



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년11월09일
(11) 등록번호 10-2323671
(24) 등록일자 2021년11월03일

(51) 국제특허분류(Int. Cl.)
G06T 5/00 (2019.01) G06K 9/00 (2006.01)
G06N 20/00 (2019.01)
(52) CPC특허분류
G06T 5/005 (2013.01)
G06K 9/00711 (2013.01)
(21) 출원번호 10-2020-0172912
(22) 출원일자 2020년12월11일
심사청구일자 2020년12월11일
(65) 공개번호 10-2021-0086487
(43) 공개일자 2021년07월08일
(30) 우선권주장
1020190177627 2019년12월30일 대한민국(KR)
(56) 선행기술조사문헌
KR1020130036514 A
박상민, LSTM-AutoEncoder 이상감지 모델,
InSpace, 2019.03.21,
<https://inspaceai.github.io/2019/03/21/Anomaly-Detection-With-LSTM-AE/>
Nguyen Long Thinh, Approved by Prof. Kim,
Yong-Guk, Autoencoder Framework for Anomaly
Detection in videos, Department of Computer
Science and Engineering Sejong University,
June 2019

(73) 특허권자
세종대학교산학협력단
서울특별시 광진구 능동로 209 (군자동, 세종대학교)
(72) 발명자
김용국
경기도 성남시 분당구 동판교로 156, 911동 503호(삼평동, 봇들마을9단지금호어울림아파트)
뉴엔 통 턴
서울특별시 광진구 능동로 209, 세종대학교
대양AI센터 705호 (군자동)
(74) 대리인
송인호, 최관락

전체 청구항 수 : 총 10 항

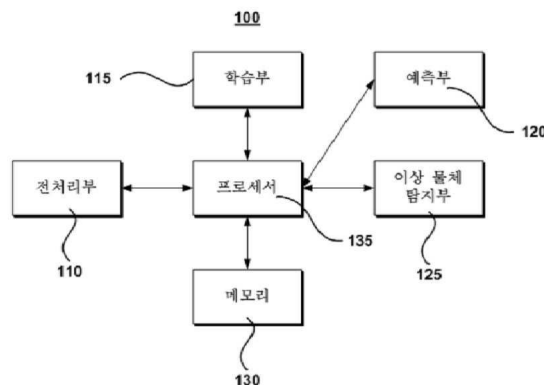
심사관 : 김광식

(54) 발명의 명칭 동영상내의 이상 물체 탐지 방법 및 그 장치

(57) 요약

동영상내의 이상 물체 탐지 방법 및 그 장치가 개시된다. 동영상내의 이상 물체 탐지 방법은 인페인팅 패턴이 적용된 입력 배치를 각각 학습된 오토인코더 모델에 적용하여 복원 배치를 재건하고, 상기 재건된 복원 배치를 컴바인하여 출력되는 복원 프레임을 이용한 공간 영역 재건 에러와 복수의 연속적인 프레임들을 학습된 LSTM 오토인코더 모델에 적용하여 시간 영역 특징점을 추출한 후 복원하여 출력되는 시간 영역 복원 프레임들을 이용한 시간 영역 재건 에러를 융합하여 이상 물체가 위치하는 영역을 추정할 수 있다.

대표도 - 도1



(52) CPC특허분류

G06N 20/00 (2021.08)

G06T 2207/10016 (2013.01)

G06T 2207/20081 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711093571
과제번호	2019-0-00231-001
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보보호핵심원천기술개발(R&D)
연구과제명	공공 인프라 안전을 위한 인공지능 기반 영상보안 기술 및 시스템 개발
기여율	1/1
과제수행기관명	세종대학교 산학협력단
연구기간	2019.04.01 ~ 2026.12.31

명세서

청구범위

청구항 1

동영상의 현재 프레임에 대해 복수의 인페인팅(inpainting) 패턴을 각각 적용하여 입력 배치를 생성하는 배치부;

상기 인페인팅 패턴이 적용된 입력 배치를 각각 학습된 오토인코더 모델에 적용하여 복원 배치를 재건하고, 상기 재건된 복원 배치를 컴바인하여 복원 프레임을 출력하는 제1 예측부;

복수의 연속적인 프레임들을 학습된 LSTM 오토인코더 모델에 적용하여 시간 영역 특징점을 추출한 후 복원하여 시간 영역 복원 프레임들을 출력하는 제2 예측부-상기 연속적인 프레임들은 상기 현재 프레임을 포함하는 복수의 연속적인 프레임들임; 및

상기 복원 프레임과 상기 현재 프레임을 이용하여 공간 영역 재건 에러를 도출하며, 상기 연속적인 프레임들과 상기 시간 영역 복원 프레임들을 이용하여 시간 영역 재건 에러를 도출하고, 상기 공간 영역 재건 에러와 상기 시간 영역 재건 에러를 융합하여 이상 물체가 위치하는 영역을 추정하는 이상 물체 탐지부를 포함하는 이상 물체 탐지 장치.

청구항 2

제1 항에 있어서,

상기 오토인코더 모델은,

상기 인페인팅 패턴이 미적용된 정상적인 프레임을 이용하여 학습되는 것을 특징으로 하는 이상 물체 탐지 장치.

청구항 3

제1 항에 있어서,

상기 제1 예측부는 상기 학습된 오토인코더 모델에 상기 인페인팅 패턴이 적용된 입력 배치들을 적용함으로써 상기 인페인팅 패턴에 대한 비정상적인 특징 벡터를 생성하는 것을 특징으로 하는 이상 물체 탐지 장치.

청구항 4

제3 항에 있어서,

상기 이상 물체 탐지부는,

상기 비정상적인 특징 벡터에 의해 상기 공간 영역 재건 에러가 증가됨에 따라 상기 시간 영역 재건 에러와의 융합시 상기 공간 영역 재건 에러가 높은 영역을 이상 물체가 존재하는 영역으로 추정하는 것을 특징으로 하는 이상 물체 탐지 장치.

청구항 5

제1 항에 있어서,

상기 인페인팅 패턴은 제로(0)을 포함하는 서로 상이한 패턴을 가지는 것을 특징으로 하는 이상 물체 탐지 장치.

