논문 2023-18-25

국내 주행환경을 고려한 자율주행 라이다 데이터 셋 구축 및 효과적인 3D 객체 검출 모델 설계

(Construction of LiDAR Dataset for Autonomous Driving Considering Domestic Environments and Design of Effective 3D Object Detection Model)

이 진 희, 이 재 근, 이 주 현, 김 제 석, 권 순*
(Jin-Hee Lee, Jae-Keun Lee, Joohyun Lee, Je-Seok Kim, Soon Kwon)

Abstract : Recently, with the growing interest in the field of autonomous driving, many researchers have been focusing on developing autonomous driving software platforms. In particular, we have concentrated on developing 3D object detection models that can improve real-time performance. In this paper, we introduce a self-constructed 3D LiDAR dataset specific to domestic environments and propose a VariFocal-based CenterPoint for the 3D object detection model, with improved performance over the previous models. Furthermore, we present experimental results comparing the performance of the 3D object detection modules using our self-built and public dataset. As the results show, our model, which was trained on a large amount of self-constructed dataset, successfully solves the issue of failing to detect large vehicles and small objects such as motorcycles and pedestrians, which the previous models had difficulty detecting. Consequently, the proposed model shows a performance improvement of about 1.0 mAP over the previous model.

Keywords : Dataset, 3D Object Detection, Autonomous Driving

Ⅰ. 서 론

근래에 자율 주행에 대한 관심이 증가하면서, 대부분의 연구자들은 자율주행 소프트웨어 플랫폼 개발에 집중하고 있다. 이러한 자율 주행 플랫폼은 대부분 인지, 주행 계획, 제어 등의 프로세스로 구성되어 있다. 이 중에서 특히 인지는 자율주행에서 필수적인 기술이며, 이 성능은 주행 계획과 제어에 영향을 주게 되므로, 인지는 자율주행에서는 중요한 요소 기술임은 분명하다. 자율주행 인지 영역에서는 3차원 공간에서의 객체 인식 연구가 활발하게 진행되고 있다 [1]. 이것은 안전한 자율 주행을 위해서는, 특히 3D 공간에서 객체의 위치와 크기를 정확하게 파악하는 것이 무엇보다도 중요하기 때문이다. 그러므로 camera 보다는 3D 공간 데이터를 수집할 수 있는 LiDAR를 주로 사용한다. 이는 LiDAR로부터 수집된 포인트 클라우드 데이터는 camera에 비해 멀리 있는 객체를 검출할 수 있을 뿐 아니라, 정확도가 높은 3차원 거리 정보를 제공하기 때문이다. 그러므로 LiDAR를 사용하여 복잡한 도로 상황에서 실시간으로 객체를 분류하고, 위치 추정하여 검출까지 함으로써 3D 객체 검출 연구는 활발히 진행되고 있다.

기존의 3D 객체 검출 논문들에서는 모델에 대해서 독립적으로 평가하고 계산하는 것을 주로 다루고 있기 때문에 성능의 수치적인 향상에만 목표를 두고 있다. 그러나 우리처럼 자율주행 연구 그룹은 실질적인 자율주행 태스크를 다루므로, 자율주행 플랫폼이 주행 도중에 나타날 수 있는 큰 차량, 보행자, 이륜차 등의 객체 검출 비율 및 성능에 관심이 집중된다. 그러므로 이 논문에서는 제안하는 3D 객체 검출 모델의 성능을 정성적/정량적으로 평가하는 프로세스를 수립하고, 모델의 성능을 검증하였다.

먼저, 도로 위 객체를 잘 탐지하기 위해, 기존의 CenterPoint 모델의 Focal loss를 수정하고, DIoU loss를 추가하여 loss를 재설계한 VariFocal-based CenterPoint 모델을 제안하였다 [2-4]. 이는 객체의 위치정보를 포함한 IoU (Intersection over Union)를 함께 학습함으로써, 평가를 수행할 때 classification score와 localization estimation 간의 불일치 문제를 완화하여 모델의 검출 성능을 향상시킬 수 있다 [5]. 이렇게 개선한 모델을 Waymo [6], H3D [7], ApolloScape [8], A3D (Autonomous Driving 3D Dataset in Korea) [9]데이터 세트로 학습하고 비교 평가하였다. 실험결과, 제안한 CenterPoint 모델은 국내에 특화된 환경의 데이터인 보행자, 이륜차를 많이 포함한 자체 구축 데이터 셋인 A3D를 학습함으로써 정성적/정량적으로 기존 데이터 세트보다 좋은 결과를 보였다. 또한 자율주행 차량에 탑재하기

