

특집논문 (Special Paper)

방송공학회논문지 제29권 제6호, 2024년 11월 (JBE Vol.29, No.6, November 2024)

<https://doi.org/10.5909/JBE.2024.29.6.906>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

2D 객체 검출 정보와 대칭 지식을 통한 강인한 3D 객체 검출

오 영 민^{a)}, 김 정 옥^{b)†}

Robust 3D Object Detection Using 2D-Based Object Detection Information and Symmetry Knowledge

Youngmin Oh^{a)} and Jung Uk Kim^{b)†}

요 약

3D 객체 검출은 객체의 종류, 위치뿐만 아니라 깊이와 방향 등 실제 3D 공간상의 정보를 추정하는 연구로, 자율주행과 로봇 공학 등 다양한 응용 분야에서 중요한 역할을 한다. 그러나 LiDAR 센서의 한계로 인해 포인트 클라우드의 객체 일부 단면 정보만을 포함하게 되어 불완전한 검출 결과를 초래할 수 있다. 이를 해결하기 위해, 본 논문에서는 객체의 대칭 지식을 활용한 새로운 3D 객체 검출 네트워크를 제안한다. 우리는 객체 기반 가상의 대칭 포인트(object guided virtual symmetry point)를 도입하여, LiDAR 정보에 대칭 지식을 포함시키고 이를 네트워크가 학습하도록 설계하였다. 또한, 지식 증류 기법과 대칭 지식 가이드 손실(symmetry knowledge guide loss)을 통해 대칭 지식이 효과적으로 전달되도록 하였으며, 실험을 통해 기존 방법보다 향상된 3D 객체 검출 성능을 달성하여 제안된 방법의 우수성을 입증하였다.

Abstract

3D object detection involves estimating not only the category and position of objects but also critical spatial information such as depth and orientation in real-world 3D space. This research is vital for various applications, including autonomous driving and robotics. However, due to the limitations of LiDAR sensors, point clouds typically contain only partial cross-sectional information of objects, which can lead to incomplete detection results. To address this issue, we propose a novel 3D object detection network that leverages symmetry knowledge of objects. We introduce the concept of an object-guided virtual symmetry point to incorporate symmetry knowledge into the LiDAR data, allowing the network to learn this information effectively. Additionally, we employ knowledge distillation techniques and a symmetry knowledge guide loss to ensure the effective transfer of symmetry knowledge. Experimental results demonstrate that the proposed method achieves superior 3D object detection performance compared to existing methods, thus validating its effectiveness.

Keyword : 3D object detection, Symmetry knowledge, 2D object information

Copyright © 2024 Korean Institute of Broadcast and Media Engineers. All rights reserved.

“This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered.”

1. 서론

3D 객체 검출(3D object detection)은 객체의 종류와 위치뿐만 아니라 깊이와 방향과 같이 객체의 실제 3D 공간상의 정보를 추정하는 연구^[1-4]이다. 3D 객체 검출이 다루는 3D 공간은 실생활과 밀접하게 연결되어 있으며, 자율주행이나 로봇 공학과 같은 다양한 응용 분야^[5]에 활용될 수 있어 최근 활발한 연구가 진행되고 있다. 전통적인 3D 객체 검출 연구^[1-2]는 깊이 추정을 위해 LiDAR 센서로부터 얻은 포인트 클라우드(point cloud)를 활용하는 LiDAR 기반 방식이 주로 연구되고 있었으며, 포인트 클라우드는 객체의 깊이 정보를 포함하고 있어 단안 및 스테레오 카메라 기반의 방식과 달리 3D 객체 검출에서 높은 정확도를 제공한다. 최근 3D 객체 검출 연구^[3-4]의 경우 이러한 LiDAR 정보에 추가적으로 RGB 이미지를 통한 가상의 포인트(virtual point)를 생성하여 보다 정밀하게 검출하려는 시도가 이루어지고 있다.

그러나 포인트 클라우드 기반의 연구에서 활용되는 LiDAR 센서에는 구조적 한계가 존재한다. LiDAR 센서는 레이저를 송수신하여 반환되는 정보를 통해 포인트 클라우드를 생성하며, 이 과정에서 레이저가 객체를 투과하지 못해 그림 1과 같이 객체의 일부 단면의 정보만 포인트 클라우드에 포함되는 문제가 발생한다. 또한 LiDAR 센서가 제대로 반사되지 않거나 먼 거리의 경우 레이저 신호가 약해져 수집할 수 있는 포인트가 줄어들어 희소 포인트(sparse point) 문제가 발생한다. 이렇게 생성된 포인트 클라우드는 RGB 이미지를 기반으로 가상의 포인트를 추가하더라도,

기본적으로 기존 LiDAR 포인트에 의존하기 때문에 객체의 가려진 부분이나 객체의 세밀한 정보를 충분히 반영하지 못할 수 있다. 이는 특히 복잡한 자율주행과 같은 실생활 환경이나 먼 거리의 객체를 검출할 때 문제가 되며, 결과적으로 3D 객체 검출의 정확도에 부정적인 영향을 미친다.

