

# 교량 CCTV 화면에서의 자세 추정 기반 이상 행동 탐지

오수빈<sup>1</sup>, 강민정<sup>1</sup>, 이상민<sup>1</sup>  
<sup>1</sup>광운대학교 인공지능응용학과

sboh0020@gmail.com, minjkang0901@gmail.com, smlee5679@gmail.com

## Anomaly Detection by Human Pose Estimation On Surveillance Videos in Bridge

Su-Bin Oh<sup>1</sup>, Min-Jeong Kang<sup>1</sup>, Sang-Min Lee<sup>1</sup>  
<sup>1</sup>Dept. of Artificial Intelligence Application, Kwangwoon University

### 요 약

본 논문은 CCTV 화면에서의 다양한 이상상황 중 교량 데이터에 특화된 자세 추정 기반 이상탐지 알고리즘을 소개한다. 교량은 크게 도로, 인도 이렇게 두 구역으로 나뉘지며, 사람들의 이동방향이 한정적이라는 특징을 가지는 장소 중 하나이다. 이러한 장소적 특징을 이용하고자 사람 자세 추정을 통해 이상의 기준을 잡고 교량 데이터에 특화된 이상탐지 알고리즘을 제안한다. CCTV 영상은 이상을 정하기 어렵고 이상에 대한 레이블이 없는 데이터가 대부분이며 이상에 대한 레이블 생성시 많은 비용 발생이 필수적이다. 본 연구에서는 이러한 한계점을 극복하고자 영상 데이터를 이미지 단위가 아닌 영상 단위로 레이블이 담긴 weakly label 을 가지는 데이터를 활용한 이상탐지 모델을 이용하였다. 특히, 교량에서의 이상상황의 특징인 사람 자세 추정으로 추출한 특징을 추가하여 기존 알고리즘의 이상탐지 예측 성능을 개선하였다.

### 1. 서론

최근 산책로, 거리 등에서 발생하는 치안 관련 사회적 문제들이 점점 증가하고 있으며, 사회적 불안감이 조성되고 있다. 이러한 문제들은 인명 피해와 같은 더 큰 문제를 야기할 수 있어 빠른 조치가 필요하다. 그렇기 때문에 CCTV 설치의 중요성 또한 증가되고 있다. 하지만 CCTV 가 점차 증가할수록 사람이 모든 화면을 감시하기에는 위험 상황 판단에 대한 많은 비용이 많이 든다는 한계가 있다. 이러한 한계를 극복하기 위해 최근 인공지능을 이용한 다양한 기술들이 개발되고 있다. 본 논문에서는 다양한 상황에서의 이상탐지 연구에 대한 기여를 목표로 개선된 이상탐지 알고리즘을 제안한다. CCTV 로 확인할 수 있는 다양한 장소 중 교량 위의 CCTV 에 초점을 맞췄다.

목표하는 이상탐지 프로세스는 (그림 1)과 같다. CCTV 가 점점 증가하게 되면 (그림 1)과 같이 더 다양한 상황에서의 이상상황들이 발생하며, 이러한 독립적인 이상상황들을 하나의 모델이 탐지하기에는 한



(그림 1) 목표 이상탐지 프로세스

계가 존재한다. CCTV 영상을 이용한 이상탐지의 한계 중 하나인 이상 정의에 대한 모호함을 유발시킨다.

따라서 장소에 독립적인 모델 구축이 필요하다. 본 논문은 다양한 장소 중 교량 데이터를 이용하였고, 교량에서 발생 가능성이 높은 이상상황의 특징을 통해 이상에 대한 기준을 잡고자 하였고, 기존 이상탐지 모델에 사람 자세 추정 모델을 이용하여 이상탐지 알고리즘 개선하였다.

