



일반논문 (Regular Paper)

방송공학회논문지 제29권 제6호, 2024년 11월 (JBE Vol.29, No.6, November 2024)

<https://doi.org/10.5909/JBE.2024.29.6.1067>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

객체 탐지와 의미론적 분할을 활용한 규칙 기반 제로-샷 영상 이상 탐지

송인표^{a)}, 이장원^{a)†}

Rule-based Zero-shot Video Anomaly Detection Using Object Detection and Semantic Segmentation

Inpyo Song^{a)} and Jangwon Lee^{a)†}

요약

영상 이상 감지는 보안, 안전, 품질 관리 등 광범위한 응용 가치로 인해 주목받는 딥러닝 연구 분야이다. 그러나 기존의 딥러닝 기반 이상 탐지 방법들은 학습 데이터에 대한 높은 의존성과 탐지 결과에 대한 제한적인 설명이라는 근본적인 한계를 가진다. 본 연구에서는 이러한 한계를 극복하기 위해 객체 탐지와 의미론적 분할을 결합한 규칙 기반의 제로-샷 영상 이상 탐지 프레임워크를 제안한다. 제안된 방법은 객체와 배경 간의 관계를 명시적 규칙으로 정의하고, 사전 학습된 비전 모델을 활용하여 장면의 구조를 정확하게 파악한다. 이를 통해 도메인 특화 학습 없이도 새로운 환경에서 효과적으로 이상을 탐지할 수 있다. Shanghaitech와 NWPU Campus 데이터셋을 활용한 실험을 통해, 제안 방법이 추가 학습 없이도 기존 접근법 대비 우수한 성능을 보임을 검증하였다.

Abstract

Video anomaly detection has emerged as a prominent deep learning research field due to its extensive applications in security, safety, and quality control. However, existing deep learning-based anomaly detection methods face fundamental limitations, heavy reliance on training data and limited explainability of detection results. To overcome these challenges, we propose a rule-based zero-shot video anomaly detection framework that integrates object detection and semantic segmentation. Our approach defines explicit rules based on object-background relationships and accurately interprets scene structure using pre-trained vision models. This enables effective anomaly detection in new environments without domain-specific training. Through experiments on the Shanghaitech and NWPU Campus datasets, we demonstrate that our method achieves superior performance to existing approaches without additional training.

Keyword : Video Anomaly Detection, Object Detection, Semantic Segmentation

a) 성균관대학교 실감미디어공학과(Sungkyunkwan University)

† Corresponding Author : 이장원(Jangwon Lee)

E-mail: leejang@skku.edu

Tel: +82-2-760-0557

ORCID: <https://orcid.org/0000-0002-6601-7302>

※ 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 메타버스융합대학원(IITP-2024-RS-2023-00254129) 및 디지털분야해외석학유치지원(RS-2024-00459638)의 연구결과로 수행되었음.

· Manuscript November 11, 2024; Revised November 14, 2024; Accepted November 14, 2024.

Copyright © 2024 Korean Institute of Broadcast and Media Engineers. All rights reserved.

“This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered.”

I. 서론

영상 이상 탐지(Video Anomaly Detection)는 비정상적인 패턴이나 위험 상황을 자동으로 식별하는 컴퓨터 비전(Computer vision) 기술을 말한다. 이는 공공 안전, 산업 품질 관리, 의료 모니터링, 자율 주행 등 광범위한 분야에서 중요한 응용 가치를 지닌다^[1,2]. 특히 CCTV(Closed Circuit Television) 기반 도시 감시 시스템에서는 무단 횡단, 담장 침입, 차량의 인도 진입과 같은 다양한 위험 상황을 감지하여 사고를 예방하는 것이 중요하다.

