VII. Appendix

A. Extended Setup Details

Our onboard camera is attached to the hopper vehicle at the top edge of the windshield. The camera has a 90° field of view and is angled upwards 5° relative to the pitch of the vehicle matching established self-driving setups [28], [10]. The "hopper" vehicle is equipped with a 900° wheel rotation and an 18.4m minimum turning radius at 40kph. With a steering input of ±1 at 40kph, the hopper can change its yaw relative to the world frame by 29.96° per second.

Our DAVE2 weights were initialized according to a Xavier uniform distribution and biases were initialized to zero to match related work. We applied data augmentation during training to prevent overfitting and extend the dataset. It consumes camera images to steer the vehicle around a racetrack.

The physical size of the simulated billboard is 4.5m × 4.5m × 0.5m which corresponds with real-world billboard sizes [60].

Note that the choices of noise variance parameter values do not reflect the severity of the levels of noise the vehicle might encounter in a real-world scenario. Rather, they are a common approach to improve the training result's generalizability such that the adversarial patch generation process does not overfit to individual pixels in the image sequence.

B. Full Comparison of DeepBillboard, DBB+, and DeepManeuver

Table IV is an extended version of Table I, with the inclusion of the expected AAE. This is the measure that most closely resembles the performance metric in the DeepBillboard paper. It measures the expected average angle error calculated according to the final collection sequence. In contrast, AAE measures the average angle error for all images seen during test runs. Expected AAE is larger than test AAE in all cases. This is expected because tests are conducted in nondeterministic environments and any given test trajectory will not be identical to the collection trajectory, resulting in differing images sequences and viewing the attack surfaces from different angles and locations. All of these factors have an effect on perturbation efficacy.

Expected AAE can help explain the under- or over-optimization of the perturbation in some instances. Scenario 1 is a good example of this. At all resolutions in Scenario 1, DBB+ has success rates in the single digits with a decrease across success rate and ADOT as resolution increases, and minimal change to test sequence AAE. However, expected AAE derived from the collection sequence (see extended results in repo) shows an increase across resolutions of (0.121, 0.167, 0.200), despite this dropoff in test sequence metrics. This could be a sign of overfitting the perturbation to the collection sequence. In other words, a higher resolution provides more freedom to manipulate the pixels of the billboard, and thus DBB+ is over-optimizing the perturbation for the collection sequence at the expense of the test sequence. As another example, Scenario 3 shows an increasing expected AAE and decreasing AAE for increasing resolutions in DeepManeuver. This indicates overfitting of the perturbation to the collection sequence. The increases in resolution provide DeepManeuver with increased freedom to manipulate the billboard appearance, overfitting it to the collection sequence and resulting in suppressed success rate and AAE. ADOT shows a general downward trend as well.

Moreover, expected AAE can help explain the robustness of the dataset. Scenario 2 shows exceptionally high expected AAE across all techniques. For reference, the highest possible AAE would be ±2 and would require all unperturbed steering values to be ±1. The extremely high expected AAE values across all techniques in Scenario 2 shows a distinct weakness of the DNN in this area of the track that is not seen in any other scenario in Table IV. This is likely due to overfitting, as straight roads are the most common topology in the training set of the DNN, creating an area of the dataset manifold with low error and high variance that is susceptible to adversarial inputs. By comparison, most other scenarios produce expected AAE values between 0.2 and 0.5 across all techniques. The other notable exception aside from Scenario 2 is Scenario 4, which is a right turn maneuver on a right-handed curve road topology. This makes sense as well in terms of overfitting, as right turns are the second-most frequent road topology in the DNN training dataset.

