

논문 2022-17-33

국내 도로 환경에 특화된 자율주행을 위한 멀티카메라 데이터 셋 구축 및 유효성 검증

(Construction and Effectiveness Evaluation of Multi Camera Dataset Specialized for Autonomous Driving in Domestic Road Environment)

이 진 희, 이 재 근, 박 재 형, 김 제 석, 권 순*

(Jin-Hee Lee, Jae-Keun Lee, Jaehyeong Park, Je-Seok Kim, Soon Kwon)

Abstract : Along with the advancement of deep learning technology, securing high-quality dataset for verification of developed technology is emerging as an important issue, and developing robust deep learning models to the domestic road environment is focused by many research groups. Especially, unlike expressways and automobile-only roads, in the complex city driving environment, various dynamic objects such as motorbikes, electric kickboards, large buses/truck, freight cars, pedestrians, and traffic lights are mixed in city road. In this paper, we built our dataset through multi camera-based processing (collection, refinement, and annotation) including the various objects in the city road and estimated quality and validity of our dataset by using YOLO-based model in object detection. Then, quantitative evaluation of our dataset is performed by comparing with the public dataset and qualitative evaluation of it is performed by comparing with experiment results using open platform. We generated our 2D dataset based on annotation rules of KITTI/COCO dataset, and compared the performance with the public dataset using the evaluation rules of KITTI/COCO dataset. As a result of comparison with public dataset, our dataset shows about 3 to 53% higher performance and thus the effectiveness of our dataset was validated.

Keywords : 2D Dataset, Camera, Autonomous driving

1. 서 론

인공지능 및 딥러닝 기술 고도화가 빠르게 진행되면서, 대량의 인공지능 데이터를 가공하는데 관심이 집중되고 있다. 이는 컴퓨터 비전, 자연어 처리, 자율주행, 음성인식 등에 이용하는 딥러닝 기술들은 데이터 양에 비례하여 개발 기술의 성능이 향상되기 때문이다. 그러므로 학습 가능한 형태로 어노테이션 및 레이블링이 포함된 대량의 학습 데이터를 수집하고, 고품질의 데이터를 가공하는데 관심을 두고 있다. 결국 딥러닝을 이용한 기술들은 대규모 데이터를 이용하여 주어진 작업을 수행하기 위한 다양한 패턴을 기계가 학습하기 때문에, 만약 자체 데이터 셋을 구축한다면 그 데이터의 유효성을 검증하는 작업은 반드시 필요하다.

2022년 4월 20일부터 도로 교통법이 개정됨에 따라 자율주행 시스템을 갖춘 차량의 도로 통행이 가능해졌다. 이에 따라 국내 복잡한 도심 내 자율주행 실현을 위해 다양한 연

구기관과 스타트업에서 자율주행 기술연구가 빠르게 가속화되고 있다. 결국 도심에서 안전한 자율주행이 실현되기 위해서는 실시간으로 변화하는 도로 및 교통 상황을 정확하게 인지하고 판단하는 딥러닝 기반의 기술이 필요하다. 이러한 기술 개발과 더불어 다양한 환경 및 시나리오에서 취득한 대량의 데이터 셋 구축도 중요하게 되었다. 2020년부터 인공지능 데이터 가공 사업을 우리나라에서도 대폭 지원하여 활성화하고 있으며, 현재 2022년에도 다양한 인공지능 기술 개발을 위한 데이터 가공 사업에 전속력을 가하고 있다. 즉, 국내 도로에서의 운행 상황과 환경 조건에서 다양한 데이터를 취득하고, 도로에서 인지할 수 있는 다양한 객체 (차, 트럭, 버스, 특수차량, 보행자, 자전거, 오토바이 등)를 학습 가능한 형태인 어노테이션 및 레이블링 형태로 변환하여 모델을 학습하는 것이 무엇보다 중요하다. 이는 기존에 공개된 데이터 셋인 KITTI, Waymo는 국내 도로 환경에서 나타날 수 있는 객체 (큰 트럭, 버스, 오토바이 등) 데이터가 부족하거나 데이터가 없기 때문에 국내 도로 환경에 맞는 데이터 셋 구축은 반드시 필요한 상황이다.

이 논문에서는 다양한 국내 도로 환경에서 멀티카메라를 이용하여 이미지 데이터를 수집/정제/가공을 수행함으로써 자체 데이터 셋을 구축하고, 이 구축한 데이터 셋의 유효성을 정성적/정량적으로 검증하는 방법을 소개하고자 한다. 즉, 자

*Corresponding Author (soonyk@dgist.ac.kr)

Received: Jul. 12, 2022, Revised: Aug. 22, 2022, Accepted: Aug. 23, 2022.

J.H. Lee: DGIST (Senior Researcher)

J.K. Lee: FutureDrive1 (Researcher)

J.H. Park: DGIST (Researcher)

J.S. Kim: DGIST (Senior Researcher)

S. Kwon: DGIST (Principal Researcher)

※ 본 연구는 과학기술정보통신부에서 지원하는 대구경북과학기술원 일반사업 (22-CoE-IT-01)과 기관고유사업 (22-IT-02) 및 지역현안 해결형 R&BD 사업 (2020-DD-RD-0339) 지원을 받아 수행 되었습니다.