본 논문에서는 앞서 언급한 문제를 해결하기 위해 객체의 대칭지식을 활용한 새로운 3D 객체 검출 네트워크를 제안한다. 3D 객체 검출이 주로 활용되는 자율주행 환경에서 차량은 가장 중요한 핵심 객체 중 하나이다. 따라서 이 연구에서는 주로 차량 검출의 정확도에 집중하고자 한다. 실제로 대부분의 차량의 경우 좌우로 대칭되는 구조를 가지며, 이는 차량이 어느 위치에 있든, 어느 방향을 향하든 변하지 않는 물리적 특성이다. 그러나 LiDAR 센서는 특정 방향에서만 반사된 레이저 정보를 수집하기 때문에 객체의 일부 단면에 대한 정보만 획득하게 되며, 이로 인해 검출기가 불완전한 정보만으로 전체 객체를 추정하여 일부 부정확한 검출 결과를 초래한다. 이 문제를 해결하기 위해 (1) 객체의 대칭 지식을 활용하는 방법과 (2) 대칭 지식을 효과적으로 검출기에 안내하는 방법을 제안한다.

먼저, 객체 기반 가상의 대칭 포인트(object guided virtual symmetry point)를 도입하여 포인트 클라우드 정보에 객체의 대칭 지식을 포함시킨다. 이를 통해 3D 객체 검출 네트워크가 이러한 정보를 입력으로 활용하여 학습할 수 있도록 한다. 이때 가상의 대칭 포인트는 기존 포인트 클라우드 정보에 추가되어, LiDAR 센서의 한계로 인한 정보 부족을 보완하고, 배경보다 객체에 집중하여 학습이 이루어질 수 있도록 안내할 수 있다.

다음으로 대칭 지식을 효과적으로 검출기에 안내하여 강인한 검출을 달성하기 위해, 서로 다른 포인트 클라우드를 함께 활용하여 학습할 수 있는 네트워크 구조를 제안한다. 이 네트워크는 기존의 포인트 클라우드와 객체 기반 가상의 대칭 포인트를 모두 입력으로 받아 객체를 검출하며, 지식 증류(Knowledge Distillation) 기법을 통해 새로운 객체의 대칭 지식이 사전 학습된 검출기에 전달된다. 이를 통해 기존의 부족한 포인트로 인한 검출 오류를 줄이고, 기존의 지식과 새로운 객체의 대칭 지식을 효과적으로 활용하여 강인한 3D 객체 검출을 수행할 수 있다.

a) 경희대학교 인공지능학과(Department of Artificial Intelligence, Kyung Hee University)

b) 경희대학교 컴퓨터공학부(Department of Computer Science and Engineering, Kyung Hee University)

‡ Corresponding Author : 김정욱(Jung Uk Kim)

E-mail: ju.kim@khu.ac.kr

Tel: +82-31-201-3768

ORCID: <http://orcid.org/0000-0003-4533-4875>

* 이 논문의 연구 결과 중 일부는 한국방송·미디어공학회 2024년 하계 학술대회에서 발표한 바 있음.

** 이 논문은 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.RS-2022-00155911, 인공지능융합혁신인재양성(경희대학교)).

· Manuscript September 4, 2024; Revised October 16, 2024; Accepted October 17, 2024.

