JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 29 No. 10, pp. 35-45, October 2024
https://doi.org/10.9708/jksci.2024.29.10.035

3D Object Detection via Multi-Scale Feature Knowledge Distillation

Se-Gwon Cheon*, Hyuk-Jin Shin*, Seung-Hwan Bae**

*M. S. candidate, Vision & Learning Lab, Dept. of Electrical and Computer Engineering, Inha University, Incheon, Korea
**Associate Professor, Vision & Learning Lab, Dept. of Electrical and Computer Engineering, Inha University, Incheon, Korea

[Abstract]

In this paper, we propose Multi-Scale Feature Knowledge Distillation for 3D Object Detection (M3KD), which extracting knowledge from the teacher model, and transfer to the student model consider with multi-scale feature map. To achieve this, we minimize L2 loss between feature maps at each pyramid level of the student model with the correspond teacher model so student model can mimic the teacher model backbone information which improves the overall accuracy of the student model. We apply the class logits knowledge distillation used in the image classification task, by allowing student model mimic the classification logits of the teacher model, to guide the student model to improve the detection accuracy. In KITTI (Karlsruhe Institute of Technology and Toyota Technological Institute) dataset, our M3KD (Multi-Scale Feature Knowledge Distillation for 3D Object Detection) student model achieves 30% inference speed improvement compared to the teacher model. Additionally, our method achieved an average improvement of 1.08% in 3D mean Average Precision (mAP) across all classes and difficulty levels compared to the baseline student model. Furthermore, when integrated with the latest knowledge distillation methods such as PKD and SemCKD, our approach achieved an additional 0.42% and 0.52% improvement in 3D mAP, respectively, further enhancing performance.

▸ **Key words**: Knowledge distillation, Multi-scale feature map, Model compression, 3D object detection, Deep learning

• First Author: Se-Gwon Cheon, Corresponding Author: Seung-Hwan Bae
*Se-Gwon Cheon (segwon1000@inha.edu), Vision & Learning Lab, Dept. of Electrical and Computer Engineering, Inha University
*Hyuk-Jin Shin (shin0528@inha.edu), Vision & Learning Lab, Dept. of Electrical and Computer Engineering, Inha University
**Seung-Hwan Bae (shbae@inha.ac.kr), Vision & Learning Lab, Dept. of Electrical and Computer Engineering, Inha University
• Received: 2024. 07. 16, Revised: 2024. 10. 04, Accepted: 2024. 10. 04.

http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945

[요 약]

본 연구에서는 모델의 경량화를 위해 교사 모델의 출력 특징맵에서 3D 객체의 정보를 추출해 학생 모델의 다중 스케일 특징맵(Multi-scale feature map)에 맞게 증류하는 3D 객체 검출용 다중 스케일 특징 지식 증류 기법인 M3KD (Multi-Scale Feature Knowledge Distillation for 3D Object Detection)를 제안한다. M3KD는 지식 증류 수행 시 학생 모델과 교사 모델의 다중 스케일 특징맵 들 간 L2 손실(loss)을 사용해 특징맵 값의 차이를 줄이게 함으로써 학생 모델이 교사 모델의 백본을 모방하게 하여 학생 모델의 전체적인 정확도를 향상시키고, 기존의 이미지 분류 태스크(Task)에서 사용하는 클래스 로짓(Logits) 지식 증류를 적용해 교사 모델의 클래스 분류 로짓을 모방함으로써 학생 모델의 검출 정확도를 향상시킨다. 본 연구가 제안한 M3KD의 효과를 증명하기 위해 KITTI (Karlsruhe Institute of Technology and Toyota Technological Institute) 데이터 셋에서 실험을 진행하였으며, 이때 학습한 학생 모델이 교사 모델 대비 30%의 추론 속도 향상을 달성하였다. 또한, 정확도에서 기존의 학생 모델과 비교시 모든 클래스 및 모든 난이도에서 평균적으로 1.08%의 3D mAP (Mean Average Precision) 향상이 있음을 확인하였다. 또한 최신 지식 증류 기법인 PKD, SemCKD에 제안하는 기법을 추가로 적용하였을 시 기존 대비 0.42%, 0.52% 높은 정확도(3D mAP)를 나타내 성능 향상을 달성하였다.

▸ **주제어**: 지식 증류, 다중 스케일 특징맵, 모델 압축, 3차원 객체 검출, 딥러닝

I. Introduction

3D 객체 검출 모델(3D object detection model)은 자율주행 자동차에 있어서 핵심이 되는 기술이다 [1]. 3D 객체 검출 모델의 경우 일반적인 객체 검출 모델과 비교해 추가적인 3D 정보를 추출해야 하므로 많은 학습 파라미터(Parameter)를 가지고 있는 무거운 백본(Backbone)에서 좋은 성능을 나타낸다 [2]. 이러한 이유로 3D 객체 검출 모델의 크기는 증가하는 추세에 있으며, 따라서 필요한 연산 비용(Computational cost) 역시 증가하고 있다. 그러나 자율주행 자동차와 같은 실시간 태스크의 경우 높은 정확도를 유지하면서, 추론 속도(Inference speed)를 빠르게 해주는 모델 경량화 기술이 필수적이다.

경량화 기술 중 지식 증류는 추론 속도는 느리지만 정확도가 높은 교사 모델(Teacher model)에서 지식을 추출하고, 추출된 지식을 추론 속도가 빠르지만, 상대적으로 정확도가 낮은 학생 모델(Student model)로 전이함으로써 정확도를 높게 유지하면서 동시에 추론 속도가 빠른 학생 모델을 만드는 것을 목표로 한다. 기존의 카메라 기반 3D 객체 검출 모델로의 지식 증류 방법의 경우 라이다(LiDAR) 기반 모델을 교사 모델로 사용한 지식 증류 [3, 4]는 그 수가 많지만, 이 방법의 경우 모델의 경량화에 초점이 맞추기보다, 정확한 교사로부터 지식을 추출해 학생 모델에 증류해 정확도를 향상시키는 것에 초점이 맞춰 연구가 진행돼 경량화 측면에서는 그 연구에 부족함이 존재한다. 그중 특히 1개의 이미지만으로 추론하는 카메라(Camera) 3D 객체 검출에서의 경량화 초점 지식 증류 [5]의 경우 교사 모델의 지식을 추출할 때 모델의 복잡한 구조에 맞게 지식을 뽑아내는 데 어려움이 존재해 아직 그 수가 적다.

본 연구에서는 모델 경량화에 초점을 맞춘 단안 카메라(Monocular camera) 기반 3D 객체 검출 모델 간 지식 증류 기법인 M3KD (Multi-Scale Feature Knowledge Distillation for 3D Object Detection)를 제안한다. 제안한 기법을 위해 본 연구는 무거운 교사 모델의 백본을 가벼운 백본으로 대체해 학생 모델을 만듦으로써 모델을 경량화한다. 이후 학생 모델에 지식을 증류하는 과정에서 L2 손실(Loss) [6]을 사용해 교사 모델의 지식을 다중 스케일(multi-scale)의 특징맵에 맞도록 학생 모델에 증류한다.