체 데이터 셋은 국내 도심 환경에서 멀티 카메라 기반으로 다양한 객체를 수집/정제/가공한 데이터이며, 구축한 데이터 셋의 품질 및 유효성 검증은 YOLO V5 모델을 활용하여 자체DB, COCO, KITTI, Waymo와 같은 공개 데이터 셋을 통해 학습된 모델을 정량적으로 평가하였다. 또한 자율주행을 위한 시나리오 기반의 테스트 데이터 셋을 구축하였는데, 이는 연속된 시나리오에 기반하며 Cut-In/Out, 교차로 혼잡상황, 오토바이/자전거 위험상황, 야간 도심 주행상황 등이 포함되어 있으며, 오픈 플랫폼에서 정성적 평가를 통하여 구축한 데이터 셋의 성능을 비교 분석하였다. 2장 관련연구에서는 기존 공개 데이터 셋의 특징 소개와 2D 객체 검출 방법들을 소개하고, 3장에서는 자체 구축한 데이터 셋의 특징과 유효성 검증 모델의 구조를 설명하고, 4장에서는 자체 데이터 셋의 성능 검증을 위한 공개 데이터 셋과 자체 데이터 셋의 정량적/정성적 평가 결과를 보이고, 5장에서 결론을 맺는다.

II. 관련 연구

1. 공개 데이터 셋

COCO 데이터셋은 2014 마이크로소프트에서 객체 탐지 (Object Detection), 키포인트 탐지 (Keypoint Detection), 세그먼테이션 (Segmentation) 등 컴퓨터 비전 (Computer vision) 분야의 필요한 테스트 (task)를 목적으로 만들어진 데이터 셋이다 [1]. 실제로 객체 검출 관련 논문에서 성능 평가 목적으로 많이 사용되는 데이터 셋이며, 객체 검출 라이브러리는 COCO 데이터 셋으로 사전 학습된 모델을 제공하고 있다. COCO 2017 데이터 셋을 가장 많이 사용하고 있는데 이는 해당 118,000장의 학습 데이터 셋, 5,000장의 검증 데이터 셋, 41,000장의 테스트 데이터 셋으로 구성되었다.

KITTI 데이터 셋은 자율주행 시나리오에 필요한 Object Detection, Stereo, Optical Flow, Visual Odometry/SLAM 등의 다양한 범주 데이터를 제공한다 [2]. 그림 1과 같이 1개의 라이다, 4개의 카메라, 1개의 GPS 센서에서 수집한 멀티 센서 데이터를 제공하며 이 중에서 2D 객체 검출 데이터 셋은 7481개의 학습용 데이터와 7518개의 테스트용 데이터로 구성되어 있으며, 8개의 클래스 정보와 학습을 위한 픽셀 단위의 바운딩 박스의 크기와 위치정보가 포함되어 있다.

Waymo 데이터 셋은 1150개의 주행 시나리오가 존재하며, 각 시나리오는 다양한 조건 (낮, 밤, 새벽)와 장소 (도심 또는 교외)에서 취득하였다 [3]. 그림 1과 같이 6개 (middle-range 1개, short-range 5개)의 라이다와 5개 (전/후/측면)의 카메라로 취득한 360도 뷰의 동기화된 융합 센서로부터 데이터를 수집하며, 120만개의 3D 레이블링과 99만개의 2D 레이블링된 데이터 셋을 제공한다.

NuScense 데이터 셋은 2019년에 공개된 대규모 자율주행 데이터 셋이다 [4]. 보스턴과 싱가포르에서 운행되어, 그

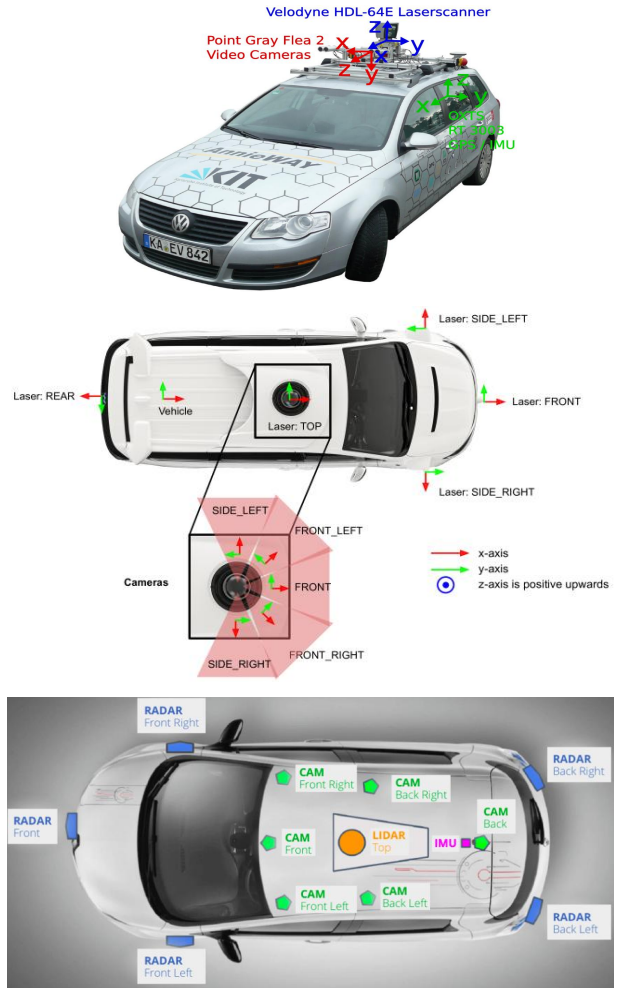


그림 1. Dataset을 위한 sensor setup: KITTI, Waymo, nuScenes Fig. 1. Sensor setup for dataset: KITTI, Waymo, and nuScenes

