

이상행동 예견 기법 개발에 관한 연구

손진환, 정희철*

경북대학교 인공지능학과

ii9696ii@knu.ac.kr, *heechul@knu.ac.kr

A Study on the Development of Abnormal Behavior Prediction

Son Jin Hwan, Heechul Jung*

Department of Artificial Intelligence,

Kyungpook National University

요 약

본 논문은 단일 장면 비디오에서 이상행동을 예견하는 딥러닝 알고리즘을 제안한다. 본 연구에서는 국내 마인즈랩에서 구축한 AI허브의 개방 데이터셋[2]인 이상행동 CCTV영상 데이터셋을 활용하여, 정상과 이상행동 장면을 분류 할 수 있는 모델을 학습한다. 본 연구는 기존 단일 프레임 폭력 인식 연구의[1] 후속으로 데이터 레이블 구성 방식에 변화를 주고 이상행동이 발생하기 전 프레임에서 이상행동을 예견하는 딥러닝 알고리즘을 제안한다. 이상행동을 예견하는 기술을 통해 미래에 일어날 사고를 예방하는 비디오 감시, 자율 주행 등과 같은 분야의 발전에 기여하기를 고대한다.

I. 서 론

최근 CCTV와 같은 단일 장면 비디오에서 방법을 위해 이상행동을 감지하는 연구들이 수행되고 있다. 일반적으로 이러한 기술은 이상 감지 검출 알고리즘을 통해 이루어진다. 이상행동은 예상되는 행동을 따르지 않는 비정상적인 사건을 감지하는 것을 의미한다. 이상행동은 실제 상황에서 제한되지 않기 때문에 매우 어렵고 광범위하다[3].

본 연구에서는 국내 마인즈랩에서 구축한 AI허브의 개방 데이터셋 CCTV 이상행동 데이터셋[2]을 활용하여 단일 장면 비디오에서 딥러닝 기반 이상행동 예견을 주제로 연구를 진행하였다. CCTV 이상행동 데이터셋은 영상 데이터셋이며, 각 프레임에 대해 정상/이상행동으로 구분되어 있다. 이를 이미지 데이터로 재구성하여 다양한 Convolutional Neural Network(CNN) 중 대표적인 아키텍처인 EfficientNet[4]을 학습시켰다. 딥러닝 모델은 PyTorch 기반으로 개발되었으며, 이상행동 데이터셋의 레이블을 활용하여 지도 학습 기반으로 진행되었다. 개발된 기술을 활용하여 자율 주행 기술 및 CCTV와 같은 단일 장면 비디오에서 방법, 사고 대응 및 모니터링 인력 절감에 기여할 수 있을 것으로 기대한다.

II. 본론

본 연구에서는 이상행동을 예견하기 위해, 국내 마인즈랩에서 구축한 AI허브의 개방 데이터셋인 CCTV 이상행동영상 데이터셋[2]을 활용하였다. 이상행동 CCTV영상 데이터셋은 12가지의 이상행동으로 정의한 총 700시간(8400클립)의 비디오 데이터셋이다. 각 클립의 길이는 5분 정도이며 UHD해상도를 가진다. 이상행동 종류, 이상 frame, 위치등에 대해 자세한 labeling이 되어 있어 있다. 이상행동 CCTV영상 데이터 세트는 매우 대규모 데이터셋이고, 촬영된 영상의 장소들이 다양하다. 이에 야외 폭력 영상 데이터셋 중 같은 장소에서 촬영된 12개의 동영상상을 사용하였다. 총 12개의 동영상 중 8개를 train, 2개를 validation, 2개를 test 데이터 셋으로 활용하였다. 이상행동 레이블의 프레임 정보를 활용하고 관심 영역을 정하여 프레임들을 분할하여 이미지 데이터셋을 새로 구축하였다.[1]



그림1. 원본 동영상



그림2. 관심영역 지정

본 연구에서는 데이터셋을 아래 그림과 같은 방식으로 분할하여 재구성하여, 이진 분류 실험을 진행하였다.

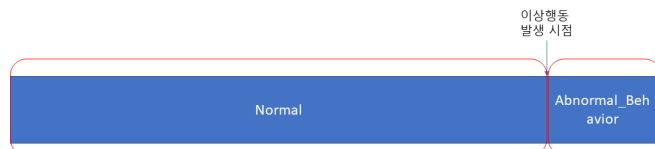


그림2. 실제 영상 레이블

i) 프레임들을 분할 한 방식 : 실제로 normal인 이상행동 이전 5초 동안의 프레임들을 이상행동으로 거짓 레이블을 부여하여 데이터셋 구축

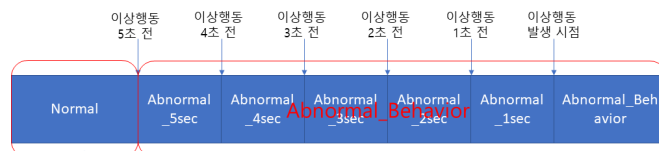


그림3. 데이터 레이블 구성 방식 (Dataset #1)

이상행동이 발생하기 전 즉, 가까운 실제로는 정상인 프레임들을 이상행동으로 거짓 레이블을 부여함으로써 이상행동이 발생하기 전 프레임에서 이상행동을 예견하기 위함이다. 추가적으로, 이상행동 발생 전 프레임들을 5초전부터 1초 간격 구간별로 나누어 각 구간에서 정상과 이상행동을 분류하여 이상행동 발생 전 예견하는 기법을 제안한다.