본 연구는 M3KD 기법을 통해 단안 카메라 3D 객체 검출 모델간 지식 증류로 경량화된 학생 모델에게 교사 모델과 같은 높은 정확도를 달성하고자 한다. 3D 객체 검출에 사용되는 데이터 세트(Dataset)의 이미지와 레이블(Label)를 사용해 크기가 작은 학생 모델을 학습할 경우 학생 모델의 작은 백본은 3D 객체 지식을 학습하기 어려워 정확도가 낮아지는 문제를 확인했다. 본 연구는 상기한 문제점을 해

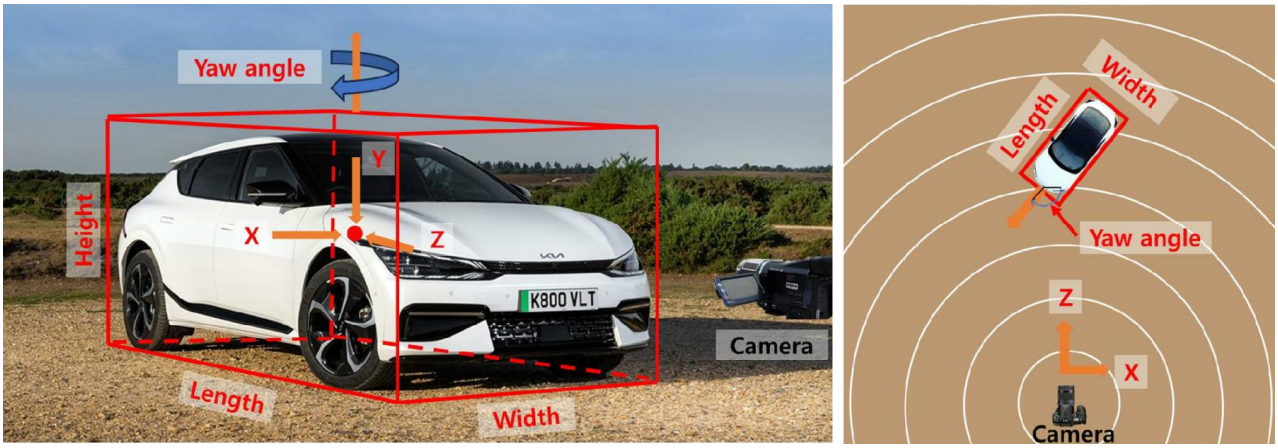

Fig. 1. 3D Object Detection Data Format on 3D Space (Left) & Top view (Right)

결하기 위해 아래의 두 가지 기법을 제안한다. 첫 번째로 학생 모델의 학습을 돕기 위해 정확도가 높은 교사 모델에서 다중 스케일 특징맵을 추출해 학생 모델의 다중 스케일 특징맵과의 L2 노름(Norm)을 계산한 다음 해당 차이를 줄이도록 해 학습을 돕는 다중 스케일 특징맵 지식 증류 기법을 제안한다. 이 기법은 각 스케일에 맞는 지식을 교사 모델에서 학생 모델로 전이하여 줌으로써 학생 모델의 전체적인 검출 정확도 향상에 기여한다. 두 번째로, 더욱 세밀한 정보를 추출하기 위해 교사 모델의 클래스 분류 로짓(Class classification logits)을 추출해 학생 모델 로짓과의 쿨백-라이블러 발산(Kullback-Leibler divergence, KLD)을 측정하고, 차이를 줄여주는 힌튼 지식 증류(Hinton KD) [7]를 클래스 분류에 적용한 클래스 KLD 지식 증류 기법을 제안한다. 이 기법은 교사 모델에서 객체 정보를 추출함으로써 학생 모델의 클래스 정확도 향상에 기여한다.

본 연구의 성능 평가를 증명하기 위해 KITTI [8] 데이터셋을 사용하여 실험을 수행한 결과, M3KD 지식 증류 기법을 사용한 학생 모델의 경우 추론 속도가 교사 모델 대비 30% 가속됨과 동시에 전체적인 클래스에서 정확도(3D mAP (Mean Average Precision))가 유지되고, 학생모델과 비교시 1.08%의 3D mAP 향상이 있음을 확인하였다. 또한 기존의 최신 지식 증류 기법인 PKD [9] 및 SemCKD [10]에 우리 기법을 추가로 적용하였을 시 기존 대비 0.42%, 0.52% 높은 3D mAP 정확도를 나타내어 성능 향상을 달성하였다.

본 연구의 주요 기여사항은 다음과 같다. i) 교사 모델과 학생 모델 간의 다중 스케일 특징맵 정렬을 통해 향상된 정확도를 가지게 하는 지식 증류 기법 제안한다. ii) 클래스 로짓을 사용한 객체 검출 정확도 향상시키는 지식 증류 기법을 제안한다. 본 논문의 구성은 다음과 같다. 2장에서는 기존의 3D 객체 검출과 지식 증류 기법에 대한 설명을 수행한다. 3장에서는 본 논문이 제안하는 기법인 다중 스케일 특징맵 지식 증류 기법과 클래스 KLD 지식 증류 기법에 대해 설명한다. 4장에서는 본 논문의 기법을 적용하였을 때의 3D 객체 검출 모델의 성능과 기존의 지식 증류 기법을 적용한 모델의 성능 비교 실험 및 각 손실을 적용하였을 때의 절제 실험과 검출 결과를 시각화한 정성적 실험을 진행한다. 5장에서는 결론 및 향후 연구과제에 관해 설명한다.

II. Related Works

1. 3D Object Detection

3D 객체 검출은 주어진 클라우드 포인트 또는 이미지와 같은 데이터에서 기존의 2D 객체 검출 [11-13]에서 예측하는 2D 정보 이외에 3D 정보를 추정하는 태스크이다. 3D 객체의 정보는 Fig. 1에서 확인할 수 있다. Fig. 1의 왼쪽 이미지에서 표현하고 있는 3D 객체 검출에서 추정해야 하는 값은 관찰자(카메라)의 현재 위치로부터 객체의 바운딩 박스(Bounding box) 중심까지의 가로, 세로, 깊이(X, Y, Z) 거리 그리고 객체 바운딩 박스의 길이, 너비, 높이(length, width, height), 또 객체의 편주각(Yaw angle) 값들을 가진다. 추가적으로, Fig. 1의 오른쪽에 표현하고 있는 방식을 탑 뷰(Top View (TV))라고 한다. 탑 뷰의 경우, 거리 정보들을 보다 표현하기 쉽도록 위에서 내려다보는 시점을 표현하며, 기존의 3D 정보에서 높이 정보에 해당하는 세로(Y)와 높이(Height) 값을 제외한 나머지 값을 표현 및 추정한다.

