Mask R-CNN 기반 Aspect Ratio를 활용한 이상행동 검출 및 영역화 방법

임현석[O], 후쉬펑[*], 곽정환(교신저자)[*]
[O]한국교통대학교 소프트웨어학과,
[*]한국교통대학교 소프트웨어학과
e-mail: {hyunseoki[O], huxufeng[*], jgwak[*]}@ut.ac.kr

Abnormal Behavior Detection and Localization Using Aspect Ratio Based on Mask R-CNN

Hyunseok Lim[O], Xufeng Hu[*], Jeonghwan Gwak(Corresponding Author)[*]
[O]Dept. of Software, Korea National University of Transportation,
[*]Dept. of Software, Korea National University of Transportation

● 요 약 ●

이상 행동을 탐지하는 딥러닝 기반 검지 시스템은 동영상 기반 데이터로부터 움직임을 보이는 객체를 추적하고 그 객체의 행동을 분석하여 정상적인 행동 범위를 벗어나는 패턴을 보이는 영역을 이상으로 탐지한다. 특히 생성적 적대 신경망(GAN)과 광학 흐름 추정(Optical flow estimation) 기법을 활용하여 움직임에 대한 특징 정보를 추출하고 이를 학습하여 행동 패턴에 대한 모델링을 수행한다. 모델 학습 및 테스트에 활용되는 데이터셋의 해상도가 낮거나 이상 행동을 표현하는 특징 정보가 부족할 경우 최종 모델 성능에 부정적 영향을 미치게 되며, 특히 광학 흐름이 표현하는 이동량 측면에서 차이가 크게 나지 않는 이상 객체의 경우 탐지가 정확하게 이뤄지지 않는다. 본 연구에서는 동영상 프레임에서 나타나는 객체의 평균 종횡비를 구하고 정상적인 비율을 벗어나는 객체에 대해서 이상 행동을 취하는 샘플로 처리하는 후처리단 모듈을 제안하여 최종적인 모델 성능을 향상시키는 방법을 고안한다.

키워드: 딥러닝(deep learning), 이상 행동(Anomaly Behavior), Semantic segmentation

I. Introduction

동영상 데이터에서 나타나는 이상 행동이란 흔히 볼 수 있는 평범한 동작 외에 정상적이지 않은 행동 패턴을 취하는 현상을 말한다. 동영상에서 이러한 행동 변화는 프레임 내 화소 값의 변화로 표현될 수 있으며, 광학 흐름 추정(Optical flow estimation) 기법은 이러한 화소의 변화를 파악할 수 있다.

정상 상황에 비해 이상 상황이 발현될 만한 행동 패턴의 가짓수는 무궁무진하기 때문에 이를 개별적으로 구분하여 분류기 학습을 하기보다 정상적인 행동 패턴만을 학습하여 그 이외의 행동 패턴을 보이는 영역을 이상 행동 영역으로 판단하는 방식을 주로 채택한다. 생성적 적대 신경망 모델[1]은 정상 행동 패턴 만을 학습하기 위한 생성 모델로서 활용될 수 있으며, 정상적인 행동 패턴에 대한 학습을 수행하고 비정상적인 행동 패턴에 대해서는 오차가 크게 발생하게 되므로 비지도 학습 기반 차이를 측정하여 이상 객체를 감지한다[2-5]. 하지만 학습에 활용되는 데이터셋의 낮은 해상도와 폐색 현상으로 인한 정보 손실은 탐지 성능을 떨어뜨리는 원인이 된다. 그러므로 Object detector를 활용하여 객체를 탐지하고 정상적인 객체에서 주로 보이는 Bounding box 영역의 가로 세로 비율에 대한 분포를 분석하여 이를 많이 벗어난 객체에 대해 이상 객체로서의 가중치를 추가한다. 객체의 Aspect ratio를 기반한 이상 객체 탐지 모듈을 모델 시스템의 후처리단에 추가하여 성능 향상의 결과를 살펴본다.

II. The Proposed Scheme

동영상 프레임에서 나타나는 객체의 지역화를 위한 Object detector로서 COCO 데이터셋으로 학습된 Mask R-CNN[6] 모델을 활용한다. Fig. 1은 이상 행동 객체가 가지는 Aspect ratio를 표현하기 위한 그림이다. 휠체어를 타고 이동하는 객체는 해당 데이터셋에서 이상 객체로 판단되기 때문에 정상 보행하는 객체의 특징과 구별되는

정보를 추출해야 한다. 일반적인 보행을 하는 사람 객체는 Mask R-CNN 모델로 예측한 지역화 Bounding box 영역의 가로 세로 비율이 1:2 비율인 반면, 휠체어를 타거나 자전거를 타는 사람에 대한 가로 세로 비율은 1:1 비율에 더 가깝게 계산된다. Fig. 2는 각 객체에 대한 비율을 표현하는 그림이다.

