악천후 환경에서 단안 3D 객체 검출 성능 향상을 위한 MonoDETR 기반 기법*

김민수○ 김승한 신민기 이서현 김정욱†

경희대학교 컴퓨터공학과
koreaminsoo@khu.ac.kr rlatmdgks4@naver.com mks5242@khu.ac.kr mgogfc5201@khu.ac.kr ju.kim@khu.ac.kr

A MonoDETR-Based Method for Improving Monocular 3D Object Detection in Adverse Weather Conditions

Kim Min Soo○, Kim Seung Han, Shin Min Ki, Lee Seo Hyun, Kim Jung Uk†
Department of Computer Science and Engineering, Kyung Hee University

요 약

자율주행 시스템의 실용성과 안전성을 확보하기 위해서는, 다양한 날씨 조건에서도 견고하게 작동할 수 있는 3D 객체 검출 기술이 필수적이다. 그러나 기존의 카메라 기반 객체 검출 모델은 대부분 맑은 날씨 데이터를 기반으로 학습되어, 안개나 비 등 악천후 환경에서는 성능이 급격히 저하되는 문제가 있다. 이를 해결하기 위해서는 일반적으로 고가의 장비가 요구되며, 상용화에는 제약이 따른다. 이에 본 연구에서는 비용 효율적인 대안으로 단일 카메라 기반의 3D 검출 모델인 MonoDETR을 활용하고, 다양한 학습 전략을 적용함으로써 악천후 상황에서의 인식 성능 개선 가능성을 실험적으로 검증한다.

1. 서 론

최근 자율주행 기술의 발전과 함께, 주변 환경을 정밀하게 인식할 수 있는 3D 객체 검출(3D object detection) 시스템의 중요성이 빠르게 커지고 있다. 그러나 실제 자율주행 환경은 비, 안개, 눈 등 다양한 악천후 조건에서 기존 객체 검출 모델의 성능이 급격히 저하되는 문제가 있다. 이러한 성능 저하는 자율주행의 안전성과 신뢰성을 심각하게 위협한다.

이러한 문제를 극복하기 위해, 라이다(LiDAR)나 스테레오 카메라(stereo camera)와 같은 고가의 장비를 활용한 연구가 진행되어 오고 있다. 하지만 비용과 시스템 복잡성 측면에서 상용화에는 제약이 존재한다.

이에 본 연구는 보다 현실적이고 경제적인 대안으로, 단안 카메라(monocular camera)를 활용한 3D 객체 검출 모델의 개선 가능성을 탐구한다. 특히 대표적인 단안 기반 모델인 MonoDETR[1]을 기반으로, 다양한 날씨 조건에서도 견고하게 작동할 수 있는 모델을 구축하고자 한다. 본 논문에서는 맑은 환경(KITTI dataset)[2]과 안개 환경(Foggy KITTI dataset)[3]에 대한 개별 학습을 수행하고, 파인튜닝(fine-tuning), 동시 학습(multi-domain), 티처-스튜던트 모델(teacher-student) 등 다양한 전략을 적용하여 악천후 상황에서의 검출 성능 향상을 목표로 한다.

2. 기존 연구

2.1 단안 카메라 기반 3D 객체 검출

단안 카메라 기반 3D 객체 검출(Monocular Camera-based 3D Object Detection)은 단일 카메라, 즉, 한 대의 카메라만을 사용하여 3D 공간에서 객체를 인식하고 검출하는 기술을 의미한다. 이는 비용과 구현 측면에서의 장점으로 인해 자율주행 분야에서 활발히 연구되고 있는 기술이다. 대표적인 기존 모델로는 SMOKE, MonoDLE, MonoFlex[4] 등이 있으며, 이들은 일반적으로 2D 이미지상의 객체 중심점(center) 근처의 시각적 특징을 이용하여 3D 위치, 크기, 방향 등을 회귀하는 center-guided detection paradigm을 따른다. 이러한 방식은 구조가 직관적이고 계산량이 적다는 장점이 있으나, 객체 간의 전역적인 깊이 관계를 반영하지 못하고, 국소적 정보에 의존하기 때문에 가려짐(occlusion)이나 원거리 객체에 대한 정확한 검출이 어렵다는 한계가 있다.