*Corresponding Author (soonyk@dgist.ac.kr)
Received: Jun. 20, 2023, Revised: Aug. 1, 2023, Accepted: Aug. 9, 2023.
J. H. Lee: DGIST (Senior Researcher)
J. K. Lee: FutureDrive (Researcher)
J. Lee: DGIST (M.S. Student)
J. S. Kim: DGIST (Senior Researcher)
S. Kwon: DGIST (Principal Researcher)
※ 본 연구는 과학기술정보통신부에서 지원하는 대구경북과학기술원 기관고유사업 (23-IT-02)과 기술사업화 역량강화사업 (2023-DG-RD-0041) 지원을 받아 수행 되었습니다.

ⓒ IEMEK J. Embed. Sys. Appl. 2023 Oct. 18(5) 203-208
ISSN : 1975-5066
http://dx.doi.org/10.14372/IEMEK.2023.18.5.203

위해, 실시간성을 고려한 TensorRT 모델로 변환하여 처리 시간도 비교하였다. 보행자와 오토바이가 상당수 포함된 A3D 데이터 셋으로 VF-CenterPoint 모델을 학습하였으며, 제안한 모델은 기존의 CenterPoint 모델보다 정량적으로 약 1.0 mAP 더 높은 성능을 보여준다.

Ⅱ. 자체 데이터 셋 구축 및 3D 객체 검출 모델 설계

우리는 우수한 딥러닝 모델 개발도 중요하지만, 학습을 위한 대규모의 데이터 셋 구축이 우선 중요하다고 판단하여, 큰 차량 및 보행자와 이륜차 등의 작은 객체가 다수 포함된 국내 주행환경에 특화된 자체 데이터 셋을 구축하였다. 또한 자율주행 차량 플랫폼에 적용하기 위해 실시간성 및 객체 검출 성능이 우수한 모델을 개발하였다.

1. 국내 환경에 특화된 자체 데이터 셋 구축

자율 주행을 위해 흔히 사용하는 공개 데이터 셋으로는 Waymo Open Dataset과 H3D, ApolloScape 데이터 셋 등이 있다. Waymo 데이터 셋은 미국 도심에서 수집한 대규모 자율주행 데이터 셋으로 다양한 시간대 (낮, 밤, 새벽 등)와 다양한 장소 (도심, 교외 등)에서 수집되었다 [6]. 학습과 검증을 위한 798,202 개의 시퀀스 정보로 구성되며, 약 16만장의 학습데이터와 4만장의 검증데이터로 되어 있다. 3D 데이터 셋은 64채널 라이다의 포인트 클라우드 데이터와 3개 클래스 (차량, 보행자, 이륜차)에 대한 어노테이션 데이터로 구성되었다.

H3D는 미국 혼다사가 미국 내 도로 상황을 수집한 데이터 셋이다 [7]. 8가지 클래스 (자동차, 보행자, 버스, 트럭, 기타 차량, 자전거, 오토바이, 동물)에 대한 3D 어노테이션과 64채널 라이다 데이터로 구성되어 있다. 3D 데이터 셋은 23,351 프레임의 학습 데이터와 6,008 프레임 검증 데이터로 구성되었다. ApolloScape는 Baidu 회사가 중국 베이징의 도로 상황을 수집한 데이터 셋이다 [8]. 64채널 라이다로 수집한 4개의 클래스 정보에 대한 3D 어노테이션을 수행하였으며, 객체 감지에 필요한 4,423장의 학습 데이터와 1,169장의 평가 데이터로 구성되었다.