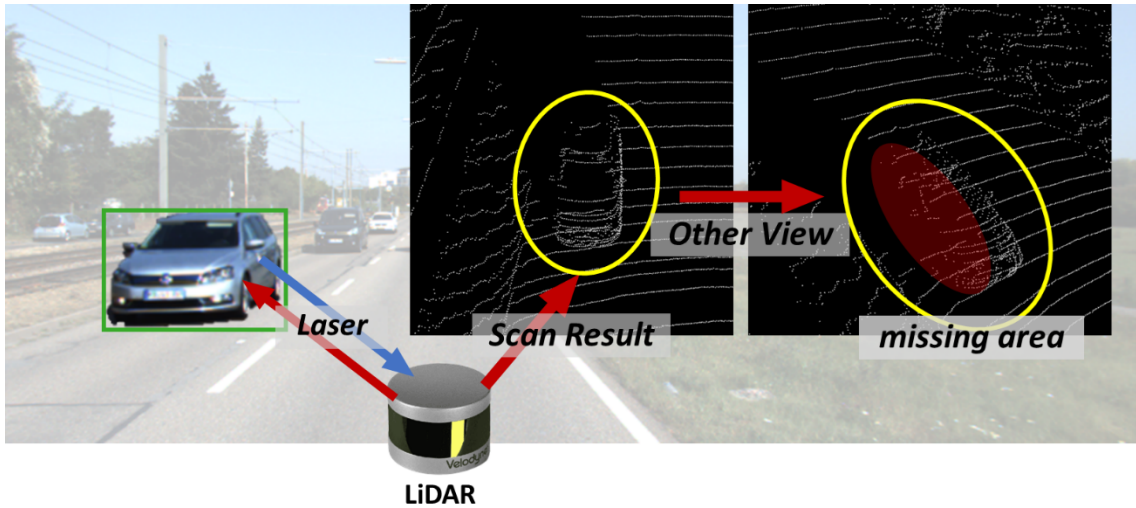


그림 1. 객체의 일부 단면의 정보만 획득하는 LiDAR 센서의 한계에 대한 예시
 Fig. 1. Example of the limitations of LiDAR sensors capturing only partial cross-sectional information of an object

II. 관련 연구

1. LiDAR 포인트 기반 3D 객체 검출 모델

LiDAR 포인트 기반의 3D 객체 검출은 깊이(depth) 정보를 포함할 수 있는 LiDAR 센서의 장점으로 인해 최근에도 활발히 연구되고 있다. 대표적인 모델 중 하나인 PV-RCNN^[1]은 3D voxel CNN(Convolutional Neural Network)과 PointNet^[6]의 장점을 결합하여 포인트 클라우드 특징을 효과적으로 학습하고 이를 통해 정밀한 3D 객체 검출을 달성한다. Voxel R-CNN^[2]은 3D 객체 검출을 위해 효과적인 복셀(voxel) 기반의 R-CNN 프레임워크를 제안한다. 이러한 연구들은 LiDAR 포인트 클라우드만을 활용하여 3D 공간에서의 객체 인식 능력을 극대화하는데 기여하고 있다.

2. 멀티모달 정보 기반 3D 객체 검출 모델

이러한 LiDAR 기반의 연구에도 불구하고, 여전히 LiDAR 센서의 한계인 포인트가 희소하다는 문제로 인해 한계가 존재한다. 따라서 MVP^[7]는 LiDAR 포인트뿐만 아니라 RGB 기반 가상 포인트를 추가하여 효과적으로 학습할 수 있도록 제안되었다. 이 방식은 RGB 정보를 활용해 LiDAR 포인트 주변에 가상 포인트를 생성함으로써, LiDAR 포인트의 희소

성 문제를 해결하였다. 그러나 가상 포인트를 그대로 활용할 경우, 생성 과정에서 발생한 다량의 노이즈가 검출에 악영향을 미칠 수 있다. 이를 해결하기 위해, VirConv^[4]는 2D로 투영된 포인트가 3D 공간보다 노이즈를 잘 구별할 수 있다는 주장을 기반으로 3D와 2D 공간에서 포인트를 학습할 수 있는 통합 네트워크를 제안하였고, 이를 통해 기존 검출기보다 우수한 검출 능력을 보였다. 그러나 VirConv 역시 LiDAR와 가상 포인트에 포함된 정보에만 의존해야 하며, 센서의 특성상 얻지 못한 정보들은 모델이 예측하는데 어려움을 주어 강력한 검출에 한계가 있다. 따라서 본 연구에서는 이러한 LiDAR의 한계를 극복하고, 효과적으로 학습할 수 있는 방법을 제안한다.

III. 제안 방법

그림 2는 제안하는 2D 객체 검출 정보와 대칭 지식을 통한 강인한 3D 객체 검출 네트워크의 학습 구조를 보여준다. 제안하는 검출기는 LiDAR 및 가상의 포인트가 포함된 포인트 클라우드(P)와 RGB 이미지 데이터(I)를 사용해 2D와 3D 공간에서의 정보를 활용한다. 여기서 포인트 클라우드 P 는 $P = \{p\}, p = [x, y, z, \alpha]$ 로 구성되며, x, y, z 는 포인트 좌표, α 는 해당 포인트의 강도를 의미한다. 먼저, 2D 검출기

