

심층 신경망을 이용한 실시간 횡단보도 보행자 검출 방법 분석

*방준호 *박민기 *송채영 *최해철†

*한밭대학교

junhobang01@gmail.com

mmkkpp22@naver.com

thdcodud1845@naver.com

choihc@hanbat.ac.kr†

Performance analysis of YOLOv5 and Faster R-CNN for real-time crosswalk pedestrian detection

Junho Bang, Min-Ki Park, Chaeyong Song, and Haechul Choi

Hanbat National University

요 약

횡단보도에서의 보행자 교통사고 방지를 위한 다양한 방법들이 연구되고 있다. 본 논문에서는 점멸 신호등 상황에서 보행자 교통사고를 감소시키기 위해 영상을 이용한 심층 신경망 기반 횡단보도 보행자 검출 방법을 소개한다. YOLOv5 와 Faster R-CNN 각각을 기반으로 다양한 버전의 횡단보도 보행자 검출기를 구현하고, 이번 실험에서 중점이 되는 이들의 수행 시간을 비교 평가하고 mAP@0.5 가 어느 정도인지 판단하여 가장 적합한 모델을 판단한다. 실험 결과 실시간 처리 측면에서 YOLOs 모델이 84 fps 를 달성함으로써 실시간 보행자 검출에 가장 좋은 성능을 보였다. 횡단보도의 상황은 상시 빠르게 변하므로 가장 빠른 처리 성능을 기록한 YOLOv5s 모델이 실시간 횡단보도 보행자 검출 시스템에 가장 적합한 것으로 판단된다.

1. 서론

점멸 신호등은 보행자의 통행량이 적은 도로에서 차량의 통행을 원활하게 하기 위해 고안된 교통 체계이다. 하지만 이에 따른 부작용으로 점멸 신호등이 요인이 된 보행자 사고가 늘어나고 있다. [1]에 따르면, 분석 대상 교차로의 일반 신호 운영 시 사고 건수는 74 건에서 점멸 운영 전환 시 121 건으로 1.7 배 증가하였으며 점멸신호 운영 시 발생한 중상자 수 51 명/100 건은 일반 신호 운영 시 43.5 명/100 건 대비 1.2 배 증가하여 경상자 증가율 1.1 배보다 높아 중상자 수의 수가 더욱 증가하여 위험한 부작용을 낳고 있다[1].

이에 따라, 본 논문에서는 횡단보도의 안전성과 효율성을 증대 시키기 위한 심층 신경망을 이용한 횡단보도 보행자 검출을 제안한다. 또한, 제안 방법을 실시간성 측면에서 분석하기 위해 객체 검출에 사용되는 대표적인 모델인 YOLOv5 의 각 모델들과,

Faster R-CNN 을 비교하였다.

본 논문의 구성은 다음과 같다. 2 절에서는 대표적인 Detecting 모델인 YOLOv5 를 바탕으로 실험을 진행을 하는 제안 방법을 설명을 하며, 3 절에서는 제안 방법에 대한 실험 결과 및 대표 모델과의 비교 분석 내용을 설명한다. 4 절에서는 실험 결과를 바탕으로 어떤 모델이 횡단보도 보행자 검출 분석에 가장 적합한지 결론짓는다.

2. 제안 방법

본 논문에서는 횡단보도에서 빠르게 걷는 보행자를 감지하기 위해, 사전 학습된 몇 가지 모델을 기반으로 사람 class 만을 포함하는 데이터셋을 추가적으로 학습시켜 사람에 대한 검출 정확도를 향상시켰다. YOLOv5 는 one-stage 모델로서 빠르고

비교적 정확하게 객체를 검출하지만 사물들의 폐색(occlusion)이 존재하는 경우 잘 구별하지 못하는 단점이 있는 것으로 알려져 있다[2]. 이와 반대로 Faster R-CNN 은 Fast R-CNN 구조와 RPN(resion proposal network) 구조를 통합함으로써 정확하고 폐색이 존재하거나 사물이 작아도 잘 검출할 수 있으나, 실시간성을 고려하고 만든 모델이 아니기에 검출 속도가 상당히 느리다. 제안 방법은 실시간성을 보장하기 위해 사전 학습된 YOLOv5 모델을 기반으로 사람 class 만을 포함하는 데이터 셋으로 학습되었으며, 무겁지만 정확한 검출 성능을 보이는 Faster R-CNN 과 실시간 처리 성능 비교 실험을 진행하였다. YOLOv5 는 상기 언급한 단점으로 정확도가 떨어지므로 0.5 이상의 mAP@0.5 가 나오는지 판별하기 위해 정확도를 측정한다. YOLOv5 학습에 사용된 데이터 셋은 coco-2017 데이터 셋으로부터 사람 class 만 포함된 8,000 장의 이미지를 무작위로 선택하여 생성하여 학습하였다.

