학습 비디오를 생성할 수 있다.

[0014] 상기 디코더는, 상기 추출된 특징에 기초하여 상기 학습 비디오에서 각 객체들을 분할한 객체 분할 지도를 생성하고, 상기 제2 기계 학습 모델은, 상기 추출된 특징에 기초하여 상기 객체 분할 지도에서 틀린 픽셀들이 어느 부분인지를 나타내는 불확실성 예측 지도를 생성할 수 있다.

[0015] 상기 불확실성 예측 장치는, 상기 디코더에서 생성한 객체 분할 지도와 해당 학습 비디오에 대한 객체 분할 정

답 지도를 비교하여 상기 디코더의 예측이 틀린 픽셀들로 구성된 객체 분할 지도를 상기 제2 기계 학습 모델의 정답 지도로 설정할 수 있다.

[0016] 상기 불확실성 예측 장치는, 상기 제2 기계 학습 모델에서 생성하는 불확실성 예측 지도가 상기 정답 지도를 닮아가도록 상기 제2 기계 학습 모델을 학습할 수 있다.

[0017] 상기 제2 기계 학습 모델의 손실 함수(L_{obs})는 아래 수학식으로 표현될 수 있다.

[0018] (수학식)

$$L_{obs} = (y_{obs} - 1) \log(1 - Observer(x, Seg(x)))$$

[0019] y_{obs} : 제2 기계 학습 모델의 정답 지도

[0020] x : 학습 비디오

[0021] $Seg(x)$: 학습 비디오가 입력되는 경우 제1 기계 학습 모델에서 예측한 객체 분할 지도

[0022] $Observer$: 제2 기계 학습 모델에서 생성되는 불확실성 예측 지도

[0023] 개시되는 일 실시예에 따른 심층 신경망의 불확실성 예측 방법은, 하나 이상의 프로세서들, 및 상기 하나 이상의 프로세서들에 의해 실행되는 하나 이상의 프로그램들을 저장하는 메모리를 구비한 컴퓨팅 장치에서 수행되는 방법으로서, 입력 데이터를 기반으로 기 설정된 태스크를 수행하도록 제1 기계 학습 모델을 학습시키는 단계; 및 상기 제1기계 학습 모델의 태스크 수행 결과에 대한 불확실성을 예측하도록 제2 기계 학습 모델을 학습시키는 단계를 포함한다.

발명의 효과

[0024] 개시되는 실시예에 의하면, 제1 기계 학습 모델의 객체 분할 결과의 불확실성을 예측하기 위한 제2 기계 학습 모델을 구성함으로써, 제1 기계 학습 모델에 대한 신뢰도를 제공할 수 있게 되고 그로 인해 심층 신경망 모델의 결정이 중요한 분야에서 그 활용도를 높일 수 있게 된다.

도면의 간단한 설명

[0025] 도 1은 본 발명의 일 실시예에 따른 심층 신경망의 불확실성 예측 장치를 나타낸 도면

도 2는 본 발명의 일 실시예에 따른 심층 신경망의 불확실성 예측 장치의 학습 과정을 나타낸 도면

도 3은 본 발명의 일 실시예에 따른 심층 신경망의 불확실성 예측 장치에서 적대적 학습 기반으로 학습 비디오를 생성하는 과정을 나타낸 도면

도 4는 본 발명의 일 실시예에 따른 심층 신경망의 불확실성 예측 방법을 나타낸 흐름도

도 5는 예시적인 실시예들에서 사용되기에 적합한 컴퓨팅 장치를 포함하는 컴퓨팅 환경을 예시하여 설명하기 위한 블록도

발명을 실시하기 위한 구체적인 내용

[0026] 이하, 도면을 참조하여 본 발명의 구체적인 실시형태를 설명하기로 한다. 이하의 상세한 설명은 본 명세서에서 기술된 방법, 장치 및/또는 시스템에 대한 포괄적인 이해를 돕기 위해 제공된다. 그러나 이는 예시에 불과하며 본 발명은 이에 제한되지 않는다.

[0027] 본 발명의 실시예들을 설명함에 있어서, 본 발명과 관련된 공지기술에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략하기로 한다. 그리고, 후술되는 용어들은 본 발명에서의 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. 상세한 설명에서 사용되는 용어는 단지 본 발명의 실시예들을 기술하기 위한 것이며, 결코 제한적이어서는 안 된다. 명확하게 달리 사

용되지 않는 한, 단수 형태의 표현은 복수 형태의 의미를 포함한다. 본 설명에서, "포함" 또는 "구비"와 같은 표현은 어떤 특성들, 숫자들, 단계들, 동작들, 요소들, 이들의 일부 또는 조합을 가리키기 위한 것이며, 기술된 것 이외에 하나 또는 그 이상의 다른 특성, 숫자, 단계, 동작, 요소, 이들의 일부 또는 조합의 존재 또는 가능성을 배제하도록 해석되어서는 안 된다.