그러나 영상 이상 탐지는 다음과 같은 여러 도전 과제를 수반한다. 첫째, 이상(Anomaly)의 정의가 맥락과 도메인에 따라 다양하게 변할 수 있다. 예를 들어, 도로에서의 보행자는 정상이지만 철도에서의 보행자는 이상 상황으로 간주된다. 둘째, 실제 환경에서 발생 가능한 모든 유형의 이상을 사전에 예측하고 학습하는 것이 불가능하다. 셋째, 이상 이벤트는 희귀하게 발생하므로, 학습 데이터를 수집하기 어렵다. 넷째, 공공 안전과 직결되는 특성상 탐지된 이상에 대한 명확한 설명이 필요한 경우가 많다.

이러한 도전 과제들을 해결하기 위해 다양한 접근 방식이 제안되어 왔다. 초기의 연구들은 주로 hand-crafted 특징을 사용한 통계적 모델링에 집중했으나^[3], 최근에는 컨볼루션 신경망(CNN)을 활용한 오토인코더(Auto-encoder)^[4,5], 생성형 적대 신경망(GAN)^[7] 등의 딥러닝 기반 방법들이 주목받고 있다. 그러나 이러한 최신 방법들도 여전히 몇 가지 근본적인 한계를 가지고 있다. 대부분의 방법들이 특정 도메인에 대한 대량의 학습 데이터를 필요로 하며, 학습 단계에서 접하지 않은 새로운 유형의 이상이나 환경 변화에 대한 적응력이 부족하다. 또한 딥러닝 모델의 'Black-box' 특성으로 인해 탐지 결과에 대한 명확한 설명을 제공하기 어렵다는 문제가 있다.

본 연구에서는 이러한 한계점들을 극복하기 위해 객체 탐지와 의미론적 분할을 활용한 규칙 기반 제로-샷(Zero-shot) 영상 이상 탐지 프레임워크를 제안한다. 제안하는 방법은 사전 학습 모델을 이용하여 객체 탐지 및 의미론적 분할을 수행하고, “인도 위의 차량”, “담장을 넘는 사람” 등과 같이 객체 정보와 위치 정보 간의 관계에 기반한 명시

적 규칙을 통해 이상 상황을 탐지한다. 사전 학습 모델 외에 추가 학습이 필요하지 않아 제로-샷 접근이 가능해지며, 이는 새로운 환경에서도 효과적으로 이상을 탐지할 수 있는 장점을 가진다. 또한, 규칙 기반 탐지로 결과에 대한 명확한 설명이 가능하다. 제안하는 방법의 주요 기여점은 다음과 같다:

규칙 기반 제로-샷 접근: 객체와 위치 관계에 기반한 명시적 규칙을 통해, 도메인 특화 학습 없이도 새로운 환경에서 적응적으로 이상을 탐지할 수 있다. 이는 기존의 딥러닝 기반 방법들이 학습 데이터의 분포를 벗어나는 경우 겪는 어려움을 해결한다.

높은 설명 가능성: 객체와 위치 관계에 기반한 명시적 규칙을 통해, 이상 상황을 탐지하므로 탐지 결과에 대해 명확한 설명이 가능하다. 이는 실제 감시 시스템을 비롯한 활용에서 사용자의 신속하고 정확한 의사결정을 지원할 수 있다.

본 연구의 효과성을 검증하기 위해 ShanghaiTech^[7], NWPU Campus^[8]와 같은 대표적인 이상 탐지 데이터셋에서 이상감지 성능을 평가하였다. 실험 결과는 제안하는 방법이 새로운 환경에서도 효과적으로 이상을 탐지할 수 있음을 보여준다.

본 논문의 구성은 다음과 같다. 2장에서는 영상 이상 탐지와 관련된 기존 연구들을 소개하고, 3장에서는 제안하는 접근법의 세부 내용을 설명한다. 4장에서는 상용 데이터셋을 이용한 실험 성능 평가 결과를 제시하고, 5장에서는 결론 및 한계점, 그리고 향후 연구 방향을 논의한다.

II. 관련 연구

영상 이상 탐지 연구는 학습 데이터의 구성과 레이블링 방식에 따라 준지도 학습(Semi-supervised learning)과 약지도 학습(Weakly-supervised learning) 방식으로 구분된다. 또한, 최근에는 설명 가능한 이상 탐지 및 학습 소요를 낮추는 제로-샷 이상 탐지에 대한 연구들이 활발히 진행되고 있다.