청구항 6

- (a) 동영상의 현재 프레임에 대해 복수의 인페인팅(inpainting) 패턴을 각각 적용하여 입력 배치를 생성하는 단계;
- (b) 상기 인페인팅 패턴이 적용된 입력 배치를 각각 학습된 오토인코더 모델에 적용하여 복원 배치를 재건하고, 상기 재건된 복원 배치를 컴바인하여 복원 프레임을 출력하는 단계;
- (c) 복수의 연속적인 프레임들을 학습된 LSTM 오토인코더 모델에 적용하여 시간 영역 특징점을 추출한 후 복원하여 시간 영역 복원 프레임들을 출력하는 단계-상기 연속적인 프레임들은 상기 현재 프레임을 포함하는 복수의 연속적인 프레임들임; 및
- (d) 상기 복원 프레임과 상기 현재 프레임을 이용하여 공간 영역 재건 에러를 도출하며, 상기 연속적인 프레임들과 상기 시간 영역 복원 프레임들을 이용하여 시간 영역 재건 에러를 도출하고, 상기 공간 영역 재건 에러와 상기 시간 영역 재건 에러를 융합하여 이상 물체가 위치하는 영역을 추정하는 단계를 포함하는 동영상내 이상 물체 탐지 방법.

청구항 7

제6 항에 있어서,
 상기 (a) 단계 이전에,
 상기 인페인팅 패턴이 미적용된 정상적인 프레임을 이용하여 상기 오토인코더 모델을 학습하는 단계; 및
 상기 연속된 프레임들을 이용하여 상기 LSTM 오토인코더를 학습하는 단계를 더 포함하는 동영상내 이상 물체 탐지 방법.

청구항 8

제6 항에 있어서,
 상기 (b) 단계는,
 상기 학습된 오토인코더 모델에 상기 인페인팅 패턴이 적용된 입력 배치들을 적용함으로써 상기 인페인팅 패턴에 대한 비정상적인 특징 벡터를 생성하는 것을 특징으로 하는 동영상내 이상 물체 탐지 방법.

청구항 9

제8 항에 있어서,
 상기 (d) 단계는,
 상기 비정상적인 특징 벡터에 의해 상기 공간 영역 재건 에러가 증가됨에 따라 상기 시간 영역 재건 에러와의 융합시 상기 공간 영역 재건 에러가 높은 영역을 이상 물체가 존재하는 영역으로 추정하는 것을 특징으로 하는 동영상내 이상 물체 탐지 방법.

청구항 10

제6 항에 따른 방법을 수행하기 위한 프로그램 코드를 기록한 컴퓨터로 판독 가능한 기록매체.

발명의 설명

기술 분야

[0001] 본 발명은 동영상 내의 이상 물체 탐지 방법 및 그 장치에 관한 것이다.

배경 기술

[0003] 이상 물체 탐지는 예상되는 동작과 관련이 없는 비정상 패턴을 인식하는 작업으로, 비디오에서의 이상 감지는 전체 비디오 시퀀스에서 거의 나타나지 않는 비정상적인 동작이나 객체를 감지하는 것이다. 즉, 이상 물체 탐지는 현재 상황과 다른 일련의 동작을 감지하는 것이다. 따라서, 한 동작이 컨텍스트에서 "정상"으로 간주될 수 있지만, 다른 컨텍스트에서는 "비정상"으로 간주될 수 있으며, 이는 컨텍스트에 따라 달라질 수도 있다.

[0004] 최근 CCTV가 기하 급수적으로 설치되고 있다. 따라서, 모든 CCTV 영상을 모니터링하고 분석하는데 필요한 인력 또한 증가하고 있으나, 모든 영상을 정밀하게 분석하는 것은 사실상 불가능하다.

발명의 내용

해결하려는 과제

[0006] 본 발명은 동영상내의 이상 물체 탐지 방법 및 그 장치를 제공하기 위한 것이다.

[0007] 또한, 본 발명은 인페인팅(inpainting) 기법을 이용하여 이상 물체 탐지의 정확도를 높일 수 있는 동영상내의 이상 물체 탐지 방법 및 그 장치를 제공하기 위한 것이다.

과제의 해결 수단

[0009] 본 발명의 일 측면에 따르면, 동영상내의 이상 물체 탐지 장치가 제공된다.

[0010] 본 발명의 일 실시예에 따르면, 동영상의 현재 프레임에 대해 복수의 인페인팅(inpainting) 패턴을 각각 적용하여 입력 배치를 생성하는 배치부; 상기 인페인팅 패턴이 적용된 입력 배치를 각각 학습된 오토인코더 모델에 적용하여 복원 배치를 재건하고, 상기 재건된 복원 배치를 컴바인하여 복원 프레임을 출력하는 제1 예측부; 복수의 연속적인 프레임들을 학습된 LSTM 오토인코더 모델에 적용하여 시간 영역 특징점을 추출한 후 복원하여 시간 영역 복원 프레임들을 출력하는 제2 예측부-상기 연속적인 프레임들은 상기 현재 프레임을 포함하는 복수의 연속적인 프레임들임; 및 상기 복원 프레임과 상기 현재 프레임을 이용하여 공간 영역 재건 에러를 도출하며, 상기 연속적인 프레임들과 상기 시간 영역 복원 프레임들을 이용하여 시간 영역 재건 에러를 도출하고, 상기 공간 영역 재건 에러와 상기 시간 영역 재건 에러를 융합하여 이상 물체가 위치하는 영역을 추정하는 이상 물체 탐지 부를 포함하는 이상 물체 탐지 장치가 제공될 수 있다.

[0011] 상기 오토인코더 모델은, 상기 인페인팅 패턴이 미적용된 정상적인 프레임을 이용하여 학습될 수 있다.

[0012] 상기 제1 예측부는 상기 학습된 오토인코더 모델에 상기 인페인팅 패턴이 적용된 입력 배치들을 적용함으로써 상기 인페인팅 패턴에 대한 비정상적인 특징 벡터를 생성할 수 있다.