림 1과 같이, 1개의 라이다, 5개의 레이더, 6개의 카메라로부터 취득한 융합센서 데이터와 그에 대한 3D 큐보이드 정보를 제공한다. BDD 100K는 UC Berkely 인공지능 연구실험실 (BAIR)에서 공개한 자율주행을 위한 비디오 데이터 셋이다. 초당 30프레임의 고화질로 취득된 100K의 비디오 시퀀스로 구성되어 있다. 영상에는 비오는 날씨, 흐린 날씨, 맑은 날씨와 같은 다양한 기상조건이 기록되어 있으며 자동차, 도로표지판, 보행자에 대한 바운딩 박스 정보를 제공한다. Tusimple은 차선인식을 위한 데이터 셋이다. 미국 내 다양한 도로환경에서 데이터를 취득하였으며, 19장의 연속된 차선 이미지 (총 14K)와 json 형식의 레이블링 데이터 (0.6K)를 제공한다. 이 논문에서는 자체 구축 데이터 셋과 유효성 검증을 위해 각각의 공개 데이터 셋 중에서 2D 객체 검출 데이터 셋을 활용하여 학습하고 비교 평가를 진행한다.

2. 2D 객체 검출 모델

객체 검출은 컴퓨터 비전 기술 분야 중에서 하나로써 주어진 이미지 내에서 관심 있는 객체를 검출하는 기술이다.

즉, 객체 검출 모델들은 객체가 있는 위치를 검출할 뿐만 아니라, 해당 객체를 분류하는 영역까지를 포함한다. 자율주행 인지 기술이 발전하고 다양한 객체를 포함한 주행환경 데이터에 관심을 가지면서 도로 상황에 나타나는 동적 객체를 검출하는 알고리즘 개발이 가속화 되고 있다. 딥러닝 기반의 객체 검출 모델은 one-stage 모델과 two-stage 모델로 구분할 수 있다. 특정 객체에 대해 어떤 물체인지 분류를 하는 classification과 객체가 있을만한 영역을 빠르게 찾는 region proposal 알고리즘이 있는데 one-stage detector는 classification과 region proposal을 동시에 수행하고, two-stage detector는 이 과정을 순차적으로 수행하여 결과를 얻는다. One-stage detector가 two-stage detector에 비해 객체 검출 정확도 성능은 낮지만, 처리 속도는 빠른 장점이 있다. 대표적인 one-stage detector로는 YOLO, SSD, RetinaNet 등이 있으며, two-stage detector는 R-CNN, Fast R-CNN, Faster R-CNN 등이 있다.

One-stage detector인 YOLO는 bounding box와 class probability를 하나의 문제로 간주하여 객체의 종류와 위치를 한 번에 예측한다 [5]. 이미지를 일정 크기의 그리드로 나눠 각 그리드에 대한 bounding box를 예측하고, bounding box의 confidence score와 grid cell의 class score의 값으로 학습한다. 간단한 처리 과정으로 속도가 매우 빠르지만 작은 객체에 대해서는 상대적으로 정확도가 낮다. SSD는 각 convolutional layer 이후에 나오는 feature map마다 bounding box의 class 점수와 offset을 구하고, NMS (Non-Maximum Suppression) 알고리즘을 통해 최종 bounding box를 결정한다 [6]. 이는 각 feature map마다 스케일이 다르기 때문에 작은 물체와 큰 물체를 모두 탐지할 수 있다는 장점이 있다. RetinaNet은 모델 학습 시 계산하는 손실 함수 (loss function)에 변화를 주어 one-stage detector들이 지닌 낮은 성능을 개선했다 [7]. One stage detector는 최대 십만 개까지 후보군 제시를 통해 학습을 진행하는데, 그 중 실제 객체인 것은 일반적으로 10개 이내이고, 다수의 후보군이 background 클래스로 분류된다. 상대적으로 분류하기 쉬운 background 후보군들에 대한 loss값을 줄여줌으로써 분류하기 어려운 실제 객체들의 loss 비중을 높이고, 그에 따라 실제 객체들에 대한 학습에 집중한다. RetinaNet은 속도 빠르면서 two-stage detector와 유사한 성능을 보인다.

Two-stage detector 중 R-CNN은 selective search를 이용해 이미지에 대한 후보영역 (Region Proposal)을 생성한다 [8]. 생성된 각 후보영역을 고정된 크기로 wrapping하여 CNN의 입력으로 사용하고, CNN에서 나온 feature map으로 SVM을 통해 분류, regressor를 통해 bounding box를 조정한다. 강제로 크기를 맞추기 위한 wrapping으로 이미지의 변형이나 손실이 일어나고 후보영역만큼 CNN을 처리하기 때문에 큰 저장 공간을 요구하고 느리다는 단점이 있다. Fast R-CNN은 각 후보영역에 CNN을 적용하는 R-CNN과

달리, 이미지 전체에 CNN을 적용하여 생성된 feature map에서 후보영역을 생성한다 [9]. 생성된 후보영역을 RoI pooling을 통해 고정 크기의 feature vector로 추출하고, feature vector에 fully connected layer를 거쳐 softmax를 통해 분류, regressor를 통해 bounding box를 조정한다. Faster R-CNN은 selective search 부분을 RPN (Region Proposal Network)로 대신 사용한다 [10]. RPN은 feature map에서 CNN 연산 시 sliding window가 anchor box로 후보영역을 예측한다. Anchor box는 미리 지정해놓은 여러 개의 비율과 크기의 bounding box를 의미한다. RPN에서 얻은 후보영역을 IoU순으로 정렬하여 NMS 알고리즘을 통해 최종 후보영역을 선택하고, 선택된 후보영역의 크기를 맞추기 위해 RoI pooling을 사용하고 이후부터는 Fast R-CNN과 동일하게 진행하는 모델이다. 이 논문에서는 one-stage detector인 YOLO 모델을 이용하여 공개 데이터 셋과의 객체 검출 성능 비교를 통해 구축 데이터 셋의 유효성을 검증하였다.

