Road Scene Understanding for Risk Anticipation from Ego-Vehicle Data

Henrique Piñeiro Monteagudo

Supervisors:
Samuele Salti (University of Bologna), Francesco Sambo and Leonardo Taccari (Verizon Connect)

SMARTHEP is funded by the European Union's Horizon 2020 research and innovation programme, call H2020-MSCA-ITN-2020, under Grant Agreement n. 956086

Research Project Overview

1st Phase — Done

Obtain a **geometric** and **semantic representation** for road scenes environments

Challenges:

- Lack of 3D GT data
- High variability in data

2nd Phase — In Progress

Predict the **future** position **of agents** and ego-vehicle on the scene

Challenges:

- Lack of GT trajectories
- Multimodality of futures

3rd Phase — To Do

Provide insights into **imminent dangers based** on built understanding

Challenges:

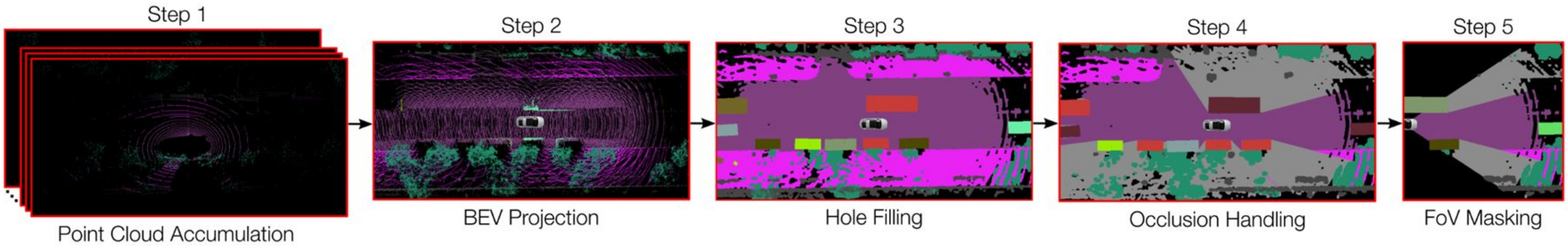
- Real-time operation
- Difficult to calibrate

Self-Supervised Bird's Eye View Segmentation → Self-Supervised Forecasting in the Bird's Eye View → Danger anticipation

Bird’s Eye View Representation

Bird’s Eye View: a 2D orthographic projection of the world along the direction of gravity. Representation with desirable properties for road scenes:

- Metric (under certain assumptions)
- Compact compared with explicit 3D like a voxel grid
- Road agents’ movement is mostly restricted to the ground plane

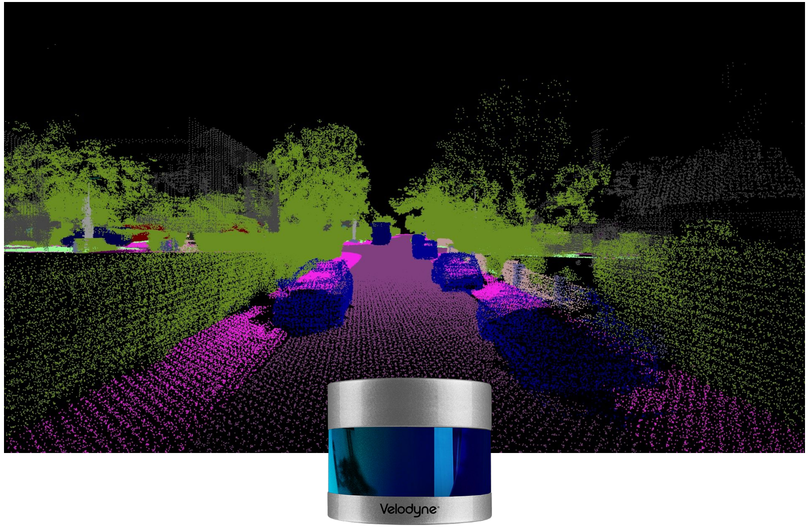


Typical pipeline to generate BEV segmentation labels to train fully supervised models. This example: from annotated point clouds in the KITTI360 dataset in *PanopticBEV, Gosala and Valada, Robotics and Automation Letters 2022*

How to supervise without 3D data?