[0013] 상기 이상 물체 탐지부는, 상기 비정상적인 특징 벡터에 의해 상기 공간 영역 재건 에러가 증가됨에 따라 상기 시간 영역 재건 에러와의 융합시 상기 공간 영역 재건 에러가 높은 영역을 이상 물체가 존재하는 영역으로 추정할 수 있다.

[0014] 상기 인페인팅 패턴은 제로(0)을 포함하는 서로 상이한 패턴을 가진다.

[0016] 본 발명의 다른 측면에 따르면, 동영상내의 이상 물체 탐지 방법이 제공된다.

[0017] 본 발명의 일 실시예에 따르면, (a) 동영상의 현재 프레임에 대해 복수의 인페인팅(inpainting) 패턴을 각각 적용하여 입력 배치를 생성하는 단계; (b) 상기 인페인팅 패턴이 적용된 입력 배치를 각각 학습된 오토인코더 모델에 적용하여 복원 배치를 재건하고, 상기 재건된 복원 배치를 컴바인하여 복원 프레임을 출력하는 단계; (c) 복수의 연속적인 프레임들을 학습된 LSTM 오토인코더 모델에 적용하여 시간 영역 특징점을 추출한 후 복원하여 시간 영역 복원 프레임들을 출력하는 단계-상기 연속적인 프레임들은 상기 현재 프레임을 포함하는 복수의 연속적인 프레임들임; 및 (d) 상기 복원 프레임과 상기 현재 프레임을 이용하여 공간 영역 재건 에러를 도출하며, 상기 연속적인 프레임들과 상기 시간 영역 복원 프레임들을 이용하여 시간 영역 재건 에러를 도출하고, 상기 공간 영역 재건 에러와 상기 시간 영역 재건 에러를 융합하여 이상 물체가 위치하는 영역을 추정하는 단계를 포함

하는 동영상내 이상 물체 탐지 방법이 제공될 수 있다.

- [0018] 상기 (a) 단계 이전에, 상기 인페인팅 패턴이 미적용된 정상적인 프레임을 이용하여 상기 오토인코더 모델을 학습하는 단계; 및 상기 연속된 프레임들을 이용하여 상기 LSTM 오토인코더를 학습하는 단계를 더 포함할 수 있다.
- [0019] 상기 (b) 단계는, 상기 학습된 오토인코더 모델에 상기 인페인팅 패턴이 적용된 입력 배치들을 적용함으로써 상기 인페인팅 패턴에 대한 비정상적인 특징 벡터를 생성할 수 있다.
- [0020] 상기 (d) 단계는, 상기 비정상적인 특징 벡터에 의해 상기 공간 영역 재건 에러가 증가됨에 따라 상기 시간 영역 재건 에러와의 융합시 상기 공간 영역 재건 에러가 높은 영역을 이상 물체가 존재하는 영역으로 추정될 수 있다.

발명의 효과

- [0022] 본 발명의 일 실시예에 따른 동영상내의 이상 물체 탐지 방법 및 그 장치를 제공함으로써, 공간 영역 예측시 인페인팅 기법을 이용하여 재건 에러를 증가시킴으로써 시간 영역 특징점과 융합을 통해 이상 물체 탐지의 정확도를 높일 수 있는 이점이 있다.

도면의 간단한 설명

- [0024] 도 1은 본 발명의 일 실시예에 따른 동영상내 이상 물체 탐지 장치의 내부 구성을 개략적으로 도시한 블록도.
- 도 2는 본 발명의 일 실시예에 따른 모델의 학습 과정을 도시한 도면.
- 도 3은 본 발명의 일 실시예에 따른 오토인코더 모델의 아키텍처를 도시한 도면.
- 도 4는 본 발명의 일 실시예에 따른 LSTM 오토인코더 모델의 아키텍처를 도시한 도면.
- 도 5는 본 발명의 일 실시예에 따른 인페인팅 패턴을 예시한 도면.
- 도 6은 본 발명의 일 실시예에 따른 인페인팅 패턴의 너비와 높이에 따른 이상 물체 탐지 결과를 도시한 도면.
- 도 7은 본 발명의 일 실시예에 따른 동영상내의 이상 물체 탐지 방법을 나타낸 순서도.
- 도 8은 본 발명의 일 실시예에 따른 인페인팅 기법을 적용한 오토인코더 모델의 공간 영역 재건을 설명하기 위해 도시한 도면.
- 도 9는 본 발명의 일 실시예에 따른 이상 물체 탐지 방법의 상세 프레임워크를 도시한 도면.

발명을 실시하기 위한 구체적인 내용

- [0025] 본 명세서에서 사용되는 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "구성된다" 또는 "포함한다" 등의 용어는 명세서상에 기재된 여러 구성 요소들, 또는 여러 단계들을 반드시 모두 포함하는 것으로 해석되지 않아야 하며, 그 중 일부 구성 요소들 또는 일부 단계들은 포함되지 않을 수도 있고, 또는 추가적인 구성 요소 또는 단계들을 더 포함할 수 있는 것으로 해석되어야 한다. 또한, 명세서에 기재된 "...부", "모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어 또는 소프트웨어로 구현되거나 하드웨어와 소프트웨어의 결합으로 구현될 수 있다.
- [0026] 이하, 첨부된 도면들을 참조하여 본 발명의 실시예를 상세히 설명한다.
- [0028] 도 1은 본 발명의 일 실시예에 따른 동영상내 이상 물체 탐지 장치의 내부 구성을 개략적으로 도시한 블록도이며, 도 2는 본 발명의 일 실시예에 따른 모델의 학습 과정을 도시한 도면이고, 도 3은 본 발명의 일 실시예에 따른 오토인코더 모델의 아키텍처를 도시한 도면이고, 도 4는 본 발명의 일 실시예에 따른 LSTM 오토인코더 모델의 아키텍처를 도시한 도면이며, 도 5는 본 발명의 일 실시예에 따른 인페인팅 패턴을 예시한 도면이고, 도 6은 본 발명의 일 실시예에 따른 인페인팅 패턴의 너비와 높이에 따른 이상 물체 탐지 결과를 도시한 도면이다.
- [0029] 도 1을 참조하면, 본 발명의 일 실시예에 따른 동영상내 이상 물체 탐지 장치(100)는 전처리부(110), 학습부(115), 예측부(120), 이상 물체 탐지부(125), 메모리(130) 및 프로세서(135)를 포함하여 구성된다.
- [0030] 전처리부(110)는 동영상의 각 프레임을 전처리하기 위한 수단이다.

