Supplementary Material – Learning from All Vehicles

Dian Chen   Philipp Krähenbühl
UT Austin

1. Detailed Infractions

In this section we report additional infraction numbers of our experiments in the main manuscript. All infractions are measured as the number of occurrences normalized per 1 kilometer traveled.

1.1. Comparison with state-of-the-art

Table 1 compares our method with baselines on the CARLA public Leaderboard [1]. Our method also leads the red light, offroad and blocked infraction numbers among all the methods.

1.2. Ablation study

Table 5 studies the effects of our key design choices. We additionally compare to a variant where we removes PointPainting [9] (LiDAR only b.b.). The rest of the backbone is the same. Both Driving Score and Route Completion of this variant are lower than the full LAV. This shows the benefit of multi-modal sensor fusion. Table 6 studies the degree to which training on other vehicles' experiences affect the driving performance. Table 7 studies different perception training schemes. Table 8 studies the effect of our iterative refinement module.

2. More Details

Figure 1 provdes an overview of architectures of M and M'. We detach the gradient of coarse trajectories predicted by M to remove any effect M_r might have on M_f during training.

Table 4 provides a list of hyperparameters. For all our experiments, we train our models on a 4 Titan Pascal GPU machine.

We use a ERFNet as our semantic segmentation architecture for PointPainting. We use the following image augmentations when we train the image-based semantic segmentation and brake prediction models: Gaussian Blur, Additive Gaussian Noise, Pixel Dropout, Multiply (scaling), Linear Contrast, Grayscale, ElasticTransformation.

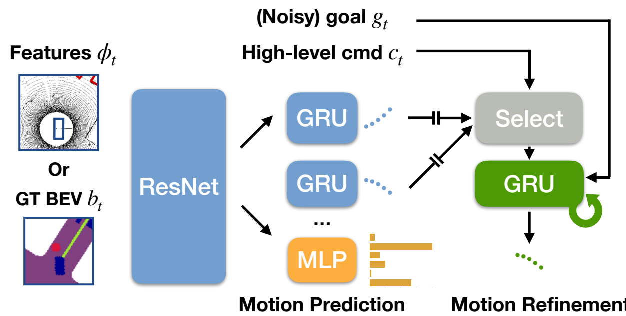


Figure 1. Overview of our motion model architectures. Motion prediction outputs future trajectories on different high-level commands for all vehicles, as well as their likelihoods. For the ego-vehicle, we additionally refine it using an iterative refinement module, conditioned on $\hat{g}_t$. We detach the gradient flow from refinement to prediction in order avoid undesired causal effect.

3. Onboard Sensors

Table 3 provide a detailed description of the sensor configurations for the ego-vehicle. We use the compass readings from IMU and GNSS readings to convert the target locations, represented in the GNSS format, to the ego-vehicle coordinate.

Note that our system does **not** rely on HD-Maps.

4. Dataset Statistics

Table 2 describes the dataset statistics on the training towns and their corresponding layouts. Our online leaderboard submission trains on all towns, whereas our local ablation models train on Town01, Town03, Town04 and Town06. They test on Town02 and Town05.

Figure 2 provides a visualization of the four routes on which the ablation models are tested.

5. License of Assets

We use the open source CARLA driving simulator [5]. CARLA is released under the MIT license. Its assets are under the CC-BY license.

Our teaser figure in the main paper uses a picture from the Waymo open dataset [7] The Waymo open dataset uses a

Rank	Method	Driving Score	Route Completion	Infraction Score	Vehicle Collisions	Pedestrian Collisions	Layout Collisions	Red light Violations	Offroad Infractions	Blocked Infractions
1	**LAV**	**61.85**	**94.46**	0.64	0.70	0.04	**0.02**	**0.17**	**0.25**	**0.10**
2	GRIAD	36.79	61.85	0.60	2.77	**0.00**	0.41	0.48	1.39	0.84
3	TransFuser+	34.58	69.84	0.56	0.70	0.04	0.03	0.75	0.18	2.41
4	Rails [2]	31.37	57.65	0.56	1.35	0.61	1.02	0.79	0.96	0.47
5	IARL [8]	24.98	46.97	0.52	2.33	**0.00**	2.47	0.55	1.82	0.94
6	NEAT [4]	21.83	41.71	0.65	0.74	0.04	0.62	0.70	2.68	5.22
7	Transfuser [6]	16.93	51.82	0.42	1.09	0.91	0.19	1.26	0.57	1.96
8	LBC [3]	8.94	17.54	**0.73**	**0.40**	**0.00**	1.16	0.71	1.52	4.69