- [0031] 또한, 제1, 제2 등의 용어는 다양한 구성 요소들을 설명하는데 사용될 수 있지만, 상기 구성 요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성 요소를 다른 구성 요소로부터 구별하는 목적으로 사용될 수 있다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제1 구성 요소는 제2 구성 요소로 명명될 수 있고, 유사하게 제2 구성 요소도 제1 구성 요소로 명명될 수 있다.
- [0033] 도 1은 본 발명의 일 실시예에 따른 심층 신경망의 불확실성 예측 장치를 나타낸 도면이다.
- [0034] 도 1을 참조하면, 심층 신경망의 불확실성 예측 장치(100)는 제1 기계 학습 모델(102) 및 제2 기계 학습 모델(104)을 포함할 수 있다.
- [0035] 제1 기계 학습 모델(102)은 주어지는 입력 데이터를 기반으로 기 설정된 태스크를 수행하도록 학습된 모델이다. 일 실시예에서, 제1 기계 학습 모델(102)은 입력되는 비디오를 기반으로 비디오 내의 객체 분할(object segmentation)을 수행하도록 학습된 모델(비디오 객체 분할 모델)일 수 있다. 즉, 제1 기계 학습 모델(102)은 비디오가 입력되는 경우, 입력된 비디오에서 객체를 검출하여 분할하도록 기 학습된 모델일 수 있다.
- [0036] 제1 기계 학습 모델(102)은 인코더(102a) 및 디코더(102b)를 포함할 수 있다. 인코더(102a)는 입력되는 비디오에서 특징을 추출할 수 있다. 디코더(102b)는 인코더(102a)로부터 출력되는 특징에 기초하여 비디오 내의 객체 부분을 예측할 수 있다. 디코더(102b)는 인코더(102a)로부터 출력되는 특징에 기초하여 비디오 내의 각 객체들을 분할한 객체 분할 지도를 생성할 수 있다. 여기서, 제1 기계 학습 모델(102)은 비디오의 이전 프레임들에서 중요한 정보들을 저장하는 메모리를 더 포함할 수 있다.
- [0037] 제2 기계 학습 모델(104)은 제1 기계 학습 모델(102)의 불확실성을 예측하기 위한 모델이다. 즉, 제2 기계 학습 모델(104)은 제1 기계 학습 모델(102)의 태스크 수행 결과가 어느 정도 불확실성을 나타내는지를 예측하기 위한 모델일 수 있다.
- [0038] 일 실시예에서, 제1 기계 학습 모델(102)이 비디오 객체 분할 모델인 경우, 제2 기계 학습 모델(104)은 제1 기계 학습 모델(102)에서 예측한 객체 분할의 결과가 어느 정도 불확실성을 갖는지(또는 어느 정도 신뢰할 만한지)를 예측하기 위한 모델일 수 있다.
- [0039] 여기서, 제2 기계 학습 모델(104)은 제1 기계 학습 모델(102)의 디코더(102b) 구조를 모방하여 생성될 수 있다. 즉, 제2 기계 학습 모델(104)은 제1 기계 학습 모델(102)의 인코더(102a)는 공유하여 사용하고, 제1 기계 학습 모델(102)의 디코더(102b)와 동일한 구조의 심층 신경망으로 구성될 수 있다.
- [0040] 제2 기계 학습 모델(104)은 제1 기계 학습 모델(102)의 인코더(102a)에서 출력되는 특징을 입력(즉, 제1 기계 학습 모델(102)의 인코더(102a)를 공유함)으로 하여 제1 기계 학습 모델(102)의 불확실성을 예측하기 때문에, 불확실성 예측에 따른 시간 비용을 줄일 수 있게 된다.
- [0041] 제2 기계 학습 모델(104)은 제1 기계 학습 모델(102)의 디코더(102b)와 스킵 커넥션(skip-connection) 되어 마련될 수 있다. 이로써, 제2 기계 학습 모델(104)이 제1 기계 학습 모델(102)의 의사 결정 과정을 관찰하도록 할 수 있다. 이때, 제2 기계 학습 모델(104)은 제1 기계 학습 모델(102)의 객체 분할 결과가 틀리는 상황을 관찰하도록 학습될 수 있다.
- [0042] 즉, 제2 기계 학습 모델(104)은 입력되는 비디오에서 객체 분할을 할 때 제1 기계 학습 모델(102)이 틀리게 예측한 픽셀(예를 들어, 객체에 해당하지 않는 픽셀을 객체에 해당하는 픽셀로 예측하거나 객체에 해당하는 픽셀을 객체에 해당하지 않는 픽셀로 예측하는 경우)들을 정답으로 사용하여 학습될 수 있다. 이하, 도 2를 참조하여 심층 신경망의 불확실성 예측 장치(100)의 학습 과정을 살펴보기로 한다.
- [0044] 도 2는 본 발명의 일 실시예에 따른 심층 신경망의 불확실성 예측 장치의 학습 과정을 나타낸 도면이다.
- [0045] 도 2를 참조하면, 제2 기계 학습 모델(104)의 학습 과정은 제1 기계 학습 모델(102)의 학습이 완료된 상태에서