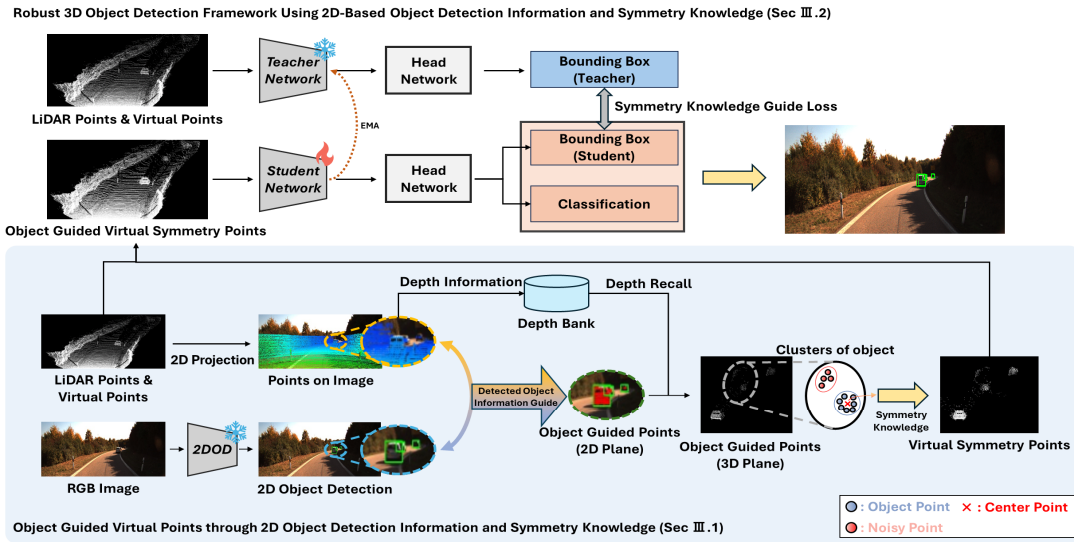


그림 2. 2D 객체 검출 정보와 대칭 지식을 통한 강인한 3D 객체 검출 구조
 Fig. 2. Architecture for robust 3D object detection using 2D-based object detection information and symmetry knowledge

를 사용해 RGB 이미지 I 에서 객체의 관심 영역을 2D Bounding-box로 추출한다. 이 2D Bounding-box는 포인트 클라우드에서 객체에 해당하는 포인트들을 선별하는 역할을 한다. 이후, 차량의 좌우 대칭 구조에 기반한 대칭 지식을 활용하여 LiDAR 센서로는 탐지되지 않은 영역의 정보를 포함하는 가상의 대칭 포인트(\hat{P})를 생성한다. 이렇게 생성된 객체 기반 가상의 대칭 포인트 \hat{P} 는 기존의 포인트 P 와 함께 제안하는 3D 객체 검출 네트워크의 입력으로 제공되어, 각 포인트 클라우드에 대한 3D Bounding-box(P_{bbox})를 예측한다. 제안하는 검출 네트워크는 기존의 포인트 클라우드를 통해 객체와 배경 정보를 학습하는 동시에, 객체 기반 가상의 대칭 포인트를 포함한 학습을 통해 기존에 놓쳤던 객체 정보를 보완한다. 이를 통해 객체 영역을 더욱 명확히 하고, 기존 포인트 클라우드만을 활용하는 검출 네트워크보다 정밀한 3D 객체 검출 성능을 구현한다.

1. 2D 객체 검출 정보와 대칭 지식을 통한 객체 기반 가상의 대칭 포인트 생성

제안하는 3D 객체 검출 네트워크의 주요 목적은 LiDAR 센서의 한계로 인해 손실되었던 객체의 가려진 정보를 활용하여 객체의 영역과 실제 3D 특성을 더욱 강조하여 학습할

수 있도록 하는 것이다. 이를 위해서는, 정확한 객체 검출과 함께 검출된 객체 정보에 대칭 특성을 지식으로 활용할 수 있는 과정이 필요하다. 더욱 정확한 객체 검출을 위해, 우리는 RGB 이미지를 활용한 2D 객체 검출 단계에서 비용 효율적이며 높은 정확도를 보이는 Deformable DETR^[8]을 도입한다. 이를 통해 2D 이미지에서 객체의 관심 영역을 나타내는 2D Bounding-box를 추출하고, LiDAR 포인트 클라우드에서 해당 객체가 위치한 포인트만을 선별할 수 있다. 특히, 주요 관심 대상인 차량은 어떤 위치나 방향에서도 좌우 대칭성을 유지하기 때문에, 이 특성을 활용하여 추출된 객체 포인트를 기반으로 가려진 영역에 가상 포인트를 생성할 수 있다. 이 과정은 객체 영역에 해당하는 포인트를 추출하는 단계와 추출된 포인트로부터 대칭 지식을 활용해 객체 기반 가상의 대칭 포인트를 생성하는 단계인 크게 두 단계로 구성된다.