C. Positive versus Negative AAE

As an extension of the discussion of results for Scenario 1 in Section III.C, Figure 7 shows the boxplot of all AAE values for all test runs of DeepManeuver on Scenario 1 using a billboard resolution of 10×10. Scenario 1 is a left turn maneuver on a straight road topology. The left turn means that, for a maneuver to be successful, we would expect to see a positive AAE. Figure 7 shows that the median value is -0.011, close to 0, with a 2nd quartile ranging from -0.105 to -0.011 and a larger 3rd quartile going from -0.011 to 0.147. The left whisker starts at -0.482 and ends at -0.105 and the right whisker starts at 0.147 and ends at 0.524. Out of a set of 37,531 angle error values, 3,769 (10.0%) are outliers, with 3,491 of those outliers being negative values. This puts the average value slightly negative, as the top of Tables I and IV show. However, Figure 7 confirms that the bulk of the test run angle error values are positive, despite there being some individual negative error values that are larger in magnitude as Figure 7 shows. This suggests the need for a performance metric in addition to AAE, perhaps showing the variance of error values or the exclusion of individual extreme values in calculation of the average angle error.

D. Fractional Factorial Study of Cut-on and Resolution

As part of our extended results, we performed two fractional factorial experiments to further explore the effects of cut-on, billboard resolution, and noise variance. Performance metrics were calculated identically to Table I and Table IV. For each technique and parameter combination, 50 perturbations were generated, and for each of those perturbations, 10 test runs were executed to determine performance metrics of success rate, ADOT, and AAE.

		DeepBillboard **`noise_var=0, cut-on=28m`**				**DBB+** **`noise_var=`$\frac{1}{15}$`, cut-on=28m`**				**DeepManeuver** **`noise_var=`$\frac{1}{15}$`, cut-on=28m`**			
	`Resol.`	**Success rate**	**ADOT**	**AAE**	**expt. AAE**	**Success rate**	**ADOT**	**AAE**	**expt. AAE**	**Success rate**	**ADOT**	**AAE**	**expt. AAE**
Scenario 1: left turn on straight road	**5×5**	0.6%	2.18	**0.000**	0.213	8.8%	2.42	-0.015	0.121	**26.4%**	* **3.06**	-0.056	0.261
	10×10	0.0%	1.99	**0.013**	0.292	5.8%	2.21	0.012	0.167	* **30.4%**	* **2.62**	-0.017	0.296
	15×15	0.0%	1.96	**0.022**	0.338	0.2%	2.01	0.011	0.200	* **21.0%**	* **2.71**	-0.033	0.351
Scenario 2: right turn on straight road	**5×5**	99.3%	2.35	-0.198	-0.955	**99.4%**	2.31	-0.201	-0.951	99.2%	* **2.50**	* **-0.204**	-0.914
	10×10	99.4%	**2.35**	-0.180	-0.998	99.6%	2.34	-0.201	-0.996	* **99.8%**	2.34	* **-0.215**	-0.862
	15×15	98.7%	2.33	-0.176	-1.072	**99.0%**	2.33	-0.181	-1.077	88.6%	* **2.35**	* **-0.186**	-0.936
Scenario 3: left turn on right curve	**5×5**	0.2%	1.09	0.012	0.363	6.4%	1.27	* **0.034**	0.306	* **54.6%**	* **1.86**	0.060	0.241
	10×10	8.8%	1.11	0.022	0.443	* **12.9%**	1.32	* **0.030**	0.477	10.8%	**1.35**	0.021	0.290
	15×15	**17.6%**	1.27	**0.032**	0.495	16.6%	1.31	0.030	0.528	3.4%	* **1.43**	0.015	0.354
Scenario 4: right turn on right curve	**5×5**	47.2%	2.55	-0.367	-0.643	99.6%	3.09	-0.766	-0.681	* **100.0%**	* **3.17**	* **-0.819**	-0.650
	10×10	15.0%	2.48	-0.283	-0.654	79.0%	2.69	-0.599	-0.649	* **94.2%**	**2.87**	* **-0.678**	-0.641
	15×15	6.3%	2.34	-0.258	-0.693	65.0%	2.50	* **-0.533**	-0.654	* **81.6%**	**2.60**	-0.515	-0.593
Scenario 5: left turn on left curve	**5×5**	1.4%	0.93	0.011	0.287	0.6%	0.90	0.009	0.279	* **9.2%**	* **1.65**	* **0.054**	0.245
	10×10	0.0%	1.22	0.025	0.405	0.0%	1.08	0.017	0.407	**1.4%**	* **1.54**	* **0.047**	0.267
	15×15	0.0%	1.27	0.026	0.548	0.0%	1.21	0.021	0.551	0.0%	* **1.38**	* **0.033**	0.356
Scenario 6: right turn on left curve	**5×5**	67.6%	1.95	-0.111	0.287	**88.2%**	**2.19**	**-0.152**	0.273	85.6%	2.17	-0.151	0.269
	10×10	8.2%	1.39	0.041	0.363	8.0%	1.40	0.041	0.349	* **31.2%**	* **1.64**	* **-0.015**	0.324
	15×15	0.0%	1.33	0.056	0.430	0.2%	1.34	0.056	0.421	* **6.0%**	* **1.40**	**0.047**	0.422