III. 구축 데이터 셋 유효성 검증

1. 자체 데이터 셋 구축

자체 구축한 데이터 셋은 국내 환경에 적합한 자율주행을 위한 멀티센서 기반 주행 데이터 셋을 제공한다. 자체 데이터 셋 구축을 위한 수집 차량은 6대의 카메라 (앞, 좌, 우, 뒤, 스테레오)를 이용하여 데이터를 취득하고, 수집된 데이터는 정제, 가공, 검수 과정 순서로 진행되어 완성된다.

다양한 환경의 데이터 취득을 위해, 취득 시나리오 및 수집 환경은 장소 (도심, 부도심, 골목, 산길, 해안도로, 고속도로, 자동차전용도로), 기상상황 (맑음, 비), 시간대 (낮, 밤)로 각각 세분화하여 다양한 환경에서 수집하며 총 300시간 이상 대량의 데이터를 취득한다. 이는 조합한 시나리오에 따라 여러 가지 상황에 대한 카메라 데이터를 수집하여 환경에 따른 데이터의 편향성을 방지하고, 자율주행을 위한 데이터 셋의 다양성을 확보하고, 특정 클래스 (특수차량, 보행자, 이륜차 등) 간의 데이터 균형이 깨지지 않도록 필요시 빈도가 높은 장소나 시간대를 조절하여 데이터를 취득하였다.

데이터 정제는 6개의 카메라에서 취득된 원시 데이터를 초당 1프레임으로 샘플링하고 각 카메라 센서 별로 객체 수와 분포를 고려하여 학습에 적합하지 않은 프레임은 제거함으로써 가공 대상 데이터를 선별한다. 카메라 센서 및 가공에 따른 정제 기준을 충족하는지 여부, 정제 완료 데이터 간의 비교를 통해 중복도 여부, 학습에 부적합한 오류 데이터 등의 유무 등을 확인하여 제거가 필요할 경우 동기화된 센서 데이터를 일괄 삭제한다. 예를 들어, 연속 5프레임 이상 객체/배경화면이 반복되거나 야간 데이터 중 빛샘점/흔들림 등 노이즈가 포함된 데이터는 정제 과정에서 필터링함으로써 구축 데이터의 품질을 향상시키기 위해 정제 과정에서



그림 2. 가공데이터 예시: 맑음/비(상), 낮/밤(하)
Fig. 2. Data annotation example: clear/rain (top), day/night (bottom)

노력하였다. 수집환경 표시를 위한 데이터 파일명 규칙도 생성하여 취득시간, 기상조건, 취득도로 유형 등의 수집 및 정보를 입력한다. 1차와 2차 정제를 수행하는데, 1차에서는 센서 데이터 동기화 및 캘리브레이션 정확성 판단 도구를 이용하여 데이터의 적합성 및 적정성 확인 후, 1Hz로 데이터를 변환하여 수집데이터를 정리하고, 2차에서는 미리 정의된 원천데이터 포맷으로 파일 및 폴더 구조를 변환하는 프로그램을 이용하여 데이터 변환을 거쳐 정제한다.

데이터 가공은 자동화 지원 어노테이션 (annotation) 도구를 활용하는데, 검출 및 가공이 미흡한 객체에 대해서는 작업자가 수동 검수 및 보정한다. 일반 차량뿐만 아니라 큰 버스와 트럭, 대형 화물차량 및 특수차량, 보행자, 이륜차 등의 객체를 포함한 어노테이션을 수행 하였으며, 이때, 클래스 분포를 균형 있게 맞추기 위해, 보행자와 이륜차 데이터를 중점적으로 수집/정제/가공하였다. 데이터를 학습하기 위한 어노테이션 파일은 json 형식이며, 이 파일에는 데이터 수집 환경과 이미지 내 존재하는 객체정보를 포함한다. 데이터 수집환경은 이미지 정보, 취득시간, 기상조건 (맑음, 폭우), 취득한 도로환경으로 구성되며, 객체정보는 클래스 (자동차, 트럭, 버스, 특수차량, 보행자, 오토바이, 자전거, dontcare)와 바운딩박스 (좌상단, 우상단, 폭, 너비), 가림정도 (0: fully visible, 1: occluded), 잘림정도 (0: non-truncated, 1: truncated)로 구성되어 있다. 각 클래스별 최소 가공 크기 기준을 정해서 라벨링을 하였으며, 전문성을 요하는 검수 기준은 라벨링 오류, 라벨링 미수행, 다른 객체 라벨링 등으로 검수 기관에서 수행하였다. 이렇게 구축한 자체 데이터 셋은 27만 장의 이미지와 어노테이션 정보로 구성되어 있다. 수집 및 정제, 가공 과정을 통해 나온 결과 예시는 그림 2와 같다.

