

딥러닝 통한 CCTV에서의 객체 위치 탐지 기술 동향

정보컴퓨터공학부 202055582 이지은

November 19, 2023

Abstract

카메라 센서인 CCTV에서 동적 객체의 위치를 파악하는 기술은 자율 주행 자동차에 시각지대의 정보를 제공해 안전성 개선에 도움이 될 것으로 기대된다. 이 글에서는 여러 연구 실험 자료를 통해 CCTV에서의 객체 위치 탐지 기술의 발전 동향을 살펴보고, 사용되는 주요 딥러닝 모델을 비교해 본다. 위치 탐지 기술의 발전 동향을 조사해 본 결과, 주로 사용되는 딥러닝 모델이 CNN에서 YOLO로 변화하며 속도, 정확성 면에서 개선되고 있음을 확인하였다. 추후 실제 CCTV에 적용해 데이터를 얻고, 이를 차량의 데이터와 통합하기 위한 연구가 필요할 것으로 생각된다.

1 서론

CCTV를 이용한 객체의 위치 탐지 및 시각화를 위한 노력은 자율 주행 자동차의 상용화를 위해 지속해서 이어지고 있다. 카메라 센서인 CCTV를 통한 객체 위치 검출 기술의 발전 동향 및 사용되는 알고리즘에 대해 연구 자료를 통해 알아보고자 한다.

현재 주목 받는 자율 주행 기술이 완전 자동화 단계에 가까워지기 위해서는 C-ITS 기술의 발전이 필수적이다. C-ITS¹란 V2X(vehicle to everything) 통신을 이용해 차량간(V2V), 차량-인프라간(V2I)의 양방향 정보 교환이 가능하도록 연결해 주는 시스템을 말하는데, 이를 통해 도로 위의 위험 요소나 돌발 상황 등에 신속하게 대응할 수 있기에 자율 주행의 핵심 기술로 주목받고 있다.

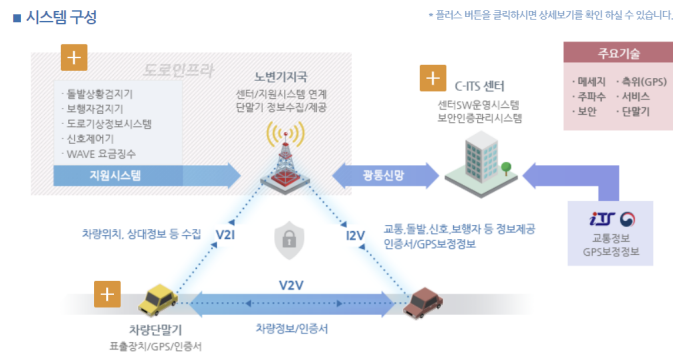


Figure 1: C-ITS 시스템 구성

¹C-ITS 시범사업 홍보관, C-ITS 소개 자료 <https://www.c-its.kr/introduction/component.do>

C-ITS의 도로 인프라 센서에는 카메라, 라이다(Lidar), 레이더(RADAR)가 포함되는데, 이들은 자동차에 탑재되어 신뢰성 있는 자율 주행을 지원한다. 주변에서 흔히 볼 수 있는 CCTV는 카메라 센서에 해당한다. 그런데 CCTV의 경우, 이미 잘 구축되어 있음에도 불구하고 단순히 모니터링에만 이용되는 등 활용도가 매우 떨어지는 문제가 있다. CCTV가 물체를 인식하고 이들의 위치에 대한 정보를 얻어낼 수 있다면 C-ITS 서비스에 다양한 형태로 활용할 수 있을 것으로 생각된다.

즉, 현재 구축된 CCTV에 이 기술을 실제로 적용한다면, 거의 모든 도로 위의 상황을 효율적으로 파악할 수 있을 것이다. 또, 자율 주행 자동차에 이 정보를 통합시킬 수 있다면 사각지대의 정보도 제공해 줄 수 있어 자율 주행 자동차의 안전성을 훨씬 높일 수 있을 것으로 기대된다.