국내 도로 상황이 반영된 데이터 셋을 구축하고 이를 학습시켜서 다양한 도로 환경에서도 강건한 객체 검출이 우리의 목표이므로, 먼저 국내 도로 환경이 반영된 A3D (Autonomous Driving 3D Dataset in Korea)라는 자체 자율주행을 위한 데이터 셋을 구축하였다 [9]. 이는 국내의 다양한 도로환경 (도심, 고속도로, 자동차 전용도로, 지방도로, 골목길 등)에서 수집한 3D 라이다 데이터 셋이다. 이 데이터 셋은 128채널 라이다를 차량의 루프 탑에 설치하고, 이를 이용하여 수집된 라이다 포인트를 가공한 데이터 셋이다. 특히, 혼잡한 도심의 교통 상황 시나리오에서 수집하였으며, 국내 도로에서 검출 가능한 큰 트럭이나 버스와 같은 차량 뿐만 아니라, 작은 객체인 보행자, 이륜차 등이 다수 포함된 데이터이며, 정제와 가공 및 검수 작업까지 완료하여 제작된

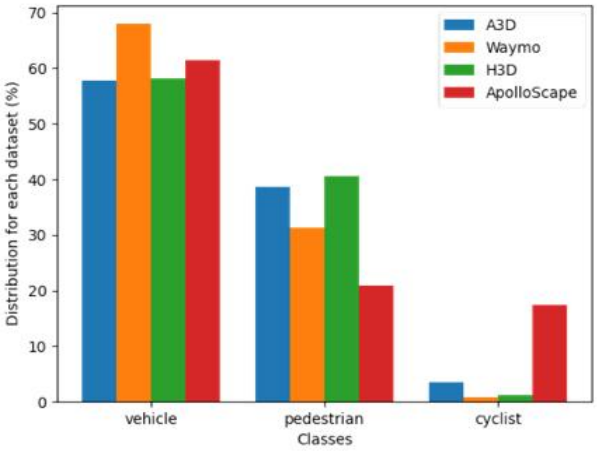

그림 1 . 공개 데이터 셋 및 자체 데이터 셋 클래스별 비율
Fig. 1. Distribution by classes for public datasets and our dataset

데이터 셋이다. A3D 데이터 셋은 723,000 개의 차량, 624,000 개의 작은 객체 (보행자, 오토바이, 자전거), 210,000 개의 큰 객체 (트럭, 버스)의 레이블 정보를 포함하고 있다. 이것은 공개 데이터 셋 중에서 가장 많이 사용되는 Waymo Open Dataset 보다 좀 더 여러 개의 클래스로 세분화하여 가공되었다. 그림 1에서 보는바와 같이, 전체 객체 비율을 비교해보면, A3D 데이터 셋은 특히 보행자는 약 40%, 이륜차는 약 5%의 높은 가공 비율을 보인다. 이는 A3D 데이터 셋이 Waymo Open Dataset과 비교해서 보행자 클래스의 가공 비율은 약 8%, 이륜차는 약 3% 더 높은 비율을 보인다. 이 논문에서는 객체 검출 모델을 비교 평가하기 위해, A3D 데이터는 약 9만 장의 1Hz 샘플링 데이터를 학습, 평가, 테스트 데이터로 구분하여 각각 68,573장, 10,698장, 10,695장씩 사용하였다. A3D 데이터 셋은 현재 국내 AI Hub에 공개되어 있다 [9].

2. 효과적인 3D 객체 검출 모델 설계

3D 객체 검출 모델인 CenterPoint [4]는 대표적인 anchor free 기반의 모델이다. CenterPoint는 VoxelNet 기반 [10-12]과 PointPillar 기반 [13, 14]의 방법으로 나누어지는데, 이 논문에서는 실시간이 가능한 1-stage PointPillar 구조의 네트워크를 사용한다. 이미지 영역의 정형화된 픽셀 데이터와 달리 포인트 클라우드는 대부분 희소하고 정렬되어 있지 않다. 따라서 VoxelNet 기반과 PointPillar 기반의 방법은 포인트 클라우드를 컨볼루션 신경망 (CNN)에 적합한 규칙적인 복셀 또는 필러로 변환하여 사용한다. VoxelNet 기반 방식은 객체 감지를 위해 포인트 클라우드의 강력한 3D 특징을 추출하는 것을 목표로 하지만, 계산 비용이 높아 inference 시간이 느린 단점이 있다. 이 문제를 해결하기 위해 VoxelNet 기반의 sparse 컨볼루션 연산을 도입하여 3D 컨볼루션의 시간 소모량을 줄일 수 있다. 그러나 이러한 기술은 아직 TensorRT 모델이 공식적으로 지원되지 않기 때문에 산업용 임베디드 시스템에 적용하기에는 여전히 어려운 문제를 가