Table 1. Comparison of our method and the state-of-the-art on the public CARLA leaderboard [1] (accessed Jan 2022). Methods are ranked by the driving score as the main metric. Driving Score, Route Completion, Infraction Score are higher the better, whereas the rest are lower the better. Infractions are measured as number of occurences per kilometer traveled. We best all other methods by a wide margin. We significantly outperform the prior best entry by **24** points on the driving score, and **25** points on the route completion. We also lead the red light, offroad and blocked infraction numbers among all the methods.

Town Name	Town Layout	Number of Frames
Town01	small, EU town	46559
Town02	small, EU town	63564
Town03	large, US town	51896
Town04	large, US town	46244
Town05	large, US town	51489
Town06	large, US town&highway	41812
Town07	small, US rural	55465
Town10	small, US city center	42747
Total		**399776**

Table 2. Number of frames and layouts of the training towns.

Count	Modality	Shape	Note
1	LiDAR	$\mathbb{R}^{L\times 4}$	Velodyne-64
1	RGB	$\mathbb{R}^{3\times 288\times 480}$	FOV=40°
3	RGB	$\mathbb{R}^{3\times 288\times 256}$	FOV=64° each 60° apart
1	IMU	–	
1	GNSS	$\mathbb{R}^2$	
1	Speedometer	$\mathbb{R}$	

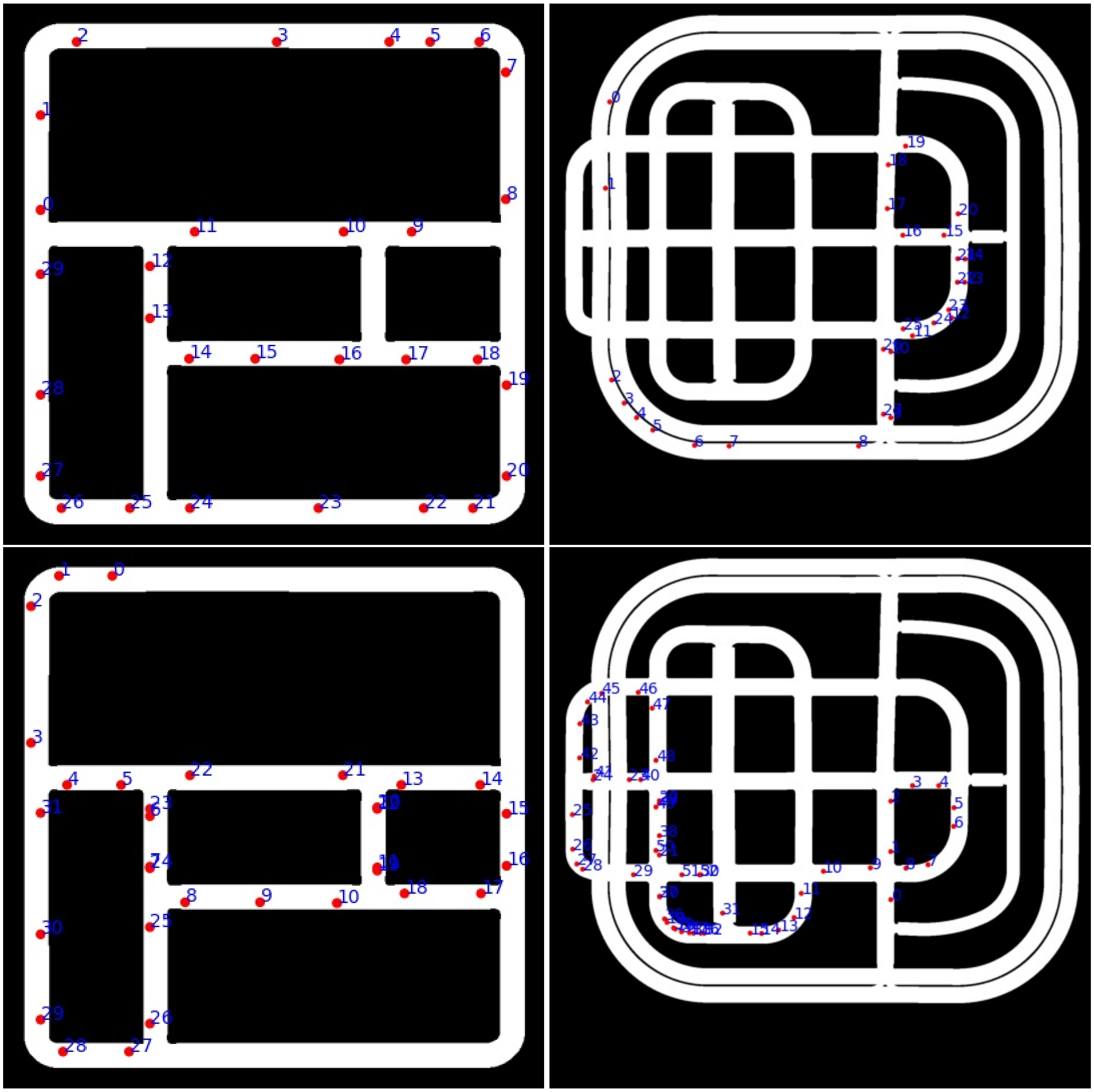
Table 3. Configuration of our ego-vehicle's on-board sensors. The four RGB cameras are mounted at $x = 1.5$m, $y = 0$m, $z = 2.4$m with respect to the ego-vehicle's centroid.