자체 구축한 데이터 셋은 도심 자율주행을 위함으므로 자동차, 버스, 자전거, 보행자가 혼재하는 복잡한 상황에서

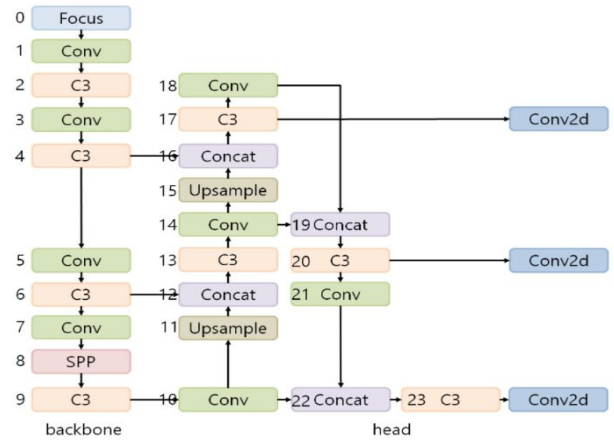


그림 3. 본 논문에서 사용한 YOLO 구조
Fig. 3. YOLO structure using in our method

도심의 주행규칙에 대한 학습을 통해 주변 환경을 예측 하는 것이 중요하다. 이를 위해 여러 상황에서 반복 실험과 안전을 위한 요소가 필수적이다. 이에 따라 이 논문에서는 추가적으로 시나리오 기반 연속 시퀀스 테스트 데이터 셋을 가공하였다. 해당 데이터 셋은 복잡한 도심, 정형/비정형 교차로, 고속도로, 자동차전용도로 등에서 카메라 (앞/좌/우/뒤)로부터 10Hz로 취득하고 이를 정제 및 가공한다. 각각의 시나리오는 170~600 프레임으로 구성되며, 예를 들어 도심 교차로 혼잡상황, Cut-In/Out, 전방 급감속 차량, 오토바이/자전거 위험상황, 분기도로/합류도로, 휴게소/주차장, 야간주행 등의 시나리오를 기반으로 연속 시퀀스 데이터 셋이다.

2. 유효성 검증 모델

2장에서 소개한 기존의 공개된 데이터 셋을 이용하여 자율주행을 위한 딥러닝 모델을 학습하고 모델을 검증하였지만, 한정된 공개 데이터 셋은 국내 도로 환경에서 검출되는 다양한 객체들을 학습하고 검증하기에는 다소 부족함이 있다. 따라서 이 논문에서는 국내 도로 환경에 특화된 6개의 멀티 카메라로 수집한 다양한 형태의 차량 및 이륜차 등의 자체 데이터 셋을 구축하고 이 데이터 셋의 유효성을 검증하기 위해 YOLO 모델을 사용하여 기존 공개 데이터 셋과의 성능을 비교 분석하였다. YOLO 모델은 자율주행 플랫폼에 적용하기 위한 검출 정확도와 실시간성 이슈에서 기본적으로 요구하는 성능을 보장하고, 추후 개선 가능성이 있으므로 이를 고려하여 이 모델을 선택하였다.

이 논문에서는 YOLOv5는 그림 3와 같은 구조의 네트워크를 이용하였다. 백본 (backbone) 네트워크는 CSPNet (Cross Stage Partial Network)을 사용하였는데, 이는 기존 DenseNet에 cross-stage block 구조를 추가하여 네트워크의 학습능력을 향상 하였다 [11, 12]. Dense block의 base layer에서 두 개의 파트로 나눠 절반의 채널 (channel) 정보만을 활용하여 컨볼루션을 수행하고, 마지막 cross-stage 계층에

표 1. 거리별 객체 height (각 클래스 pixel 기준)
Table 1. Object height by range (based on each class pixel)

Range (m)	Stereo (height)		Camera (height)	
	Person (pixel)	Car (pixel)	Person (pixel)	Car (pixel)
35	71	68	45	52
45	55	61	39	38
55	45	43	33	35
65	38	38	27	30
75	31	32	24	24
80	30	30	23	23

서 합침으로 base layer의 weight 정보는 보존하여 정확도 (accuracy)는 유지하고, 추론 (inference) 속도를 줄일 수 있다. 검출기 (detector)의 Neck은 SPP (Spatial Pyramid Pooling)와 PAN(Path Aggregation Network)을 사용하는데, 이는 FPN (Feature Pyramid Network)을 활용하여 낮은 층의 피처를 높은 층으로 전달하는 bottom-up path augmentation layer를 사용한다 [13-15]. 이를 통해 정보 흐름을 단축하고, 낮은 층의 피처 정보를 보존하여 객체의 정확한 위치정보를 추정할 수 있다. Fully connected layer의 제한된 input shape으로 인해, feature에 대한 resize 또는 cropping은 필연적으로 발생한다. 이로 인해 발생하는 기하학적 왜곡 및 원본 데이터의 손실을 방지하고자 피라미드 구조로 각각 풀링을 진행하고 생성된 벡터를 concatenation 하여 adaptive pooling을 수행하는 SPP를 이용하였다. Loss는 GIOU (Generalized IoU) loss를 이용하는데, 이는 IOU loss 함수의 한 종류이며, GT (Ground Truth)와 예측한 바운딩 박스가 overlap 되지 않아 FP (False Positive) 값이 생겨서 gradient vanishing 문제를 개선하는데 사용한다. GIOU loss는 1-GIOU로 계산되며, 두 박스 사이의 최소영역과 공통영역을 정의하여 학습과정에서 regression을 계산한다.