2 본론

여러 가지 논문을 이용해 기술 발전 동향에 대한 조사를 진행하였으며, 범위를 2018년부터 현재까지로 지정하여 어떤 연구를 통해 기술 발전이 이루어지고 있는지 알아보았다. 추가로, 사용되는 주요 딥러닝 모델간의 비교를 통해 어떤 발전이 있는지 살펴보고자 한다.

2.1 CNN을 이용한 객체 탐지 방법

2018년에는 딥러닝을 이용해서 CCTV 영상 내의 차량 및 보행자의 수를 정확히 검출하기 위한 기술 연구가 진행된 바 있다.[1] 기존의 영상처리 알고리즘은 조도 변화와 화질 열화에 약한 성능을 보였기에, 딥러닝 알고리즘 중 CNN(Convolutional Neural Network, 합성곱 신경망)을 이용해 CCTV 영상 기반 차량 및 보행자 계수 정확성을 향상하고자 하는 움직임이 있었다.

Table 1: CNN 알고리즘 적용 시 검출 성능 결과

대상	차량			보행자		
	오전	오후	밤	오전	오후	밤
총량	168	157	170	130	128	36
검출 수	168	157	170	104	100	33
검출 비율 (%)	100	100	100	80.2	78.4	91.6

위 표1에서 확인할 수 있듯이, 차량의 경우는 조도 변화나 카메라의 설치 각도에 상관없이 모두 검출되어 안정적인 성능을 보였다. 반면, 보행자의 경우는 여러 환경 요인으로 인해 아직 검출이 부족한 모습을 보였음을 알 수 있다. 하지만 조도 변화에 큰 영향을 받지 않을 수 있도록 객체 검출 기술을 발전시켰다는 데서 의의가 있다.

2.2 YOLO를 이용한 객체 탐지 방법

2021년에는 영상 분석을 위한 인공지능 기술인 YOLO(You Only Look Once) 알고리즘을 이용한 동적 객체 위치 추적에 대한 연구가 진행되었다. YOLO(Redmon,2016)는 객체 인식과 분류를 차례로 수행하는 CNN과 달리 인식과 분류를 한 번에 수행할 수 있는 모델로 실시간에 가까운 처리가 가능하다.

”CCTV 영상을 활용한 동적 객체의 위치 추적 및 시각화 방안” 논문[2]에서 YOLO 알고리즘을 이용해 학습하고, 최소 사각형 형태 및 변환 행렬 기술을 통해 웹 기반 시각화까지 성공한 모습을

보여주었다. 이 연구에서는 Bochkovskiy et al.(2020)의 YOLOv4 모델 및 Kafka 서버, Python, JSON 포맷을 이용하며, 공공 데이터 포털의 개방된 CCTV 영상 데이터를 활용해 실험이 진행되었다.

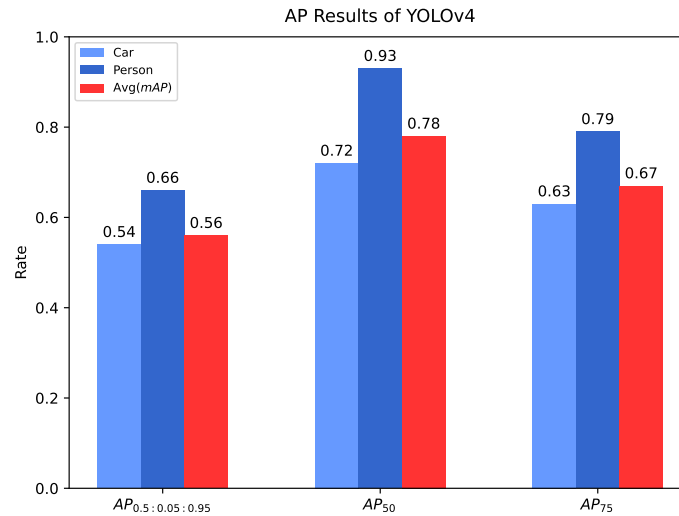


Figure 2: YOLOv4의 AP 결과

표준 정밀도인 AP의 평균값이 0.60 - 0.80 사이로 높게 나타난 모습을 확인할 수 있다. 그래프2에서도 볼 수 있듯이 평균값 중 mAP_{50} 값이 가장 높게 나타났기에, 이후 YOLO의 성능 변화 여부를 판단하기 위해 IoU의 값을 0.5로 설정한 mAP_{50} 의 값을 평균 정밀도의 값으로 고려한다.