- [0031] 학습부(115)는 전처리된 프레임을 이용하여 오토인코더 모델과 LSTM 오토인코더 모델을 학습하기 위한 수단이다.
- [0032] 오토인코더 모델은 각각의 프레임을 입력받아 공간 영역 특징점을 학습하기 위한 수단이다. 오토인코더 모델은 인코더와 디코더를 포함하는 컨볼루션 오토인코더 모델일 수 있다.
- [0033] 오토인코더 모델은 인코더단에서 입력 프레임에 대한 공간 특징 벡터를 추출하고, 디코더단에서는 공간 특징 벡터를 이용하여 프레임을 복원(재건)할 수 있다.
- [0034] 오토인코더 모델은 인코더단에서 차원을 축소하여 특징 벡터(잠재 벡터(latent vector))를 생성하며, 디코더단에는 이를 활용하여 입력 프레임과 최대한 유사하도록 복원하는 과정을 거친다. 학습 과정에서 오토인코더 모델은 입력 프레임과 복원 프레임의 차이가 최소가 되도록 가중치를 조정하는 과정을 반복적으로 수행할 수 있다.
- [0035] 오토인코더 모델의 경우, 인코더와 디코더는 대칭적인 구조를 가질 수 있다. 따라서, 오토인코더 모델은 인코더에서 생성된 특징 벡터를 이용하여 원하는 복원 프레임을 생성할 수 있다.
- [0036] 도 2에 오토인코더 모델의 학습 과정이 예시되어 있다. 본 발명의 일 실시예에 따르면, 학습 과정에서 트레이닝 데이터는 비정상적인 물체가 포함되지 않은 프레임들을 이용하여 학습될 수 있다. 예를 들어, 오토인코더 모델은 컨볼루션 오토인코더 모델일 수 있다. 도 3에는 오토인코더 모델의 아키텍처가 도시되어 있다.
- [0037] 오토인코더 모델의 인코더(E)는 주어진 입력(X(프레임))에 대한 잠재 변수(z)를 예측하고, 디코더(D)는 잠재 변수로부터 신호를 복원한다. 따라서, 인코더(E)와 디코더(D)는 함께 학습될 수 있다. 학습의 목적은 신호 복원의 정확도를 최대화하는 것으로, 다른 말로 표현하면 재건 에러를 최소화하는 것이다. 이를 수학적 식으로 나타내면 수학적 식 1과 같다.

수학적 식 1

$$\min_{\xi}(E, D) = \min |D(E(X)) - X|^2$$

- [0038]
- [0039] LSTM 오토인코더 모델은 컨볼루션 LSTM 레이어를 적층한 구조를 포함한다.
- [0040] LSTM(long Short-Term Memory) 신경망은 표준 RNN 또는 기타 변형과 비교하여 실제 응용에서 시간적 종속성의 상관 관계를 기억하는 성공적이고 강력한 모델임이 입증되었다. 컨볼루션 LSTM은 LSTM의 확장으로 LSTM의 행렬 곱셈을 컨볼루션 LSTM이 비디오 데이터 처리에 적합하도록 만든 다음 시공간 상관 관계를 잘 포착하고 일부 응용 분야에서 LSTM을 능가한다.
- [0041] 시간 t에서 컨볼루션 LSTM 셀의 활성화는 다음과 같이 제공된다.

$$i_t = \sigma(w_{xi} \otimes x_t + w_{hi} \otimes h_{t-1} + b_i)$$

$$f_t = \sigma(w_{xf} \otimes x_t + w_{hf} \otimes h_{t-1} + b_f)$$

$$o_t = \sigma(w_{xo} \otimes x_t + w_{ho} \otimes h_{t-1} + b_o)$$

$$c_t = f_{xf} \cdot c_{t-1} + i_t \cdot \tanh(w_{xc} \otimes x_t + w_{hc} \otimes h_{t-1} + b_c)$$

$$h_t = o_t \cdot \tanh(c_t)$$

- [0042]
- [0043] 여기서, x_t 는 시간 t에서 입력을 나타내고, i_t , f_t , o_t 는 입력 게이트, 망각 게이트 및 출력 게이트를 나타낸다. 또한, c_t, c_{t-1}, h_t 와 h_{t-1} 는 시간 t와 (t-1)에서 메모리와 출력 활성화를 나타낸다. σ 와 \tanh 는 시그

모이드와 비선형 하이퍼볼릭 탄젠트를 나타낸다. $D \times h \times w$ 크기의 입력 특징맵을 위해 컨볼루션 LSTM 모델은 $d_m \times h \times w$ 크기의 메모리 맵을 출력할 수 있다. 여기서, d_m 은 메모리에 의해 학습된 시간 특징(temporal feature)의 개수를 나타낸다. 본 발명의 일 실시예에 따른 LSTM 오토인코더 모델은 컨볼루션 LSTM의 장점을 이용하기 위해, 복수의 컨볼루션 LSTM 레이어를 적층하는 구조로 설계된다.