✅

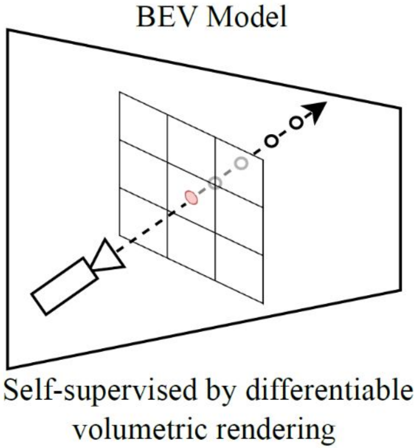
❌

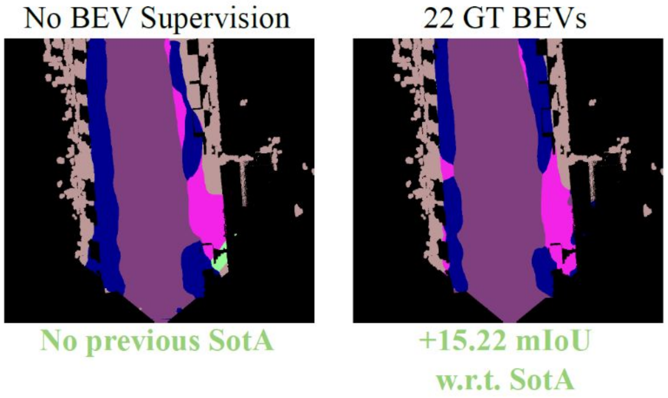


Our proposal: *RendBEV*

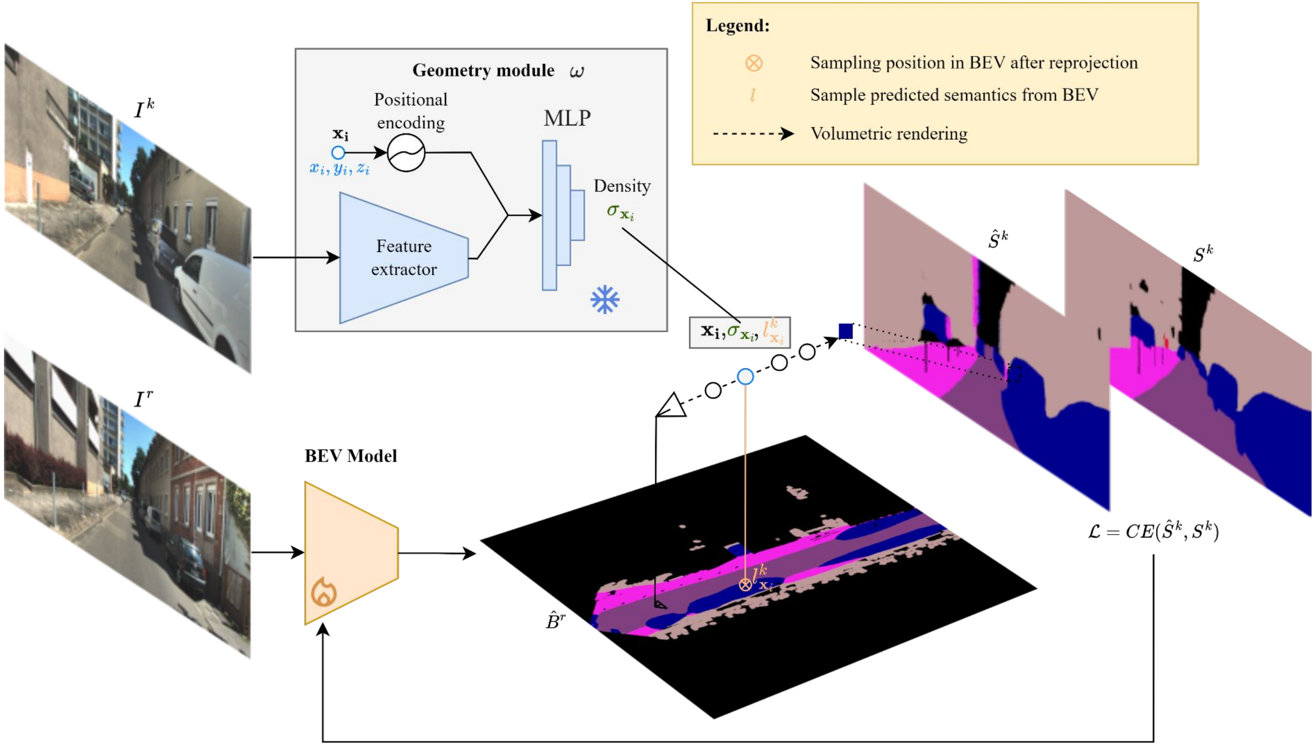
Key ideas:

- Neural fields to extract 3D scene geometry
- Shift supervision to perspective view by sampling from BEV and rendering future frames





The *RendBEV* method



Piñeiro et al., *RendBEV: Semantic Novel View Synthesis for Self-Supervised Bird's Eye View Segmentation*, *under review*

RendBEV – Quantitative Results

We beat an unsupervised baseline (IPM)

BEV (%)	Method	Road	Sidewalk	Building	Terrain	Person	2-Wheeler	Car	Truck	mIoU
0	IPM [15]	58.39	23.07	12.55	32.47	0.58	**1.44**	11.61	**5.16**	18.16
	RendBEV(ours)	**68.34**	**33.27**	**33.26**	**44.60**	**1.23**	0.72	**32.37**	3.39	**27.15**

RendBEV – Quantitative Results

We beat an unsupervised baseline (IPM) and provide SotA results when used as pretraining with 1% of the data

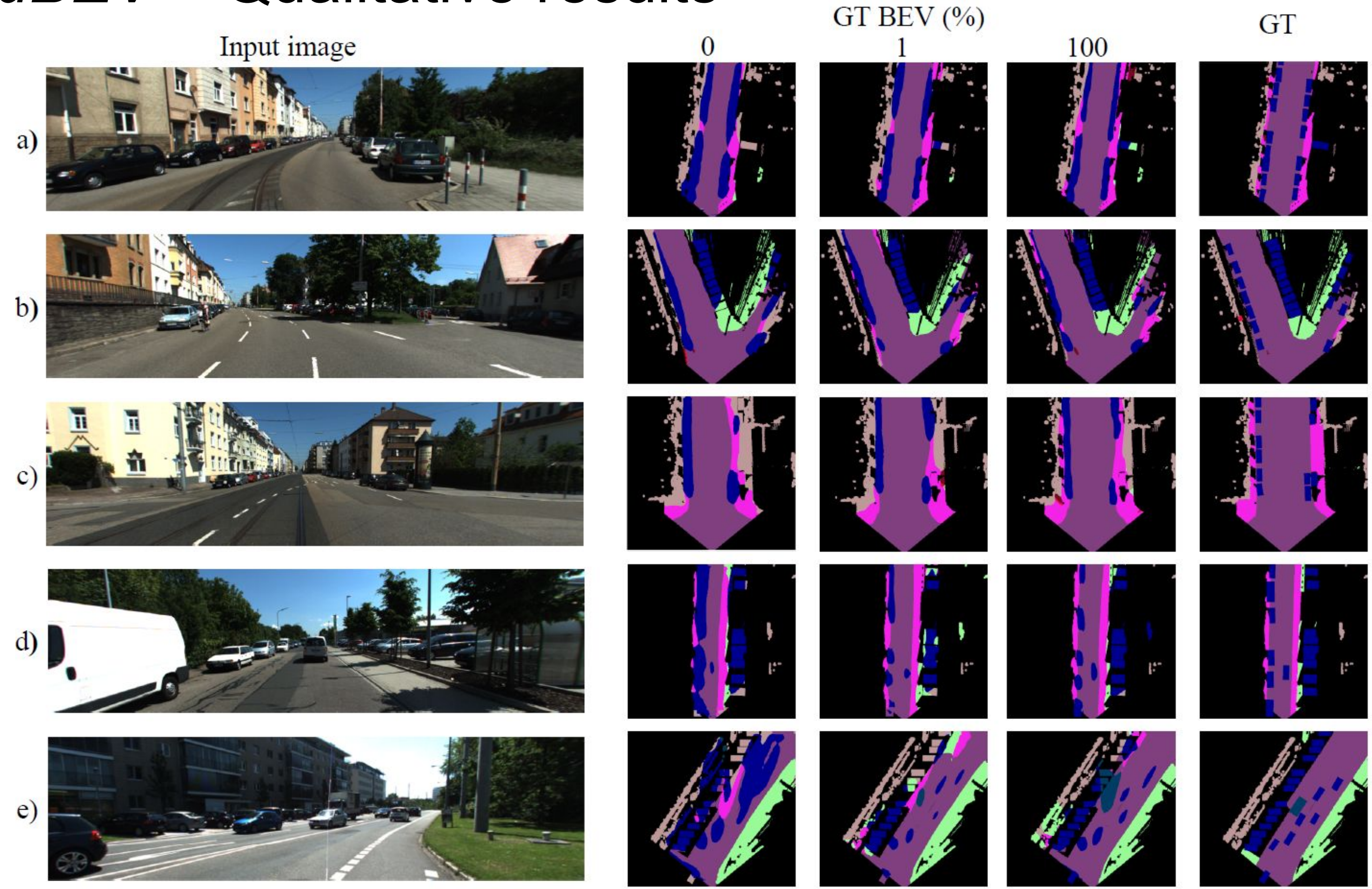
BEV (%)	Method	Road	Sidewalk	Building	Terrain	Person	2-Wheeler	Car	Truck	mIoU
0	IPM [15]	58.39	23.07	12.55	32.47	0.58	**1.44**	11.61	**5.16**	18.16
	RendBEV(ours)	**68.34**	**33.27**	**33.26**	**44.60**	**1.23**	0.72	**32.37**	3.39	**27.15**
1	SkyEye [5]	70.69	31.13	32.38	40.08	0.00	0.00	29.08	3.95	25.91
	RendBEV(ours)	**74.76**	**40.20**	**41.07**	**46.40**	**1.67**	**3.94**	**35.78**	**5.90**	**31.22**