그림 1. 대표적인 준지도학습 복원 기반 이상 감지 프레임워크
 Fig. 1. Representative semi-supervised anomaly detection framework using reconstruction

1. 준지도 학습 영상 이상 탐지

준지도 학습 기반 접근법은 정상 데이터만으로 구성된 학습 데이터셋을 사용하여 모델이 정상 패턴을 학습하고, 이를 기반으로 정상 및 비정상 데이터가 섞여 있는 검증 및 평가 데이터에서 이상을 탐지한다. 준지도 학습 기반 접근법들을 주로 이미지 복원(Reconstruction) 혹은 이미지 예측(Prediction) 기반 프레임워크를 활용하여, 입력 영상과 복원/예측된 영상 간의 차이를 이상 점수로 활용한다. 관련 연구로 초기에는 컨볼루션 신경망 기반 오토인코더를 활용한 방법들^[4,5]이 제안되었으며, 이후 연구에서는 메모리 모듈을 도입하여^[9,10] 정상 패턴의 특징을 저장하고 참조하는 방식이 시도되었다. 또한, 적대적 신경망(Adversarial network)^[7]과 적대적 학습(Adversarial learning)^[11]을 통해 정상과 비정상 패턴의 구분을 심화하는 방법들이 제안되었다. 이러한 준지도 학습 기반 방법들은 그림 1과 같이 복원 혹은 예측 이미지와 정답 이미지의 오차 시각화를 통해 이상이 탐지된 영역을 특정할 수 있다는 장점이 있으나, 그 이상의 의미론적 해석에는 한계가 있다.

2. 약지도 학습 영상 이상 탐지

약지도 학습 이상 탐지 환경에서는 학습 데이터와 평가 데이터 모두 정상 및 비정상 데이터를 포함되며, 학습 시에는 비디오 단위의 이상 상황 포함 여부 정보만이 제공된다. 이는 프레임 단위의 상세한 레이블링이 없이도 모델을 학습할 수 있어, 준지도 학습과 비슷한 수준의 레이블링 소요를 가지면서도 비교적 높은 정확도를 보인다. 대표적인 방법으로는 다중 인스턴스 학습(Multiple Instance Learning)^[1], 특징 강도(Feature magnitude) 기반 학습^[12], 자기 학습(Self training)^[13] 등이 있다. 하지만, 모델 구조가 복잡하고 이상 탐지 영역을 특정할 수 없는 등 이상 상황에 대한 설명이 제한적이라는 한계가 있다.

3. 설명 가능한 영상 이상 탐지

최근에는 이상 탐지 결과의 해석과 설명을 시도하는 연구들이 이루어지고 있다. Singh 등은 이미지를 격자 형태 영역으로 분할하고 각 영역의 외관(Appearance)과 움

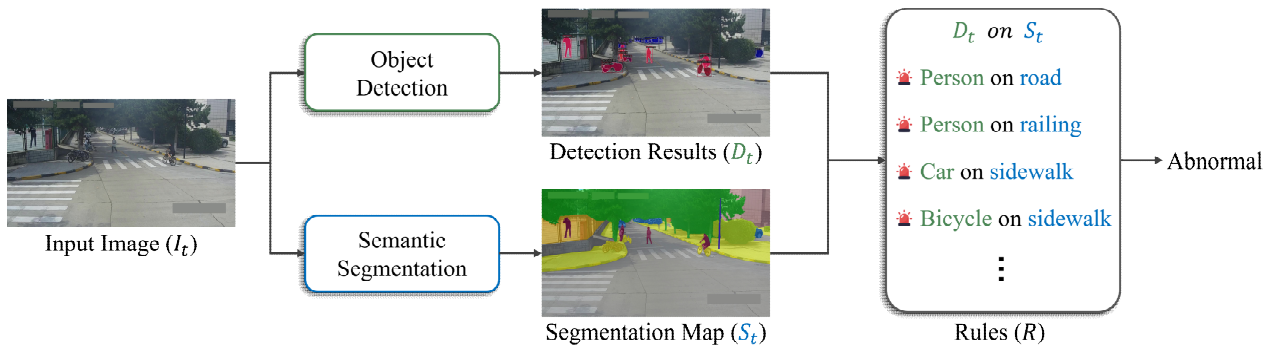


그림 2. 객체 인식과 의미론적 분할을 활용한 규칙 기반 영상 이상 탐지
 Fig. 2. Rule based anomaly detection using object detection and semantic segmentation