customized non-commercial license[1]. Part of our codebase uses the official ResNet implementation. Its codes are under the MIT license.

[1]https://waymo.com/intl/en_us/dataset-download-terms/

Stage	Hyperameter	Values
Privileged Motion	batch size	512
	learning rate	3e-4
	others weight λ_{other}	0.5
	command weight λ_{cmd}	0.1
Perception & Distill.	batch size	32
	learning rate	3e-4

Table 4. List of hyperparameters.



(a) Town02 (b) Town05

Figure 2. Visualization of the test routes in unseen towns of the ablation models.

	Driving Score	Route Completion	Infraction Score	Vehicle Collisions	Pedestrian Collisions	Layout Collisions	Red light Violations	Offroad Infractions	Blocked Infractions
LAV	**45.20** $\pm$ 6.35	**91.55** $\pm$ 5.61	**0.49** $\pm$ 0.06	**0.92** $\pm$ 0.42	0.00 $\pm$ 0.00	**0.33** $\pm$ 0.50	0.28 $\pm$ 0.28	**0.27** $\pm$ 0.01	**0.01** $\pm$ 0.02
Ego-vehicle only	38.56 $\pm$ 1.86	84.76 $\pm$ 5.12	0.46 $\pm$ 0.02	1.17 $\pm$ 0.50	0.00 $\pm$ 0.00	1.82 $\pm$ 0.06	0.34 $\pm$ 0.20	0.37 $\pm$ 0.09	0.09 $\pm$ 0.08
No distillation	28.23 $\pm$ 2.27	81.05 $\pm$ 6.04	0.36 $\pm$ 0.04	2.08 $\pm$ 0.34	0.00 $\pm$ 0.00	7.87 $\pm$ 0.15	**0.21** $\pm$ 0.04	1.01 $\pm$ 0.13	0.05 $\pm$ 0.05
LiDAR only b.b.	26.37 $\pm$ 2.62	74.96 $\pm$ 4.21	0.31 $\pm$ 0.04	6.51 $\pm$ 3.03	0.00 $\pm$ 0.00	0.02 $\pm$ 0.03	0.26 $\pm$ 0.23	1.56 $\pm$ 0.71	0.09 $\pm$ 0.04

Table 5. Driving performance ablation of the key components of our approach on test towns. Infractions are measured as number of occurrences per kilometer traveled. Mean and standard deviation are computed over three runs. All models are the same despite the ablated option.