- [0044] 본 발명의 일 실시예에 따른 LSTM 오토인코더 모델의 입력과 반복 변환은 입력과 상태에 대한 가중치 행렬의 곱셈을 요구하는 LSTM(Long Short-Term Memory)에서 사용하는 일반적인 선형 변환 대신 2차원 컨볼루션 변환이다. LSTM 오토인코더 모델은 2D 이미지 형태로 시각적인 정보를 해석하고 순차적인 데이터(시퀀스)를 이해하도록 설계될 수 있다. 배치 정규화 레이어가 LSTM 레이어 뒤에 추가될 수 있다. 배치 정규화 레이어는 다음 레이어로 이동하기 전에, 평균이 0이고 분산이 1을 가지도록 입력의 분포를 변경할 수 있다. 이를 통해 수렴 속도가 빨라져서 학습 과정이 빨라지고 별도의 레이어가 다른 레이어와 독립적으로 학습하도록 할 수 있다.
- [0045] LSTM 오토인코더 모델의 최종 계층은 시간 분산 컨볼루션 계층으로, 컨볼루션 LSTM 계층의 출력에서 중요한 시각적 특징을 추출하고 시그모이드 활성화 함수를 사용하여 비디오 프레임을 출력한다.
- [0046] LSTM 오토인코더 모델의 목표는 주어진 과거와 현재 프레임을 이용하여 시각적 프레임을 예측하는데 도움을 줄 수 있는 움직임(장면 내 물체의 움직임)으로 인한 중요한 변화를 캡처하는 것이다.
- [0047] 오토인코더 모델이 일정한 형태의 정규화를 사용하여 입력의 분해를 최적화할 수 있는 공간 특징을 학습하는 것에 비해, LSTM 오토인코더 모델은 학습 가능한 매개 변수가 적은 디코더를 갖도록 설계된다. 디코더는 인코더에 즉각적인 피드백을 제공할 수 있으나, 공간적인 경우와 같이 인코더의 실수를 수정하는 기능은 없다.
- [0048] 학습 과정 중 오류는 주로 인코더에 기인하며, 더 유의한 특징맵을 생성하도록 제약될 수 있다. 구성 가능한 파라미터 t 는 시간 블록의 길이를 나타낸다. LSTM 오토인코더 모델의 아키텍처는 도 4에 도시된 바와 같다.
- [0049] 예측부(120)는 제1 예측부와 제2 예측부를 포함한다.
- [0050] 제1 예측부는 학습된 오토인코더 모델을 이용하며, 제2 예측부는 LSTM 오토인코더 모델을 이용한다.
- [0051] 제1 예측부는 인페인팅 패턴이 적용된 입력 배치를 각각 학습된 오토인코더 모델에 적용하여 복원 배치를 재건하고, 재건된 복원배치를 컴바인(combined)하여 복원 프레임을 출력하기 위한 수단이다. 인페인팅 패턴의 일 예는 도 5에 도시된 바와 같다.
- [0052] 예를 들어, 샘플 프레임 X1, X2가 주어지는 경우를 가정하며, X1은 정상 프레임(normal frame)이고, X2는 비정상 프레임(abnormal frame)이라고 가정하기로 한다.
- [0053] 본 발명의 일 실시예에 따르면, 오토인코더 모델이 X1에 대해 입력 분포를 재구성하고, X2에 대해서는 입력 분포를 잘못 재구성하기를 원한다. 그러나, 오토인코더 모델의 학습 과정에서 훈련 세트를 통해 신호를 복원하도록 훈련되었기 때문에, 비정상 유형 이미지를 포함한 다른 이미지를 생성하지 못하도록 방지하지 않는다.
- [0054] 예를 들어, 보행자 이미지를 정확하게 재건하도록 훈련된 오토인코더 모델은 자전거 이미지를 정확하게 재건하지 않도록 훈련되지는 않는다. 이는 오토인코더 모델이 다른 종류의 이미지에 대해 작동하는 압축 전략을 수용함으로써 학습 손실을 줄일 수 있음을 나타낸다.
- [0055] 결과적으로 컨볼루션 오토인코더 모델은 일반적인 이미지를 나타내는 것이 보장되지 않으며, 일부 쿼리 이미지가 통과하여 불량 신호에 의한 작은 재구성 오류가 초래될 수 있음을 의미한다. 이는 오토인코더 모델에만 국한되는 것은 아니며 신경망을 사용하는 모든 생성 모델에 적용된다.
- [0056] 따라서, 본 발명의 일 실시예에 따르면, 제1 예측부는 예측 과정에서 인페인팅(inpainting) 기법을 이용할 수 있다. 즉, 제1 예측부는 동영상의 프레임에 대해 복수의 인페인팅 패턴을 적용한 후 이를 학습된 오토인코더 모델에 적용할 수 있다. 인페인팅 패턴은 도 5에 도시된 바와 같다. 예를 들어, 인페인팅 패턴의 각 픽셀값은 제로(0)로 설정될 수 있다.
- [0057] 제1 예측부는 인페인팅 패턴이 적용된 프레임을 학습된 오토인코더 모델에 적용하여 인페인팅 패턴에 대한 비정상 특징 벡터를 구축할 수 있다. 비정상 특징 벡터는 수학적 2와 같이 나타낼 수 있다.

수학식 2

$$\Gamma_{vs}[D(E(\Psi_{vs}(X))) - D(E(X))]^2$$

[0058]

[0059] 여기서, Ψ_{vs} 는 프레임에 제로(0)를 포함하는 인페인팅 패턴을 오버라이팅하는 오퍼레이션을 나타내고, Γ_{vs} 는 제로(0)를 포함하는 인페인팅 패턴을 제외하고 오버라이팅하는 오퍼레이션을 나타낸다. 그런 다음 제로가 아닌 영역을 프레임 특징으로 단순 결합(concatenation)한다.

[0060] 오토인코더 모델은 인페인팅 패턴내에서 보간을 위한 단서로 입력(X)을 사용할 수 없기 때문에 이전 학습을 기초로 인페인팅 패턴내에서 외삽해야만 한다. 예를 들어, 인페인팅 패턴에 의해 자전거가 숨겨진 경우, 오토인코더 모델은 장면에 조절된 일반적인 보행자를 구성하고 이를 자전거와 비교할 수 있다. 이로 인해, 입력된 자전거 객체와 보행자 객체 사이의 재구성 에러가 커지게 된다.

[0061] 인페인팅 패턴의 너비는 구성 가능한 매개 변수이다. 도 6에는 인페인팅 패턴의 너비와 높이를 다르게 하여 실험한 결과가 제시되어 있다. 본 발명의 일 실시예에 따르면 인페인팅 패턴의 너비가 3인 경우 최적화된 결과가 도출되는 것을 알 수 있다.

[0062] 다시 정리하면, 본 발명의 일 실시예에 따른 제1 예측부는 동영상내의 프레임에 인페인팅 패턴을 적용하여 학습된 오토인코더 모델을 적용하여 복원 프레임을 출력할 수 있다.