직임을 분석하여 인간이 이해할 수 있는 형태의 설명과 함께 이상 상황을 탐지한다^[14]. Gu 등은 대규모 언어 모델과 트랜스포머 구조를 활용하여 탐지된 이상 유형에 대한 설명을 제공한다^[15]. 그러나 이러한 방법론들도 새로운 환경이나 이상 유형에 대한 추가 학습이 필요하거나 이미지 레벨의 데이터에서의 적용에 그친다는 한계가 있다. 특히 감시 시스템을 비롯한 실제 활용 시에 요구되는 직관적인 설명 제공은 여전히 해결해야 할 과제로 남아 있다.

이러한 한계점들을 극복하기 위해, 본 연구에서는 객체 탐지와 의미론적 분할을 결합한 규칙 기반의 제로-샷 비디오 이상 감지 방법론을 제안한다. 제안하는 방법은 학습 없이도 새로운 환경에서 효과적으로 동작하며, 객체와 배경 간의 관계에 기반한 직관적인 설명을 제공할 수 있다.

III. 제안 기법

본 장에서는 객체 탐지와 의미론적 분할을 결합한 규칙 기반 이상 탐지 프레임워크를 제안한다. 제안하는 방법은 “차량이 보도 위에 있다” 또는 “사람이 물 위에 있다”와 같이, 객체와 배경 간의 관계를 직관적인 규칙으로 정의하고 이를 기반으로 이상을 탐지한다. 이러한 규칙 기반 접근은 새로운 환경에서도 추가 학습 없이 적용 가능하며(제로-샷), 탐지 결과에 대해 명확한 설명을 제공할 수 있다는 장점이 있다.

1. 객체 인식 및 의미론적 분할

제안하는 프레임워크는 각 입력 프레임에 대해 객체 탐지와 의미론적 분할을 수행한 후, 사전 정의된 규칙에 기반하여 이상을 탐지한다. 이를 위해 먼저 최신 비전 모델인 InternImage를 기반으로 객체 탐지와 의미론적 분할을 수행한다^[16].

$$\begin{aligned} I_t &\rightarrow \text{Object Detection}(I_t) \rightarrow D_t \\ I_t &\rightarrow \text{Semantic Segmentation}(I_t) \rightarrow S_t \end{aligned} \quad (1)$$

식 (1)과 같이 t 번째 프레임의 이미지 입력 I_t 는 객체 탐지와 의미론적 분할 과정을 거쳐 객체 탐지 결과(Detection results) D_t 와 의미론적 분할 결과(Segmentation map) S_t 가 검출된다.

$$\begin{aligned} d_t^i &= (b_t^i, c_t^i, s_t^i) \in D_t \\ \text{where } b_t^i &= (x_1, y_1, x_2, y_2) \end{aligned} \quad (2)$$

t 번째 프레임의 객체 탐지 결과 D_t 는 객체들의 집합으로 식 (2)와 같이, 바운딩 박스(Bounding box) 좌표 $b_t^i = (x_1, y_1, x_2, y_2)$, 객체 클래스 c_t^i , 탐지 신뢰도 점수 s_t^i 로 구성되는 각 객체의 객체 탐지 결과 d_t^i 의 집합이다.

$$S_t(x, y) \in \{1, \dots, K\} \quad (3)$$

t 번째 프레임의 의미론적 분할 결과 S_t 는 각 픽셀 위치 (x, y) 에 대해 식 (3)과 같이 정의된다. 여기서 K 는 총 클래스 수이며, $S_t(x, y)$ 는 위치 (x, y) 에서의 배경 클래스를 나타낸다.