3D 객체 검출의 경우 주어지는 입력 데이터에 따라 크게 라이다 기반 3D 객체 검출 [14]과 카메라 기반 3D 객체 검출 [15, 16] 두 가지로 분류된다. 본 연구에서 사용하

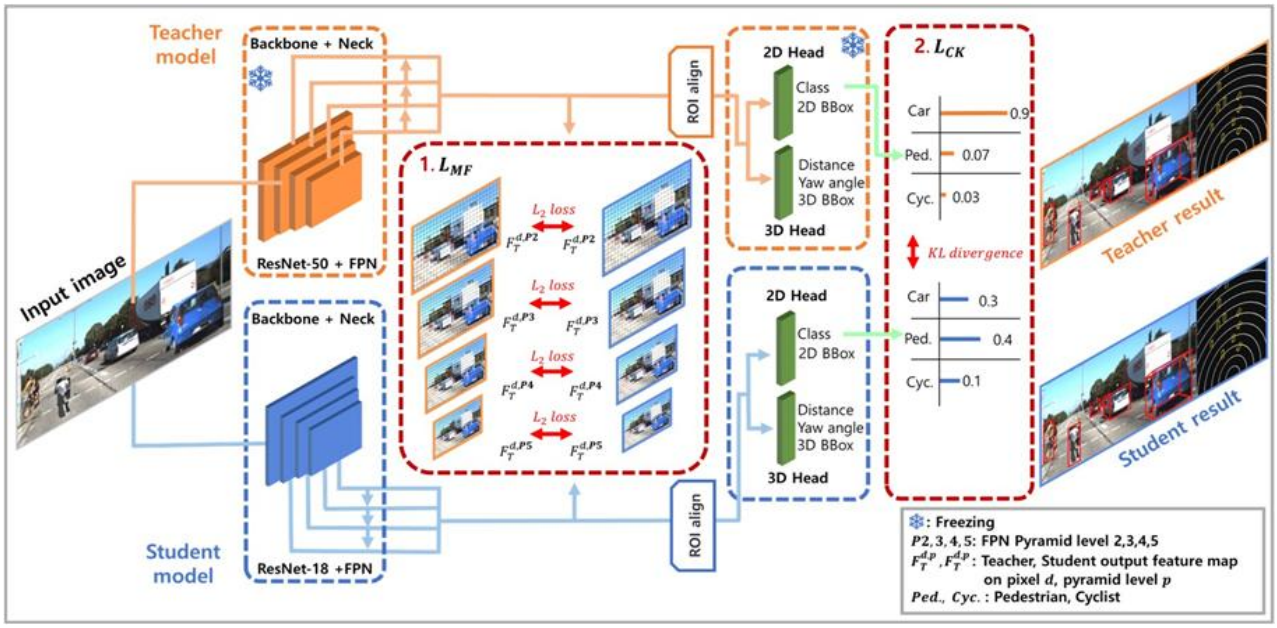

Fig. 2. Multi-Scale Feature Knowledge Distillation for 3D Object Detection (M3KD) Framework

는 단안 카메라 기반 3D 객체 검출의 경우 한 장의 이미지를 입력받아 3D 정보를 예측하는 태스크이다. 카메라 기반 3D 객체 검출 모델 중 M3d-rpn [15]은 3D 객체 검출 모델이 3D 정보를 사용하지 않아 정확도가 떨어진다는 문제를 지적하고 추가적인 거리별 컨볼루션 레이어(Convolution layer)를 사용해 정확도를 높이는 방법을 제안하였다. MonoRCNN [16]은 이미지에서 보이는 자동차의 높이가 가장 강건하고 추론하기 쉬운 거리 정보를 담고 있다고 제안하며 이미지의 높이로부터 거리를 추정하는 3D 객체 검출 모델을 제안하였다. 이와 같이, 카메라 기반 3D 객체 검출의 경우는 3D 정보를 보다 정확하게 유추해 정확도를 향상시키는 것에 초점을 맞추어 연구가 진행되고 있으며 이를 위해 더 좋은 특징을 추출할 수 있는 무거운 백본을 사용하는 추세에 있고 연산 비용 역시 증가하고 있다. 본 연구는 무거워지는 백본을 경량화하기 위한 기법인 M3KD를 제안한다. M3KD는 단안 카메라 기반 3D 객체 검출 기법인 MonoRCNN [16]을 교사 모델로 하여 교사 모델의 무거운 백본을 가벼운 백본으로 대체하고, 지식 증류에 M3KD 손실을 사용하여 빠른 추론 시간을 가지면서 높은 정확도를 동시에 가지는 효율적인 모델을 만드는 기법을 제안한다.

2. Knowledge Distillation

지식 증류는 고성능이지만 추론속도가 느린 교사 모델에서 지식을 추출한 다음 저성능이지만 추론속도가 빠른 학생 모델로 증류하여 추론속도가 빠르고 동시에 고성능인 학생 모델을 학습하는 방법이다. 지식 증류를 처음 제안한 힌튼은 교사 모델의 출력값을 소프트 레이블(Soft label)로 하여 학생 모델이 모방하는 지식 증류 [7] 방법을 제안하였다. 최근 제안된 지식 증류 방법으로 PKD [9]는 피어슨 상관 계수(Pearson correlation coefficient)를 활용해 교사와 학생 모델 간의 특징맵 간 선형 상관 관계를 측정하고 측정된 값을 모방하여 성능과 효율성이 향상된 객체 검출기를 위한 새로운 지식 증류 프레임워크(Framework)를 제안함으로써 특징맵 간의 구조 정보 및 관계 전달을 향상시켜 우수한 검출 정확도를 달성하였다.

SemCKD [10]는 교차 계층(Cross layer) 상호 작용(Interaction)을 활용하여 교사와 학생 모델 간의 특징 표현을 정렬하여 지식 전달을 원활하게 하여 모델 성능을 향상시키는 방법을 제안하였다. 이 방법은 학생의 특징이 교사의 특징과 의미론적으로 일치하도록 보장함으로써 특징맵의 의미론적 불일치 문제를 효과적으로 해결하였다. 카메라 기반 3D 객체 검출 모델로의 지식 증류 [3-5]는 활발히 제안됐지만 대부분 라이다 교사 모델의 정확한 지식을 증류해 학생 모델의 정확도를 올리려는 것에 그 초점이 맞추어져 있으며, 경량화에 초점이 맞춰진 기법은 그 수가 적었다. 경량화에 초점이 맞춰진 FD3D [5]의 경우, 교사 모델 외에 지식 증류 전용 헤드(Head)를 추가하고 추가적인 어텐션(Attention) 모델을 사용하는 등 지식 증류 과정이 복잡하다는 단점이 존재한다.

본 연구에서 제안하는 M3KD 기법은 학생 모델을 가벼운 모델로 들어 추론 속도를 향상시키고, 별도의 추가적인 지식 증류 전용 헤드 없이 간편하게 지식을 추출해 증류할 수 있는 경량화 초점 지식 증류 기법을 제안한다.