LiDAR와 같은 포인트 클라우드는 3D LiDAR 좌표계에서 미터 단위로 표시되는 반면, RGB 이미지에서 얻어지는 검출 정보는 픽셀 단위로 표현된다. 이 두 정보를 통합하여 다루기 위해서는 좌표계를 변환하는 과정이 필요하다. 2D 객체 검출 기반 3D 객체 검출 네트워크^[9]는 포인트를 이미지 좌표계로 변환하고, 이를 2D 객체 검출 정보에 투영한다. 이 과정에서 생성된 2D 객체 검출 정보에 대응하는 포인트는 객체 기반 포인트(object guided point)로 정의되며,

이를 P^{2D} 로 표기한다. 이러한 포인트들은 2D 공간 내에 존재하게 되며, 3D에서 2D로 변환하는 과정에서 남는 깊이는 깊이를 은행(depth bank)에 저장된다. 이후 이 깊이를 2D 공간의 객체 기반 포인트를 다시 3D 공간인 LiDAR 좌표로 변환하는 데 활용된다. 이를 통해 우리는 객체에 해당하는 영역의 포인트 정보를 나타내는 3D 공간상의 객체 기반 포인트(P^{3D})를 얻을 수 있다.

앞서 3D 공간상의 객체 기반 포인트 P^{3D} 를 추출하였지만, 해당 포인트는 LiDAR 센서의 한계로 인해 객체의 가려진 영역이나 포인트의 희소성 문제를 완전히 해결하지는 못한다. 이를 극복하기 위해, 우리는 각 객체에 대한 포인트 클라우드에 대칭 지식을 적용하는 방법을 제안한다. 먼저, 검출된 각 객체에 대해 포인트들을 클러스터링하여, 해당 객체 내부에서 실제로 객체에 속하는 포인트(P_{object}^{3D})와 노이즈에 해당하는 포인트(P_{noise}^{3D})를 구분한다. 이를 위해 k-means 클러스터링 기법을 도입하여 객체 기반 포인트를 효과적으로 분할하고, 객체 클러스터에서 객체의 중심 좌표(p_{center})를 추출한다. 이 중심 좌표 p_{center} 를 기반으로 대칭 지식을 포인트 클라우드에 적용하여, 가상의 대칭 포인트 \hat{P} 를 생성한다. 생성된 가상 포인트는 기존에 LiDAR 센서로는 탐지되지 않았던 객체의 가려진 영역을 보완하는데 사용되며, 객체의 전체적인 형태에 대한 정보를 추가적으로 학습에 활용할 수 있게 된다.

2. 2D 객체 검출 정보와 대칭 지식을 통한 강인한 3D 객체 검출 네트워크

제안하는 3D 객체 검출 네트워크는 2D 객체 검출 정보와 대칭 지식을 활용하여, 기존의 포인트 클라우드와 객체 기반 가상의 대칭 포인트를 결합해 대칭 지식을 네트워크에 효과적으로 전달하는 것이 핵심이다. 이를 위해, 네트워크 구조는 기존의 포인트와 가상의 대칭 포인트를 모두 학습할 수 있는 teacher-student 네트워크 구조가 제안되었으며, 그림 2에서 확인할 수 있다. 먼저, student 네트워크는 사전 학습되지 않은 검출 네트워크로써 가상의 대칭 포인트에서 3D Bounding-box(p_{bbox}^s)와 class 정보를 예측하며, 3D 객체 검출^[2-3]에서 주로 사용되는 L_{OD} 를 통해 학습하게 된다. L_{OD} 는 아래와 같이 표기된다:

$$L_{OD} = L_{reg} + L_{cls}, \quad (1)$$

여기서 L_{reg} 와 L_{cls} 는 regression loss와 classification loss를 의미한다.

다음으로 teacher 네트워크는 사전 학습된 네트워크로써 기존의 포인트 클라우드 정보에서 3D Bounding-box(p_{bbox}^t)를 예측하게 된다. 앞서 student 네트워크에서 학습된 대칭 지식은 EMA(exponential moving average) 기법을 통해 teacher 네트워크에 전달되어 기존 검출 네트워크의 지식을 잃지 않고, 대칭 지식을 추가할 수 있다. 하지만 기존의 포인트와 가상의 대칭 포인트를 통한 학습 지식 간에는 차이가 존재하며, 우리는 이러한 차이를 모델이 이해하고 대칭 지식을 안정적으로 전달하기 위해 학습 정도를 조절할 수 있는 대칭 지식 가이드 손실(symmetry knowledge guide loss) L_{skg} 를 제안하며, 아래와 같이 계산된다:

$$L_{skg} = \|p_{bbox}^s - p_{bbox}^t\|_2^2. \quad (2)$$

마지막으로, 제안하는 2D 객체 검출 정보와 대칭 지식을 통한 강인한 3D 객체 검출 네트워크는 두 손실을 더해서 학습되며, 아래와 같이 정의된다:

$$L_{Total} = L_{OD} + L_{skg} \quad (3)$$

이를 통해 제안하는 네트워크는 학습 단계에서 teacher 모델의 부족한 지식을 지속적으로 파악하고 보완할 수 있으며, 기존의 지식에 대칭 지식을 통합하여 놓치고 있던 객체의 부족한 정보를 보완한다. 이는 객체의 일부 단면만을 사용하는 것이 아니라 온전한 정보를 네트워크에 전달할 수 있도록 하여, 강인한 3D 객체 검출을 실현할 수 있다.