2. 객체-배경 관계 분석

객체와 배경의 관계를 분석하기 위해, 먼저 각 객체의 바운딩 박스 주변 영역에서 가장 빈번하게 나타나는 배경 클래스를 식 (4)에 나타난 배경 클래스 추출 함수 B 에 따라 추출한다.

$$B(b_t^i, S_t) = \operatorname{argmax}_c \{ \{(x, y) \in A(b_t^i) | S_t(x, y) = c\} \} \quad (4)$$

여기서, argmax_c 는 가장 많이 등장하는 배경 클래스 c 를 반환하는 함수이며, $A(b_t^i)$ 는 바운딩 박스 좌표 b_t^i 에 의해 정의되는 영역(Area)을 의미하며 식 (5)와 같이 나타낼 수 있다.

$$A(b_t^i) = \{(x, y) | x_1 \leq x \leq x_2, y_1 \leq y \leq y_2\} \quad (5)$$

이 과정을 통해 각 객체가 위치한 주요 배경 영역을 식별

할 수 있으며, 이를 이용하여 다음 단계에서 정의될 규칙 기반 이상 탐지를 위한 핵심 정보가 된다.

3. 규칙 기반 이상 탐지

이상 탐지를 위한 규칙 집합 R 은 식 (6)과 같이 객체 클래스 c_t^i 와 해당 객체가 존재하면 안 되는 배경 클래스의 집합 $K_{forbidden}$ 의 쌍으로 정의된다.

$$R = \{(c_t^i, K_{forbidden})\} \quad (6)$$

예시로 다음과 같은 직관적인 규칙을 적용할 수 있다.

(car, {sidewalk, water}), (person, {road, railing})

위와 같이 규칙을 적용하면 차가 보도나 물 위에 있는 경우, 사람이 차도나 난간 위에 있는 경우를 이상 상황으로 정의하는 규칙을 세울 수 있다. 이러한 규칙을 기반으로 프레임 t 에 대한 이상 탐지 여부 AD_t 는 식 (7)과 같이 정의된다.

$$AD_t = \begin{cases} 1 & \text{if } (c_t^i, B(b_t^i, S_t)) \in R \text{ for } d_t^i \in D_t \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

이를 통해 각 프레임의 이상 탐지 여부는 프레임 내의 모든 객체에 대해, 객체의 클래스와 해당 객체가 위치한 배경 클래스의 조합이 규칙 집합 R 에 정의된 금지된 조합과 일치할 경우 이상으로 판단한다.

IV. 실험 결과

본 장에서는 제안한 규칙 기반 제로-샷 이상 탐지 프레임워크의 성능을 검증하기 위한 실험과 그 결과를 논의한다. Shanghaitech와 NWPU Campus 데이터셋을 활용하여 기존 방법론과의 비교 실험을 통해 제안 방법의 효과성을 검증한다.

1. 데이터셋

실험에는 영상 이상 탐지 분야의 대표적인 벤치마크인 Shanghaitech와 NWPU Campus 데이터셋을 사용하였다. 두 데이터셋은 CCTV 환경에서 촬영된 영상으로 구성되어 있어 실제 도시 감시 시스템의 이상 탐지 성능을 평가하기에 적합하다. 본 실험에서는 단순 달리기나 장난과 같은 경미한 행동을 제외하고, 보도 위의 차량 진입이나 차도 위의 보행자와 같이 교통 안전 및 공공 안전과 직결되는 심각한 이상 상황을 중심으로 평가 데이터를 구성하였다. 이러한 선별 과정을 통해 Shanghaitech와 NWPU Campus 각각 기존 평가 데이터의 약 60%를 실험에 사용하였다.