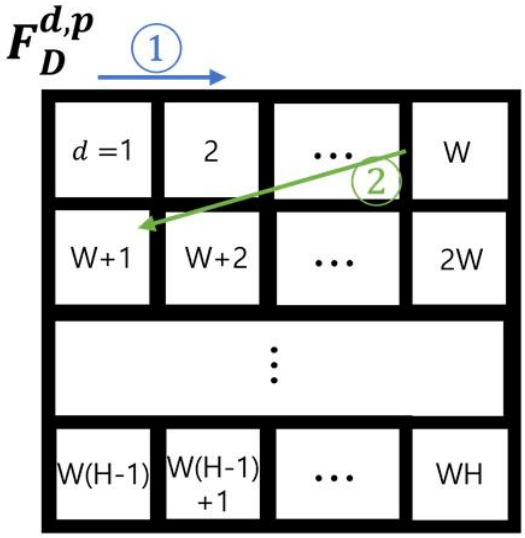

Fig. 3. Explanation of Coordinate Order from Feature Map

III. Methodology

본 장에서는 먼저 3D 객체 감지 모델 구조를 간단히 설명한 다음 본 논문이 제안하는 다중 스케일 특징맵 손실(Multi-scale feature map loss)을 설명한 후, 클래스 분류 쿨백-라이블러(KLD) 손실 방법을 설명해 본 발산 손실 방법을 설명해 본 연구가 제안하는 방법을 순차적으로 설명한다. 본 연구에서 제안하는 M3KD 방법의 주요 아키텍처(Architecture)는 Fig. 2에 표현되어 있다. Fig. 2에서 M3KD는 교사 모델과 학생 모델로 구성되며, 프리징 된 교사 모델에서 학생 모델로 지식전이를 통해 학생 모델의 학습을 수행한다. 이때 교사와 학생 모델 중간 특징맵에 FPN (Feature Pyramid Network)[17] 피라미드간의 차이를 줄여주는 다중 스케일 특징맵 손실(L_{MF})을 적용하여 다양한 객체 크기에 효과적인 지식 수행을 가능하게 하고, 출력 로짓에 클래스 분류 쿨백-라이블러 발산 손실(L_{CK})을 적용하여 학생모델에게 보다 세밀한 객체 검출 정보를 증류한다.

1. Camera-based 3D Object Detection Model Structure and Knowledge Distillation

단안 카메라 기반 3D 객체 검출은 일반적으로 크게 백본과 헤드의 두 부분으로 나누어진다. 백본은 이미지에서 특징맵을 추출하고, 헤드는 추출된 특징맵을 사용하여 클래스, 2D 바운딩 박스, 3D 바운딩 박스, 거리 및 편주각을 예측하는 완전 연결 층(Fully connected layer)으로 구성된다. 지식 증류 과정에는 추론 속도가 느리지만 정확도가 높은 교사 3D 객체 검출 모델과 추론 속도가 빠르지만, 정확도가 낮은 학생 3D 객체 검출 모델을 필요로 한다. 지식 증류에 필요한 교사 모델을 생성하기 위해 먼저 교사 모델을 MonoRCNN [16]에서 제안하는 손실을 통해 학습시킨다. 이후 학습된 교사 모델을 프리징(Freezing)시킨 후 본 연구가 3.2, 3.3장에서 제안하는 M3KD 기법을 사용해 학생 모델에 지식 증류를 수행한다.

2. Multi-scale Feature Map KD Method

다중 스케일의 지식 증류를 위해 다중 해상도의 특징 맵을 생성하는 FPN [17]에서 지식 증류를 수행한다. FPN은 네트워크의 다양한 계층에서 추출된 특징을 통합하는 구조를 갖는다. 피라미드 레벨은 이러한 다양한 해상도의 특징 맵을 정의하며, 각 레벨은 입력 이미지의 특정 스케일을 반영한다. 일반적으로, 상위 레벨은 더 큰 스케일(작은 해상도)에서 추출된 특징을, 하위 레벨은 더 작은 스케일(큰 해상도)에서 추출된 특징을 사용하여, 전반적인 객체 검출 성능을 향상시킨다.

먼저 본 연구는 한 개의 피라미드 레벨에서 지식 증류를 수행하기 위해 L2 손실 [6]을 사용해 Mean Squared Error (MSE) 계산을 진행한다. L_{MSE}의 수식은 다음과 같다.

$$L_{MSE}(F_T^{d,p}, F_S^{d,p}) = \frac{L_2(F_T^{d,p}, F_S^{d,p})}{W^p H^p}. \quad (1)$$

$$\text{where,} \quad L_2(F_T^{d,p}, F_S^{d,p}) = \sum_{d=1}^{W^p \times H^p} (F_T^{d,p} - F_S^{d,p})^2$$

W^p 와 H^p는 FPN의 피라미드 레벨 p에서의 특징맵의 너비와 높이이다. 특징맵의 너비와 높이이며, $L_2(\cdot)$는 입력값에 대한 모든 픽셀에서의 제곱값의 차이의 합을 구해주는 함수이다. $F_T^{d,p}$ 및 $F_S^{d,p}$는 각각 교사 및 학생 모델에 의해 추출된 피라미드 레벨 p에서의 출력 특징맵이 픽셀 d 위치에서 가지는 값이다. 이때 d는 픽셀의 좌표 위치이며 d좌표의 순서는 Fig. 3에서 확인할 수 있다. Fig. 3을 보면 d좌표는 특징맵 F에서 $d=(1,1)$부터 시작한다. Fig. 3에서 ① 과정을 진행하여 값이 증가함에 따라 너비가 1 증가하고 너비가 W의 값을 가질 때마다 ② 과정을 진행하여 높이 좌표를 1 증가시킨 뒤 너비 좌표를 1로 초기화하여, 전체 $d=(W,H)$ 영역에서의 특징맵 차이값을 구한다.

이후 MSE 손실을 모델 FPN의 모든 피라미드 레벨의 특징맵에 적용해 L_{MF} 손실을 구한다. 수식은 다음과 같다.

$$L_{MF} = \sum_{p=1}^{N_p} (L_{MSE}(F_T^{d,p}, F_S^{d,p})). \quad (2)$$

N_P는 백본의 피라미드 레벨의 총 개수이다. L_{MF}는 L_{MSE}의 값을 모든 피라미드에서 구하고 더해줌으로써 구할 수 있다. 이를 통해 모든 피라미드 레벨에 특징맵에서 학생 모델이 교사 모델을 모방하도록 함으로써 다중 스케일 지식의 증류를 수행한다.

3. Class KL Divergence KD Method

본 연구는 지식 증류를 수행할 때 학생 모델이 보다 원활한 검출을 진행하기 위해 교사 모델의 클래스 분류 지식을 전이한다. 본 연구는 클래스 분류 지식을 추출해 학습하기 위해 기존의 클래스 분류 모델에서 사용하는 힌튼의 지식 증류 [7]를 적용하였다. 적용하는 L_{CK}식은 다음과 같다.

$$L_{CK} = \sum_{i}^{C} P(T,i) \cdot \ln \frac{\mathrm{P(T,i)}}{\mathrm{P(S,i)}}. \quad (3)$$

$$\text{where,} \quad P(D,i) = \frac{e^{Z^{i_D}/t}}{\sum_{j=1}^{C} e^{Z^{j_D}/t}}$$

$P(\cdot)$는 소프트 맥스 함수이며 Z^{i_D}, Z^{j_D}는 교사 모델 T 또는 학생 모델 S를 포함하는 검출기 $\{T,S\} \in D$에서의 i, j클래스에 대한 클래스 로짓값이고 C는 총 클래스 개수이다. t는 온도(Temperature) 값으로 로짓값의 분포를 평탄하게 만드는 데 사용하며 큰 온도 값일수록 더 평탄한 분포를 가지게 된다. 이를 통해 학생 모델이 교사 모델의 로짓을 모방하도록 함으로써 보다 정확한 검출 지식을 증류한다.