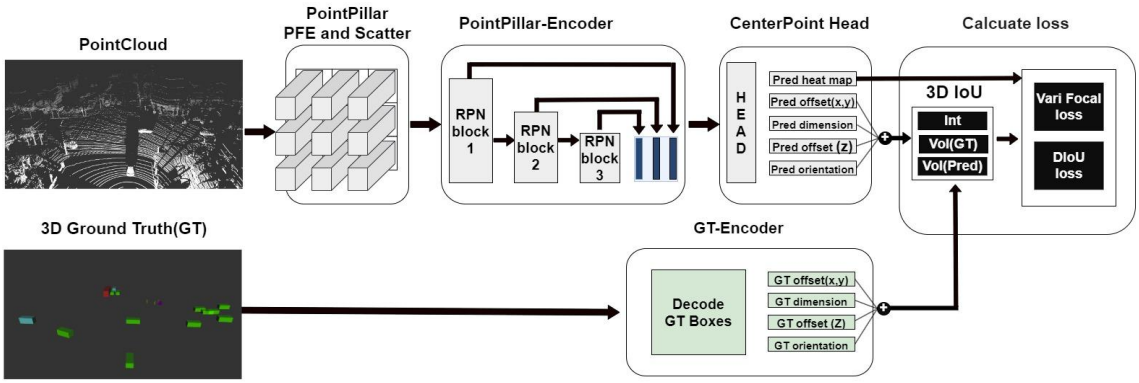

그림 2. VariFocal CenterPoint 전체 구조
Fig. 2. The overall architecture of VariFocal CenterPoint

지고 있다. 우리는 실시간성 이슈를 고려하여 Pillar 기반의 CenterPoint 모델을 사용하였다. 그림 2에서와 같이, 이는 Pillar Feature Encoding (PFE), Region Proposal Network (RPN), Head로 구성된다. 먼저 PFE를 통하여 sparse하고 불규칙한 포인트를 CNN이 적용 가능한 형태인 pillar 구조로 변환한다. 그 이후에, ConvNet과 DeConvNet이 결합된 RPN block의 결과물인 pillar feature로 인코딩을 수행한다. 마지막으로, CenterNet [15] 기반의 방법과 유사하게, 네트워크로부터 heatmap, 오프셋 (x, y), Z축, 차원 (width, length, height), heading을 inference하여, 3D 객체 생성에 필요한 classification score와 bounding box 정보를 예측한다. 그림 2에서와 같이, 제안하는 VariFocal 기반의 CenterPoint 전체 아키텍처는 이전의 CenterPoint 모델과 유사하다.

우리는 실험을 통하여, PointPillar Encoder와 Focal loss로 구성된 기존의 CenterPoint 모델은 종종 부정확한 위치에 바운딩 박스들을 그리는 것을 확인하였다. 이러한 바운딩 박스는 대부분 false-positive로써, classification score는 상대적으로 높지만, IoU가 현저히 낮은 경향을 보인다. 그러므로 이러한 low-quality 박스들에 대한 예측은 CenterPoint의 전체적인 성능에 부정적인 영향을 줄 수 있다. 우리는 inference 하는 동안에 추가적인 계산 비용을 발생시키지 않고, CenterPoint 모델의 검출 성능을 높이기 위해서, 학습할 때에 classification score와 IoU를 동시에 학습하게 하는 VariFocal 기반의 CenterPoint 모델을 제안한다. 이것은 기존의 Focal loss를 재설계한 것으로, classification score와 IoU가 결합된 Classification-IoU joint representation을 통하여 학습시킨다. 결과적으로, loss 레벨에서 confidence score와 localization estimation 간의 불일치 문제를 완화하므로, 모델의 검출 성능을 더욱 향상시킨다. 즉, 2D 모델에서 사용된 VariFocal loss 기법과 같이, 기존 방법처럼 GT (Groud Truth) label을 hard label 형태인 0이나 1의 값으로 구분하지 않고, 우리는 soft label 형태인 0과 1 사이의 IoU aware GT values인 IoU^{ω}로 대체하여 계산한다 [16]. 여기서 ω는 IoU weights를 의미한다. 제안하는 loss function을 적용하면, 학습하는 동안에 IoU가 더 큰 positive sample에 집중하게 되어 전체적으로 검출 성능이 향상된다. 또한 우리는 positive sample box와 predicted box의 3D 바운딩 박스 교차영역은 최대화하고, 두 박스 사이의 3D normalized distance는 최소화 하는 DIoU loss 방법을 적용하여 검출 성능을 더욱 향상시켰다 [3].