IV. 실험 결과

1. KITTI 데이터셋

KITTI 데이터셋^[5]은 3D 객체 검출 연구에서 가장 널리 사용되며, 이미지뿐만 아니라 LiDAR 및 카메라 정보까지

표 1. KITTI 검증 데이터셋에서의 검출 결과

Table 1. Detection results on KITTI validation dataset

Method	Modality	Car 3D AP (R40)			Car BEV AP (R40)		
		Easy	Moderate	Hard	Easy	Moderate	Hard
PV-RCNN ^[1]	LiDAR	92.57	84.83	82.69	95.76	91.11	88.93
Voxel R-CNN ^[2]		92.38	85.29	82.86	95.52	91.25	88.99
Focals Conv ^[3]	LiDAR+RGB	92.86	85.85	85.29	95.45	91.51	91.21
VirConv ^[4]		93.05	88.00	85.60	95.94	91.86	89.28
Ours	LiDAR+RGB	95.32	88.25	85.64	96.15	91.89	91.32

표 2. 포인트 정보에 따른 차량 3D AP 검출 결과

Table 2. Car 3D AP detection results based on point information

Method	Input Point Cloud	Car 3D AP (R40)		
		Easy	Moderate	Hard
VirConv	LiDAR & Virtual Points	93.05	88.00	85.60
	Object Guided Points	83.55	73.64	69.82
	Object Guided Virtual Symmetry Points	93.11	88.05	85.62
Ours	-	95.32	88.25	85.64

표 3. 지식 증류 기법에 따른 차량 3D AP 검출 결과

Table 3. Car 3D AP detection results based on knowledge distillation

Method	Knowledge Distillation	Car 3D AP (R40)		
		Easy	Moderate	Hard
VirConv	-	93.05	88.00	85.60
Ours	Train Teacher with GT	93.07	86.15	83.44
	Train Teacher with EMA	95.32	88.25	85.64

제공한다. 본 논문에서는 기존 연구^[1-4]에 따라 7,481개의 학습 데이터와 3,769개의 검증 데이터를 통해 성능을 비교한다.

성능 평가를 위해서 우리는 3D average precision(AP)를 채택한다. 3D AP는 Precision-Recall 곡선 아래의 면적을 계산한 값으로, 기존 연구^[1-4]에 따라 40 Recall에 대한 3D AP (R40)와 bird-eye view에 대한 BEV AP (R40)을 사용하며, 자동차 객체에 대해 IoU 임계값 0.7에 대해 성능 평가를 진행한다. 우리는 객체의 크기와 가려짐 정도에 따라 세 가지 난이도(‘Easy’, ‘Moderate’, ‘Hard’)로 나눠서 비교를 진행한다. ‘Easy’ 난이도는 객체가 온전히 이미지 내부에 있고 가려지지 않은 상태, ‘Moderate’ 난이도는 객체가 일부 가려지거나 이미지 경계를 넘어간 상태, ‘Hard’ 난이도는 객체가 심하게 가려지거나 절반 이상이 이미지 경계를 벗어난 상태를 나타낸다. 제안하는 네트워크는 Adam 옵티마이저를 사용하여 60 epoch에 걸쳐 배치 크기 2의 단일 RTX 4090 GPU에서 훈련한다. 우리는 0.01의 학습률을 사

용하며, OpenPCDet toolbox^[10]를 활용하여 실험을 진행한다. 논문의 모든 실험 결과는 같은 환경에서 진행되었으며, 모두 KITTI 검증 데이터에서 진행된다.

2. KITTI 검증 데이터셋에서의 결과

본 논문에서는 KITTI의 검증 데이터셋에서의 차량 3D AP와 BEV AP 결과를 통해 성능 비교를 진행한다. 실험은 우리의 네트워크의 기반인 VirConv^[4] 모델 중 포인트 클라우드 정보만을 활용한 가장 기본 모델인 VirConv-L을 기준으로 진행하였다. 표 1에서 우리가 제안하는 방법의 경우 3D뿐만 아니라 BEV에서 모두 향상된 성능을 보이며, 특히 “Easy” 난이도에서 3D AP가 2.27%, “Hard” 난이도에서 BEV AP 2.04% 증가하였다. 이는 우리의 대칭 지식을 활용할 수 있는 네트워크가 기존의 LiDAR 포인트의 부족한 정보를 보충하며, 강인한 검출에 효과적으로 동작하고 있음을 보여준다.

