POINTFUSION

DEEP SENSOR FUSION FOR 3D BOUNDING BOX ESTIMATION

ZAQI MOMIN

AGENDA

INTRODUCTION

METHODOLOGY

EXPERIMENTS

RESULTS

SUMMARY

INTRODUCTION

- PointFusion is a generic 3D object detection method.
- Existing methods use complex multi-stage pipelines or sensor-specific assumptions.
- PointFusion uses both image and 3D point cloud data for application-agnostic 3D object detection.

- Image Data is processed by Convolutional Neural Network (CNN).
- 3D Point Cloud Data is processed by PointNet architecture.
- Fusion Network combines CNN and PointNet outputs.
- Predicts multiple 3D bounding box hypotheses.
- Uses input 3D points as spatial anchors.

METHODOLOGY

EXPERIMENTS

- Training involves resizing and shifting ground truth 2D bounding boxes by 10%.
- Input image crops resized to 224×224; max 400 input 3D points sampled.
- Evaluation uses top 300 2D detector boxes per image.
- 3D detection score = 2D detection score × predicted 3D bounding box scores.

DATASETS

KITTI DATASET
SUN-RGBD DATASET
AP3D DATASET

METRICS

A predicted 3D box is a true positive if its 3D IoU with a ground truth box is above a threshold.
KITTI thresholds: 0.7 (Car), 0.5 (Cyclist, Pedestrian).
SUN-RGBD threshold: 0.25 for all classes.

ARCHITECTURE

Image Processing: ResNet-50 pretrained on ImageNet.
Point Cloud Processing: Original PointNet without batch normalization.
2D Object Detector: Faster-RCNN pretrained on MS-COCO and fine-tuned.

RESULTS

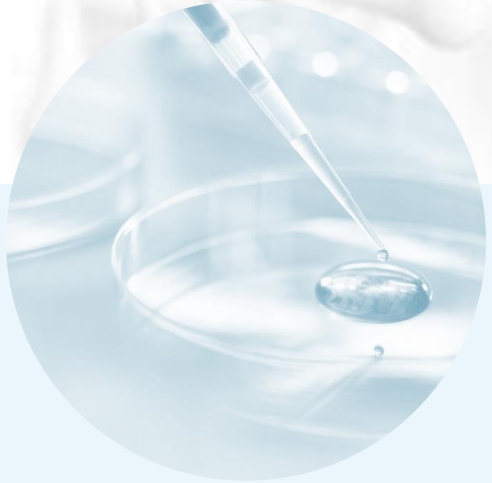
Method	Input	Easy	Mod.	Hard
3DOP [2]	Stereo	12.63	9.49	7.59
VeloFCN [18]	3D	15.20	13.66	15.98
MV3D [3]	3D + rgb	71.29	62.68	56.56
rgb-d	3D + rgb	7.43	6.13	4.39
Ours-global-no-im	3D	28.83	21.59	17.33
Ours-global	3D + rgb	43.29	37.66	32.23
Ours-dense-no-im	3D	62.13	42.31	34.41
Ours-dense	3D + rgb	71.53	59.46	49.41
Ours-final	3D + rgb	74.71	61.24	50.55
Ours-final (all-class)	3D + rgb	77.92	63.00	53.27

SUMMARY

- PointFusion Network accurately estimates 3D object bounding boxes from image and point cloud.
- Raw point cloud data processed directly using PointNet model and avoids lossy preprocessing like quantization or projection.
- Dense Fusion Network combines image and point cloud representations and predicts multiple 3D box hypotheses relative to input 3D points (spatial anchors), also automatically learns to select the best hypothesis.
- Demonstrated effectiveness across different datasets without relying on dataset and sensor-specific assumptions.

FUTURE WORK

- Integrating the 2D detector with the PointFusion network into a unified end-to-end 3D detection system.
- Extending the model to include temporal aspects for joint detection and tracking in video and point cloud streams.



THANK YOU

ZAQI MOMIN
ZAQIMOMIN@IITGN.AC.IN|+91 9998142340