OCTraN: 3D Occupancy Convolutional Transformer Network in Unstructured Traffic Scenarios

Aditya Nalgunda Ganesh, Dhruval Pobbathi Badrinath, Harshith Mohan Kumar, Priya SS, Surabhi Narayan
{adityang5, dhruvalpb, hiharshith18, sspriya147}@gmail.edu surabhinarayan@pes.edu
Department of Computer Science, PES University, Bengaluru

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Goal

Train **3D Occupancy** prediction networks using **depth boosting** for generating ground truth data.

Motivations

Existing autonomous driving datasets:

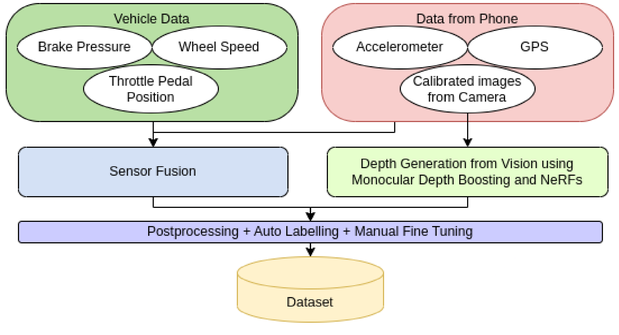
- Lack complex dense unstructured traffic.

Contributions:

- Depth dataset generation pipeline.
- **OCTraN**, a transformer architecture that uses iterative-attention producing 3D occupancy features.

Dataset Generation

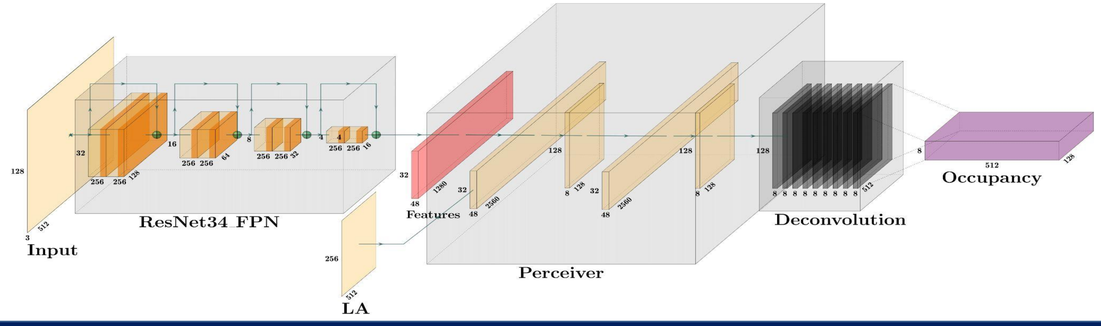
Collected a dataset spanning 114 minutes and 165K frames in Bengaluru, India in day and night lighting conditions.



OCTraN Architecture

OCTraN consists of:

- Standard convolutional backbone.
- Feature-space transformer.
- Convolution transpose.

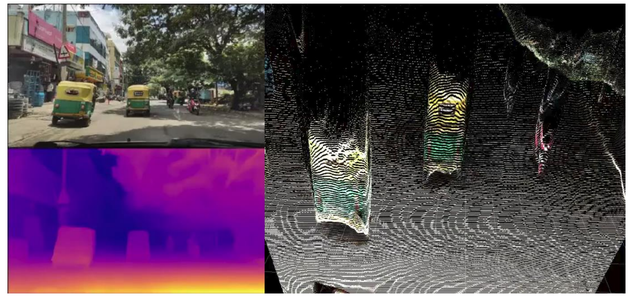


Results

Ablation Study of our proposed architecture comparing the optimal hyperparameters and the achieved IoU scores.

Method	Dataset	Hyperparameters									IoU (%)
		CH	CDH	LMP	LR	BS	MF	D	LH	LDH	
OCTraN-B	KITTI	1	64	0	0.001	1	1000	1	8	32	18.818
OCTraN-V0	KITTI	8	32	0.5	0.0001	2	500000	4	8	32	**34.669**
OCTraN-V1	KITTI	8	32	0.5	0.0001	4	500000	8	4	32	28.408

RGB frame and corresponding disparity map from our dataset. Projection through camera calibration



OCTraN prediction (right) reduces disparity error of LiDAR (left) by predicting in-depth space.