Table 2: YOLO 알고리즘 적용 및 좌표계 변환 후 탐지된 객체 위치 비교 결과

n	P'	P^*	$Err(m)$
1	37.39810236995146, 127.11284371695545	37.39810236995146, 127.11284458460097	0.27
2	37.398056022983745, 127.11284453412071	37.398056022983745, 127.1128445986055	0.06
...			
25	37.39790503296632, 127.11296219917062	37.397905021241262, 127.11296124944073	0.44
26	37.3979793496272, 127.11291813053069	37.397977934962716, 127.11291835477248	0.02
Avg			0.15

위 표2에서 P' 는 변환된 공간 좌표계의 점을 의미하고, P^* 은 수직 교차점, Err 는 두 점 사이의 거리를 나타낸다. 오차의 평균(Average)이 각각 0.15m의 수치를 나타내었다는 것으로 보아 객체의 위치가 거의 정확히 탐지되었음을 알 수 있다. 카메라와 멀리 떨어진 영역에서의 심한 왜곡은 보완 사항이지만, 위치 동기화의 정확도를 높이고 이를 웹 시각화까지 시도했다는 점에서 의의가 있다.

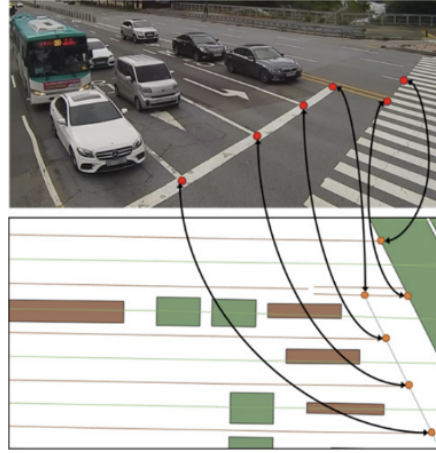


Figure 3: 좌표계 변환 - 두 좌표계에 대한 대응점 정의

2.3 YOLO와 TensorRT를 결합한 객체 탐지 방법

2022년에는 YOLO와 더불어 TensorRT를 결합하여 객체 위치 추적에 대한 연구가 진행되었다. TensorRT는 모델 최적화 엔진으로 양자화, 그래프 최적화 등을 통해 연산을 최적화함으로써 딥러닝 모델의 추론 속도를 높이는 데 도움을 준다.

”C-ITS를 위한 CCTV 영상의 실시간 동적 객체 탐지 가속화” 논문[3]을 보면, YOLO와 TensorRT를 함께 사용하여 동적 객체 탐지의 추론에 드는 시간을 눈에 띄게 단축했음을 알 수 있다. 이 연구에서는 YOLOv5s 모델, FP32 모델(YOLOv5s + TRT32), FP16 모델(YOLOv5s + TRT16), PyTorch를 이용하였으며, 공공 데이터 포털의 공개 데이터 6,000건을 8:1:1의 비율로 무작위로 나누어 각각을 학습, 검증, 테스트 데이터로 사용하였다.