2. 실험 설정

객체 인식과 의미론적 분할을 위해 InternImage 모델을 사용하였으며, 80개의 객체 클래스를 가지는 COCO 데이터셋과 150개 클래스를 포함하는 ADE20K 데이터셋에서 사전학습된 가중치를 적용하였다^[17,18]. 객체 인식 결과 중 탐지 신뢰도 점수가 0.3 이상인 객체만을 규칙 기반 이상 탐지에 활용하였다. 본 실험에서 실제로 사용된 규칙 설정은 (person, {road, water, fence}), (skateboard, {sidewalk}), (car, {sidewalk}), (bus, {sidewalk}), (truck, {sidewalk}), (bicycle, {sidewalk}), (motorcycle, {sidewalk})이다. 성능 비교를 위해 복원 기반 이상 탐지 방법인 MNAD^[10]를 기준 모델로 선정하였다. 기존 방법들의 경우 이상 예측값이 0.5 이상일 때 이상으로 판단하였다. 제로-샷 환경에서의 공정한 평가를 위해 교차 검증(Cross validation) 방식을 채택하였다. 구체적으로, Shanghaitech에서 학습 후 NWPU Campus에서 평가하거나, 그 반대의 경우를 실험하였다. 평가 지표는 전체 예측 중 정확한 예측의 비율을 나타내는 정확도(Accuracy), 실제 이상 중 정확히 탐지된 비율을 의미하는 재현율(Recall), 이상으로 예측된 것 중 실제 이상인 비율을 나타내는 정밀도(Precision), 그리고 정밀도와 재현율의 조화평균인 F1 점수(F1 score)를 사용하였다.

표 1. NWPU Campus 학습, Shanghaitech 평가 이상 감지 실험
 Table 1. Anomaly detection results on Shanghaitech dataset with model trained on NWPU Campus

Method	Acc	Recall	Precision	F1 Score
MNAD w/o memory	50.00	59.78	40.59	48.35
MNAD	52.92	60.20	58.99	59.59
Ours	72.76	74.11	60.68	61.23

표 2. Shanghaitech 학습, NWPU Campus 평가 이상 감지 실험
 Table 2. Anomaly detection results on NWPU Campus dataset with model trained on Shanghaitech

Method	Acc	Recall	Precision	F1 Score
MNAD w/o memory	53.15	71.90	46.48	56.46
MNAD	57.90	70.87	60.44	65.24
Ours	66.51	78.45	63.07	68.32

3. 이상 감지 실험 결과

표 1과 표 2의 교차 검증 결과에서 확인할 수 있듯이, 제안한 규칙 기반 방법은 기존의 복원 기반 이상 탐지 모델인 MNAD와 비교하여 전반적으로 우수한 성능을 보였다. 특히 MNAD가 학습에 상당한 시간을 소요한 반면, 제안 방법은 별도의 학습 없이도 안정적인 성능을 유지하였다. Shanghaitech 데이터셋에서는 19.84%, NWPU 데이터셋에서는 8.61%의 정확도 향상을 달성하였다. 재현율, 정밀도, F1 점수 모두에서의 향상은 이상 상황 검출의 정확성과 신

뢰성이 모두 개선되었음을 입증한다.

그림 3은 Shanghaitech와 NWPU Campus 데이터셋에서의 이상 탐지 결과를 시각화한 것이다. 제안한 방법이 객체와 배경 간의 관계를 효과적으로 분석하여 이상 상황을 정확하게 탐지함을 보여준다. 또한 이상 상황에 대한 직관적인 설명 제공을 통해 실제 감시 시스템에서의 활용 가능성을 입증하였다.

V. 결론 및 향후 연구

본 논문에서는 객체 탐지와 의미론적 분할을 결합한 규칙 기반의 제로-샷 영상 이상 탐지 프레임워크를 제안하였다. 기존의 딥러닝 기반 이상 탐지 방법들이 직면한 학습 데이터 의존성과 설명 가능성의 한계를 극복하기 위해, 객체와 배경 간의 관계를 명시적 규칙으로 정의하고 이를 기반으로 이상을 탐지하는 새로운 방법론을 개발하였다. 제안된 방법은 도메인 특화 학습 없이도 새로운 환경에서 효과적으로 이상을 탐지할 수 있으며, 탐지 결과에 대해 직관적인 설명을 제공할 수 있다는 장점을 가진다. Shanghaitech와 NWPU Campus 데이터셋을 활용한 실험적 검증을 통해 제안 방법의 효과성을 입증하였다. 특히 별도 학습 없이 새로운 환경에서도 안정적인 성능을 유지함을 확인하였다. 본 연구는 실제 감시 시스템에서 핵심적으로 요구되는