$$PL_{vf} = -\{q\times\log(p)+(1-q)\times\log(1-p)\}\times q, \quad (1)$$

$$NL_{vf} = -\log(1-p)\times(p)^{\beta}\times(1-\sigma)^{\tau}, \quad (2)$$

$$L_t = A\times L_{vf} + B\times L_{reg} + C\times L_{DIoU}. \quad (3)$$

위에서 설명한 내용을 수식으로 정리해보면 다음과 같다. 수식 (1)과 (2)는 VariFocal loss를 계산하는 수식으로, 각각 positive 샘플과 negative 샘플에 대해 계산한다. q는 Classification-IoU Joint 점수를 나타내고, p는 IoU를 포함한 예측 점수를 나타내며, σ는 가우시안 필터를 나타낸다. 수식 (2)에서 β와 τ는 가중치 매개변수이며, 이는 CenterPoint와 동일하게 표현한다. 수식 (1)은 y(y=[0, IoU^{ω}])에서 positive 샘플의 계산식이다. 교차 엔트로피 부분은 $-\{q*\log(p)+(1-q)*\log(1-p)\}$로 정의되며, 스케일 계수는 q로 표시된다. 수식 (1)을 적용하면 IoU가 높은 고품질 positive 샘플의 loss 값은 가중치를 낮추어 계산되는 반면, IoU가 낮은 저품질 샘플의 loss 값은 증가하게 된다. 따라서 제안하는 loss 함수는 많은 수의 저품질 샘플이 loss에 영향을 미치는 것을 방지할 수 있다. 결과적으로 제안한 VF-CenterPoint는 저품질 positive 샘플에 집중하여 학습할 수 있으므로 모델이 바운딩 박스에 대한 보다 정확한 위치를 생성할 수 있다. 수식 (2)는 y가 0인 negative 샘플의 계산을 나타내며, 이는 Focal loss 수식과 동일하다. 이 논문에서는 성능을 개선하기 위해 DIoU loss [3]의 개념도 적용했다. DIoU loss는 예측된 박스와 GT 박스 사이의 중첩 영역을 최대화하고, 두 박스의 3D 거리를 최소화하는 것을 목표로 한다. 수식 (3)은 total loss를 계산하는 수식으로, L_{vf}는 confidence score와 positive 샘플 사이에서 계산되는 VariFocal classification loss를 의미하고, L_{reg}는 $L1$ regression loss를 나타낸다. LDIoU는 예측된 박스와 GT 박스 사이에서 계산된 DIoU loss를 의미한다. A는

classification loss의 가중치 매개변수이고, B와 C는 각각 regression loss 및 DIoU loss의 가중치 매개변수들이다. 기존 설정에 따라 A는 1로 설정되고 B와 C는 2로 정의한다.

Ⅲ. 실험 결과

우리는 Waymo, H3D, ApolloScape, A3D 데이터 셋에서 기존의 CenterPoint 모델과 제안하는 VariFocal CenterPoint 모델에 대해 동적 객체 검출 성능을 정량적/정성적으로 비교하였다. 또한 모델의 학습 과정과 데이터 증강 기법은 기존 설정과 동일하게 세팅하였다. 모든 모델은 Intel (R) Xeon (R) CPU E5-2630 v4와 GeForce RTX 2080 Ti가 장착된 데스크탑 PC에서 학습하였고, 테스트 환경은 Intel CPU i9-11900K과 GeForce GTX 1070 Ti가 탑재된 PC에서 수행하였다. 각 데이터 셋을 학습시키는 동안 Waymo Open Dataset은 36 epoch, ApolloScape는 100 epoch, H3D와 A3D는 36 epoch를 설정하였다. VariFocal loss의 학습 가중치는 각각 β=2 와 τ=4로 설정한다. 평가 과정에서 이전 CenterPoint 모델과 제안된 VF-CenterPoint 모델은 모두 TensorRT로 변환하였다. 두 모델 모두 실제 자율 주행 차량에 적용하기 위해 TensorRT로 변환된 모델들로 정확도 성능을 비교하기 위해 각각 실험하였다. 그 후에 각 데이터 셋을 사용하여 모델을 평가하여 정확도와 지연 시간 측면에서 성능을 비교하였다.

