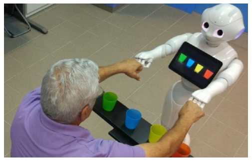
Cross-Modality Learning for 3D Object Detection

Jinhyung (David) Park
Faculty Advisor: Professor Kris Kitani

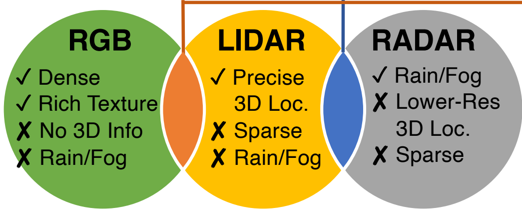
Motivation



Assistive Robots

Autonomous Driving

Autonomous systems heavily rely on **3D perception** to interact with surroundings.



Different modalities have their own strengths and weaknesses. To improve 3D detection, we fuse modalities to leverage these **mutually beneficial relationships**.

Challenges & Core Directions

Challenges in multimodal learning:

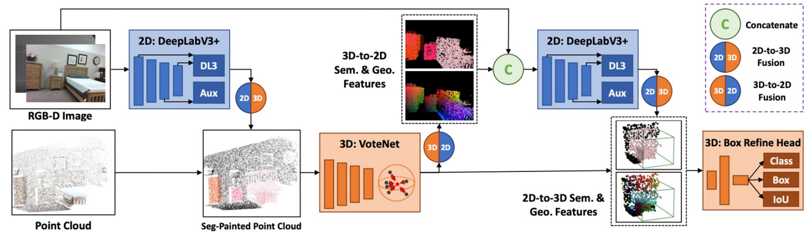
- Model overfits to a single modality.
- Representation differences between modalities inhibit training.

Core idea: Task-level correction and cross-modality consistency strengthen feature learning and improve task performance.

References

[1] **Jinhyung D. Park**, Xinshuo Weng, Yunze Man, and Kris Kitani. Multi-modality task cascade for 3d object detection. BMVC, 2021.
[2] Yu-Jhe Li, **Jinhyung Park**, Matthew O' Toole, and Kris Kitani. Modality-Agnostic Learning for Radar-Lidar Fusion in Vehicle Detection. CVPR 2022.
[3] **Jinhyung Park**, Chenfeng Xu, Yiyang Zhou, Masayoshi Tomizuka, and Wei Zhan. DetMatch: Two Teachers are Better Than One for Joint 2D and 3D Semi-Supervised Object Detection. arXiv:2203.09510, 2022.

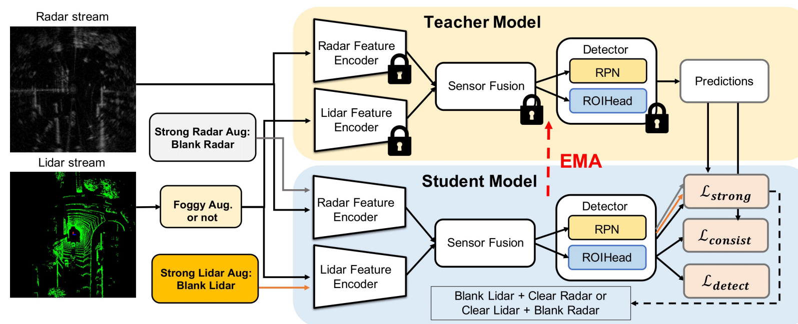
Multi-Modality Task Cascade for 3D Object Detection[1]



Joint 3D detection & 2D semantic segmentation.

We propose **recursive, task-level cascade** of alternating modalities: 2D->3D->2D->3D.

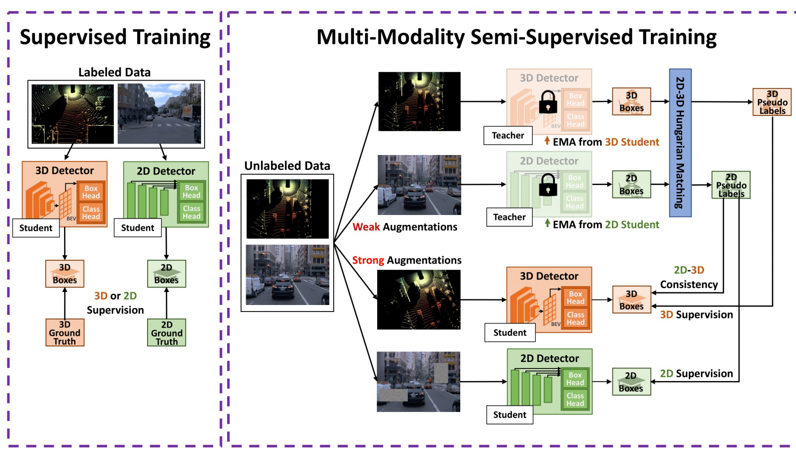
Modality-Agnostic Learning for Radar-Lidar Fusion in Vehicle Detection[2]



LiDAR-Radar fusion model robust to total sensor failure of one modality.

Extends Teacher-Student mutual learning to encourage model **to hallucinate features** of the missing modality.

Joint LiDAR & RGB Semi-Supervised Learning for 3D Object Detection[3]



Semi-supervised learning leveraging paired RGB & LiDAR frames in autonomous driving.

Hungarian matching of 2D & 3D detections generates cleaner, cross-modality consistent pseudo-labels.