## 2. 관련 연구

### 2-1. CCTV 데이터를 이용한 이상탐지

CCTV 데이터의 이상탐지를 위해 이상에 대한 레이블이 필요하다. 하지만 실제로 영상 데이터의 프레임 단위로 레이블링이 되어있는 데이터는 충분하지 않으며, 이를 위해 전처리 과정을 진행하기에는 시간과 비용이 많이 소모된다[1]. 이런 한계를 극복하기 위해 레이블 정보가 적은 데이터를 활용하는 다양한 이상탐지 알고리즘이 등장하였다. 본 논문에서는 프레임 단위가 아닌 영상 단위로 이상에 대한 레이블링된 데이터를 이용하여 deep multiple instance learning(MIL) 방법을 사용한 이상탐지 알고리즘을 활용했다[2]. 해당 알고리즘은 영상 단위로 정상, 이상에 대해 데이터를 구분하고, 일정 길이로 나뉜 비디오를 instance 로써 학습시켜 각각의 정상, 이상 집단에 이상치 점수를 부여한다. 결론적으로 정의된 이상치 점수는 순위 손실 함수에 의해 학습되며 이상의 집단에서 제일 이상치 점수가 높은 구간을 이상상황이라고 탐지한다. 하지만 이러한 탐지방법은 비디오 구간에서의 상황인지는 가능하지만 데이터의 특성이 달라지면 모델이 빠르게 데이터에 적응하기에는 어려운 한계가 존재한다.

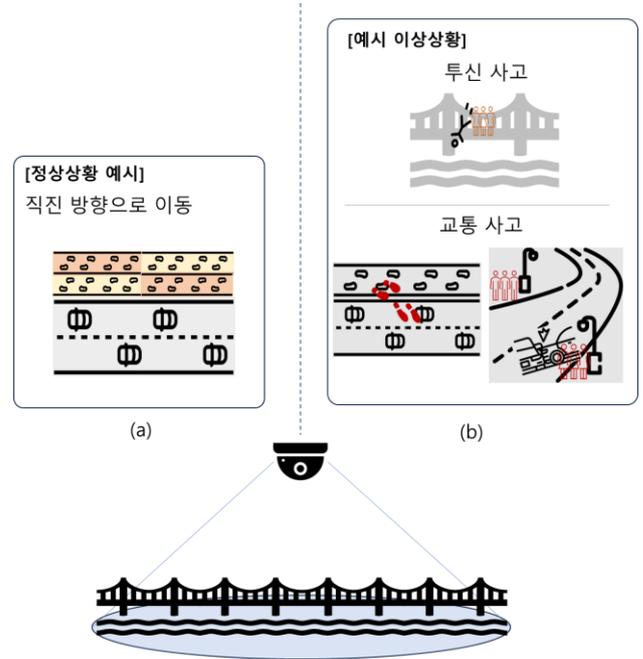
### 2-2. Human pose estimation

일반적인 human pose estimation(HPE)은 사람의 key point 를 통해 자세를 추정하는 기법이다. 본 논문은 교량 위의 특정 상황에 대한 이상을 정의할 때 사람의 움직임, 화면 상에서의 위치 등의 특징을 기반으로 이상을 정의하였다. 특히 여러 HPE 기법 중 각각의 특징을 모듈화하여 다양한 모델 구성이 가능한 detectron2 를 통해 특징을 추출하고자 하였다[3]. Detectron2 는 다양한 네트워크와 모델을 지원하며, 빠른 속도와 높은 정확도를 가진다.

본 논문에서는 ResNet-50 네트워크를 이용한 mask R-CNN 베이스 모델을 사용하였고, feature pyramid network(FPN)를 통해 여러 스케일에서 특징을 추출하였다.

## 3. 제안 방법론

교량 위에서의 정상 및 이상상황은 (그림 2)와 같은 특징을 가지고 있다.