[0063] 제2 예측부는 복수의 연속적인 프레임들을 학습된 LSTM 오토인코더 모델에 적용하여 시간 영역 특징점을 추출한 후 복원하여 시간 영역 복원 프레임들을 출력한다.

[0064] LSTM 오토인코더 모델은 학습 과정을 참조하여 설명한 바와 동일하며, LSTM 오토인코더 모델에 적용되는 복수의 연속적인 프레임들은 인페인팅 패턴이 적용되지 않는 정상적인 프레임들일 수 있다.

[0065] 이상 물체 탐지부(125)는 제1 예측부의 출력 결과인 복원 프레임과 입력 프레임을 이용하여 공간 영역 재건 에러를 도출하고, 제2 예측부의 출력 결과인 시간 영역 복원 프레임들과 입력된 연속적인 프레임들을 이용하여 시간 영역 재건 에러를 도출한 후 이를 융합하여 이상 물체가 위치하는 영역을 추정한다.

[0066] 공간 영역 재건 에러, 시간 영역 재건 에러는 각각 원래 프레임과 예측 프레임 사이의 거리를 측정하는 것이므로, 이상 물체 탐지부(125)는 최종 융합 에러에서 추출된 특징으로 분류기(SVM: support vector machine)를 학습하여 이상 정도를 결정할 수 있다. 여기서 분류기는 단일 클래스(one-class) 분류기일 수 있다.

[0067] 본 발명의 일 실시예에서는 분류기 학습을 위해 각 픽셀의 에러값의 거리를 평균값을 계산한다.

[0068] 융합 에러를 R이라 하며, m x n차원을 가지는 프레임에 대한 제1 분류 특징은 수학식 3과 같다.

수학식 3

$$\|F1\| = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |R_{ij}|^2}$$

[0069]

[0070] 여기서, i, j는 픽셀을 나타낸다.

[0071] 이상 강도를 나타내는 제1 분류 특징이 있으므로, 제2 분류 특징은 프레임 내의 이상 픽셀의 위치를 추출할 수 있다. 픽셀의 인덱스 개수를 선택하고, 값들의 평균값을 계산한 후 이를 이용하여 임계값을 결정할 수 있다.

[0072] 이와 같이 계산된 임계값은 단일 클래스 분류기를 학습하기 위한 제2 분류 특징으로 사용된다. 분류기는 정규 데이터 특징으로 훈련되어 정규 분포를 다른 것과 분리할 수 있는 초평면을 형성하도록 학습된다. 분류기는 임

력 특징이 주어지면, 이상 정도는 분리된 초평면까지의 거리로 계산된다. 정규화를 통해 모든 점수가 0에서 1 범위 내에 있도록 max-min에 의해 거리 값이 조절된다. 메모리(130)는 본 발명의 일 실시예에 따른 동영상내의 이상 물체 탐지 방법을 수행하기 위해 필요한 명령어들을 저장한다.

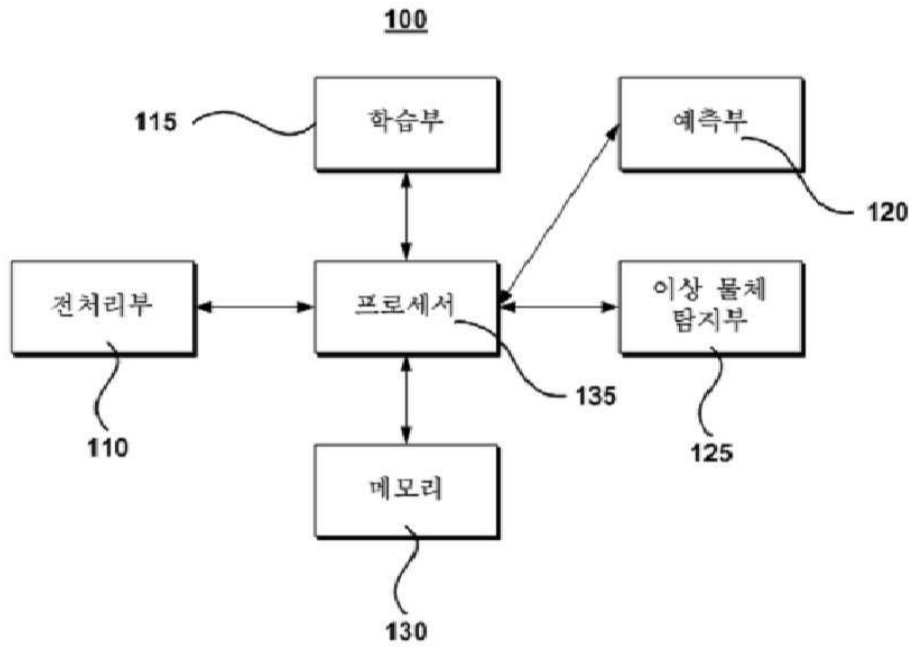
- [0073] 프로세서(135)는 본 발명의 일 실시예에 따른 동영상내의 이상 물체 탐지 장치(100)의 내부 구성 요소들(예를 들어, 전처리부(110), 학습부(115), 예측부(120), 이상 물체 탐지부(125), 메모리(130) 등)을 제어하기 위한 수단이다.
- [0075] 도 7은 본 발명의 일 실시예에 따른 동영상내의 이상 물체 탐지 방법을 나타낸 순서도이고, 도 8은 본 발명의 일 실시예에 따른 인페인팅 기법을 적용한 오토인코더 모델의 공간 영역 재건을 설명하기 위해 도시한 도면이고, 도 9는 본 발명의 일 실시예에 따른 이상 물체 탐지 방법의 상세 프레임워크를 도시한 도면이다.
- [0076] 단계 710에서 이상 물체 탐지 장치(100)는 트레이닝 데이터 셋을 이용하여 오토인코더 모델과 LSTM 오토인코더 모델을 각각 학습한다. 여기서, 오토인코더 모델과 LSTM 오토인코더 모델은 콘볼루션 오토인코더 모델, 콘볼루션 LSTM이 적층된 오토인코더 모델일 수 있다.
- [0077] 오토인코더 모델과 LSTM 오토인코더 모델은 이미 진술한 바와 같이, 정상적인 프레임들(즉, 이상 물체가 존재하지 않는 프레임들)을 대상으로 학습될 수 있다. 이는 도 1을 참조하여 설명한 바와 동일하므로 중복되는 설명은 생략하기로 한다.
- [0078] 단계 715에서 이상 물체 탐지 장치(100)는 동영상내의 현재 프레임(입력 프레임)에 대해 복수의 인페인팅(inpainting) 패턴을 각각 적용하여 입력 배치를 생성한다.
- [0079] 단계 720에서 이상 물체 탐지 장치(100)는 인페인팅 패턴이 적용된 입력 배치를 각각 학습된 오토인코더 모델에 적용하여 복원 배치를 재건하고, 재건된 복원 배치를 결합하여 복원 프레임을 출력한다. 이에 대한 일련의 과정을 보다 상세히 도시하면 도 8에 도시된 바와 같다.
- [0080] 단계 725에서 이상 물체 탐지 장치(100)는 복수의 연속적인 프레임들을 학습된 LSTM 오토인코더 모델에 적용하여 시간 영역 특징점을 추출한 후 복원하여 시간 영역 복원 프레임들을 출력한다.
- [0081] 단계 730에서 이상 물체 탐지 장치(100)는 복원 프레임과 현재 프레임을 이용하여 공간 영역 재건 에러를 도출하며, 연속적인 프레임들과 시간 영역 복원 프레임들을 이용하여 시간 영역 재건 에러를 도출하고, 공간 영역 재건 에러와 시간 영역 재건 에러를 융합하여 이상 물체가 위치하는 영역을 추정한다. 상술한 바와 같이, 공간 영역 예측시 인페인팅 기법을 적용하여 비정상 특징 벡터를 생성하여 재건 에러를 증가시킨 후 시간 영역 예측 결과와 융합하여 이상 물체를 탐지하는 일련의 과정에 대한 프레임워크를 보다 상세히 도시하면 도 9와 같이 나타낼 수 있다.
- [0083] 본 발명의 실시예에 따른 장치 및 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 컴퓨터 판독 가능 매체에 기록되는 프로그램 명령은 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 분야 통상의 기술자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media) 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.
- [0084] 상술한 하드웨어 장치는 본 발명의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.
- [0085] 이제까지 본 발명에 대하여 그 실시예들을 중심으로 살펴보았다. 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명이 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 구현될 수 있음을 이해할 수 있을 것이다. 그러므로 개시된 실시예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 한다. 본 발명의 범위는 전술한 설명이 아니라 특허청구범위에 나타나 있으며, 그와 동등한 범위 내에 있는 모든 차이점은 본 발명에 포함된 것으로 해석되어야 할 것이다.

