

주차 관련 이동체 객체 검출 문제 해결을 위한 딥러닝 모델

우혁, 신창섭
올뉴스시스템즈

hwu@allnewsystems.com, shincs@allnewsystems.com

A Deep Learning Model for Solution of Parking Related Moving Object Detection Problem

Hyuk Wu, Chang Sub Shin
Allnewsystems

요 약

본 논문은 자율 주차 시스템에서 위험 이동체 감지를 위한 딥러닝 모델을 제안한다. YOLOR 모델을 기반으로 Learning rate 를 변경하고, Grid mask 방식을 추가하는 등의 기법을 통해 모델을 설계 및 학습했다. 해당 모델을 이용하여 세단, 어린이, 유모차 등의 14 개 클래스를 대상으로 하는 2022 인공지능 온라인 경진대회에서 0.54mAP 의 성능으로 1 위를 달성했다.

Abstract

This paper proposes a deep learning model for detecting dangerous moving objects in an autonomous parking system. Based on the YOLOR model, the model was designed and trained by changing the learning rate and adding a grid mask method. We achieved first place with a performance of 0.54mAP in the 2022 artificial intelligence online contest targeting 14 classes such as sedans, children, and strollers.

I. 서 론

자율 주행 시스템은 미래의 모빌리티 서비스 혁신하기 위한 잠재력을 가지는 기술로 인식되고 있고, 영상 처리 및 객체 인식(Object Detection) 기술에 대해서 많은 연구가 이루어지고 있다.

객체 인식은 자율 주행을 위한 핵심 기술 중에 하나이며, 자율 주행을 위한 객체 인식은 주행 환경에서 안전을 담보로 수행한다는 점에서 다양한 환경에서 강건하면서도 실시간 성능을 만족해야 한다. 또한 자원 활용에도 많은 제약이 따른다. 그래서 아직까지도 많은 연구가 필요한 분야이다.

이러한 자율 주행을 위한 객체 인식 기술은 위험 이동체 감지 경보 시스템이나 자율 주차 시스템 등에 활용할 수 있다. 그 중에서도 자율 주차 시스템에서는 근접한 객체들에 대해 주행 환경 인식이 필요하다. 주차 중인 차량 근처에 어린이나 유모차 등의 주의가 필요한 객체가 근접거리에서 갑자기 나타났을 때 이를 신속히 인지하고 안전을 위한 대처가 필요 하기 때문이다. 이를 위하여 주차 중인 차량 근처의 이동 객체를 정확히 탐지하고 인지하는 인공지능 기술 개발이 필요하다.

본 논문에서는 이러한 문제를 해결하기 위해 YOLOR[1] 모델을 기반으로 Learning rate 를 변경하고 Grid mask 방식을 추가하는 등의 기법을 통해 모델을 설계 및 학습했다. 해당 모델을 통해 정보통신산업진흥원에서 주관한 2022 인공지능 온라인 경진대회에서 주차 시에 차량 카메라의 이미지에서 이동체 객체 검출 문제에서 주어진 환경 내에서 0.54mAP 의 성능으로 1 위를 달성했다.

II. 본론

i) YOLOR

딥러닝 베이스의 오브젝트 디텍션 모델은 크게 원 스테이지와 두 스테이지 모델로 나눌 수 있다. 본 논문에서는 수행 시간의 제약을 고려하여 원 스테이지 계열의 모델인 YOLOR 을 선택했다. YOLOR 은 YOLO[2-4]를 베이스로 하는 모델로서 COCO[5] 데이터셋을 대상으로 실시간 객체 인식 분야에서 SOTA 를 달성했다. YOLOR 은 명시적, 암시적 지식을 학습함으로써 모델이 객체 인식뿐만 아니라 다양한 분야의 인지, 인식 기능을 수행할 때에도 보다 일반적이고 강건한 지식을 배울 수 있도록 했다. 명시적, 암시적 지식을 Kernel space alignment 와 Prediction refinement 를 통해서 통합하여 모델의 정확도와 일반성을 제고했다. 또한, YOLOR-D6 은 베이스 모델인 Scaled-YOLOv4[6]의 비슷한 정확도를 가진 모델인 Y4-P7 에 비하여 약 2 배정도의 빠른 추론 속도를 달성했다.

ii) 모델 설계

실시간성을 만족시키기 위해 YOLOR 의 모델 중에서 가장 작은 모델인 YOLOR-CSP 모델을 사용했다. 가로, 세로 640 픽셀 크기의 입력 해상도를 가진다. 추론 속도는 배치 1 을 기준으로 106FPS 이다.

본 논문에서는 한 개의 T4 GPU 에서 36 시간 안에 학습을 모두 완료하기 위해서 Learning rate 를 변경했다.