4. Overall Loss

본 연구에서 제안하는 전체적인 훈련 손실은 다음과 같다.

$$L_{M3KD} = w_{MF}L_{MF} + w_{CK}L_{CK} + w_{2D}L_{2D} + w_{3D}L_{3D}. \quad (4)$$

L_{MF}, w_{MF} 는 앞서 3.2, 장에서 설명한 Multi-scale Feature Map 손실과 가중치이며, L_{CK}, w_{CK} 는 3.3, 장의 Class KL Divergence 손실과 가중치이다. L_{2D}, w_{2D} 는 각각 2D 이미지에서 사용되는 객체 검출을 위한 손실과 가중치로 이때 가중치는 바운딩 박스 회귀 손실과, 클래스 분류 손실로 이루어진다. 바운딩 박스 회귀 손실의 경우 Faster R-CNN [18]의 RPN 및 박스 회귀 손실을 사용하고, 클래스 분류 손실의 경우 교차 엔트로피를 사용하였다. 다음으로 L_{3D}, w_{3D} 는 3D 정보를 추출하는 손실과 가중치로써 이때 손실은 본 연구의 베이스라인인 MonoRCNN [16]에서 제안한 요(Yaw) 각도 및 깊이(depth) 손실을 사용하였다.

5. Implementation Details

본 연구의 지식 증류를 실험하기 위해 사용할 3D 객체 검출 기법으로는 MonoRCNN [16]을 사용하였다. 교사 모델의 경우 백본으로 ResNet-50 [2]을 사용하는 MonoRCNN에서 제공하는 학습된(Pretrained) 모델을 사용하였으며, 학생 모델의 경우 교사 모델에서 백본을 ResNet-18 [2]로 변경해 가벼운 모델을 설계하였다. 사용하는 교사 모델과 학생 모델인 ResNet-18, ResNet-50은 컨볼루션 레이어는 4개의 피라미드 레벨(p)의 특징맵($F_T^{d,p}$)을 사용한다. 지식 증류 성능을 비교하기 위한 방법으로는 현재 카메라 기반 3D 객체 검출 모델을 위한 관련 연구 사례가 적고 카메라 기반 3D 객체 검출기법인 FD3D [5] 기법도 단안 카메라가 아닌 다중(Multi) 카메라 3D 객체 검출을 위한 지식 증류 방법인 관계로 지식 증류의 성능의 비교를 위해 단안 카메라 3D 객체 검출용 지식 증류 기법을 직접 구현할 필요가 있다. 본 연구는 단안 카메라용 3D 객체 검출 지식 증류 기법의 비교 평가를 위해 최신 지식 증류 기법중 2D 객체 검출을 위한 지식 증류 기법인 PKD [9] 와 이미지 지식 증류 기법인 SemCKD [10]를 본 연구에 알맞도록 변환하여 사용하였다. PKD의 경우 교사 모델의 특징맵과 학생 모델의 특징맵에 피어슨 상관 계수를 적용하여 L2 거리를 측정하도록 하였으며, SemCKD의 경우 교사 모델의 1, 2번째 계층과 학생 모델의 1번째 계층의 상호작용을 사용하여 학생 모델이 학습하도록 적용하였다.

IV. Experiments

이번 장에서는 본 연구가 제안한 M3KD 지식 증류로 학습한 학생 모델의 성능 평가를 KITTI validation 데이터 세트를 사용해 진행한다. 수행한 실험의 설정과 비교 모델을 설명하고 제안한 M3KD의 세부 방법들에 대한 절제 실험(Ablation studies)과 정성적인(Qualitative) 실험을 진행한다.

1. Experiment Setting

본 연구에서 제안하는 M3KD 방법의 효과를 증명하기 위해 KITTI 데이터 세트[8]를 사용하였다 KITTI 데이터 세

Table 1. 3D object detection accuracy (AP_{TV}, AP_{3D}) on Car class comparison on KITTI validation dataset (Orange and Blue are highest and second-highest AP scores of all student models, respectively.)

Backbone	KD method	Car AP_{TV} (↑) (Easy	Norm.	Hard)	Car AP_{3D} (↑) (Easy	Norm.	Hard)	Time (↓)
ResNet-50	Teacher (Vanilla)	27.41%	21.14%	17.48%	19.13%	14.69%	12.42%	61.9 ms
ResNet-18	Student (Vanilla)	26.17%	18.74%	15.07%	17.71%	13.19%	10.43%	47.7 ms
	PKD [9]	26.98%	19.39%	15.71%	19.31%	13.60%	11.19%	47.7 ms
	SemCKD [10]	24.25%	18.34%	14.91%	17.15%	12.57%	10.42%	47.5 ms
	M3KD (Ours)	27.64%	20.69%	17.09%	20.30%	15.01%	12.42%	47.7 ms

Table 2. 3D object detection accuracy (AP_{3D}) on pedestrian, cyclist class comparison on KITTI validation dataset (Orange and Blue are highest and second-highest AP scores of all student models, respectively.)

Backbone	KD method	Pedestrian AP_{3D} (↑) (Easy	Norm.	Hard)	Cyclist AP_{3D} (↑) (Easy	Norm.	Hard)	Time (↓)
ResNet-50	Teacher (Vanilla)	6.85%	5.58%	4.58%	4.38%	2.48%	2.57%	61.9 ms
ResNet-18	Student (Vanilla)	7.05%	5.09%	3.94%	2.70%	1.52%	1.29%	47.7 ms
	PKD [9]	9.57%	6.50%	5.29%	3.59%	2.01%	1.54%	47.7 ms
	SemCKD [10]	8.33%	5.84%	4.79%	4.93%	2.68%	2.68%	47.5 ms
	M3KD (Ours)	7.38%	5.61%	4.51%	3.55%	2.02%	1.83%	47.7 ms

트는 7481개의 훈련(Training) 이미지에 자동차, 보행자 및 자전거 세 가지 클래스에 대한 2D 및 3D 바운딩 박스 주석(Annotation)으로 이루어진 데이터 세트이다. 이때 객체별로 이미지 내의 객체 잘림(Truncation), 폐색(Occlusion) 또는 거리에 따라 easy(쉬움), norm.(normal(보통)), hard(어려움)의 3가지 난이도로 세부 분류가 되어 있다. 본 연구는 단안 카메라 객체 검출의 성능을 평가하기 위해 KITTI 검증 분할(Validation split) [19]을 사용하였으며 이를 통해 훈련 데이터를 3,712장의 학습 셋과 3,769장의 검증 셋으로 나누어 실험을 진행하였다. 지식 증류에는 Titan V GPU 4대를 사용하였으며 배치 사이즈는 16으로 설정하였다, 학습률(Learning rate)은 0.002를 사용하였다. 학습은 총 30000번 반복 수행(Iteration)을 진행하였으며 15000번, 20000번, 25000번 반복 수행마다 학습률을 10분의 1로 감소시켰다. 학습을 안정화하는 클립(Clip) 값은 3으로 설정하였다. L_{MF}, L_{CK} 의 계수인 w_{CK}, w_{MF}, w_{CK}는 각각 0.5, 0.1을 적용하였으며, L_{2D}, L_{3D}의 계수인 w_{2D}와 w_{3D}는 모두 1.0을 적용하였다.