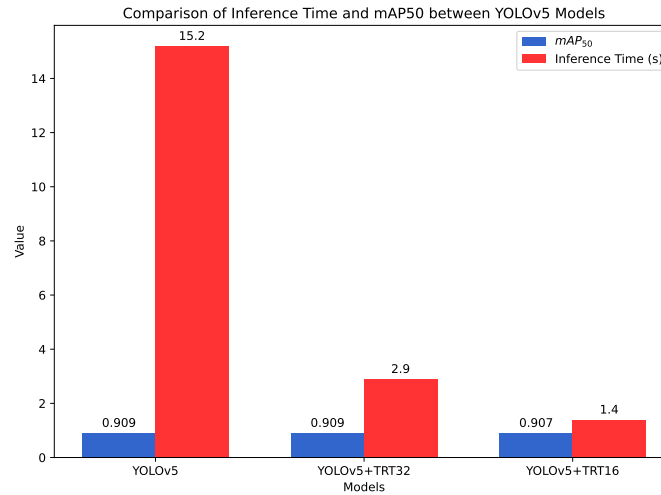


Figure 4: 객체 검출 모델의 평균 정밀도 및 추론 시간 비교

위 그래프4의 결과를 참고해 계산하면, TensorRT 모델의 평균 mAP_{50} 값은 0.908, 평균 추론 시간은 2.2초로 나타났음을 알 수 있다. 즉, 성능 면에서는 YOLO 모델과 YOLO + TensorRT 모델의 차이가 크지 않지만, 추론 시간의 차이는 크게 나타났음을 알 수 있다. YOLO 모델만 이용해도 90% 이상의 정확성을 보일 수 있지만 모델 최적화 엔진을 함께 사용함으로써 동적 객체 탐지 시간을 단축할 수 있음을 확인하였다는 점에서 의의가 있다.

2.4 CNN과 YOLO의 비교

CNN(Convolutional Neural Network, 합성곱 신경망)은 필터를 통하여 이미지에 대한 정보를 추출하고, 그 정보를 바탕으로 이미지를 인식하는 딥러닝 알고리즘 중 하나이다. 반면 YOLO(You Only Look Once)는 이미지 내에 있는 객체의 위치를 한 번만 보고도 파악할 수 있는 알고리즘이다.[4] 따라서 인식과 분류를 한 번에 실행하는 YOLO가 CNN에 비해 처리 속도가 빠르다. 또한 최소한의 배경 오류를 이용하기에 정확도 또한 YOLO가 높다고 할 수 있다. 이는 다음 표3를 보고 확인할 수 있다.

객체 검출 기술 발전을 확인하기 위해 발전한 형태인 Faster R-CNN과 YOLOv5를 비교하였다. 비교 내용은 다음과 같다.

Table 3: Faster R-CNN과 YOLOv5의 비교

	Faster R-CNN	YOLOv5
Precision(P)(%)	75.7	83.37
Training Loss	0.166	0.017
$mAP_{0.5-0.95}$ (%)	41.5	61.29
Inference speed(s)	0.070	0.013
Training time/epoch(s)	114	18
Model Size(MB)	244.58	51.1

위 표3에서 값은 각 모델의 평균 수치를 이용해 나타내었다. Faster R-CNN의 내용은 ResNet50, VGG16, MVGG16, Mobile-Net V2, Inception V3 모델의 평균 값이며, YOLOv5의 내용은 Ys, Ym, Yl의 평균 값이다.[5]

표를 보면 YOLOv5가 Faster R-CNN에 비해 7.67% 높은 mAP 값을 나타냈으며, 추론 시간 또한 YOLOv5가 Faster R-CNN에 비해 5.4배 적게 소요된다는 것을 알 수 있다. 이외에도 학습 손실이나 모델 크기 면에서도 YOLOv5가 낮은 값을 나타내며 더 좋은 성능을 보였다.

이를 통해 2018년의 CNN을 이용한 객체 탐지 방법보다 2022년의 YOLOv5를 이용한 객체 탐지 방법에서 성능 개선이 있음을 확인할 수 있다. 즉, 비교 결과를 통해 CCTV에서의 객체 위치 검출 기술이 발전하고 있음을 유추해 볼 수 있다.

3 결론

CCTV에서의 객체 위치 검출 기술의 발전 동향에 대해 조사해 본 결과, CNN과 YOLO가 핵심 기술로 사용되고 있음을 알 수 있었다. 조사 결과를 다시 한 번 표4로 정리하면 다음과 같다.