수행될 수 있다. 먼저, 불확실성 예측 장치(100)는 적대적 학습 기반으로 학습 데이터를 생성할 수 있다. 제1 기계 학습 모델(102)이 비디오 객체 분할 모델인 경우, 학습 데이터는 비디오 형태일 수 있다. 학습 비디오는 원래의 비디오에 사람 눈에 보이지 않는(invisible) 노이즈가 추가된 것일 수 있다.

[0046] 이에 따라, 학습 비디오는 제1 기계 학습 모델(102)에서 객체 분할이 잘 되지 못하도록 하는 역할을 하게 된다. 즉, 제1 기계 학습 모델(102)은 입력되는 비디오에서 객체들을 잘 분할하도록 학습된 상태이기 때문에, 제1 기계 학습 모델(102)에서 객체 분할에 대한 예측이 잘 안되도록 하기 위해 적대적 학습 기반의 학습 비디오를 생성하여 기 학습된 제1 기계 학습 모델(102)로 입력하게 된다. 제1 기계 학습 모델(102)의 입장에서 학습 비디오는 적대적 공격 샘플이 되게 된다. 학습 비디오를 생성하는 과정에 대한 자세한 설명은 후술하기로 한다.

[0047] 학습 비디오가 입력되는 경우, 제1 기계 학습 모델(102)의 인코더(102a)는 학습 비디오에서 특징을 추출할 수 있다. 추출된 특징은 각각 제1 기계 학습 모델(102)의 디코더(102b)와 제2 기계 학습 모델(104)로 입력되게 된다.

[0048] 제1 기계 학습 모델(102)의 디코더(102b)는 추출된 특징에 기반하여 학습 비디오에서 각 객체들을 분할한 객체 분할 지도를 생성할 수 있다. 불확실성 예측 장치(100)는 디코더(102b)에서 예측하는 객체 분할 지도와 해당 학습 비디오에 대한 객체 분할 정답 지도를 비교하여 디코더(102b)의 예측이 틀린 픽셀과 디코더(102b)의 예측이 맞춘 픽셀들을 구분할 수 있다.

[0049] 일 실시예에서, 불확실성 예측 장치(100)는 디코더(102b)에서 예측한 객체 분할 지도와 객체 분할 정답 지도를 비교하여 디코더(102b)의 예측이 틀린 픽셀은 1로 설정하고, 디코더(102b)의 예측이 맞는 픽셀은 0으로 설정할 수 있다. 이때, 불확실성 예측 장치(100)는 디코더(102b)의 예측이 틀린 픽셀들로 구성된 객체 분할 지도를 제2 기계 학습 모델(104)의 정답 지도로 설정할 수 있다.

[0050] 한편, 제1 기계 학습 모델(102)의 인코더(102a)에서 추출된 학습 비디오의 특징은 제2 기계 학습 모델(104)로도 입력된다. 제2 기계 학습 모델(104)은 추출된 특징에 기초하여 불확실성 예측 지도를 생성할 수 있다. 여기서, 불확실성 예측 지도는 제1 기계 학습 모델(102)의 디코더(102b)에서 예측한 객체 분할 지도에서 틀린 픽셀들이 어느 부분인지를 나타내는 것일 수 있다.

[0051] 즉, 제2 기계 학습 모델(104)은 제1 기계 학습 모델(102)의 디코더(102b)에서 예측한 객체 분할 지도에서 어느 부분이 틀린 픽셀인지를 예측하도록 학습될 수 있으며, 이를 위해 인코더(102a)에서 추출된 특징을 기초로 불확실성 예측 지도를 생성하게 된다. 이에 따라, 디코더(102b)의 예측이 틀린 픽셀들로 구성된 객체 분할 지도가 제2 기계 학습 모델(104)의 정답 지도로 설정되게 된다.

[0052] 이때, 불확실성 예측 장치(100)는 제2 기계 학습 모델(104)에서 출력하는 불확실성 예측 지도가 정답 지도(즉, 디코더(102b)의 예측이 틀린 픽셀들로 구성된 객체 분할 지도)를 닮아가도록(그 차이가 최소화 되도록) 제2 기계 학습 모델(104)을 학습할 수 있다. 제2 기계 학습 모델(104)의 손실 함수(L_{obs})는 아래 수학적 식 1로 나타낼 수 있다.