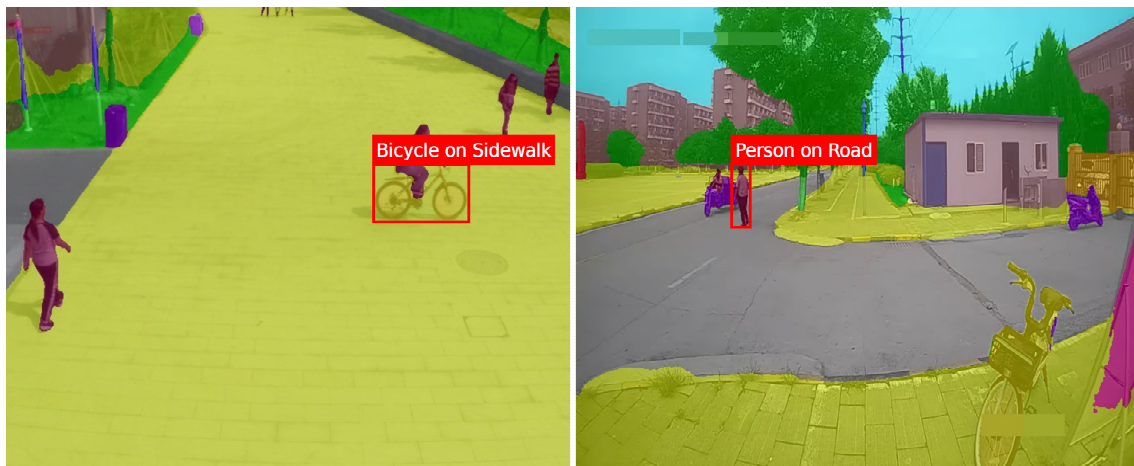


그림 3. Shanghaitech (좌), NWPU Campus (우) 이상 감지 시각화
 Fig. 3. Anomaly detection visualization on Shanghaitech (left), and NWPU Campus (right)

설명 가능성과 새로운 환경에 대한 적응력을 동시에 만족시키는 실용적인 이상 탐지 프레임워크를 제시했다는 점에서 학술적 의의가 있다.

그러나 본 연구는 방화나 강도와 같이 특정 영역과의 관계성이 없는 이상 행동에 대해서는 탐지가 제한된다는 한계를 가진다. 이러한 한계를 극복하기 위해 후속 연구에서는 시각적 특징 기반의 이상 탐지 방법론과의 통합을 진행할 예정이다. 또한 대규모 언어 모델을 활용한 규칙 정의의 자동화를 통해 시스템의 실용성을 더욱 향상시키는 연구를 수행하고자 한다.

참 고 문 헌 (References)