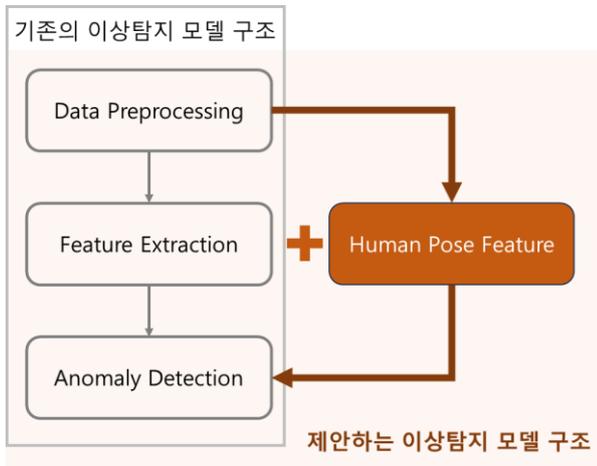


(그림 2) 교량에서의 이상 및 정상상황 예시

(그림 2-a)는 교량 위 정상상황에 대한 예시이다. 그림처럼 교량에서의 정상상황은 한정되어 있다. 도로에 비해 인도가 좁기 때문에 길거리보다는 사람의 행동이 한정되어 있다. 또한 대부분 한 개의 인도만 존재하므로 사람들이 걸어가는 방향성이 직진으로 한정되어 있다. (그림 2-b)는 교량 위 이상상황에 대한 예시이다. 투신 사고와 같은 이상상황일 경우 사람이 교량 위를 걷어가다가 이동하지 않고 주변을 돌아다니거나, 투신 사고를 막기위해 사람들이 몰리는 현상들이 발생한다. 교통 사고의 경우 크게 인사 사고와 차량 사고로 나눌 수 있다. 인사 사고의 경우 사람이 한 방향으로 이동을 하다가 갑자기 방향을 트는 경우 발생할 가능성이 높으며, 이런 상황 또한 주변에 사람들이 몰려들 가능성이 있다. 차량 사고의 경우 사고가 발생하면 사람이 앞의 경우들과 마찬가지로 사람들이 몰려드는 상황, 도로 위에서 사람들이 나와있는 상황 등 다양한 상황들이 발생한다.

이처럼 교량 위 정상 및 이상에 대한 상황들에 따라 사람들이 평소 교량을 건너는 상황과는 다른 움직임을 보이며, 이러한 영향력이 이상탐지 시에 영향이 아예 없다고 보기는 어렵다. 이러한 교량 데이터 특징에 집중하여 사람 자세 추정을 통해 이상상황에 대한 기준을 규정하고자 하였다. 본 논문은 기존 알고

리즘의 시공간적 특징을 추출하는 네트워크 구조에 추가로 사람 자세 추정에 대한 특징을 함께 고려한 알고리즘을 제안한다.



(그림 3) 제안 방법론 구조

(그림 3)은 제안 방법론의 구조이다. 기존 이상탐지는 영상 데이터를 짧은 영상단위로 전처리한 후 각 짧은 영상을 instance 로써 특징을 추출하여 해당하는 영상에 대한 이상탐지를 진행한다. 본 논문에서는 사람 행동 추정에 대한 특징을 이용하기 위해 짧은 영상단위의 전처리된 데이터에 대해 이미지 단위로 추가 전처리 후 짧은 영상에 대한 사람 자세 추정의 특징을 추가한 모델 구조를 제안한다. 이때 영상에 등장하는 사람에 대한 17 가지의 키포인트의 x, y 좌표와 추정 값의 신뢰점수가 특징 값이 된다. 영상에 여러 사람이 등장할 경우 각각의 키포인트마다 신뢰도가 높은 사람의 값으로 할당해 보다 정확한 키포인트를 바탕으로 특징을 구성한다.

이와 같은 모델 구조를 활용하여 영상 단위로 정의된 레이블을 그대로 유지할 수 있다. 또한 시공간에 대한 특징만 추출하는 기존 모델과 다르게 사람들의 자세를 추정하는 특징이 추가되어 교량 데이터의 특징상 이상에 대한 정의를 더 강건하게 해준다.