Table 4: CCTV에서의 객체 위치 탐지 기술의 발전 동향

발표 시점	2018.06	2021.06	2022.08
딥러닝 모델	CNN	YOLOv4	YOLOv5s (+TensorRT)
개발 프레임워크	Darknet	Darknet	Pytorch
학습 데이터(개)	17,792	5,250	4,800
실험 데이터(개)	798	153	600
평가 내용	시간대별 검출률(Recall)	정밀도, 위치 정합도	TRT 모델 정밀도, 추론 시간
성능 평가	평균 : 0.98	mAP_{50} :0.79, 오차:0.15m	mAP_{50} :0.908, 추론 시간:2.2s

CNN을 이용한 경우, 육안으로 분석한 내용을 기준으로 객체 검출 여부만을 측정하였기에 검출률 결과를 재현율로만 나타낼 수 있었다. 따라서 YOLO와의 성능 비교가 어려웠기에 CNN과 YOLO를 비교해보는 작업도 수행하였다. 이를 통해 YOLO가 정밀도, 시간 측면에서 모두 발전되었음을 확인할 수 있었다. 다음은 CNN과 YOLO의 비교 내용이다. 이 그래프5에서는 데이터의 범위가 크기 때문에 그 차이를 뚜렷하게 표시하고자 로그 스케일을 사용해 값을 나타내었다.

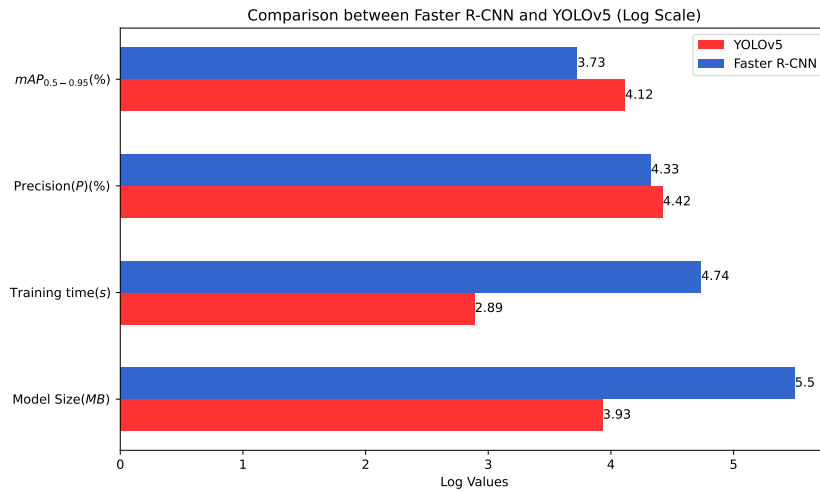


Figure 5: Faster R-CNN과 YOLOv5의 비교

최신 기술에서는 실시간 탐지에 더 뛰어난 YOLO를 이용함으로써 성능 개선이 이루어진 것으로 보인다. CCTV에서의 객체 위치 탐지 기술은 실시간 처리가 핵심 요소이므로 앞으로도 처리 속도가 빠르며 정확도도 높은 YOLO의 개발이 이어질 것으로 생각된다. 또한, 추후 해당 데이터를 자율 주행 데이터와 통합한다면 자율 주행의 안전성에 도움을 주는 효율적인 방안이 될 것으로 기대된다.

References

- [1] 이태희, 김기주, 윤경수, 김광주, and 최두현, “CCTV 영상 기반 딥러닝을 이용한 차량 및 보행자 계수 방법,” *한국지능시스템학회 논문지*, vol. 28, no. 3, pp. 219–224, 2018.

- [2] 박상진, 조국, 임준혁, and 김민찬, “CCTV 영상을 활용한 동적 객체의 위치 추적 및 시각화 방안,” *지적과 국토정보*, vol. 51, no. 1, pp. 53–65, 2021.
- [3] 한영석 and 정재윤, “C-ITS를 위한 CCTV 영상의 실시간 동적 객체 탐지 가속화,” *한국전자거래학회지*, vol. 27, no. 3, pp. 87–94, 2022.
- [4] 이용환 and 김영섭, “객체 검출을 위한 CNN과 YOLO 성능 비교 실험,” *반도체디스플레이기술학회지*, vol. 19, no. 1, pp. 85–92, 2020.
- [5] K. Ahmed, “Smart Pothole Detection Using Deep Learning Based on Dilated Convolution,” *MDPI*, 2021.