[0053] (수학적 식 1)

$$L_{obs} = (y_{obs} - 1) \log(1 - Observer(x, Seg(x)))$$

[0054]

[0055] y_{obs} : 제2 기계 학습 모델의 정답 지도

[0056] x : 학습 비디오

[0057] $Seg(x)$: 학습 비디오가 입력되는 경우 제1 기계 학습 모델에서 예측한 객체 분할 지도

[0058] $Observer$: 제2 기계 학습 모델에서 생성되는 불확실성 예측 지도

[0059] 이러한 과정으로 제2 기계 학습 모델(104)을 학습하게 되면, 학습이 완료된 제2 기계 학습 모델(104)은 소정의 비디오가 제1 기계 학습 모델(102)로 입력되어 제1 기계 학습 모델(102)에서 객체 분할 지도를 생성할 때, 제1 기계 학습 모델(102)의 객체 분할 지도에서 각 픽셀들에 대한 불확실성(즉, 각 픽셀 중 틀리게 예측될 수 있을 확률)을 예측할 수 있게 된다.

[0061] 도 3은 본 발명의 일 실시예에 따른 심층 신경망의 불확실성 예측 장치에서 적대적 학습 기반으로 학습 비디오

를 생성하는 과정을 나타낸 도면이다.

[0062] 도 3을 참조하면, 불확실성 예측 장치(100)는 학습 데이터가 되는 비디오(원래 비디오, original video)에서 객체에 해당하는 영역 부근에 노이즈에 해당하는 공격 조각을 추가하여 학습 비디오를 생성할 수 있다. 즉, 불확실성 예측 장치(100)는 기 학습된 제1 기계 학습 모델(102)이 학습 비디오의 객체 부근에서 객체 분할 예측이 틀리도록 객체에 해당하는 영역 부근에 공격 조각을 추가하여 학습 비디오를 생성할 수 있다.

[0063] 이때, 불확실성 예측 장치(100)는 원래 비디오에 공격 조각이 포함된 마스크를 합성하여 학습 비디오를 생성할 수 있다. 마스크는 공격 조각에 해당하는 픽셀의 값은 1이고, 공격 조각에 해당하지 않는 픽셀의 값은 0을 가질 수 있다. 여기서, 공격 조각의 크기는 원래 비디오 내 객체의 개수와 각 객체의 크기에 기초하여 결정될 수 있다. 이와 같이, 원래 비디오에서 객체에 해당하는 영역 부근에 공격 조각을 추가함으로써, 지역적인 적대적 공격(Local Adversarial Attack)을 진행하게 된다.

[0064] 불확실성 예측 장치(100)는 원래 비디오에 공격 조각이 포함된 마스크를 합성하여 학습 비디오를 생성한 후, 학습 비디오를 제1 기계 학습 모델(102)에 입력할 수 있다. 이때, 불확실성 예측 장치(100)는 제1 기계 학습 모델(102)의 손실이 최대가 되도록 학습 비디오를 생성할 수 있다.

[0065] 즉, 불확실성 예측 장치(100)는 원래 비디오에서 어떤 부분에 노이즈에 해당하는 공격 조각을 추가해야 제1 기계 학습 모델(102)에서 객체 분할 예측이 틀리도록 할 수 있는지를 고려하여 학습 비디오를 생성할 수 있다. 불확실성 예측 장치(100)는 노이즈가 추가된 비디오 프레임을 제1 기계 학습 모델(102)에 통과시켜 노이즈가 추가된 비디오 프레임에 대한 기울기를 찾을 수 있다. 이때, 왜곡의 양을 적게 만들기 위해 실수 값을 곱해준 뒤, 공격 조각이 포함된 마스크를 곱하여 제1 기계 학습 모델(102)의 손실이 최대화 되는 섭동(perturbation)을 생성할 수 있다. 불확실성 예측 장치(100)는 하기 수학적 식 2를 통해 학습 비디오를 생성할 수 있다.

[0066] (수학적 식 2)

$$\tilde{x} = x + LAA(Seg, x_N)$$

[0067]

[0068] x : 원래 비디오

$$LAA(Seg, x_N) = \epsilon(\nabla_{x_N} L(Seg(x_N), y))\Omega(x)$$

[0069]

[0070] x_N : 노이즈가 추가된 비디오 프레임

[0071] $\Omega(x)$: 공격 조각이 추가되는 영역

[0072] ϵ : 기 설정되는 실수 값

[0074] 도 4는 본 발명의 일 실시예에 따른 심층 신경망의 불확실성 예측 방법을 나타낸 흐름도이다. 도시된 흐름도에서는 상기 방법을 복수 개의 단계로 나누어 기재하였으나, 적어도 일부의 단계들은 순서를 바꾸어 수행되거나, 다른 단계와 결합되어 함께 수행되거나, 생략되거나, 세부 단계들로 나뉘어 수행되거나, 또는 도시되지 않은 하나 이상의 단계가 부가되어 수행될 수 있다.