3. 유효성 평가 기준

자체 데이터 셋의 성능평가 기준은 COCO, KITTI 데이터 셋을 참고하였다. COCO 데이터 셋은 mAP을 사용하였는데, 이는 모든 클래스에 대해 IOU가 0.5부터 0.95까지를 0.05씩 증가하면서 각각의 AP를 구한 후의 평균값이다. KITTI 데이터 셋은 가림/잘림 정도와 객체의 높이 (픽셀) 값에 따라 평가 난이도 (easy, moderate, hard)를 분류한다. IOU 임계값을 차량은 0.7 이상, 보행자와 이륜차는 0.5 이상으로 기준을 정하여 mAP를 계산하였다. 자체 데이터 셋은 표 1과 같이 카메라로부터 거리에 따른 객체의 높이 (픽셀) 값을 계산하여 바운딩 박스의 가공 기준을 설정하였다. 차량은 30 픽셀 미만, 보행자와 이륜차는 45 픽셀 미만의 GT/Inference 바운딩 박스는 평가에서 제외되며, 동일한 조건으로 공개 데이터 셋도 평가를 진행하였다. 평가 기준은 COCO와 KITTI로 구분되며, 평가 난이도는 easy와 hard로

구분한다. 이것은 객체가 완전히 보이고 (non-occluded) 잘림이 없는 경우 (non-truncated)는 easy로, 객체에 가림 (occluded)이 존재하고 잘림이 있는 경우 (truncated)는 hard로 구분됨을 의미한다.

IV. 성능 평가

1. 평가 실험 환경

자체 데이터 셋의 유효성을 검증하기 위해, 구축한 데이터 셋 275,400장 이미지를 8:1:1 비율로 나누어 학습, 검증, 평가를 진행하였다. 데이터를 학습시키기 위한 라벨 포맷은 가공 포맷인 json 형식을 YOLO 라벨 포맷인 txt 형식으로 변환 (좌표계, 파라미터, 데이터 형식 변경) 하였다. 이와 동일하게 공개 데이터 셋도 학습과 검증 데이터를 나누어 진행하였는데, 121,408장의 COCO 이미지를 9.5:0.5 비율로 학습과 검증 데이터를 구분하였으며, 7481장의 KITTI 학습용 이미지 데이터를 8.5:1.5 비율로 나누어 학습과 검증 데이터를 구분하였다. 각 데이터 셋의 이미지 개수와 객체 수, 클래스 수는 표 2와 같다. 특히, COCO 데이터 셋은 80개 클래스를 포함한 모든 이미지를 학습하고, 자율주행 시 영향을 주는 동적 객체 6개 클래스에 대해서만 평가를 수행하였다. COCO, KITTI, 자체 데이터 셋 평가는 차량 (차/트럭/버스), 이륜차 (오토바이/자전거), 보행자의 3개 클래스 그룹으로 구분하여 수행하였다.

유효성 평가 과정에서 CPU (Intel xeon CPU E5-2650 v4)와 GPU (NVIDIA GeForce TITAN_X)를 사용하였다. 학습은 평가 GPU와 동일한 GPU 8개를 사용하고 batch 사이즈는 8로 정의하고, epoch은 자체 DB 128, COCO DB는 300, KITTI DB는 150으로 정의하였다. 검증 및 평가 과정에서 한 개의 GPU에 대해 1 배치 사이즈를 정의하고, 후처리 (NMS)에 대한 파라미터 IOU 임계값은 0.6, Confidence 임계값은 0.6으로 설정하였다.

2. 실험 결과

YOLOv5 모델을 이용한 공개 데이터 셋 대비 자체 데이터 셋에 대한 유효성 평가 결과는, COCO 기준 평가는 표 3, KITTI 기준 평가는 표 4와 같다. 표 3에서 보는 바와 같이, COCO 평가 기준으로 데이터 셋을 비교하면, 자체 데이터 셋은 COCO 데이터 셋보다 차량은 약 13%, 이륜차는 약

표 2. 데이터 셋 비교
Table 2. Comparison of dataset

	#images	#objects	#classes
COCO	121,408	14,073	6(3)
KITTI	7,481	39,597	7(3)
Our_DB	247,860	1,868,254	6(3)

표 3. 검증용 데이터 셋 성능 비교 (COCO 기준)

Table 3. Performance comparison of validation dataset (based on COCO)

	mAP	AP (0.5~0.95)		
		Veh.	Cyc.	Ped.
COCO	0.607	0.614	0.521	0.687
KITTI	0.671	0.825	0.667	0.520
Our_DB	0.651	0.747	0.605	0.601

표 4. 검증용 데이터 셋 성능 비교 (KITTI 기준)

Table 4. Performance comparison of validation dataset (based on KITTI)

	Diff.	AP (0.7)			AP (0.5)		
		Veh.	Cyc.	Ped.	Veh.	Cyc.	Ped.
KITTI	Easy	93.70	38.32	56.83			
	Mod.	82.95	24.81	40.41			
	Hard	70.82	20.72	35.66			
Our_DB	Easy	96.01	74.97	81.82			
	Hard	91.60	73.67	76.33			

표 5. 시나리오 테스트 데이터 셋 성능 비교

Table 5. Performance comparison on the scenario test dataset

	Scen.	AP (0.5)	
		Veh.	Ped.
COCO YOLOv5	1	0.631	0.439
	2	0.624	0.467
	3	0.410	0.450
Our_DB YOLOv5	1	0.776	0.754
	2	0.939	0.855
	3	0.754	0.868