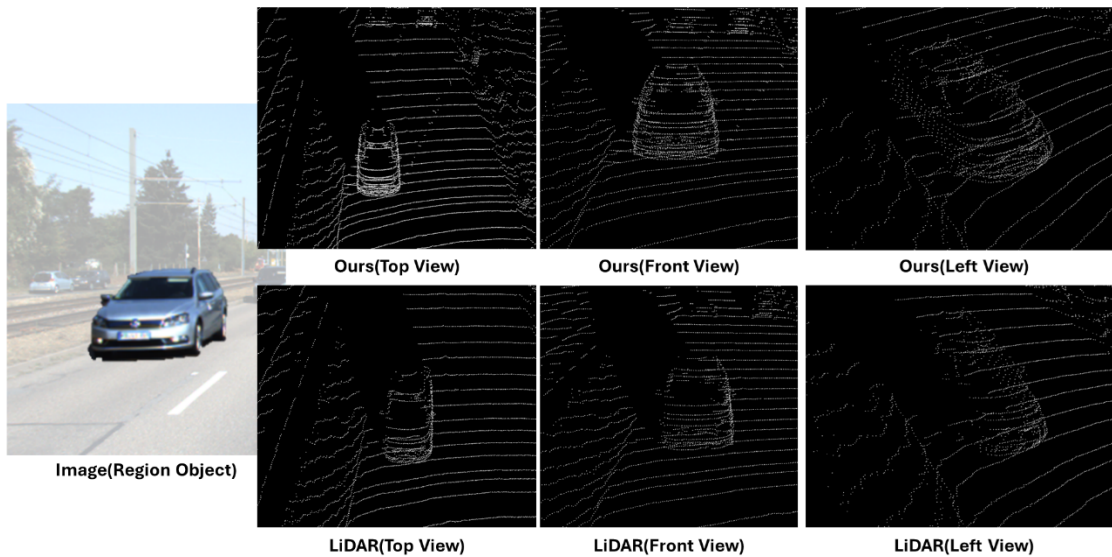


그림 3. 객체 기반 가상의 대칭 포인트 시각화
 Fig. 3. Object guided virtual symmetry points visualization

3. 객체 기반 가상의 대칭 포인트의 효과성

우리는 입력 포인트 정보의 변화에 따라 차량 3D AP 검출 성능 비교 실험을 진행하였다. 표 2에서 볼 수 있듯이, 대칭 지식이 포함되지 않은 단순한 객체 기반 포인트를 사용한 경우, 기존 포인트 클라우드를 활용한 검출보다 성능이 저하되었다. 반면, 우리가 제안하는 객체 기반 가상의 대칭 포인트만 활용한 경우, 모든 난이도에서 기존 방법보다 검출 성능이 향상되었다. 이를 통해 객체의 대칭 지식을 활용한 가상의 포인트가 LiDAR 센서의 한계를 극복하는데 효과적임을 확인할 수 있었으며, 가상의 포인트만 활용하는 것보다 우리가 제안하는 지식 증류 기법을 통한 학습 네트워크가 기존 검출기에서의 부족한 지식을 보완하며, 새로운 지식을 통한 효과적인 학습하여 더욱 강인한 3D 객체 검출 성능을 발휘할 수 있음을 보여준다.

4. 제안하는 3D 객체 검출 네트워크의 효과성

우리는 지식 증류 기법에 따라 차량 3D AP 검출 성능 비교 실험도 진행하였다. 표 3에서 GT를 통한 학습 방법은 teacher와 student 모델 모두 고정(freeze)시키지 않은 상태로 GT와 학습하는 방식으로, 서로 가중치를 공유하여 업데

이트된다. 이러한 방식은 기존의 포인트 클라우드와 대칭 포인트와 같이 서로 다른 정보를 모두 구별 없이 동일하게 학습하여 부족한 지식을 제대로 보완하지 못하고 오히려 난이도가 올라갈수록 검출에 악영향을 주게 된다. 하지만 EMA와 대칭 지식 가이드 손실 함수를 통한 학습은 기존의 지식에서 부족한 부분을 보완하며, 대칭 지식을 효과적으로 전달하여 모든 난이도에서 강인한 검출을 달성할 수 있음을 보여준다.

5. 객체 기반 가상의 대칭 포인트 시각화

그림 3은 제안하는 객체 기반 가상의 대칭 포인트의 시각화한 결과를 보여준다. 기존의 LiDAR 센서로 취득된 포인트와 비교했을 때, 제안하는 방법은 객체의 가려진 부분에 대해 실제 차량의 대칭성을 반영한 가상의 포인트를 생성함을 확인할 수 있다. 이러한 대칭 포인트는 성능뿐만 아니라 시각화 측면에서도, 기존의 LiDAR 센서와 비교해 볼 때 차량의 대칭 특성을 보다 세밀하게 반영하고 있다는 것을 보여준다.