[0075] 도 4를 참조하면, 불확실성 예측 장치(100)는 제1 기계 학습 모델(102)을 학습할 수 있다(S 101). 일 실시예에서, 불확실성 예측 장치(100)는 비디오를 제1 기계 학습 모델(102)에 입력하고, 입력된 비디오에서 각 객체를 분할한 객체 분할 지도를 생성하도록 제1 기계 학습 모델(102)을 학습시킬 수 있다.

[0076] 다음으로, 불확실성 예측 장치(100)는 제1 기계 학습 모델(102)의 불확실성을 예측하기 위한 제2 기계 학습 모델(104)을 생성할 수 있다(S 103). 일 실시예에서, 불확실성 예측 장치(100)는 제1 기계 학습 모델(102)의 디코더(102b)와 동일한 구조의 심층 신경망으로 제2 기계 학습 모델(104)을 생성할 수 있다.

[0077] 다음으로, 불확실성 예측 장치(100)는 제2 기계 학습 모델(104)을 학습하기 위한 학습 비디오를 생성할 수 있다(S 105). 일 실시예에서, 불확실성 예측 장치(100)는 원래 비디오에서 객체에 해당하는 영역 부근에 노이즈에 해당하는 공격 조각을 추가하여 학습 비디오를 생성할 수 있다. 이때, 불확실성 예측 장치(100)는 제1 기계 학습 모델(102)의 손실이 최대가 되도록 학습 비디오를 생성할 수 있다.