기존의 모델은 긴 Epoch 를 사용하기에 모델이 수렴하기에 긴 시간이 소요된다. 이는 안정적인 학습을 위해 초기에 Learning rate 를 조금씩 올리는 Warm-up[7]과 Learning rate 값을 몇 번씩 초기값으로 돌려 학습하는 Cosine annealing with Restarts schedule[8]을 사용하기 때문이다. 해당 문제를 해결하기 위해 기존의 Warm-up 과 Restarts schedule 을 바꾸어 Exponential decay 방법으로 학습을 진행했다. Exponential decay 는 Learning rate 를 지수적으로 감소시키는 학습 방법으로 초기에 큰 값으로 줄어들이지만, 0 이 되지는 않게 한다. Learning rate 를 초기에 빨리 감소시키는 방법으로 짧은 학습 시간을 극복하여 최적의 Hyperparameter 를 찾음으로써 정확도를 제고했다.

Grid mask[9]는 입력 이미지의 일부를 랜덤하게 지워서 정보 손실을 유발하는 데이터 증강 기법이다. 다른 데이터 증강 기법들과는 달리 작은 크기의 일정한 패턴을 사용하여 간격을 띄워 정보 손실을 유발한다. 이러한 방법은 연속적으로 과도한 정보 손실을 피하여 정보의 균형을 유지하여 모델을 보다 강건하게 만든다. Grid mask 기법을 이용하여 일부 가려진 객체들에 대한 정확도를 제고했다.

iii) 모델 학습

베이스 라인 모델은 COCO 데이터셋으로 Pre-training 한 모델을 사용했다. 입력 해상도는 최소 가로, 세로 640*640 픽셀에서 최대 Full high definition (FHD) 해상도인 1920*1080 까지 사용하여 학습했다. 또한 학습간, Mosaic augmentation 과 Multi-scale training[10]을 사용했다. 이러한 학습법을 통해 한정적인 GPU 메모리를 다소 극복했다. 24,650 장의 트레이닝 셋에 대하여 36 시간 동안 39 Epoch 동안 학습을 진행했다. 35 Epoch 이후는 Backbone 과 Neck 은 학습하지 않고 Head 만 학습했다. 19,520 장의 테스트 셋에 대하여 0.54mAP 를 달성했다. 학습 결과는 그림 1 과 같다.



[그림 1] 주차 관련 이동체 인식 결과

III. 결론 및 향후 연구

본 논문은 자율 주차 시스템에서 위험 이동체 객체 검출 문제 해결을 위한 딥러닝 모델을 제안한다. YOLOR 모델을 베이스로 학습 시간 단축을 위하여 Exponential decay learning rate 를 사용하여 학습 시간을 단축하고, Grid mask 기법을 통해 일부분이 가려진 객체들에 대하여 정확도를 제고하였다. 세단, 어린이, 유모차 등의 14 개 클래스를 대상으로 약 2 만 장의 테스트 셋에 대해서 0.54mAP 를 달성했다. 향후 더 다양한 클래스에 대해 탐지할 수 있는 기법과 비나 눈이 오는 상황에

대해서도 같은 성능을 도출할 수 있는지에 대해서 연구할 예정이다.

ACKNOWLEDGMENT

이 논문은 2022 년도 과학기술정보통신부의 재원으로 정보통신산업진흥원의 지원을 받아 수행된 연구임. (A1504-22-1003)

참 고 문 헌

- [1] Wang, C. Y. "You only learn one representation: Unified network for multiple tasks," arXiv preprint arXiv:2105.04206, 2021.
- [2] Redmon, Joseph. "You only look once: Unified, real-time object detection," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779-788, 2016.
- [3] Redmon, Joseph. "YOLO9000: better, faster, stronger," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7263-7271, 2017.
- [4] Redmon, Joseph. "Yolov3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.
- [5] Lin, T. Y. "Microsoft coco: Common objects in context," European conference on computer vision, Springer.Cham, pp. 740-755, 2014.
- [6] Wang, C. Y. "Scaled-yolov4: Scaling cross stage partial network," Proceedings of the IEEE/cvf conference on computer vision and pattern recognition, pp. 13029-13038, 2021.
- [7] He, T. "Bag of tricks for image classification with convolutional neural networks," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 558-567, 2019.
- [8] Loshchilov, I. "Sgdr: Stochastic gradient descent with warm restarts," arXiv preprint arXiv:1608.03983, 2016.
- [9] Chen, P. "Gridmask data augmentation," arXiv preprint arXiv:2001.04086, 2020.
- [10] Singh, B. "Sniper: Efficient multi-scale training," Advances in neural information processing systems, 2018.