Fig. 1. Frame with difficult object detection

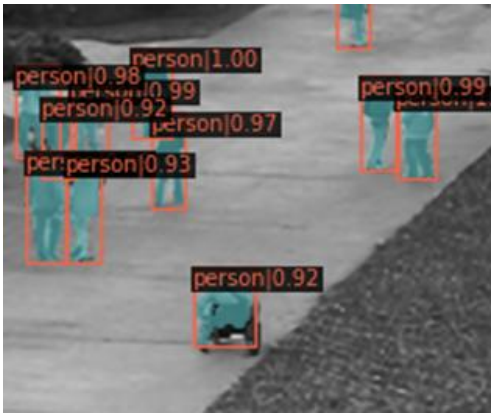

Fig. 2. Detection and Segmentation results by the Mask R-CNN

Fig. 3은 Mask R-CNN 모델을 기반으로 해당 데이터셋 프레임의 Segmentation 결과 중 관심 객체의 Aspect ratio가 정상 객체와 크게 다른 객체만 남긴 결과이다. 이러한 결과는 최종 Anomality map에 가중치로서 추가된다.

Fig. 3. Area with unusual aspect ratio

III. Experiment results

동영상 프레임 중 이러한 Aspect ratio 모듈이 가장 잘 적용될 수 있는 시퀀스를 선정하고 Aspect ratio 모듈의 적용 여부에 따른 최종 Frame-level AUC, EER 값을 확인한다. 본 연구에서 진행된 실험은 UCSD Ped1 데이터셋을 기반으로 수행되었으며 Aspect ratio 모듈의 적용 유무에 따른 최종 AUC, EER 값을 측정하였다.

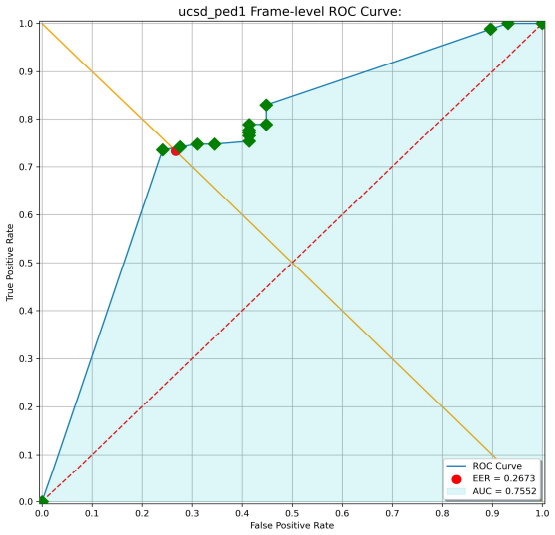

Fig. 4. Experimental results without the aspect ratio module

Fig. 4는 Aspect ratio 모듈 없이 테스트된 실험의 ROC 곡선이며 75.52%의 AUC를 보이고 있다.

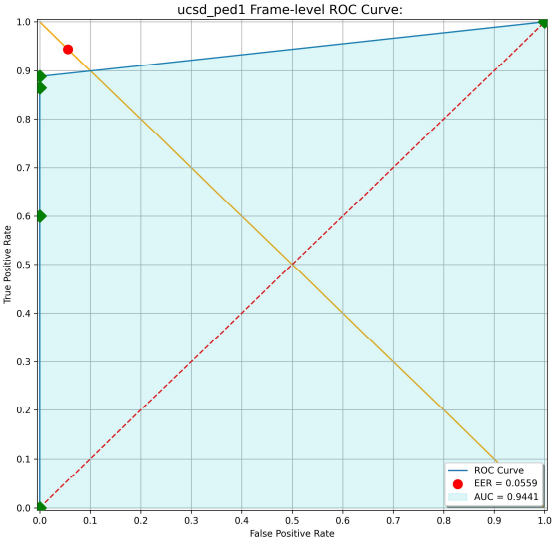

Fig. 5. Experimental results with the aspect ratio module

Fig. 5는 Aspect ratio 모듈이 적용된 상태에서 테스트된 실험의 ROC 곡선이며 94.41%의 AUC를 보이면서 Aspect ratio 모듈이 추가된 후처리 단계가 모델의 최종적인 성능 향상에 도움이 되고 있음을 나타내고 있다.