- [0078] 다음으로, 불확실성 예측 장치(100)는 학습 비디오를 이용하여 제2 기계 학습 모델(104)을 학습할 수 있다(S 107). 불확실성 예측 장치(100)는 학습 비디오를 이용하여 제1 기계 학습 모델(102)의 객체 분할 결과가 틀리는 상황을 관찰하도록 제2 기계 학습 모델(104)을 학습할 수 있다.
- [0079] 구체적으로, 불확실성 예측 장치(100)는 학습 비디오를 제1 기계 학습 모델(102)의 인코더(102a)로 입력하여 학습 비디오에서 특징을 추출할 수 있다. 이때, 추출된 특징은 각각 제1 기계 학습 모델(102)의 디코더(102b)와 제2 기계 학습 모델(104)로 입력될 수 있다.
- [0080] 여기서, 제1 기계 학습 모델(102)의 디코더(102b)는 추출된 특징에 기반하여 학습 비디오에서 각 객체들을 분할한 객체 분할 지도를 생성할 수 있다. 불확실성 예측 장치(100)는 디코더(102b)에서 예측한 객체 분할 지도와 객체 분할 정답 지도를 비교하여 디코더(102b)의 예측이 틀린 픽셀들로 구성된 객체 분할 지도를 제2 기계 학습 모델(104)의 정답 지도로 설정할 수 있다.
- [0081] 그리고, 제2 기계 학습 모델(104)은 인코더(102)에서 추출된 특징을 기초로 불확실성 예측 지도를 생성하는데, 불확실성 예측 장치(100)는 제2 기계 학습 모델(104)에서 출력하는 불확실성 예측 지도가 정답 지도(즉, 디코더(102b)의 예측이 틀린 픽셀들로 구성된 객체 분할 지도)를 닮아가도록 제2 기계 학습 모델(104)을 학습할 수 있다.
- [0082] 개시되는 실시예에 의하면, 제1 기계 학습 모델(102)의 객체 분할 결과의 불확실성을 예측하기 위한 제2 기계 학습 모델(104)을 구성함으로써, 제1 기계 학습 모델(102)에 대한 신뢰도를 제공할 수 있게 되고 그로 인해 심층 신경망 모델의 결정이 중요한 분야에서 그 활용도를 높일 수 있게 된다.
- [0084] 도 5는 예시적인 실시예들에서 사용되기에 적합한 컴퓨팅 장치를 포함하는 컴퓨팅 환경(10)을 예시하여 설명하기 위한 블록도이다. 도시된 실시예에서, 각 컴포넌트들은 이하에 기술된 것 이외에 상이한 기능 및 능력을 가질 수 있고, 이하에 기술된 것 이외에도 추가적인 컴포넌트를 포함할 수 있다.
- [0085] 도시된 컴퓨팅 환경(10)은 컴퓨팅 장치(12)를 포함한다. 일 실시예에서, 컴퓨팅 장치(12)는 심층 신경망의 불확실성 예측 장치(100)일 수 있다.
- [0086] 컴퓨팅 장치(12)는 적어도 하나의 프로세서(14), 컴퓨터 판독 가능 저장 매체(16) 및 통신 버스(18)를 포함한다. 프로세서(14)는 컴퓨팅 장치(12)로 하여금 앞서 언급된 예시적인 실시예에 따라 동작하도록 할 수 있다. 예컨대, 프로세서(14)는 컴퓨터 판독 가능 저장 매체(16)에 저장된 하나 이상의 프로그램들을 실행할 수 있다. 상기 하나 이상의 프로그램들은 하나 이상의 컴퓨터 실행 가능 명령어를 포함할 수 있으며, 상기 컴퓨터 실행 가능 명령어는 프로세서(14)에 의해 실행되는 경우 컴퓨팅 장치(12)로 하여금 예시적인 실시예에 따른 동작들을 수행하도록 구성될 수 있다.
- [0087] 컴퓨터 판독 가능 저장 매체(16)는 컴퓨터 실행 가능 명령어 내지 프로그램 코드, 프로그램 데이터 및/또는 다른 적합한 형태의 정보를 저장하도록 구성된다. 컴퓨터 판독 가능 저장 매체(16)에 저장된 프로그램(20)은 프로세서(14)에 의해 실행 가능한 명령어의 집합을 포함한다. 일 실시예에서, 컴퓨터 판독 가능 저장 매체(16)는 메모리(랜덤 액세스 메모리와 같은 휘발성 메모리, 비휘발성 메모리, 또는 이들의 적절한 조합), 하나 이상의 자기 디스크 저장 디바이스들, 광학 디스크 저장 디바이스들, 플래시 메모리 디바이스들, 그 밖에 컴퓨팅 장치(12)에 의해 액세스되고 원하는 정보를 저장할 수 있는 다른 형태의 저장 매체, 또는 이들의 적합한 조합일 수 있다.
- [0088] 통신 버스(18)는 프로세서(14), 컴퓨터 판독 가능 저장 매체(16)를 포함하여 컴퓨팅 장치(12)의 다른 다양한 컴포넌트들을 상호 연결한다.
- [0089] 컴퓨팅 장치(12)는 또한 하나 이상의 입출력 장치(24)를 위한 인터페이스를 제공하는 하나 이상의 입출력 인터페이스(22) 및 하나 이상의 네트워크 통신 인터페이스(26)를 포함할 수 있다. 입출력 인터페이스(22) 및 네트워크 통신 인터페이스(26)는 통신 버스(18)에 연결된다. 입출력 장치(24)는 입출력 인터페이스(22)를 통해 컴퓨팅 장치(12)의 다른 컴포넌트들에 연결될 수 있다. 예시적인 입출력 장치(24)는 포인팅 장치(마우스 또는 트랙패드 등), 키보드, 터치 입력 장치(터치패드 또는 터치스크린 등), 음성 또는 소리 입력 장치, 다양한 종류의 센서 장치 및/또는 촬영 장치와 같은 입력 장치, 및/또는 디스플레이 장치, 프린터, 스피커 및/또는 네트워크 카드와 같은 출력 장치를 포함할 수 있다. 예시적인 입출력 장치(24)는 컴퓨팅 장치(12)를 구성하는 일 컴포넌트로서 컴퓨팅 장치(12)의 내부에 포함될 수도 있고, 컴퓨팅 장치(12)와는 구별되는 별개의 장치로 컴퓨팅 장치(12)와