RendBEV – Quantitative Results

We beat an unsupervised baseline (IPM) and provide SotA results when used as pretraining with 1% or 100% of the data

BEV (%)	Method	Road	Sidewalk	Building	Terrain	Person	2-Wheeler	Car	Truck	mIoU
0	IPM [15]	58.39	23.07	12.55	32.47	0.58	**1.44**	11.61	**5.16**	18.16
	RendBEV(ours)	**68.34**	**33.27**	**33.26**	**44.60**	**1.23**	0.72	**32.37**	3.39	**27.15**
1	SkyEye [5]	70.69	31.13	32.38	40.08	0.00	0.00	29.08	3.95	25.91
	RendBEV(ours)	**74.76**	**40.20**	**41.07**	**46.40**	**1.67**	**3.94**	**35.78**	**5.90**	**31.22**
100	TIIM [23]	63.08	28.66	13.70	25.94	0.56	6.45	33.31	8.52	22.53
	VED [14]	65.97	35.41	37.28	34.34	0.13	0.07	23.83	8.89	25.74
	VPN [17]	69.90	34.31	33.65	40.17	0.56	2.26	27.76	6.10	26.84
	PON [21]	67.98	31.13	29.81	34.28	2.28	2.16	37.99	8.10	26.72
	Simple-BEV [7]	70.66	35.50	34.67	41.18	1.04	2.11	38.24	12.42	29.48
	PoBEV [6]	70.14	35.23	34.68	40.72	2.85	5.63	39.77	14.38	30.42
	SkyEye [5]	72.82	38.27	40.86	**45.86**	**3.59**	**7.74**	41.37	9.74	32.53
	RendBEV(ours)	**74.83**	**40.98**	**41.80**	45.63	3.47	6.09	**45.55**	**16.74**	**34.39**

RendBEV – Qualitative results



Next steps

Expand **RendBEV**

- Experiment in more datasets and study generalization capabilities
- Architectural tweaks, adapt to different camera intrinsics

Target: journal paper in coming months

Advance towards **anticipation**

- Predict future position of objects in the Bird's Eye View
- Integrate with our self-supervision framework