Vehicles Range	Driving Score	Route Completion	Infraction Score	Vehicle Collisions	Pedestrian Collisions	Layout Collisions	Red light Violations	Offroad Infractions	Blocked Infractions
$\leq$ 5m	**46.06** ± 1.70	88.77 ± 1.01	0.51 ± 0.02	1.27 ± 0.22	0.00 ± 0.00	**0.05** ± 0.09	0.43 ± 0.11	0.69 ± 0.09	0.11 ± 0.10
$\leq$ 15m	45.20 ± 6.35	**91.55** ± 5.61	0.49 ± 0.06	0.92 ± 0.42	0.00 ± 0.00	0.33 ± 0.50	0.28 ± 0.28	**0.27** ± 0.01	**0.01** ± 0.02
$\leq$ 25m	37.42 ± 3.09	89.56 ± 5.61	**0.61** ± 0.12	**0.85** ± 0.26	0.00 ± 0.00	0.61 ± 0.12	**0.23** ± 0.13	0.43 ± 0.07	0.06 ± 0.10

Table 6. Driving performance in test towns of models trained with different range of other vehicles. All models are the same except for other vehicles' maximum range used during training.

Perception Training	Driving Score	Route Completion	Infraction Score	Vehicle Collisions	Pedestrian Collisions	Layout Collisions	Red light Violations	Offroad Infractions	Blocked Infractions
None	8.47 ± 0.83	9.34 ± 0.35	**0.90** ± 0.07	2.37 ± 2.08	0.00 ± 0.00	**0.00** ± 0.00	**0.19** ± 0.32	**0.15** ± 0.26	**0.00** ± 0.00
Joint	28.36 ± 2.11	79.58 ± 4.99	0.34 ± 0.02	1.65 ± 0.72	0.00 ± 0.00	7.75 ± 1.70	0.45 ± 0.32	0.49 ± 0.03	0.24 ± 0.21
Staged	**45.20** ± 6.35	**91.55** ± 5.61	0.49 ± 0.06	**0.92** ± 0.42	0.00 ± 0.00	0.33 ± 0.50	0.28 ± 0.28	0.27 ± 0.01	0.01 ± 0.02

Table 7. Driving performance in test towns of models with different perception training scheme. All models are the same except for perception training.

Refinement Iteration	Driving Score	Route Completion	Infraction Score	Vehicle Collisions	Pedestrian Collisions	Layout Collisions	Red light Violations	Offroad Infractions	Blocked Infractions
$K = 0$	12.69 ± 2.86	35.85 ± 2.91	0.42 ± 0.03	9.15 ± 3.88	0.00 ± 0.00	9.50 ± 2.02	0.33 ± 0.36	4.11 ± 2.11	1.41 ± 1.22
$K = 1$	21.30 ± 1.10	85.90 ± 2.46	0.25 ± 0.01	2.09 ± 0.10	0.00 ± 0.00	5.58 ± 0.28	0.35 ± 0.26	0.93 ± 0.08	0.03 ± 0.03
$K = 5$	**45.20** ± 6.35	**91.55** ± 5.61	**0.49** ± 0.06	**0.92** ± 0.42	0.00 ± 0.00	**0.33** ± 0.50	**0.28** ± 0.28	**0.27** ± 0.01	**0.01** ± 0.02

Table 8. Driving performance ablation on the effect of motion refinement. All models are the same except for number of refinement iterations.

References

[1] Carla autonomous driving leaderboard (accessed november 2021). https://leaderboard.carla.org/leaderboard/, 2021. 1, 2

[2] Dian Chen, Vladlen Koltun, and Philipp Krähenbühl. Learning to drive from a world on rails. In *ICCV*, 2021. 2

[3] Dian Chen, Brady Zhou, Vladlen Koltun, and Philipp Krähenbühl. Learning by cheating. In *CoRL*, 2019. 2

[4] Kashyap Chitta, Aditya Prakash, and Andreas Geiger. Neat: Neural attention fields for end-to-end autonomous driving. In *ICCV*, 2021. 2

[5] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. Carla: An open urban driving simulator. In *CoRL*, 2017. 1

[6] Aditya Prakash, Kashyap Chitta, and Andreas Geiger. Multi-modal fusion transformer for end-to-end autonomous driving. In *CVPR*, 2021. 2

[7] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al. Scalability in perception for autonomous driving: Waymo open dataset. In *CVPR*, 2020. 1

[8] Marin Toromanoff, Emilie Wirbel, and Fabien Moutarde. End-to-end model-free reinforcement learning for urban driving using implicit affordances. In *CVPR*, 2020. 2

[9] Sourabh Vora, Alex H Lang, Bassam Helou, and Oscar Beijbom. Pointpainting: Sequential fusion for 3d object detection. In *CVPR*, 2020. 1