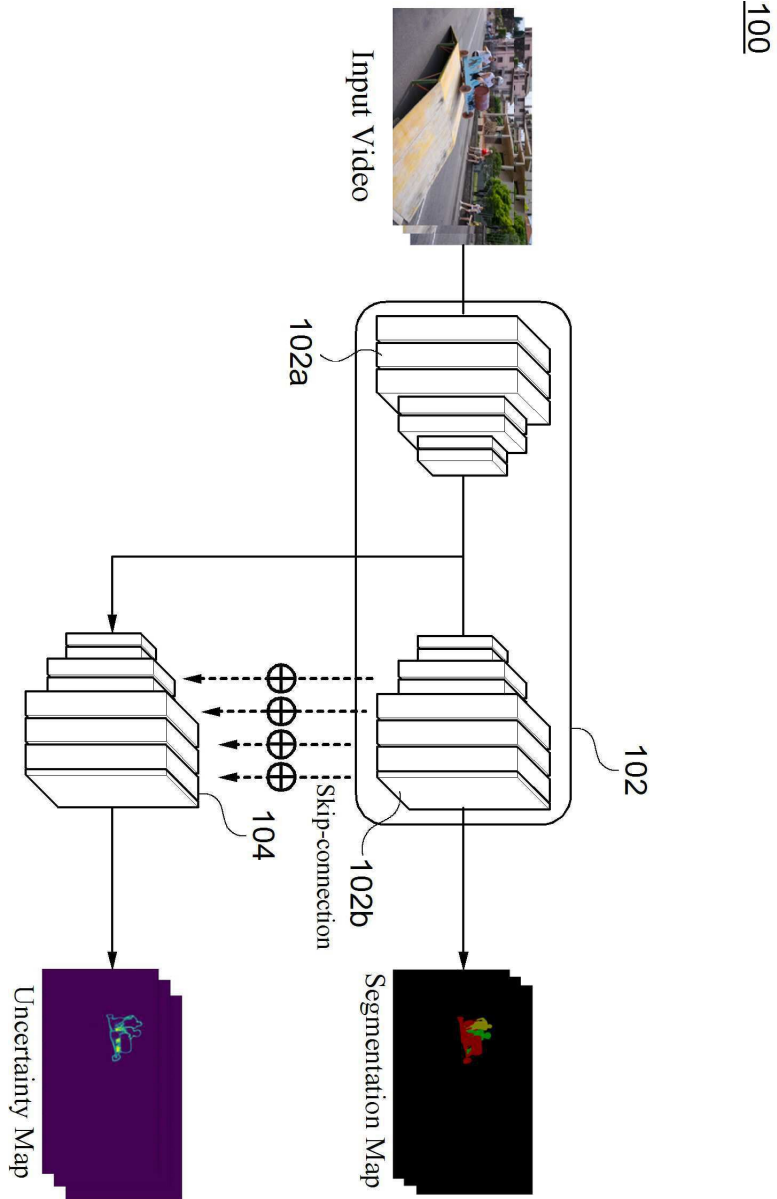
연결될 수도 있다.

[0091] 이상에서 본 발명의 대표적인 실시예들을 상세하게 설명하였으나, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자는 상술한 실시예에 대하여 본 발명의 범주에서 벗어나지 않는 한도 내에서 다양한 변형이 가능함을 이해할 것이다. 그러므로 본 발명의 권리범위는 설명된 실시예에 국한되어 정해져서는 안 되며, 후술하는 특허 청구범위뿐만 아니라 이 특허청구범위와 균등한 것들에 의해 정해져야 한다.

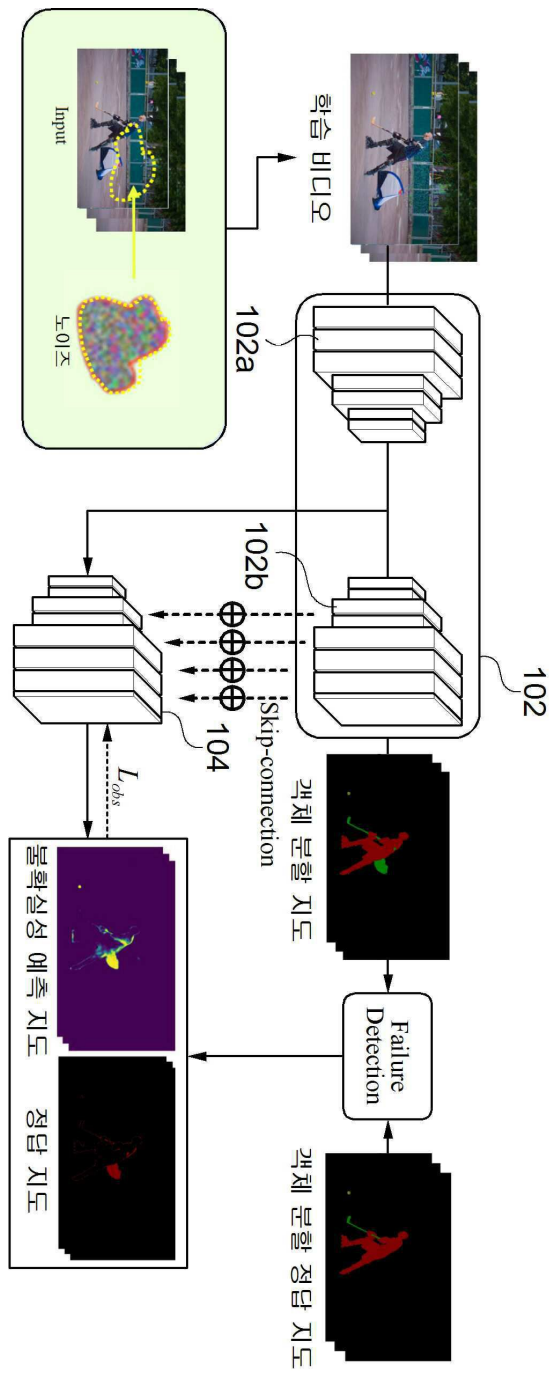
부호의 설명

[0093] 10 : 컴퓨팅 환경
12 : 컴퓨팅 장치
14 : 프로세서
16 : 컴퓨터 관독 가능 저장 매체
18 : 통신 버스
20 : 프로그램
22 : 입출력 인터페이스
24 : 입출력 장치
26 : 네트워크 통신 인터페이스
100 : 심층 신경망의 불확실성 예측 장치
102 : 제1 기계 학습 모델
102a : 인코더
102b : 디코더
104 : 제2 기계 학습 모델