Target: conference paper in coming months

Collaborations: viewpoint shift dataset

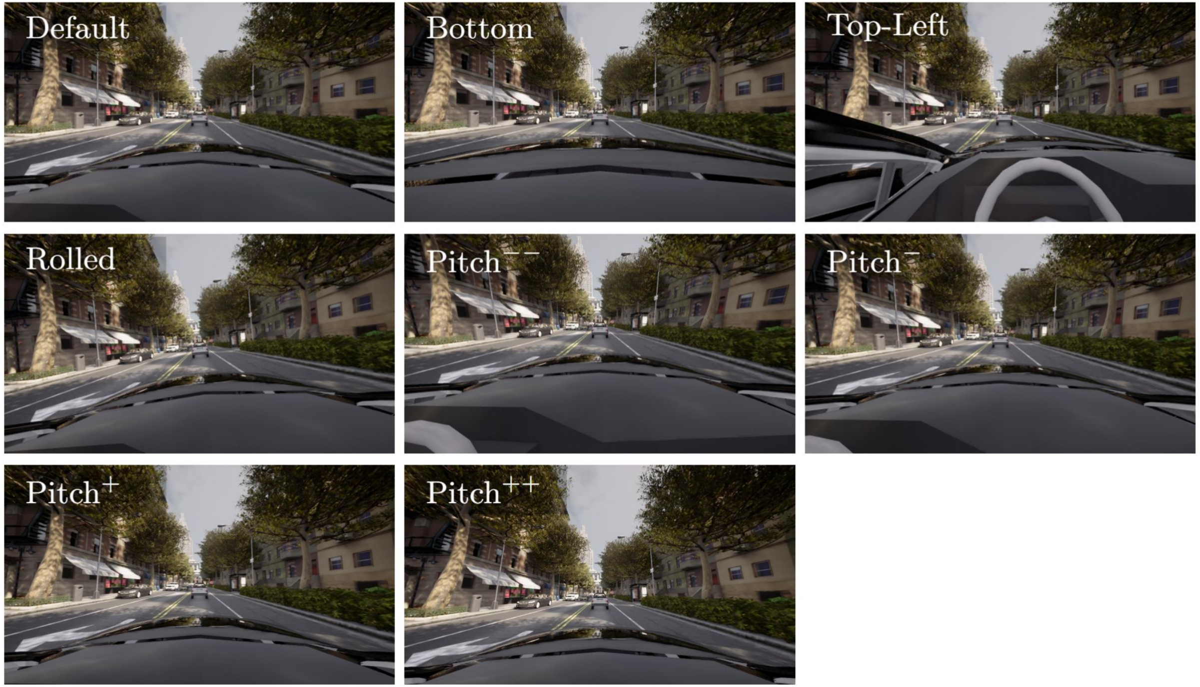
VisDepth: novel dataset with viewpoint shifts in dashcams and evaluation methodology to quantify impact of different camera positions and orientations on monocular depth estimation performance

Dataset available at: [ViewpointDepth]

[*A New Dataset for Monocular Depth Estimation Under Viewpoint Shifts*], Pjetri et al., presented at ECCV [Vision-Centric Autonomous Driving Workshop]

Collaboration: VS-Sim, Synthetic CARLA dataset

- Synthetic dataset generated with the CARLA simulator
- Evaluation of BEV semantic segmentation models against viewpoint shifts
- Work in progress, more soon!



VS-Sim: A Synthetic Dataset for Viewpoint Shift Robustness

Secondment: Anomaly detection with HEP data at UoM

- Hands-on experience with HEP data and software (my first time with ROOT 😁)
- Anomaly detection chats and literature review
- Testing autoencoder in Baler on data

Caterina and me at UoM

Conclusion

Main takeaways:

- Developed a method capable of performing **BEV semantic segmentation** with **no explicit BEV supervision** for the first time
- Collaborated on other projects to produce **new datasets for relevant tasks**
- Pivoting towards future prediction and danger anticipation

Thank you for your attention!

Rendering (maybe add a NeRF or BtS video or something)

$$\alpha_i = \exp(1 - \sigma_{\mathbf{x}_i}\delta_i)$$

$$T_i = \prod_{j=1}^{i-1}(1 - \alpha_j)$$

$$\hat{c} = \sum_{i=1}^{m} T_i \alpha_i c_{\mathbf{x}_i}$$