8% 높은 성능을 보이고, 보행자는 약 8% 낮은 성능을 보였다. KITTI 데이터 셋보다 차량은 약 7%, 이륜차는 약 6% 낮은 성능을 보였으나, 보행자는 약 8% 높은 성능을 보였다. 표 4와 같이, KITTI 평가 기준으로 데이터 셋을 비교할 때, 자체 데이터 셋은 KITTI 데이터 셋보다 Easy 기준의 경우에 차량은 약 3%, 이륜차는 약 37%, 보행자는 약 25%의 높은 성능을 보였으며, Hard 기준의 경우에 차량은 약 21%, 이륜차는 약 53%, 보행자는 약 41%의 높은 성능을 보였다. 두 가지 기준의 검증 결과에서 보는 바와 같이, 차량 보다는 이륜차와 보행자에서 다른 공개 데이터 셋보다 자체 데이터 셋이 높은 성능을 보였는데, 이는 자체 구축한 학습 데이터 셋 중에서 작은 객체인 이륜차 비율이 4% (66,281개), 보행자 비율이 36% (586,800개)로 다른 공개 데이터 셋보다 많기 때문이다.

시나리오 중심 테스트 데이터 셋에 대한 유효성 평가 결과는 표 5에서 보는 바와 같다. 시나리오 1은 도심 골목길 주행 상황이고, 시나리오 2는 도심 교차로에 혼잡한 상황이

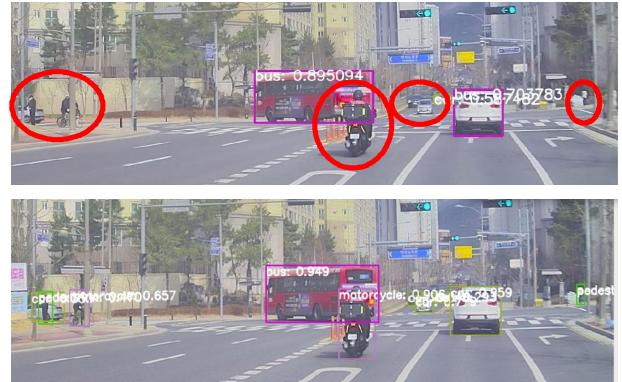


그림 4. Open platform YOLOv5 모델과 성능 비교: COCO YOLOv5 (위), Our_DB YOLOv5 (아래)

Fig. 4. Performance comparison on

고, 시나리오 3은 야간 도심 주행 상황의 연속 데이터 셋이다. 오픈 소스 플랫폼의 YOLOv5 모델은 COCO 데이터 셋으로 학습된 pre-trained 모델이고, Our_DB YOLOv5 모델은 자체 구축한 데이터 셋으로 학습된 모델이다. 표 5는 정량적 결과로, 오픈소스 플랫폼 모델에 비해 자체 데이터를 학습한 모델이 모든 시나리오에서 차량 약 30%, 보행자 약 35% 높은 성능을 보였다. 이륜차는 COCO 데이터 셋과 자체 데이터 셋의 어노테이션 기준이 다르기 때문에, 표 5 결과는 자체 어노테이션 기준을 적용해서 평가하므로 이륜차 클래스는 평가 항목에서 제외하였다.

그림 4는 정성적으로 평가한 결과로, YOLOv5 모델을 이용하였으며, 위에 이미지는 COCO 데이터 셋으로 학습한 결과이고, 아래 이미지는 자체 데이터 셋으로 학습한 결과이다. 빨간색으로 표시된 부분이 객체를 검출하지 못한 영역을 나타내는데, 자체 데이터 셋을 학습시킨 결과에서는 작은 객체인 보행자, 이륜차뿐만 아니라 멀리서 작게 보이는 차량들도 제대로 검출하는 것을 확인하였다. 이는 기존 공개 데이터 셋으로 학습한 모델은 도심이나 교차로 등에서 트럭, 보행자, 이륜차 등에 대해 미검출 하는 경우가 많았는데, 자체 구축한 데이터 셋을 학습한 모델은 국내 도로환경에서 나타나는 다양한 객체에 대해서도 검출이 가능함을 확인하였다. 자체 데이터 셋은 AI 허브에 추후 공개 예정이다.

V. 결론

이 논문은 국내 자율주행 환경에 적합한 멀티 카메라 기반의 자체 데이터 셋을 구축하고 유효성을 검증하였다. 해당 데이터 셋은 도심이나 교차로 등과 같은 복잡한 도로 환경에서 나타날 수 있는 다양한 동적 객체에 대해 멀티 카메라 데이터로 구성된다. 구축한 데이터 셋의 품질 및 유효성을 검증하기 위해 YOLOv5 모델을 이용하여, COCO, KITTI와 같은 공개 데이터 셋을 통해 학습시킨 결과와 자

체 데이터 셋을 학습시킨 결과를 정성/정량적으로 성능을 비교 평가하였다. 자체 구축 데이터 셋은 공개 데이터 셋에 비해 작은 동적 객체나 큰 차량에 대해 높은 성능을 보임으로, 구축한 데이터 셋의 유효성을 확인하였다. 또한 도심이나 교차로 등의 복잡한 환경에 대해 시나리오 기반의 테스트 셋을 구축하여 평가하였다. 오픈 소스 플랫폼의 pre-trained 된 YOLOv5 모델과 성능을 비교하니 자체 데이터 셋을 학습한 YOLOv5 모델이 정성/정량적으로 우수한 성능을 보였다. 이 데이터 셋은 향후 국내 AI 허브에 공개될 예정이다.