부호의 설명

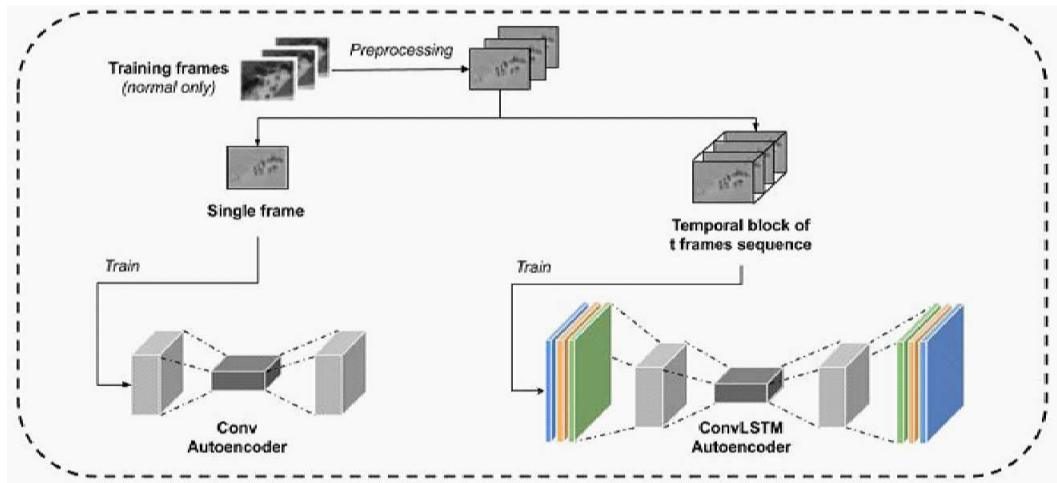
- [0087] 100: 이상 물체 탐지 장치
- 110: 전처리부
- 115: 학습부
- 120: 예측부
- 125: 이상 물체 탐지부
- 130: 메모리
- 135: 프로세서

도면

도면1



도면2



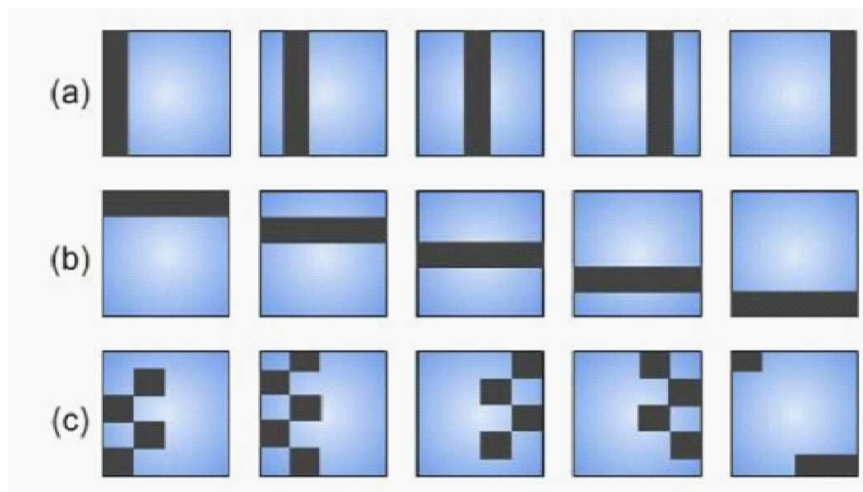
도면3

Conv. Auto.		
Input	-	(160, 240, 1)
Conv. ⁶	128 fil, 7x7, st=4	(20, 30, 64)
BN	-	(40, 60, 128)
Activation	ReLU	(t, 40, 60, 128)
Conv.	64 fil, 5x5, st=2	(20, 30, 64)
BN	-	(20, 30, 64)
Activation	ReLU	(20, 30, 64)
Conv.	32 fil, 3x3, st=1	(20, 30, 32)
BN	-	(t, 20, 30, 32)
Activation	ReLU	(t, 20, 30, 32)
Deconv. ⁷	64 fil, 3x3, st=2	(40, 60, 64)
BN	-	(t, 40, 60, 64)
Activation	ReLU	(t, 40, 60, 64)
Deconv.	1 fil, 11x11, st=4	(160, 240, 1)