각 모델의 객체 검출 정확도에 대한 평가는 차량, 보행자, 이륜차 총 세 가지 클래스에 대한 3D IoU로 계산한다. 우리는 Waymo, KITTI 등 널리 알려진 평가 방법을 따라, IoU threshold를 클래스마다 각각 차량은 0.7, 보행자와 이륜차는 각각 0.5로 설정하였다. 다음에 나오는 표에서 CP는 CenterPoint 모델을 의미하고, V_CP는 우리가 제안한 모델인 VariFocal CenterPoint를 의미한다.

표 1에서와 같이 3D 평가 기준에서 Waymo 데이터 셋에서 VariFocal CenterPoint는 기존의 CenterPoint 보다 차량 약 0.7%, 보행자 약 1.7%, 이륜차 약 0.8%로 전체 mAP는 약 1.1% 더 나은 성능을 달성하였다. Apollo scape 데이터 셋에서는 VariFocal CenterPoint가 기존의 CenterPoint 보다 차량은 약 0.7%, 보행자 약 2%, 이륜차 약 1% 높아서 전체 mAP는 약 1.2% 정도로 더 높은 성능을 보였다. H3D 데이터 셋에서는 7개 클래스로 세분화하여 평가하였으며, VariFocal CenterPoint가 기존의 CenterPoint 보다 모든 클래스에서 더 높은 성능을 보였으며, 차량 약 2%, 보행자 약 0.6%, 이륜차 약 3%로 전체 mAP는 약 1.7% 더 나은 성능을 달성하였다. 국내에서 수집한 자체 데이터 셋인 A3D도 멀티클래스로 세분화하여 평가하였으며, VariFocal CenterPoint의 검출 성능은 차량과 보행자, 이륜차가 거의 동일하여 mAP도 약 0.04% 더 나은 성능을 달성하였다. 실시간 검출 속도 차이는 0.3~1.5ms 이내로 두 모델이 거의 유사하였다. 이를 통해 제안한 모델인 VariFocal CenterPoint가 모든 데이터 셋에서 평균적으로 3D 객체 검출 성능이 약 1.1% 더 우수한 것을 확인하였다.

표 1. 네 개의 데이터 셋에 따른 모델들 간의 3D 평가 성능 비교
Table 1. Performance comparison of 3D detection models with four dataset

3D	Class	Model	Dataset			
			Waymo	Apollo.	H3D	A3D
mAP		CP	69.88	65.10	39.24	48.31
		V_CP	**70.92**	**66.27**	**40.95**	**48.35**
AP	Car	CP	73.84	68.70	51.80	62.51
		V_CP	**74.53**	**69.52**	**52.70**	**62.52**
	Bus	CP	-	-	7.44	66.02
		V_CP	**-**	**-**	**9.24**	**67.15**
	Truck	CP	-	-	28.57	60.74
		V_CP	**-**	**-**	**28.19**	**59.78**
	Other	CP	-	-	13.41	13.33
		V_CP	**-**	**-**	**16.23**	**15.36**
	Bicyc.	CP	72.48	69.88	65.39	11.95
		V_CP	**73.22**	**70.87**	**67.70**	**9.99**
	Motor biike	CP	-	-	59.98	70.37
		V_CP	**-**	**-**	**63.85**	**70.40**
	Ped.	CP	63.32	56.71	48.11	53.22
		V_CP	**65.00**	**58.41**	**48.71**	**53.27**

표 2. 네 개의 데이터 셋에 따른 모델들 간의 BEV 평가 성능 비교
Table 2. Performance comparison of BEV detection models with four dataset

BEV	Class	Model	Dataset			
			Waymo	Apollo.	H3D	A3D
mAP		CP	72.50	71.09	46.71	51.82
		V_CP	**74.09**	**71.66**	**47.43**	**52.35**
AP	Car	CP	82.01	77.07	68.29	76.76
		V_CP	**83.94**	**78.97**	**68.30**	**76.22**
	Bus	CP	-	-	20.01	74.12
		V_CP	**-**	**-**	**16.42**	**74.13**
	Truck	CP	-	-	39.94	68.95
		V_CP	**-**	**-**	**38.98**	**68.90**
	Other	CP	-	-	15.85	14.50
		V_CP	**-**	**-**	**19.43**	**16.94**
	Bicyc.	CP	63.47	71.50	68.13	10.49
		V_CP	**64.38**	**72.11**	**69.99**	**12.19**
	Motor biike	CP	-	-	60.06	73.23
		V_CP	**-**	**-**	**64.11**	**73.32**
	Ped.	CP	72.03	64.10	54.69	44.70
		V_CP	**72.03**	**64.52**	**54.78**	**44.73**