- [1] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6479-6488. doi: <https://doi.org/10.1109/CVPR.2018.00678>
- [2] Y. Yao, M. Xu, Y. Wang, D. J. Crandall, and E. M. Atkins, "Unsupervised traffic accident detection in first-person videos," in 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2019, pp. 273-280. doi: <https://doi.org/10.1109/IROS40897.2019.8967556>
- [3] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz, "Robust real-time unusual event detection using multiple fixed-location monitors," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 3, pp. 555-560, March 2008. doi: <https://doi.org/10.1109/TPAMI.2007.70825>
- [4] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis, "Learning temporal regularity in video sequences," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 733-742. doi: <https://doi.org/10.1109/CVPR.2016.86>
- [5] Y. S. Chong and Y. H. Tay, "Abnormal event detection in videos using spatiotemporal autoencoder," in International Symposium on Neural Networks, Springer, Cham, 2017, pp. 189-196. doi: https://doi.org/10.1007/978-3-319-59081-3_23
- [6] M. Sabokrou, M. Khalooei, M. Fathy, and E. Adeli, "Adversarially learned one-class classifier for novelty detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3379-3388. doi: <https://doi.org/10.1109/CVPR.2018.00356>
- [7] W. Liu, W. Luo, D. Lian, and S. Gao, "Future frame prediction for anomaly detection - a new baseline," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6536-6545. doi: <https://doi.org/10.1109/CVPR.2018.00684>
- [8] C. Cao, Y. Huang, Y. Yang, L. Wang, Z. Wang, and T. Tan, "A new comprehensive benchmark for semi-supervised video anomaly detection and anticipation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 13937-13947. doi: <https://doi.org/10.1109/CVPR52729.2023.01953>
- [9] D. Gong, L. Liu, V. Le, B. Saha, M. R. Mansour, S. Venkatesh, and A. v. d. Hengel, "Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 1705-1714. doi: <https://doi.org/10.1109/ICCV.2019.00179>
- [10] H. Park, J. Noh, and B. Ham, "Learning memory-guided normality for anomaly detection," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 14372-14381. doi: <https://doi.org/10.1109/CVPR42600.2020.01438>
- [11] S. Akcay, A. Atapour-Abarghouei, and T. P. Breckon, "Ganomaly: Semi-supervised anomaly detection via adversarial training," in Computer Vision - ACCV 2018, Springer, Cham, 2019, pp. 622-637. doi: https://doi.org/10.1007/978-3-030-20893-6_39
- [12] Y. Tian, G. Pang, Y. Chen, R. Singh, J. N. Veigas, and H. Hassani, "Weakly-supervised video anomaly detection with robust temporal feature magnitude learning," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 4975-4986. doi: <https://doi.org/10.1109/ICCV48922.2021.00493>
- [13] S. Li, F. Liu, and L. Jiao, "Self-training multi-sequence learning with transformer for weakly supervised video anomaly detection," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, no. 2, 2022, pp. 1433-1441. doi: <https://doi.org/10.1609/aaai.v36i2.20028>
- [14] A. Singh, M. J. Jones, and E. G. Learned-Miller, "Eval: Explainable video anomaly localization," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 17393-17403. doi: <https://doi.org/10.1109/CVPR52729.2023.01795>
- [15] Gu, Zhaopeng, et al. "Filo: Zero-shot anomaly detection by fine-grained description and high-quality localization," Proceedings of the 32nd ACM International Conference on Multimedia, 2024, pp. 2041-2049. doi: <https://doi.org/10.1145/3664647.3680685>
- [16] W. Wang et al., "Internimage: Exploring large-scale vision foundation models with deformable convolutions," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 2794-2805. doi: <https://doi.org/10.1109/CVPR52729.2023.01385>
- [17] T. Y. Lin et al., "Microsoft coco: Common objects in context," in Computer Vision - ECCV 2014, Springer, Cham, 2014, pp. 740-755. doi: https://doi.org/10.1007/978-3-319-10602-1_48
- [18] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba, "Scene parsing through ade20k dataset," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 633-641. doi: <https://doi.org/10.1109/CVPR.2017.544>

저 자 소 개



송 인 표

- 2023년 : 한국항공대학교 기계공학과 공학사
- 2023년 ~ 현재 : 성균관대학교 실감미디어공학과 석사과정
- ORCID : <https://orcid.org/0009-0005-1281-9149>
- 주관심분야 : Human Pose Estimation, Anomaly Detection, Knowledge Distillation



이 장 원

- 2006년 : 성균관대학교 정보통신공학부 학사
- 2008년 : 성균관대학교 전자전기컴퓨터공학과 석사
- 2018년 : Indiana University Intelligent and Interactive Systems 박사
- 2018년 ~ 2020년 : ObjectVideo Labs at Alarm.com Research Scientist
- 2020년 ~ 2023년 : 한국항공대학교 항공전자정보공학부 교수
- 2023년 ~ 현재 : 성균관대학교 실감미디어공학과 교수
- ORCID : <https://orcid.org/0000-0002-6601-7302>
- 주관심분야 : Computer Vision for Robotics, Human-Robot Interaction