도면4

ConvLSTM Auto. ⁰	Configuration	Output Shape
Input	-	(t, 160, 240, 1)
TDC ¹	128 fil ⁴ , 11x11, st ⁵ =4	(t, 40, 60, 128)
BN ²	-	(t, 40, 60, 128)
Activation	ReLU	(t, 40, 60, 128)
ConvLSTM 1	64 fil, 3x3, st=1	(t, 20, 30, 64)
ConvLSTM 2	48 fil, 3x3, st=1	(t, 20, 30, 48)
ConvLSTM 3	32 fil, 3x3, st=1	(t, 20, 30, 32)
ConvLSTM 4	16 fil, 3x3, st=1	(t, 20, 30, 16)
ConvLSTM 5	64 fil, 3x3, st=1	(t, 20, 30, 64)
TDD ³	64 fil, 5x5, st=2	(t, 40, 60, 128)
BN	-	(t, 40, 60, 128)
Activation	ReLU	(t, 40, 60, 128)
TDD	1 fil, 11x11, st=4	(t, 160, 240, 1)

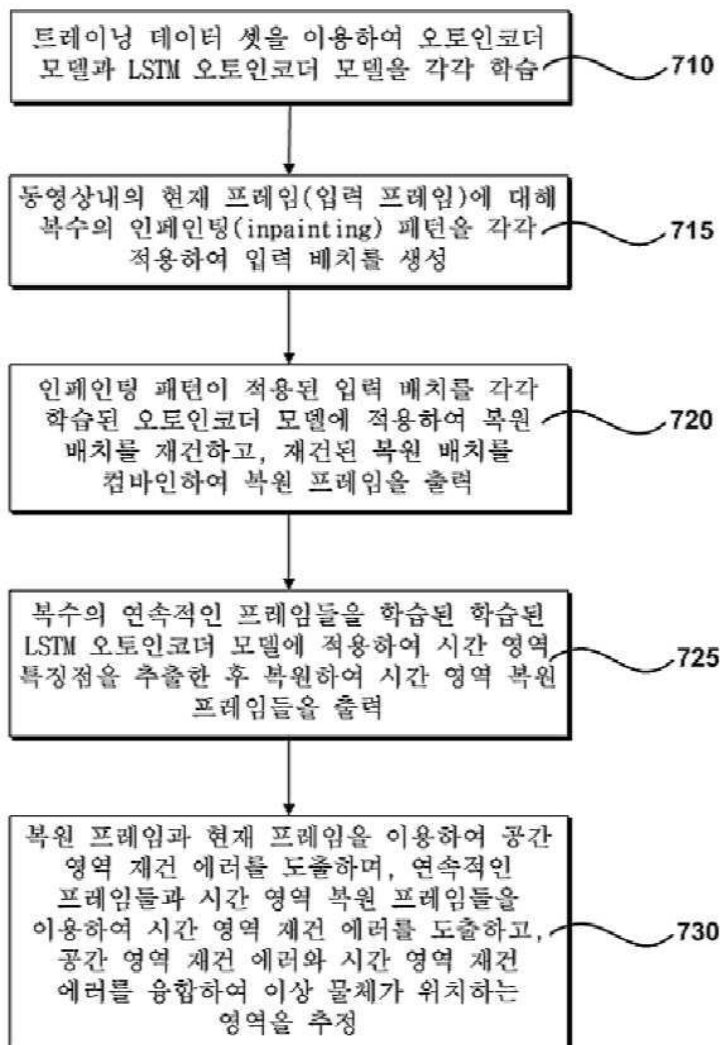
도면5



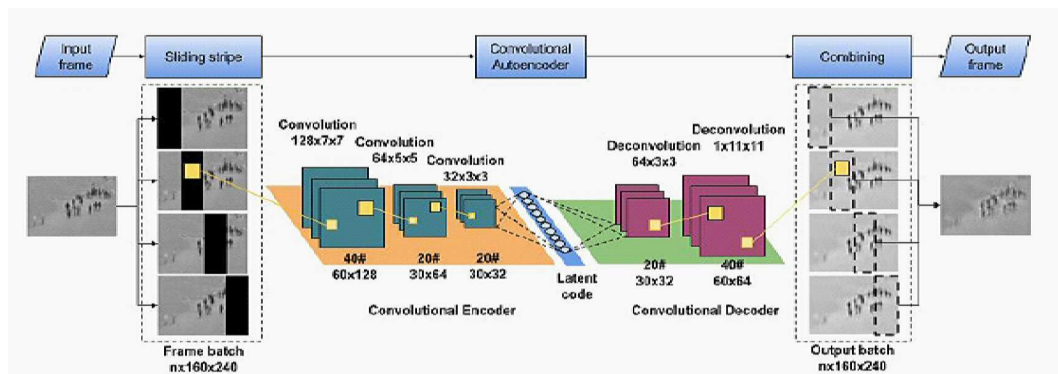
도면6

$\Psi_{vs}(X)$	Peds1 appearance pixel AUC	Peds2 appearance pixel AUC
vs_3	92.0%	91.7%
vs_7	91.0%	89.36%
vs_{10}	89.2%%	84.36%
hs_3	91.7%	91.1%
hs_7	88.6%	83.7%
hs_{10}	87.7%	86.1%

도면7



도면8



도면9