Table 3. 3D objecet detection accuracy (mAP_{3D}) on Car, Pedestrian and Cyclist classes comparison on KITTI validation dataset (Orange and Blue are highest and second-highest mAP scores of all student models, respectively. "Avg." represents average across all difficulty levels. "Va." represents Vanilla model. all score unit is percentage (%))

KD method	mAP_{3D} (↑) (Easy	Norm.	Hard.)	Avg.
Teacher (Va.)	10.12	7.58	6.52	8.07
Student (Va.)	9.15	6.60	5.22	6.99
PKD	10.82	7.37	6.01	8.07
SemCKD	10.14	7.03	5.96	7.71
M3KD (Ours)	10.41	7.55	6.25	8.07

2. Comparison Result

비교 실험에는 KITTI 검증(Validation) 셋을 사용하였으며 성능 평가 지표(Metrics)로는 바운딩 박스의 X, Z 좌표만을 고려하는 AP_{TV} (Top view average precision)를 자동차 클래스에서 진행하였고, X, Y, Z 좌표를 모두 고려하는 AP_{3D} (3D average precision)성능을 자동차, 보행자, 자전거 각각의 클래스에서 진행하였으며, AP_{3D}를 모든 클래스에서 평균내는 mAP_{3D}를 사용해 평가를 진행하였다. 또, 경량화 된 정도를 측정하기 위해 방법으로 이미지

Table 4. Ablation study on KITTI validation dataset. Each loss applied separately, expressed with circle. ("Back" represents used backbone; R-50 and R-18 are ResNet-50 and ResNet-18 respectively. Orange and Blue are highest and second-highest AP scores of all R-18 student models, respectively. "Avg." represents average across all difficulty levels. all score unit is percentage (%))

Back	Method	L_{MF}	L_{CK}	Car AP_{3D} (↑) (Easy	Norm.	Hard)	Ped. AP_{3D} (↑) (Easy	Norm.	Hard)	Cyc. AP_{3D} (↑) (Easy	Norm.	Hard)	Avg.
R-50	Vanilla			19.13	14.69	12.42	6.85	5.58	4.58	4.38	2.48	2.57	8.07
	M3KD	○	○	19.45	14.82	12.46	7.68	5.79	4.74	4.10	2.27	2.42	8.19
R-18	Vanilla			17.71	13.19	10.43	7.05	5.09	3.94	2.70	1.52	1.29	6.99
	M3KD	○		19.95	14.58	12.19	7.85	6.00	4.64	3.34	1.85	1.93	8.04
			○	18.11	13.13	10.86	9.00	6.42	5.28	3.89	2.07	2.02	7.86
		○	○	20.30	15.01	12.42	7.38	5.61	4.51	3.55	2.02	1.83	8.07
	PKD			19.31	13.60	11.19	9.57	6.50	5.29	3.59	2.01	1.54	8.07
		○	○	19.59	14.53	12.20	8.91	6.32	4.90	5.65	2.11	2.19	8.49
	Semckd			17.15	12.57	10.42	8.33	5.84	4.79	4.93	2.68	2.68	7.71
		○	○	19.80	14.05	11.79	8.34	5.97	4.74	4.25	2.59	2.50	8.23

한 장당 추론 시간을 측정하여, ms (Millisecond)로 측정해 Time 칸에 표시하였다. 제안한 방법의 비교 대상으로는 학습 데이터 세트의 참값(Ground truth)으로 학습한 바닐라(Vanilla) 학생 모델과 PKD [9], SemCKD [10] 방법으로 지식 증류한 학생 모델을 사용해 비교 평가를 진행하였다. 실험의 결과는 Tables 1-3에서 확인할 수 있다. 먼저 자동차 클래스의 성능 수치를 나타낸 Table 1을 보면 제안하는 M3KD 기법으로 학습한 모델이 PKD, SemCKD로 학습한 학생 모델 대비 자동차 클래스의 easy 난이도의 AP_{TV} 성능에서 0.66%, 3.39% 높은 수치를 달성하였다. 또 AP_{3D} 성능에서는 각각 0.99%, 3.15% 더 높은 AP_{3D}를 달성하였다. 또한, 기존 KD를 사용하지 않은 vanilla 학생 모델 대비 AP_{3D}가 easy 난이도에서 2.59% 향상됨을 보임으로써 제안한 기법의 효과를 확인할 수 있었다. 특히 무거운 백본을 가지는 교사 모델보다도 1.17% 만큼 뛰어난 성능을 나타낸 것을 확인해 자동차 클래스에서 우수한 성능을 확인 할 수 있다. 그다음으로 보행자와 자전거 클래스의 성능 수치를 나타낸 Table 2를 본다면 AP_{3D}성능에서 본 논문이 제안하는 기법이 기존 학생모델 대비 AP 보행자와 자전거에 대한 easy 난이도에서 각각 0.33%, 0.85% 향상한 것을 확인할 수 있었다. 다른 기법인 PKD, SemCKD와 비교시 자전거에서는 어려움 난이도에서 PKD 기법 대비 0.29% 높은 수치를 확인할 수 있었다. 다만 보행자에서는 타 기법 대비 다소 떨어지는 점이 있음을 확인하였다. 모델의 추론 속도를 보면 교사 모델의 전체 추론 시간의 경우는 61.9ms이고 학생 모델의 추론 시간은 약 47.7ms로 학생 모델의 30%의 추론 시간 감소를 확인했다. 마지막으로 전체 클래스에서 mAP_{3D} 를 측정한 Table 3을 본다면 기존 PKD 기법 대비 보통 난이도와 어려움 난이도에서 각각 0.18%, 0.24% 만큼 이득이 있는 것을 확인하였으며, SemCKD기법과 비교시, 모든 난이도에서 mAP_{3D} 점수에 있어 이득이 있음을 확인 할 수 있었다. 또한 모든 클래스와 난이도의 mAP_{3D} 평균을 측정한 "Avg." 수치에서 교사 모델과 동등한 성능인 8.07%를 달성하여 빠른 속도와 동시에 성능이 유지됨을 확인하였으며, 기존의 학생 모델의 6.99%수치와 비교시 1.08%의 성능향상이 있어 모든 클래스에서 성능이 증가함을 확인하였다.