References

- [1] T. Y. LIN, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, "Microsoft COCO: Common Objects in Context," in Proc. European Conference on Computer Vision, Springer, Cham, pp. 740-755, 2014.
- [2] A. Geiger, P. Lenz, C. Stiller, R. Urtasun, "Vision Meets Robotics: The Kitti Dataset," The International Journal of Robotics Research, Vol. 32, No. 11, pp. 1231-1237, 2013.
- [3] P. Sun, H. Kretschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, V. Vasudevan, W. Han, J. Ngiam, H. Zhao, A. Timofeev, S. Ettinger, M. Krivokon, A. Gao, A. Joshi, Y. Zhang, J. Shlens, Z. Chen, D. Anguelov, "Scalability in Perception for Autonomous Driving: Waymo Open Dataset," in Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 2446-2454, 2020.
- [4] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, O. Beijbom, "Nuscenes: A Multimodal Dataset for Autonomous Driving," in Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 11621-11631, 2020.
- [5] A. Bochkovskiy, C. Y. Wang, H. Y. M. Liao, "Yolov4: Optimal Speed and Accuracy of Object Detection," arXiv preprint arXiv:2004.10934, 2020.
- [6] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, A. C. Berg, "Ssd: Single Shot Multibox Detector," in Proc. European Conference on Computer Vision, Springer, Cham, pp. 21-37, 2016.
- [7] T. Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollar. "Focal Loss for Dense Object Detection," in Proc. IEEE International Conference on Computer Vision, pp. 2980-2988, 2017.
- [8] R. Girshick, J. Donahue, T. Darrell, J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," in Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 580-587, 2014.
- [9] R. Girshick, "Fast r-cnn," Proceedings of the IEEE International Conference on Computer Vision, pp. 1440-1448, 2015.
- [10] S. Ren, K. He, R. Girshick, J. Sun, "Faster r-cnn: Towards Real-time Object Detection with Region Proposal Networks," Advances in Neural Information Processing Systems 28 2015.
- [11] C. Y. Wang, H. Y. M. Liao, Y. H. Wu, P. Y. Chen, J. W. Hsieh, I. H. Yeh, "CSPNet: A New Backbone that can Enhance Learning Capability of CNN," in Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 390-391, 2020.
- [12] G. Huang, Z. Liu, L. v. d. Maaten, K. Q. Weinberger, "Densely Connected Convolutional Networks," in Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700-4708, 2017.
- [13] K. He, X. Zhang, S. Ren, J. Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 37, No. 9, pp. 1904-1916, 2015.
- [14] S. Liu, L. Qi, H. Qin, J. Shi, J. Jia, "Path Aggregation Network for Instance Segmentation," in Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 8759-8768, 2018.
- [15] H. K. Kim, K. Y. Yoo, J. H. Park, H. Y. Jung, "Deep Learning Based Gray Image Generation from 3D LiDAR Reflection Intensity," IEMEK J. Embed. Sys. Appl., Vol. 14, No. 1, pp. 1-9, 2019 (in Korean).

Jin-Hee Lee (이진희)



2005 Computer Science from Korea National Open University (B.S.)

2007 Computer and Information Engineering from Inha University (M.S.)

2015 Computer and Information Engineering from Inha University (Ph.D.)

2015~Division of Automotive Technology from Daegu Gyeongbuk Institute of Science & Technology (DGIST). (Senior researcher)
Field of Interests: Autonomous driving, Deep learning, Sensor fusion
Email: jhlee07@dgist.ac.kr

Jae-Keun Lee (이재근)

2015 Mechanical and ICT Convergence Engineering from Sunmoon University (B.S.)

2021~2022 Division of Automotive Technology from Daegu Gyeongbuk Institute of Science & Technology (DGIST). (Researcher)

2022~FutureDrive. (Researcher)

Field of Interests: Autonomous driving, 3D Vision, Deep learning

Email: lej8104@futuredrive.net

Jaehyeong Park (박재형)

2013 Information and Communication Engineering from Daegu University (B.S.)

2015 Information and Communication Engineering from Daegu University (M.S.)

2015~Division of Automotive Technology from Daegu Gyeongbuk Institute of Science & Technology (DGIST). (Researcher)

Field of Interests: Autonomous driving, Deep learning, Stereo camera

Email: stillrunning@dgist.ac.kr

Je-Seok Kim (김세석)

2009 Mechanical Engineering from Yeungnam University (B.S.)

2012 Automotive Engineering from Hanyang University (M.S.)

2015 Automotive Engineering form Hanyang University (Ph.D.)

2020~Division of Automotive Technology from Daegu Gyeongbuk Institute of Science & Technology (DGIST). (Senior researcher)

Field of Interests: Automotive driving, Planning algorithm

Email: jeseok@dgist.ac.kr

Soon Kwon (권순)

2003 Electric & Electronic Engineering from Korea University (B.S.)

2006 Electronic Engineering from Seoul National University (M.S.)

2006~Division of Automotive Technology from Daegu Gyeongbuk Institute of Science & Technology (DGIST). (Principal researcher)

Field of Interests: Autonomous driving, Deep learning, Sensor fusion

Email: soonyk@dgist.ac.kr