Fig. 6은 Aspect ratio 적용 유무에 따라 실제 영상 프레임에서 객체가 탐지되는 결과를 보여주고 있으며, 왼쪽의 빨간색으로 표현된 부분은 이상 행동 객체에 대한 Ground-truth 영역이며, 오른쪽의 보라색 부분은 Ground-truth의 빨간색 영역과 Aspect ratio 모듈을 통해 최종 구현된 Anomality map이 표현하는 파란색 영역이 합쳐진 부분으로 보라색으로 표현되고 있으며, 이상 객체를 적절히 탐지하고 있음을 나타내고 있다.

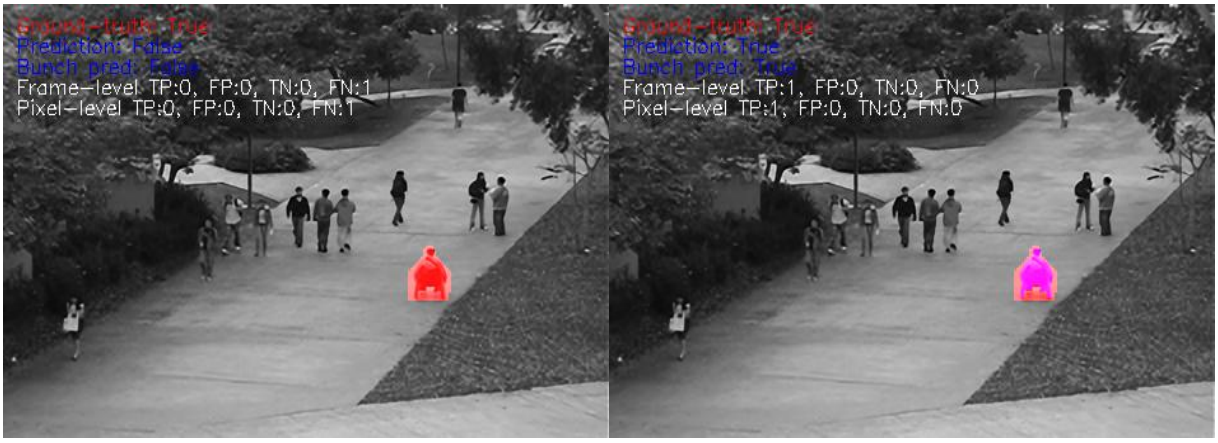

Fig. 6. Detection result of not applying aspect ratio module (left), Detection result of applying aspect ratio module (right)

IV. Conclusions

본 연구에서는 생성 모델의 패턴 학습에 기반하여 정상 행동의 공간적 움직임 특징들을 학습하고, 이상 행동이 나타나는 공간에서 광학 흐름 연산과의 차이를 활용해 이상 객체에 대한 지역화를 수행하되 광학 흐름 패턴의 차이 만으로는 탐지하기 힘든 이상 객체를 강인하게 탐지하기 위해서 후처리단에 Aspect ratio 모듈을 추가하여 최종 이상 객체 탐지 성능을 향상시켰다.

ACKNOWLEDGEMENT

This was supported by Korea National University of Transportation in 2021 and "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(MOE) (2021RIS-001).

REFERENCES

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," Proc. Adv. Neural Inf. Process. Syst., pp. 26722680, 2014

[2] M. Ravanbakhsh, M. Nabi, E. Sangineto, L. Marcenaro, C. Regazzoni and N. Sebe, "Abnormal event detection in videos using generative adversarial nets," 2017 IEEE International Conference on Image Processing (ICIP), pp. 1577-1581, 2017, DOI: 10.1109/ICIP.2017.8296547

[3] M. Ravanbakhsh, E. Sangineto, M. Nabi and N. Sebe, "Training Adversarial Discriminators for Cross-Channel Abnormal Event Detection in Crowds," 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1896-1904, 2019, DOI: 10.1109/WACV.2019.00206

[4] W. Liu, W. Luo, D. Lian and S. Gao, "Future Frame Prediction for Anomaly Detection - A New Baseline," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6536-6545, 2018, DOI: 10.1109/CVPR.2018.00684

[5] H. Lim, J. Gwak, “Generative optical flow based abnormal object detection method using a spatio-temporal translation network ,” Journal of the Korea Society of Computer and Information, vol. 26, no. 4, pp. 11-19, Apr. 2021.

[6] K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask R-CNN," 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2980-2988, doi: 10.1109/ICCV.2017.322.