2.2 MonoDETR

MonoDETR은 기존 center-guided 방식의 한계를 보완하기 위해 제안된 단안 카메라 기반 3D 객체 검출 모델로, Transformer 기반의 구조를 채택하고 있다. MonoDETR은 DETR에서 사용하는 object query 개념을 도입하여, 객체를 anchor 없이 직접 예측할 수 있다. 또한 foreground depth map을 예측한 후 이를 활용한 depth-guided cross-attention을 통해 전역적인 장면 정보에 기반한 3D 검출이 가능하다. 특히, 별도의 NMS(Non-Maximum Suppression)나 anchor 없이 end-to-end로 학습되며, 물체가 얼마나 가까운지를 계산하는 depth encoder와 물체가 어떤 종류인지를 계산하는 visual encoder를 병렬로 구성

* “본 연구는 과학기술정보통신부 및 정보통신기획평가원의 2025년도 SW중심대학사업의 결과로 수행되었음 “(2023-0-00042)

하여 깊이(depth)와 시각 정보(appearance)를 효과적으로 통합한다. 이 구조는 복잡한 장면에서도 객체 간 깊이 관계를 학습할 수 있어 일반적인 center-guided 모델 대비 더 우수한 성능과 일반화 능력을 보인다.

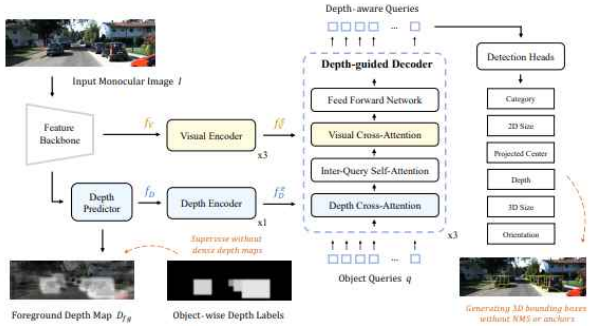


그림 1 MonoDETR의 전체 파이프라인

3. 문제 정의

기존의 단안 3D 객체 검출 모델들은 대부분 맑은 날씨 조건의 데이터셋을 기반으로 학습되어 왔다. 하지만 실제 자율주행 환경에서는 안개, 비, 야간 등 다양한 악천후 상황이 빈번하게 발생하며, 이로 인해 모델의 성능이 급격히 저하되는 문제가 있다.

실제로 본 연구에서 수행한 초기 실험 결과, 맑은 환경(KITTI)에서 학습된 모델은 안개 환경(Foggy KITTI)에서 현저한 성능 저하를 보였으며, 반대로 안개 환경에서 학습된 모델 역시 맑은 환경에서의 검출 성능이 낮게 나타났다. 표 1과 그림 2는 기존 단일 환경 기반 학습 방식이 날씨 조건 변화에 대한 일반화 능력이 부족하다는 점을 보여주는 결과이다.

표 1 단일 환경 학습 모델의 성능 결과표

상황	맑음			안개		
난이도	Easy	Mod.	Hard	Easy	Mod.	Hard
only Clear	21.1497	15.7086	12.7789	10.2880	6.3928	4.9368
only Foggy	4.7991	3.5814	2.8888	19.9485	13.8913	11.1456

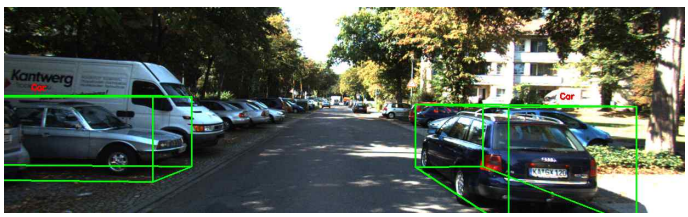


그림 2 안개 사진 학습 모델의 맑은 환경에서의 성능

따라서 다양한 기후 조건에서도 견고한 성능을 유지할 수 있도록 학습 데이터 구성 및 모델 학습 전략의 개선이 필요하다. 본 논문에서는 이를 해결하기 위해, 날씨 조건을 고려한 데이터 활용 및 구조적 대응 방안을 실험적으로 탐색한다.

4. 해결 방안

4.1 파인튜닝 모델

파인튜닝(Fine-tuning)은 기존에 일반 환경에서 학습된 모델을 바탕으로, 새로운 환경의 데이터에 대해 추가 학습을 수행하여 모델을 특화시키는 방법이다. 본 실험에서는 먼저 맑은 날씨 이미지(KITTI Dataset)를 기반으로 MonoDETR을 50 epoch 동안 학습한 후, 해당 가중치를 초기값으로 사용하여 안개 이미지(Foggy KITTI)를 추가로 50 epoch 학습하는 방식으로 파인튜닝을 수행하였다.