3. Ablation Studies

제안하는 M3KD의 각 기법에 대해 절제 실험을 진행하였다. 절제 실험은 학생인 바닐라 모델에서 본 연구가 제안한 방법인 L_{MF}, L_{CK}를 추가하였을 때의 각각의 클래스에 대한 AP_{3D}를 KITTI 검증 셋에서 측정하였다. 실험에 대한 결과는 Table. 4에서 확인할 수 있다. Table. 4를 보게 될 경우 L_{MF} 적용해 기존의 바닐라 학생 모델과 비교해 볼 시 모든 난이도의 평균을 나타낸 "Avg."에서 mAP_{3D}가 6.99%에서 8.04%로 증가해 전체적인 난이도와 클래스의 AP_{3D}에서 향상이 있는 것을 확인할 수 있다. 이는 교사 모델에서 다중 특징맵의 지식을 학생 모델에 직접 증류함으로써 다양한 크기의 특징들을 더 세밀하게 학습한 결과이다. 다음으로 L_{CK}를 적용 시 모든 난이도의 평균에서 mAP_{3D}가 6.99%에서 7.86%로 성능 개선이 이루어진 것을 확인할수 있었다 특히 보행자 클래스의 모든 난

Fig. 4. KITTI Qualitative Results Examples. We visualize M3KD and PKD methods on the KITTI validation dataset. The orange, yellow, and green boxes in the image represent the 2D and top view projections of the predicted 3D bounding boxes of M3KD, PKD, GTs, respectively. Points side of boxes are 3D IoU and top view IoU points. The radius difference between two adjacent white circles is 5 meters.

이도인 쉬움, 보통, 어려움 난이도에서 각각 9.00%, 6.42%, 5.28%를 달성하여 높은 AP3D를 가지는 것을 확인할 수 있었다. 이는 교사 모델에서 클래스 정보를 학생 모델에게 학습시켜 검출 성능을 올림으로써 작은 자전거 클래스와 같은 난이도가 높은 객체를 잘 잡아낸 결과이다. 마지막으로 L_{MF}와 L_{CK}를 모두 적용한 M3KD 기법의 경우 모든 난이도의 평균에서 mAP_{3D}가 평균적으로 8.07%을 달성해 뛰어난 성능을 달성한 것을 확인할 수 있었다. 추가적으로 경량화 하지 않은 ResNet-50 백본 모델과 SemCKD PKD에 우리의 M3KD 기법을 적용하여 보았다. 먼저 ResNet-50 백본 모델에 지식 증류를 사용 할 시 모든 난이도에서 mAP_{3D}가 평균이 8.07%에서 8.19%로 성능 향상이 있음을 확인하였다. SemCKD와 PKD의 경우 기존 SemCKD의 8.07% 대비 M3KD를 적용한 SemCKD는 8.49%로 0.42% 향상이 있었고, 기존 PKD의 7.71% 대비 M3KD를 적용한 PKD 8.23%로 0.52% 향상이 있어, 본 연구가 제안한 M3KD 기법이 다른 KD 기법과 같이 적용하였을 때에도 효과가 있음을 확인하였다.

4. Qualitative Results

제안된 M3KD와 PKD [9]의 비교에 대한 정성적인 평가를 진행하였다. 정성적인 평가로는 결과값인 3D 바운딩 박스를 2D 이미지와 탑 뷰에 정사영하여 4개 결과의 각 왼쪽과 오른쪽에 시각화 했으며, 바운딩 박스 옆에 각각 3D IoU (Intersection over Union)과 Top Veiw IoU로 정량적 수치를 표시하였다. 본 연구가 제안한 M3KD 결과값 주황색으로, 비교 대상인 PKD의 결과값을 노란색으로 표시하였고, KITTI 데이터 세트의 참값을 초록색으로 표시하였다. 해당하는 이미지는 Fig. 4에서 확인할 수 있다. 이때 결과를 비교 해보면, (a)와 (b)의 왼쪽 2D 이미지에 자동차의 3D IoU를 살펴보면 노란색으로 표시된 PKD의 값 대비 M3KD가 전반적으로 높은 값을 가지는 것을 알 수 있다. 그뿐만 아니라, Top View 이미지에서도 (a)와 (b)는 M3KD가 PKD에 비해서 높은 TV IoU를 가지는 것을 알 수 있다. (a)의 가장 먼거리에 존재하는 자동차에 대해서 3D IoU와 Top View IoU는 M3KD가 각각 0.514와 0.528로 PKD의 0.306, 0.312에 비해서 0.208, 0.216 높은 값을 가진다. 이것은 본 연구가 제안하는 L_{CK}로 인해 식별이 어려운 객체도 잘 검출해 낸 결과이다. 그 외에도 (b)와 같이 일반적인 경우의 자동차 클래스에 대해서도 PKD와 비교했을 때 3D IoU와 Top View IoU는 0.194와 0.195 높은 값을 가진다. 반면, (c)와 (d)의 시각화 내용은 Table 4에서 보여준 내용처럼 보행자와 자전거 항목에서 M3KD가 PKD에 비해서 3D IoU와 TV IoU 수치가 조금 떨어지는 것을 알 수 있다. 보행자 항목에 해당하는 (c)에 경우에는 3D IoU와 TV IoU 각각 0.029와 0.036 정도의 근소한 수치 차이가 존재한다는 것을 알 수 있다. 또한, 자전거 항목에 해당하는 (d)에 경우에도 3D IoU와 TV IoU 각각 0.07와 0.02 정도의 근소한 수치 차이가 존재한다는 것을 알 수 있다.

V. Conclusion

기존의 3D 객체 검출 모델을 위한 지식 증류 방법의 경우 모델의 정확도 향상에 초점을 맞춘 연구가 주류였으며, 경량화에 초점을 맞춘 연구는 그 개수가 적었다. 본 연구는 경량화 초점의 단안 카메라 3D 객체 검출 모델 간의 지식 증류 방법인 M3KD 지식 증류를 제안했다. M3KD는 학생 모델의 무거운 백본을 가벼운 백본으로 대체한 학생 모델을 만들어 경량화했다. 이후, 지식 증류를 수행하여 학생 모델의 원활한 학습을 유도했다. 이를 위해, 먼저 교사 모델의 백본 특징맵과 학생 모델 백본 특징맵의 L2 손실값을 줄임으로 학생 모델이 교사 모델을 모방하도록 유도하는 다중 스케일 특징맵 지식 증류(Multi-scale Feature Map KD) 기법을 제안했다. 추가적으로, 교사 모델의 클래스 지식을 학습하기 위해 학생 모델의 클래스 로짓에 쿨백-라이블러 발산을 사용해 교사 모델의 클래스 로짓을 모방하도록 하는 클래스 분류 쿨백-라이블러 발산 지식 증류 기법을 적용하였다. 본 연구는 이 방법을 통해 교사 모델보다 30% 빠른 추론 속도를 가지는 학생 모델을 만듦과 동시에 KITTI 데이터 세트의 전체적인 난이도에서의 mAP_{3D} 정확도는 교사모델의 높은 정확도를 유지함을 확인하였다. 또한, 최신 지식 증류 기법인 PKD와 SemCKD를 사용해 비교 실험을 통해 mAP_{3D} 의 전체적인 난이도에서 PKD와는 동일한 성능 향상을 나타냄을 보였고, SemCKD 와 비교 시 0.36% 높은 성능을 확인하였다. 추후 실험으로는 다른 지식 증류 기법 대비 부족하였던 자전거와 보행자 클래스에서의 성능 향상 방법에 대한 연구와, 학생 모델의 백본과 헤더를 더욱 경량화하여 임베디드 보드(Embedded Board)에 맞는 학생 모델에 대한 지식 증류 기법을 개발하는 것을 연구목표로 하고자 한다. 제안한 M3KD 지식 증류 방법이 자율주행과 같은 실시간성과 정확도가 동시에 필요한 효율적인 3D 객체 검출 모델의 개발에 이바지할 것으로 기대한다.