TABLE IV: All techniques on 3 road topologies at 3 resolutions. Metrics are the success rate, AAE under test, and ADOT. The best-performing values per row are in bold. Statistically significant best-performing values are starred.

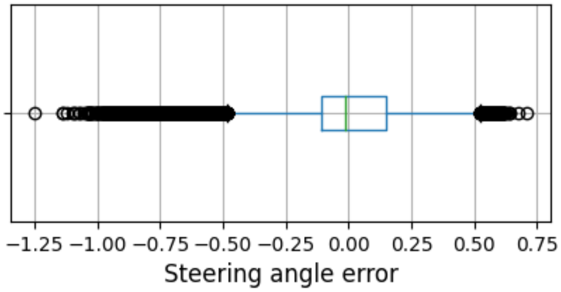


Fig. 7: All AAE for DeepManeuver on Scenario 1 using a 10×10 resolution.

		DBB+ **noise_var=1/15**		**DeepManeuver** **noise_var=1/15**	
		Success rate	**ADOT**	**Success rate**	**ADOT**
resol.=5	**cut-on=20m**	0.0%	1.91	**70.2%**	**2.88**
	cut-on=24m	1.2%	2.25	**62.6%**	**2.92**
resol.=15	**cut-on=20m**	0.0%	1.91	**0.6%**	**2.07**
	cut-on=24m	0.0%	1.93	**0.2%**	**1.97**

TABLE V: Fractional factorial experiment exploring the effects of cut-on and resolution on DBB+ and DeepManeuver.

		DBB+ **cut-on=24m**		**DeepManeuver** **cut-on=24m**	
		Success rate	**ADOT**	**Success rate**	**ADOT**
resol.=5	**noise_var=1/25**	0.8%	2.02	**45.2%**	**2.75**
	noise_var=1/15	1.2%	2.25	**62.6%**	**2.92**
	noise_var=1/5	14.6%	2.24	**92.2%**	**3.15**
resol.=15	**noise_var=1/25**	0.0%	1.91	0.0%	**1.94**
	noise_var=1/15	0.0%	1.93	**0.2%**	**1.97**
	noise_var=1/5	0.0%	1.97	**8.4%**	**2.16**

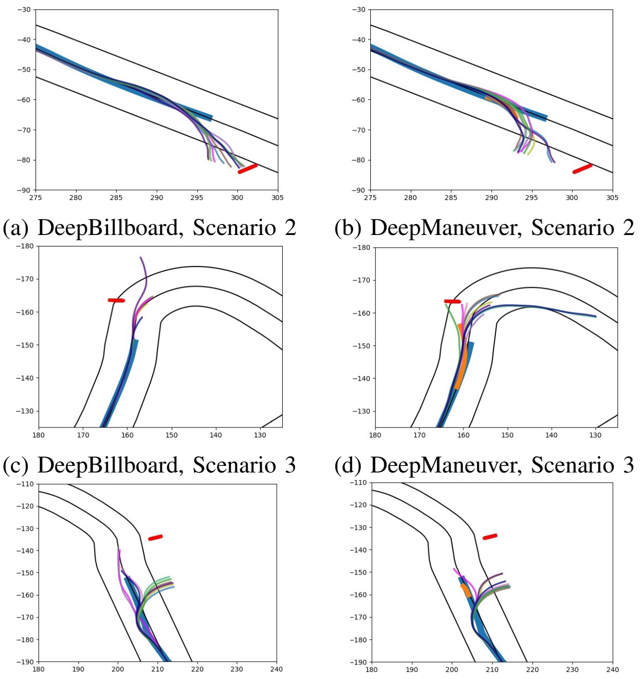
TABLE VI: Fractional factorial experiment exploring the effects of noise variance and resolution on DBB+ and DeepManeuver.

