

210724(목) - 데이터 수집 관련 과제

AI Hub

이상행동 CCTV 영상

- 폭행 행동 관련 영상 913개 (78시간 이상 분량)

Feature

- 폭행 발생 시점, 다양한 각도의 영상 제공, 계절에 따른 인상 착의 영상 제공
- 이상행동(폭행) 시작점, 종료시점, 연령, 성별, 객체의 위치 자료 포함

Dataset from Papers

Real-world Anomaly Detection in Surveillance Videos

- UCF-Crime dataset
 - 아동학대, 폭행 관련 데이터 (총 200개 이상의 영상)

Vision-based Fight Detection from Surveillance Cameras

- Surveillance Camera Fight Dataset
 - 150개의 Fight 영상 & 150개의 non-Fight 영상 (각각 2초 정도)

RWF2000 - A Large Scale Video Database for Violence Detection

- Raw Surveillance Videos from YouTube
 - 5초(30fps) 분량이며, 총 2,000개 이상의 영상과 300,000 frame이상의 violent 자료 보유

Youtube

- 아동 학대 관련 영상
- 키워드: ['虐待儿童'(중국어로 아동학대), 'CCTV 아동학대','child abuse','児童虐待'('일본어로 아동학대'),'아동폭행']

수집방법

- 위 키워드에 해당하는 유튜브 영상 제목과 링크 크롤링
- pytube 라이브러리를 사용하여 해당 영상 다운로드
 - 문제점 1. 아동학대 cctv 영상과 관련이 적은 영상링크도 크롤링 된다 (ex. 아동학대 문제 인터뷰 등)
 - ⇒ 제목 전처리/ 필터링 하여 관련성이 높은 영상 추출
 - ⇒ 최종으로 사람이 검수
- google에서 아동학대 gif 다운로드

전처리 하여 폭행이 일어나는 시점의 아동학대 자료 / non-child abuse 자료로 분리