ACKNOWLEDGEMENT

This work was supported in part by the National Research Foundation of Korea (NRF) grants funded by the Korea government (MSIT) (No. NRF-2022R1C1C1009208) and funded by the Ministry of Education (No.2022R1A6A1A03051705); supported in part by Institute of Information & communications Technology Planning & Evaluation (IITP) grants funded by the Korea government (MSIT) (No.2022-0-00448/RS-2022-II220448: Deep Total Recall, 30%, No.RS-2022-00155915: Artificial Intelligence Convergence Innovation Human Resources Development (Inha University))

REFERENCES

[1] Mao Jiageng, "3D object detection for autonomous driving: A comprehensive survey," International Journal of Computer Vision, Vol. 131, No. 8, pp. 1909-1963, August 2023. DOI: 10.1007/S11263-023-01790-1

[2] He Kaiming, "Deep residual learning for image recognition," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, Las Vegas, NV, USA, June 2016. DOI: 10.1109/CVPR.2016.90

[3] Zhou Shengchao, "UniDistill: A Universal Cross-Modality Knowledge Distillation Framework for 3D Object Detection in Bird's-Eye View," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5116-5125, Vancouver, BC, Canada, June 2023. DOI: 10.1109/CVPR52729.2023.00495

[4] Chong Zhiyu, "Monodistill: Learning spatial features for monocular 3d object detection," arXiv preprint arXiv:2201.10830 Vol. abs/2201.10830, 2022.

[5] Zeng Jia, "Distilling Focal Knowledge from Imperfect Expert for 3D Object Detection," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 992-1001, Vancouver, BC, Canada, June 2023. DOI: 10.1109/CVPR52729.2023.00102

[6] Chen Defang, "Knowledge distillation with the reused teacher classifier," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11923-11932, New Orleans, LA, USA, June 2022. DOI: 10.1109/CVPR52688.2022.01163

[7] G. Hinton, O. Vinyals, and J. Dean, "Distilling the Knowledge in a Neural Network," arXiv, March 2015. DOI: 10.48550/arXiv.1503.02531

[8] Andreas Geiger, Philip Lenz, and Raquel Urtasun, "Are we ready for autonomous driving? the KITTI vision benchmark suite," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3354-3361, Providence, RI, USA, June 2012. DOI: 10.1109/CVPR.2012.6248074

[9] Cao Weihan, "Pkd: General distillation framework for object detectors via pearson correlation coefficient," Advances in Neural Information Processing Systems 35, pp. 15394-15406, New Orleans, LA, USA, November 2022.

[10] Wang Can, "SemCKD: Semantic calibration for cross-layer knowledge distillation," IEEE Transactions on Knowledge and Data Engineering, Vol. 35, No. 8, pp. 6305-6319, June 2023: 6305-6319. DOI: 10.1109/TKDE.2022.3171571

[11] Seung-Hwan Bae, "Deformable Part Region Learning and Feature Aggregation Tree Representation for Object Detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 45, pp. 10817-10834, September 2023. DOI: 10.1109/TPAMI.2023.3268864

[12] Seung-Hwan Bae, "Deformable part Region Learning for object detection," Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 36, No. 1, pp. 95-103, 2022. DOI:10.1609/AAAI.V36I1.19883

[13] Seong-Ho Lee, and Seung-Hwan Bae, "AFI-GAN: Improving feature interpolation of feature pyramid networks via adversarial training for object detection," Pattern Recognition, Vol. 138, pp. 1-14, June 2023. DOI: 10.1016/J.PATCOG.2023.109365

[14] Lang Alex H, "Pointpillars: Fast encoders for object detection from point clouds," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12697-12705, Long Beach, CA, USA, June 2019. DOI: 10.1109/CVPR.2019.01298

[15] Brazil, Garrick, and Xiaoming Liu, "M3d-rpn: Monocular 3d region proposal network for object detection," Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9286-9295, Seoul, Korea, October 2019. DOI: 10.1109/CVPR.2019.01298

[16] Shi Xuepeng, "Geometry-based distance decomposition for monocular 3d object detection," Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 15152-15161, Montreal, QC, Canada, October 2021. DOI: 10.1109/ICCV48922.2021.01489

[17] Lin Tsung-Yi, "Feature pyramid networks for object detection," Proceedings of the IEEE Conference on Computer Vision and Pattern recognition, pp. 2117-2125, Hawaii, USA, July 2017. DOI: 10.1109/CVPR.2017.106

[18] Ren Shaoqing, "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 39, No. 6, pp. 1137-1149, June 2017, DOI: 10.1109/TPAMI.2016.2577031

[19] Xiaozhi Chen, Kaustav Kundu, Ziyu Zhang, Huimin Ma, Sanja Fidler, and Raquel Urtasun, “Monocular 3d object detection for autonomous driving,” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2147-2156, Las Vegas, NV, USA, June 2016. DOI: 10.1109/CVPR.2016.236

Authors

Se-Gwon Cheon received the B.S. degree in electrical engineering from Inha University in 2022, and is currently pursuing the M.S. degree with the Department of Electrical Computer Engineering at Inha University, Korea. His current research interest is Knolwedge distillation and 3D object detection.

Hyuk-Jin Shin received the B.S. degree with Department of Computer Science, Chung-Buk National University, South Korea. He is currently pursuing the M.S. - Ph.D. integrated course degree. His research interests are object detection and knowledge distillation.

Seung-Hwan Bae received the BS degree in information and communication engineering from Chungbuk National University, in 2009 and the MS and PhD degrees in information and communications from the Gwangju Institute of Science and Technology (GIST), in 2010 and 2015, respectively. He was a senior researcher at Electronics and Telecommunications Research Institute (ETRI) in Korea from 2015 to 2017. He was an assistant professor in the Department of Computer Science and Engineering at Incheon National University, Korea from 2017 to 2020. He is currently an Associate Professor with the Department of Electrical and Computer Engineering at Inha University, His research interests include object tracking, object detection, generative model learning, continual learning, on-device ML, etc.