The first of these two studies studies the correlative effects of cut-on and resolution for two techniques, DBB+ and DeepManeuver. Table V shows the significant effects of cut-on and resolution. Although DeepManeuver performance is significantly better for all combinations, the change from resolution of 5×5 to 15×15 shows a 69.6pp and 62.4pp reduction in effectiveness. Similarly, ADOT is decreased by almost a meter when increasing resolution. Increasing resolution from 5×5 to 15×15 is a significant increase in perturbation degrees of freedom. Rather than 25 grid squares, it can manipulate 225 grid squares and that leads to a tendency to over-optimize. A decrease in steps available to PGD might be needed to accompany an increase in resolution in order to overcome this issue. This could be a way to optimize the technique and decrease the time to generate perturbations in future work.

DBB+ suffers from a shortened image and action sequence much more so than DeepManeuver. In Tables I and IV, DBB+ reaches 8.8% success rate for resolution=5×5 and 0.2% success rate for resolution=15×15 on Scenario 1. However, in Table V only one combination (cut-on=24m, resolution=5×5) has a success rate greater than 0.0% and an ADOT value above 1.93. This is likely because it has significantly fewer images to use to calculate the gradient to direct PGD, resulting in a perturbation that is underfit to the images captured during a given test run.

E. Fractional Factorial Study of Noise Variance

Table VI presents the second part of the fractional factorial study, showing the correlative effects of noise variance and resolution for two techniques, DBB+ and DeepManeuver. Table VI keeps cut-on at a constant 24m, the median value cut-on, for both techniques. As in Table V, all best-performing values belong to DeepManeuver. This could again be a feature of the earlier cut-on used here than in Tables I and IV. For example, the Scenario 1 row shows that DBB+ achieves an 8.8% and 0.2% success rate for resolutions 5×5 and 15×15, respectively. A 4m decrease in cut-on reduces those values to 0.8% and 0.0%, respectively.



(a) DeepBillboard, Scenario 2 (b) DeepManeuver, Scenario 2

(c) DeepBillboard, Scenario 3 (d) DeepManeuver, Scenario 3

(e) DeepBillboard, Scenario 6 (f) DeepManeuver, Scenario 6

Fig. 8: Perturbations generated by DeepBillboard and DeepManeuver on Topologies 1-3 for billboard resolution=10×10. Billboard is shown in red, the original trajectory is thick blue, DeepManeuver collection sequence is thick orange, and test trajectories are thin lines.

Conversely, DeepManeuver performance improves significantly with a shorter cut-on and a lower resolution, especially when coupled with a higher noise variance. In the rows pertaining to resolution=5×5, we can best see the impact of noise variance. Higher noise variance shows a drastic increase in success rate by 47pp and ADOT by 0.4m when increasing noise variance from $\frac{1}{25}$ to $\frac{1}{5}$. A similar, though not nearly as drastic, improvement in performance metrics can be seen in the DBB+ column. Here the highest success rate is 14.6%, increased by 13.8pp from 0.8% for the lowest noise variance. ADOT showed similar values for noise variances $\frac{1}{15}$ and $\frac{1}{5}$, increased about 0.23 ±0.02m from the lowest noise variance of $\frac{1}{25}$.

Higher resolution cut all success rates down to single digit percentages, although higher noise variances brought the success rate from 0.0% up to 8.4%. Still, ADOT was reduced to to 2.02m on average across all noise variances, which is not enough to have the vehicle depart from the road surface of an approximately 8m wide road. Overall, DeepManeuver shows a clear advantage over DBB+ in that it can produce perturbations that force the desired maneuver under a variety of parameter combinations, cutting down the search space needed to tune the technique.