표 2에서와 같이 BEV (Bird Eye's View) 평가 기준에서 Waymo 데이터 셋에서 VariFocal CenterPoint는 기존의 CenterPoint 보다 차량 약 1.9%, 보행자 약 1.9%, 이륜차 약 0.9%로 전체 mAP는 약 1.5% 더 나은 성능을 달성하였다. Apollo scape 데이터 셋에서는 VariFocal CenterPoint가 기존의 CenterPoint 보다 차량은 약 1.9%, 보행자 약 0.4%, 이륜차 약 0.6% 높아서 전체 mAP는 약 0.6% 정도로 더 높은 성능을 보였다. H3D 데이터 셋에서는 7개 클래스로 세분화

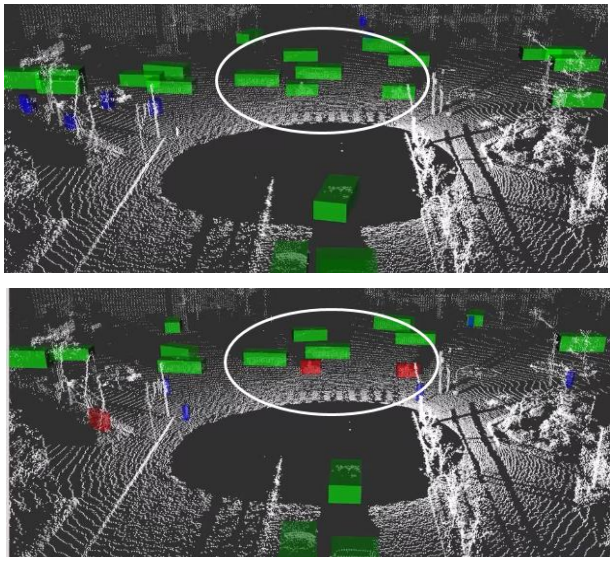

그림 3. CenterPoint와 VariFocal CenterPoint의 정성적 결과 예시
Fig. 3. Example of qualitative results: CenterPoint and VariFocal CenterPoint

하여 평가하였으며, VariFocal CenterPoint가 기존의 CenterPoint 보다 버스와 트럭을 제외한 나머지 클래스에서 더 높은 성능을 보였으며, 차량 약 1%, 보행자 약 0.1%, 이륜차 약 3%로 전체 mAP는 약 0.7% 더 나은 성능을 달성하였다. 국내에서 수집한 자체 데이터 셋인 A3D도 멀티클래스로 세분화하여 평가하였으며, VariFocal CenterPoint의 검출 성능은 차와 트럭을 제외한 나머지 클래스에서 더 높은 성능을 보였다. 차량은 약 1.5%, 보행자 약 0.03%, 이륜차 약 1% 높아서 전체 mAP는 약 0.5% 정도로 더 높은 성능을 보였다. 실시간 검출 속도 차이는 1ms 이내로 두 모델이 거의 유사하였다. 이를 통해 제안한 모델인 VariFocal CenterPoint가 모든 데이터 셋에서 평균적으로 BEV 객체 검출 성능이 약 0.9% 더 우수한 것을 확인하였다.

그림 3은 A3D 데이터 셋으로 학습한 VariFocal CenterPoint와 CenterPoint 모델을 정성적으로 평가한 예시이다. 그림에서와 같이 흰색 원으로 표시된 부분을 보면, CenterPoint 모델은 이륜차를 오검출하는 반면 VariFocal CenterPoint 모델은 이륜차로 검출하고 있다. 제안하는 모델이 특히 보행자와 이륜차 등의 작은 객체를 잘 검출하는 것을 확인할 수 있었다. 자체 구축한 데이터 셋인 A3D는 현재 AI 허브에 공개되어 있다 [7].

