Understanding Core 3D Object Detection Concepts

3D Object Detection is the process of identifying and localizing objects in a 3D space. Unlike 2D detection, which provides bounding boxes in image space, 3D detection adds information about depth, orientation, and size in a 3D environment.

# Input Data Formats

- RGB: Standard 2D image with color.

- Depth Map: Per-pixel depth from a sensor.

- Point Cloud: A set of 3D points (x, y, z).

- RGB-D: Combination of color and depth (e.g., RealSense, Kinect).

# 3D Bounding Box Format

[class, center (x, y, z), dimensions (w, h, l), orientation θ]

# Common Representations

|  |  |  |
| --- | --- | --- |
| Representation | Description | Example Use Case |
| Point Cloud | Raw 3D points from sensors | VoteNet, PointNet |
| Voxel Grid | 3D grid of volumes | VoxelNet |
| Bird’s Eye View | Top-down 2D projection | Autonomous vehicles |
| RGB-D Image | 2D image + aligned depth | Indoor detection (SUN RGB-D) |

# Main Approaches

|  |  |  |
| --- | --- | --- |
| Type | Description | Example Models |
| Point-based | Process raw point clouds | PointNet, VoteNet |
| Voxel-based | Convert to voxel grids | VoxelNet, SECOND |
| BEV-based | Top-down projection | CenterPoint, PIXOR |
| Fusion | Combine RGB and depth | AVOD, 3D-SIS |

# Evaluation Metrics

- 3D IoU (Intersection over Union): Overlap between predicted and ground truth boxes.

- mAP@0.25 or mAP@0.5: Mean Average Precision at IoU thresholds.

- Precision / Recall

# Recommended Reading

- VoteNet: https://arxiv.org/abs/1904.09664

- PointNet: https://arxiv.org/abs/1612.00593

- SUN RGB-D Dataset: http://rgbd.cs.princeton.edu/