총 12개의 동영상 중 8개를 train, 2개를 validation, 2개를 test 데이터

셋으로 활용하였고, 이상행동 발생하기 5초전까지를 normal, 이상행동 발생하기 5초 전부터 normal과 같은 수의 프레임만큼 abnormal로 분할하여 재 구축된 이미지의 개수는 총 train 21778 장, validation 3904 장, test 이미지 3582 장이다. 이렇게 구축된 데이터셋을 활용하여 본 연구에서는 EfficientNet b0, b5 모델을 학습을 시켰으며, 손실 함수로는 이진 분류를 위해 cross entropy loss를 사용하였고, 입력 이미지 사이즈는 224x224, 배치사이즈는 32, 학습률은 0.05, 총 30에폭으로 설정하였다.

III. 본론

위에서 언급한 바와 같이 EfficientNet b0, b5 모델을 사용하여 정상/폭력을 이진 분류하는 태스크를 학습하였으며, 이에 대한 분류 정확도는 다음 표와 같다.

Model	Dataset #1	
	Test	Val
EfficientNet b0	83.8	74.4
EfficientNet b5	90.3	82.8

표1. 정상/이상행동 이진 분류 인식률

이상행동 발생 이전 프레임 5초부터 1초 간격으로 시간대별로 나누어 이상행동 예견 실험을 진행하였다. 5_sec는 이상행동 발생 5초전~4초전 동안의 프레임들이고 1_sec는 이상행동 발생 1초전~이상행동 발생까지의 프레임들이다. 동영상은 30FPS이므로 1초당 30장, 두 개의 test동영상과 valid동영상이므로 각 실험은 입력 60장의 이미지로 진행되었다. 이상행동 발생 시간 전 구간에서 이상행동으로 분류할 확률을 예견율이라고 표현하였다. T는 test셋 V는 validation셋이다.

Model	5_sec	4_sec	3_sec	2_sec	1_sec
	T	T	T	T	T
EfficientNet b0	56.7	50	50	63.3	50
EfficientNet b5	95	61.7	85	100	58.3

표2. 이상행동 시간대별 예견률

추가적으로 영상 데이터셋을 이상행동 발생 이전 5초동안을 이상행동으로 그 전 5초동안의 프레임을 정상으로 분할하여 소규모 데이터셋을 구축 및 20에폭으로 학습하고 위와 같은 시간대별 예견률을 테스트 하였다.

Model	5_sec	4_sec	3_sec	2_sec	1_sec
	T	T	T	T	T
EfficientNet b0	3.3	15.0	0	20	10
EfficientNet b5	0	26.7	38.3	1.7	0

표3. 이상행동 시간대별 예견률

시간이 지나면서 이상행동에 가까워 질수록 예견율이 높아짐을 기대하였으나, 테스트셋의 양이 적은 점과, 시계열 모델을 사용하지 않고 프레임을 분류한 점 등의 이유로 생각했던 원하는 결과를 얻지는 못하였다. 추후 object detector, 모델변경, processing 및 data augmentation등 다른 기법들을 더 적용하여 추가적인 연구를 진행할 계획이다. b0모델에 비해 b5모델이 더욱 깊은 모델이며 일반적으로 b0모델보다 깊은 b5모델이 더 효과

가 있는 것으로 보인다.

IV. 결론

본 논문에서는 단일 장면의 비디오에서의 이상 감지에 대한 개념을 설명하고, 사용 가능한 국내 데이터 셋을 활용하여 단일 프레임 기반의 폭력 감지를 위한 데이터셋으로 재구축하였다. 이를 EfficientNet b0와 b5 모델을 학습시켰으며, 5초부터 1초 간격으로 시간대별로 나누어 예견율을 측정하였다. 이를 통해 이상행동에 근접할수록 그 특징을 띄고 있어 예견율이 올라가는 것을 측정하기 위함이었으나, 본 연구에서는 원하는 결과를 얻지는 못했다. 추후 연구에서 실험을 확장 성능을 높이고, precision과 recall을 함께 측정하여, 신뢰할 수 있는 이상행동 예견 시간을 측정하는 연구를 계획하고 있다. 이상행동을 사전에 감지하여 미래에 일어날 사고를 예방하는 비디오 감시, 자율 주행 등과 같은 분야의 발전 접목하기를 기대한다.

ACKNOWLEDGMENT

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학CT연구 센터지원사업의 연구결과로 수행되었음" (IITP-2022-2020-0-01808*)
 "이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임" (No. 2020R1C1C100742311).

참 고 문 헌

- [1] 손진환, and 정희철. "딥러닝 기반 단일 프레임 폭력 인식 기법 개발에 관한 연구." 한국통신학회 학술대회논문집 (2022): 1431-1432.
- [2] 이상행동 CCTV영상 데이터셋 마인즈랩 2019
- [3] Bharathkumar Ramachandra, Michael J. Jones, Ranga Raju Vatsavai, "A Survey of Single-Scene Video Anomaly Detection" in IEEE Transactions on Pattern Analysis and Machine Intelligence, 25 November 2020
- [4] Tan, Mingxing, and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." International conference on machine learning. PMLR, 2019.