Ⅳ. 결 론

이 논문에서는 보행자, 이륜차를 많이 포함한 국내에 도로 환경에 특화된 3D 라이다 데이터 셋인 A3D를 자체적으로 구축하고, 도로 위 객체를 잘 탐지하기 위해, 기존의 CenterPoint 모델의 focal loss를 수정하고, DIoU loss를 추가하여 loss를 재설계한 VariFocal CenterPoint 모델을 제안하였다. 이는 객체의 위치정보를 포함한 IoU를 함께 학습함으로써, 평가를 수행할 때 classification score와 localization estimation 간의 불일치 문제를 완화하여 모델의 검출 성능을 향상시켰다. 제안한 모델을 Waymo Open Dataset, H3D, ApolloScape, A3D 데이터 셋으로 학습하여 기존 모델과 비교 평가하였으며, 우리가 제안한 모델이 기존 모델에 비해 전체적으로 mAP가 약 1% 정도 객체 검출 성능이 우수한 것을 확인하였다.

References

[1] J. H. Lee, J. K. Lee, J. H. Park, J. S. Kim, S. Kwon, “Construction and Effectiveness Evaluation of Multi Camera Dataset Specialized for Autonomous Driving in Domestic Road Environment,” IEMEK J. Embed. Sys. Appl., Vol, 17, No. 5, pp. 273-280, 2022 (in Korean).

[2] T. Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, “Focal Loss for Dense Object Detection,” in International Conference on Computer Vision (ICCV), pp. 2980-2988, October 2017.

[3] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, “Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression,” in 34nd AAAI Conference on Artificial Intelligence, Vol. 34, No. 07, 2020.

[4] T. Yin, X. Zhou, P. Krähenbühl, “Center-based 3D Object Detection and Tracking,” in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 11784-11793, 2021.

[5] S. Wu, X. Li, X. Wang, “IoU-aware Single-stage Object Detector for Accurate Localization,” Image and Vision Computing, Vol. 97, 103911, 2020.

[6] P. Sun, H. Kretzschmar, 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 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2446-2454, 2020.

[7] A. Patil, S. Malla, H. Gang, Y. T. Chen, “The H3D Dataset for Full-Surround 3D Multi-Object Detection and Tracking in Crowded Urban Scenes,” International Conference on Robotics and Automation (ICRA), pp. 9552-9557, 2019.

[8] Y. Ma, X. Zhu, S. Zhang, R. Yang, W. Wang, “TrafficPredict: Trajectory Prediction for Heterogeneous Traffic-Agents,” in 33nd AAAI Conference on Artificial Intelligence, pp. 6120-6127, 2019.

[9] AI Hub, “A3D: Korean Autonomous Driving Dataset,” https://aihub.or.kr.

[10] Y. Zhou, O. Tuzel, “VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection,” in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4490-4499, 2018.

[11] W. Zheng, W. Tang, S. Chen, L. Jiang, “CIA-SSD:

Confident IoU-Aware Single-Stage Object Detector From Point Cloud," in 35nd AAAI Conference on Artificial Intelligence, Vol. 35, No. 4, pp. 3555-3562, 42021.
[12] Y. Hu, Z. Ding, R. Ge, W. Shao, W. Huang, "AFDetV2: Rethinking the Necessity of the Second Stage for Object Detection from Point Clouds," in 36nd AAAI Conference on Artificial Intelligence, 2022.
[13] A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, O. Beijbom, "PointPillars: Fast Encoders for Object Detection From Point Clouds," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 12697-12705, 2019.
[14] G. Shi, R. Li, C. Ma, "PillarNet: Real-Time and High-Performance Pillar-Based 3D Object Detection," in European Conference on Computer Vision (ECCV), pp. 35-52, 2022.
[15] X. Zhou, D. Wang, P. Krähenbühl, "Objects as Points," in arXiv cs.CV 1904.07850, 2019
[16] H. Zhang, Y. Wang, F. Dayoub, N. Sünderhauf, "VarifocalNet: An IoU-aware Dense Object Detector," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8514-8523, 2021.

Jin-Hee Lee (이 진 희)

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 (이 재 근)

2021 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)
2022~Artificial Intelligence from Kyungpook National University (KNU)
Field of Interests: Autonomous driving, 3D Vision, Deep learning
Email: lejk8104@gmail.com

Joohyun Lee (이 주 현)

2022 Computer Science and Engineering from Konkuk University (B.S)
2022~Interdisciplinary Engineering from Daegu Gyeongbuk Institute of Science & Technology (DGIST)
Field of Interests: Autonomous driving, 3D Vision, Deep learning
Email: jhlee0714@dgist.ac.kr

Je-Seok Kim (김 제 석)

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 (권 순)

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