그림 3 파인튜닝 모델 구조도

표 2 파인튜닝 모델의 성능 결과표

상황	맑음			안개		
난이도	Easy	Mod.	Hard	Easy	Mod.	Hard
fine tuning	19.3353	12.6116	10.4234	24.3539	17.1487	13.8724

학습 결과, 단일 환경 학습 모델에 비해 전반적인 성능 향상이 관찰되었다. 특히 안개 환경에서의 성능은 3개 난이도(Easy, Moderate, Hard) 모두에서 기존 안개 환경으로만 학습한 모델보다 더 높은 정확도를 기록하였다.

하지만 이 과정에서 맑은 환경에서 학습했던 내용이 일부 덮여서 성능이 저하되는 현상인 망각(forgetting)[5] 문제가 발생할 수 있음도 관찰되었다. 안개 환경에 특화된 학습이 진행될수록 맑은 날씨 조건에서의 정확도가 소폭 하락하는 경향이 나타났다.

4.2 동시 학습 모델

동시 학습(Multi-domain)은 서로 다른 환경의 데이터를 하나의 모델에서 동시에 학습시키는 전략으로, 모델이 다양한 도메인에 대해 일반화된 표현을 학습하도록 유도하는 방법이다. 본 실험에서는 내부 코드를 수정하여 하나의 MonoDETR 모델에 맑은 날씨(KITTI) 이미지와 안개 날씨(Foggy KITTI) 이미지를 함께 입력하고, 두 환경을 동시에 학습하는 방식으로 모델을 구성하였다.

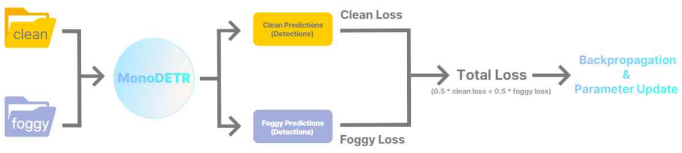


그림 4 동시 학습 모델 구조도

표 3 동시 학습 모델의 성능 결과표

상황	맑음			안개		
난이도	Easy	Mod.	Hard	Easy	Mod.	Hard
multi domain	19.2354	14.6725	11.9970	20.4995	15.3475	12.4296

학습 결과, 맑은 환경에서의 성능은 단일 환경 학습이나 파인튜닝 모델 대비 소폭 상승하는 경향을 보였다. 그러나 안개 환경에서는 파인튜닝 모델 대비 상대적으로

낮은 성능을 기록하였다.

4.3 티처-스튜던트 모델

티처-스튜던트(Teacher-Student)[6] 학습은 고성능의 사전 학습된 모델(Teacher)로부터 예측값(soft label)을 전달받아, 저성능 모델(Student)이 이를 모방하며 학습하는 지식 증류(Knowledge Distillation) 기법이다. 본 실험에서는 동시학습 기반 구조를 유지하면서, 안개 환경에서의 성능 부족을 보완하기 위해 해당 기법을 적용하였다.

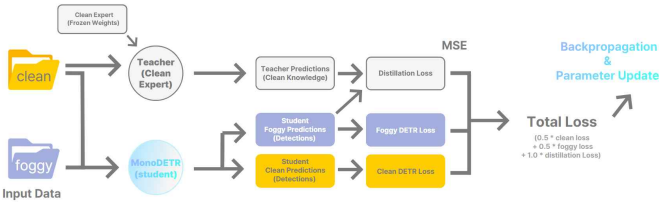


그림 5 티처-스튜던트 모델 구조도

먼저, 맑은 환경(KITTI)에서 학습된 MonoDETR 모델을 Teacher로 설정하고, Student 모델이 Clean과 Foggy 이미지를 모두 처리하도록 하였다. 이때 Teacher는 Clean 이미지에 대한 예측값을 생성하여 Student의 Foggy 예측에 대한 지도 신호로 사용하였다. 학습은 다음과 같은 손실 함수 조합으로 이루어졌다.