3. 실험 조건

실험에는 대표적인 심층 신경망 기반 객체 검출 모델인 YOLOv5 는 각기 다른 버전인 (s, l, m, x)을 실시간성과 정확도를 측정하고 Faster R-CNN 에 대해 각각 실시간성을 측정하였다. YOLOv5 모델의 각 버전은 s, l, m, x 순서로 검출 성능이 좋아지지만 실시간성이 나빠지는 trade-off 가 발생한다. 이때, 실시간성을 측정한 평가 척도로 FPS 를 사용하였으며 정확도는 mAP@0.5 를 기준으로 평가되었다. YOLOv5 모델들과 Faster R-CNN 의 사전 학습 모델별 실시간성을 평가하기 위한 평가 비디오의 해상도는 FHD(1920x1080) 이다. 또한 YOLOv5 의 mAP@0.5 를 평가하기 위한 데이터 셋으로는 2,511 장의 VOC dataset 을 사용하였다. 실험 환경은 아래와 같다.

- OS: Ubuntu Linux 20.02 LTS
- GPU: NVIDIA RTX 3060 (12GB)

4. 실험 결과

[표 1]은 YOLOv5 4 개의 모델에 대한 실시간 성능 및 객체 검출 정확도 평가 결과이다. Size 는 weights 파일의 크기를 나타낸 것으로, [표 1]에서 YOLOv5 는 S 에서 X 모델로 갈수록 mAP@0.5 가 증가하나 처리 가능한 프레임의 수가 현저히 떨어지는 것을 확인할 수 있다. 또한, [표 2]는 Faster R-CNN 에 대한 실시간 성능 평가 결과로, 사전 학습된 Faster R-CNN 은 복잡한 two-stage 네트워크 구조로 인하여 실시간성이 보장되지 않았다[2]. 따라서 one-stage YOLOv5 모델들에 비해 상당히 낮은 실시간 처리 성능을 나타냈다[3].

표 1 YOLOv5 모델들의 weights-size, 성능 비교

Model	model	Size(MB)	mAP@0.5	FPS
YOLOv5	S	14.5	0.755	84.0
YOLOv5	M	42.2	0.846	68.2
YOLOv5	L	92.9	0.861	44.8
YOLOv5	X	691.1	0.875	29.0

표 2 Faster R-CNN 의 weights-size, 성능 비교

Model	Network	Size(MB)	FPS
Faster R-CNN	R50-FPN	167.3	8.9
Faster R-CNN	X101-FPN	420.8	8.8



[그림 2] 보행자 검출 사진

5. 결론

점멸 신호등은 통행량이 적은 도로에서 통행을 원활하기 위한 효율성 있는 제안이다. 하지만 점멸 신호등으로 인해 일반 신호 대비 사고 건수가 1.7 배가 증가하는 것에 대해 대책이 필요하다. 따라서 심층 신경망을 이용한 보행자 검출을 제안하게 되었다. 이를 위하여 사람을 인식하여 횡단보도 보행자가 대기하고 있는 것을 판단하거나, 무단 횡단에 대한 사고 예방을 위하여 빠른 시간에 [그림 2]와 같이 사람이 어디에 위치해 있는지를 판단하여 알림을 주는 것이 필요하다. 이를 위해 실시간성이 잘 보장되는 모델을 이용하는 것이 매우 중요하다. 본 논문에서는 이를 위한 사전 실험의 결과로, 실시간 보행자 탐지에 가장 빠르게 처리할 수 있는 YOLOv5s 사용하는 것이 가장 적합한 모델임을 확인하였다.

참고문헌

[1] 삼성교통안전문화연구소, “야간 점멸신호 교차로 사고실태 및 운영기준 개선방안”, 2017

[2] Petru Soviany, Radu Tudor Ionescu, “Optimizing the Trade-off between Single-Stage and Two-Stage Deep Object Detectors using Image Difficulty Prediction”, Department of computer Science, University of Bucharest, Romania, 2018

[3] Fardad Dadboud, Vaibhav Patel, Varun Mehta, Miodrag Bolic, “Single-Stage UAV Detection and Classification with YOLOV5: Mosaic Data Augmentation and PANet”, 2021 17th IEEE International Conference on Advanced Video and Based Surveillance(AVSS), 2021

[4] Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”, Microsoft Research, 2015