추가로, 제안하는 객체 기반 가상의 대칭 포인트는 객체의 방향이나 위치와 무관하게 객체의 대칭 특성을 반영하여 생성된다는 것을 확인할 수 있다. 예시로, 그림 3에서

보여지는 차량은 LiDAR 센서의 위치 대비 정면에 존재하지 않지만, 온전한 대칭 포인트를 생성하는 것을 확인할 수 있다. 이는 제안하는 가상 포인트를 생성하는 과정이 3D 공간에서 진행되며, 노이즈를 구별할 수 있는 클러스터링 기법과 중심 좌표 기반의 포인트 생성 기술이 효과적으로 작동함을 보여준다.

V. 결론

본 논문에서는 LiDAR 센서의 한계로 인해 발생하는 3D 객체 검출의 문제를 해결하기 위해 객체의 대칭 지식을 활용한 새로운 3D 객체 검출 네트워크를 제안한다. 제안된 방법은 LiDAR 포인트 클라우드에서 누락된 객체의 가려진 정보를 보완하여, 더욱 강인한 3D 객체 검출을 가능하게 한다. 다양한 실험을 통해 제안된 방법의 효과성을 입증하였으며, 공식 데이터셋의 모든 난이도에서 기존 연구보다 향상된 결과를 보여주었다. 이를 통해 자율주행 등 다양한 실생활 응용 분야에서의 활발히 적용되는 것을 기대할 수 있다.

참고 문헌 (References)

- [1] S. Shi, C. Guo, L. Jiang, Z. Wang, J. Shi, X. Wang and, H. Li, "PV-RCNN: Point-Voxel Feature Set Abstraction for 3D Object Detection," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, USA, pp.10526-10535, 2020.
doi: <http://dx.doi.org/10.1109/cvpr42600.2020.01054>
- [2] J. Deng, S. Shi, P. Li, W. Zhou, Y. Zhang, and H. Li, "Voxel R-CNN: Towards High Performance Voxel-based 3D Object Detection," Proceedings of the AAAI Conference on Artificial Intelligence, pp. 1201 - 1209, 2021.
doi: <https://doi.org/10.1609/aaai.v35i2.16207>
- [3] Y. Chen, Y. Li, X. Zhang, J. Sun, and J. Jia, "Focal Sparse Convolutional Networks for 3D Object Detection," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, USA, pp.5428-5437, 2022.
doi: <https://doi.org/10.1109/cvpr52688.2022.00535>
- [4] H. Wu, C. Wen, S. Shi, X. Li, and C. Wang, "Virtual Sparse Convolution for Multimodal 3D Object Detection," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, Canada, pp.21653-21662, 2023.
doi: <http://dx.doi.org/10.1109/cvpr52729.2023.02074>
- [5] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Providence, USA, pp.3354-3361, 2012.
doi: <http://dx.doi.org/10.1109/cvpr.2012.6248074>
- [6] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space," Advances in Neural Information Processing Systems, pp.5105-5114, 2017.
- [7] T. Yin, X. Zhou, and P. Krähenbühl, "Multimodal Virtual Point 3D Detection," Advances in Neural Information Processing Systems, pp.16494-16507, 2021.
- [8] X. Zhu, W. Su, L. Lu, B. Li, X. Wang and J. Dai, "Deformable DETR: Deformable Transformers for End-to-End Object Detection," arXiv preprint arXiv:2010.04159, 2020.
doi: <https://doi.org/10.48550/arXiv.2010.04159>
- [9] Y. Oh, J. U. Kim and K. R. Park, "Effective 3D Object Detection using 2D Object Detection information," The Korean Institute of Broadcast and Media Engineers Summer Conference, 2023.
- [10] OpenPCDet: OpenPCDet Toolbox for LiDAR-based 3D Object Detection, <https://github.com/open-mmlab/OpenPCDet>.

저 자 소 개



오 영 민

- 2023년 2월 : 호서대학교 자동차CT공학과 학사
- 2023년 3월 ~ 현재 : 경희대학교 인공지능학과 석사과정
- ORCID : <https://orcid.org/0009-0003-8949-9927>
- 주관심분야 : 컴퓨터 비전, 3D 객체 검출, 멀티모달 AI

저 자 소 개



김 정 욱

- 2016년 2월 : 아주대학교 전자공학과 학사
- 2018년 2월 : KAIST 전기 및 전자공학부 석사
- 2022년 8월 : KAIST 전기 및 전자공학부 박사
- 2022년 9월 ~ 현재 : 경희대학교 컴퓨터공학부 교수
- ORCID : <https://orcid.org/0000-0003-4533-4875>
- 주관심분야 : 컴퓨터 비전, 머신러닝, 멀티모달 AI