## 4. 실험

### 4-1. 데이터

학습 데이터로는 UCF-crime 을 사용하였다. 해당 데이터는 총 13 가지의 이상 상황과 정상 상황에 대한 CCTV 데이터들을 포함한다. 본 논문에서는 289 개의 정상 데이터와, 395 개의 이상 데이터를 사용하였다. 검증 데이터로는 교량에서의 정상, 이상 영상을 사용하였고, 교량에서의 제안 방법론의 검증을 진행하였다.

### 4-2. 실험 결과

(그림 4)는 정상 상황에 대한 모델 결과이다. 해당 상황은 사람 한 명이 교량을 건너는 상황이며, 위의 지표를 보면 이상 상황이 아닌 정상 상황임을 잘 인지하는 것을 확인할 수 있다[4]. (그림 5)는 교량 위 이상 상황에 대한 모델 결과이며, 아이와 함께 교량을 건너는 영상이다[5]. 하지만 해당 영상은 후에 아이를 안고 투신 시도를 하는 비정상 상황에 대한 영상이다. 인도가 잘 보이지 않는 교각에서 아이와 같이 걸어가고 있는 모습 같지만, 모델에서 두 사람이 교각에 가까워질 때 이상 점수가 높아지는 결과를 보여줬다. 이에 대한 이상 점수를 나타내는 그래프는 3d conv feature embedding 과 제안 방법인 추정된 사람의 키포인트로 산출된다. 이때 키포인트는 높은 신뢰도의 키포인트 위치 값을 사용한다. 해당 프레임이 정상상황이라고 3d conv feature 를 통해 산출되어도 키포인트 값을 통해 정상상황 중 이상과 가까운 경우의 점수를 부여하여 교각과 가까워지는 상황에 대한 이상점수를 추정할 수 있다.



(그림 4) 정상 상황에 대한 모델 검증



(그림 5) 이상 상황에 대한 모델 검증

## 5. 결론

CCTV 영상에 대한 이상탐지는 아직까지 많은 어려움이 존재한다. 본 논문은 모든 상황에 대한 이상탐지를 위한 특정 장소에 효과적인 이상탐지 모델을

제안한다. 교량에 대한 특징을 이용해 사람 자세 추정 특징을 추가하여 새로운 방법론은 제안하였다. 실험을 통해 교량 장소에 대한 사람 자세 특징 추출이 이상탐지에 좋은 영향을 줄 수 있다는 결과를 도출하였다. 이처럼 이상의 정의가 어려운 CCTV 영상에 대해 특정 장소에 특화된 모델들을 독립적으로 구현함으로써 새로운 이상탐지 모델의 구조를 통해 전반적인 성능 향상에 기여가 될 것이라 생각한다. 추가 연구로는 많은 사람들의 키포인트 의미를 담아 이상점수를 보완하여 보다 특화된 독립적인 이상탐지 모델 개발을 진행할 예정이다.

### 참고문헌

- [1] Nawaratne, R., Alahakoon, D., De Silva, D., & Yu, X. Spatiotemporal anomaly detection using deep learning for real-time video surveillance. *IEEE Transactions on Industrial Informatics*, 16, 1, 393-402. 2019.
- [2] Sultani, W., Chen, C., & Shah, M. Real-world anomaly detection in surveillance videos. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 6479-6488.
- [3]<https://github.com/facebookresearch/detectron2>  
(Available: October 4, 2023)
- [4] TBS 시민의방송. (2015, Feb 16). [tbstv] 한강다리 'CCTV 감시 시스템' 확대[Video]. YouTube. <https://www.youtube.com/watch?v=qFYoOx8qTl0>.
- [5] 연합뉴스 Yonhapnews. (2021 Oct 25). 다리 위 아찔한 투신 시도...버스에서 나타난 '슈퍼맨' / 연합뉴스 (Yonhapnews)[Video]. YouTube. <https://www.youtube.com/watch?v=ApXltj-sBpU>