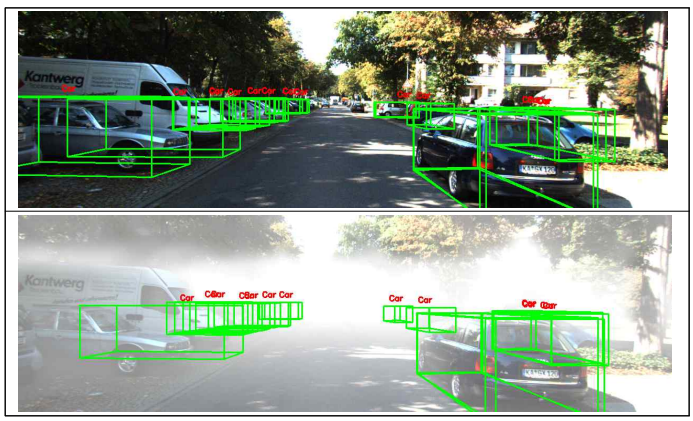
$$\text{Total Loss} = 0.5 \times \text{Clean Loss} + 0.5 \times \text{Foggy Loss} + 1.0 \times \text{Distillation Loss}$$

이때 Distillation Loss는 Teacher와 Student 간의 예측값 차이를 기반으로 계산된 Multi-Layer MSE 손실을 사용하였다. 구체적으로 DETR의 중간 레이어들(auxiliary layers)과 최종 레이어의 classification logits 및 bounding box 예측값에 대해 각각 MSE를 계산하여 합산하였다. 이를 통해 맑은 환경에서의 일반화된 표현을 안개 환경 학습에 효과적으로 전이시키고자 하였다.

표 5 티처-스튜던트 모델의 성능 결과표

상황	맑음			안개		
난이도	Easy	Mod.	Hard	Easy	Mod.	Hard
teacher	21.0327	16.1129	13.4071	22.6718	16.2694	13.5120
student						

표 6 티처-스튜던트 모델의 시각화 결과



학습 결과, 전체적으로 성능 향상이 관찰되었으며, 특히 기대했던 안개 환경에서의 정확도 개선이 확인되었다.

5. 결론 및 향후 연구

본 연구에서는 단안 카메라(monocular camera)를 기반으로 한 3D 객체 검출 모델인 MonoDETR을 활용하여, 악천후 환경에서의 성능 저하 문제를 해결하기 위한 다양한 학습 전략을 실험적으로 탐색하였다. 먼저, 맑은 환경에서 학습된 모델이 안개 환경에서 성능이 크게 저하되고, 그 반대의 경우에도 성능 하락이 발생하는 것을 통해 단일 도메인 학습의 한계를 확인하였다.

이에 대한 대응 전략으로 파인튜닝(fine-tuning), 동시학습(multi-domain training), 그리고 티처-스튜던트(teacher-student) 구조를 적용하였다. 파인튜닝은 안개 환경에 대한 높은 적응력을 보여주었으나, 기존 학습 내용의 일부가 손실되는 망각 문제를 동반하였다. 동시학습은 두 도메인을 동시에 고려함으로써 맑은 날 환경에서는 일정 수준의 성능 향상을 이끌어냈지만, 안개 환경에서는 상대적으로 경쟁력이 부족한 결과를 보였다. 마지막으로 티처-스튜던트 모델은 맑은 날 환경에서 학습된 고성능 모델의 정보를 활용하여 안개 환경에서의 인식 성능을 보완할 수 있음을 실험적으로 확인하였다.

향후 연구에서는 눈, 야간 등 더 다양한 악천후 조건을 포함한 실험을 통해 모델의 범용성과 일반화 성능을 검증하려 한다.

참 고 문 헌

[1] Zhang, Renrui, et al. “Monodetr: Depth-guided transformer for monocular 3d object detection.“ Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

[2] Geiger, Andreas, Philip Lenz, and Raquel Urtasun. “Are we ready for autonomous driving? the kitti vision benchmark suite.“ 2012 IEEE conference on computer vision and pattern recognition. IEEE, 2012.

[3] Oh, Youngmin, et al. “MonoWAD: Weather-Adaptive Diffusion Model for Robust Monocular 3D Object Detection.“ European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024.

[4] Zhang, Yunpeng, Jiwen Lu, and Jie Zhou. “Objects are different: Flexible monocular 3d object detection.“ Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

[5] Goodfellow, Ian J., et al. “An empirical investigation of catastrophic forgetting in gradient-based neural networks.“ arXiv preprint arXiv:1312.6211 (2013).

[6] Hinton, Geoffrey, Oriol Vinyals, and Jeff Dean. “Distilling the knowledge in a neural network.“ arXiv preprint arXiv:1503.02531 (2015).