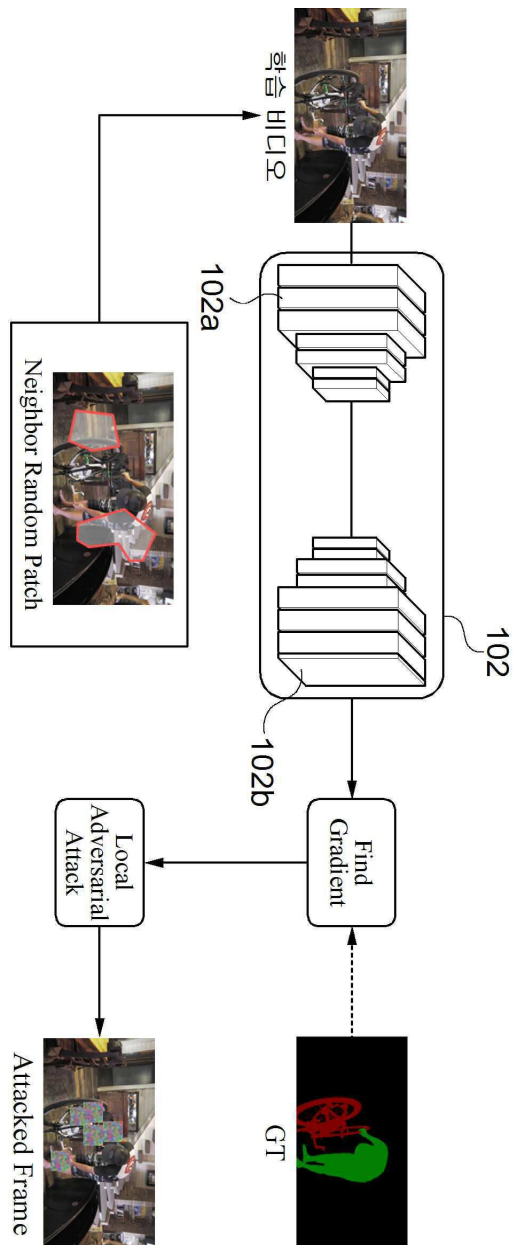
도면
도면1



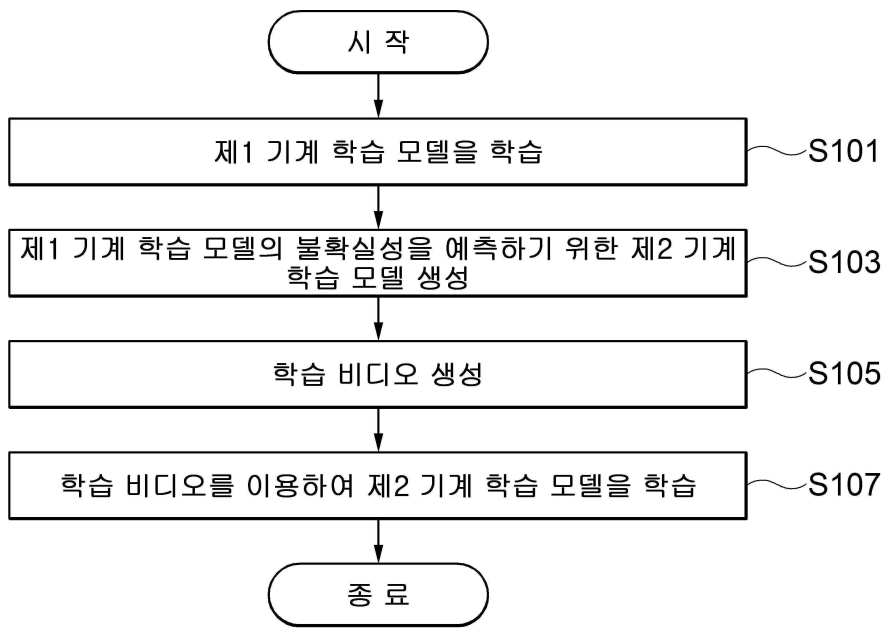
도면2



도면3



도면4



도면5

10

