A. Supplementary Material

Section A.1 of this supplementary material gives a deeper insight into the noise factor distribution of simulated data generated using Unreal Engine[4] for the targeted synthetic data augmentation case studies of parking slot detection and traffic lane detection. Since simulated data used for the third case study of monocular depth estimation was sampled from the publicly available virtual KITTI [8] dataset with source details provided in the main paper, we do not include any additional statistics here. Example simulated images used for the depth estimation task are shown in Fig. 15 in Section A.5.

Qualitative results of targeted synthetic data augmentation are also included in this supplementary material in Sections A.2, A.3 and A.4 for the tasks of slot detection, lane detection and depth estimation respectively. Fig. 10 shows qualitative results of cross-dataset generalization experiments from the paper for the task of parking slot detection. Note the significant improvement in the number of true positives and their confidence scores as we move from left to right with the leftmost column showing results from the model trained on 100% real data, middle column showing results from the best model trained on a mix of real and simulated data (A + S) and right most column showing results from the best model trained on a mix of real and sim2real data (A + G). Fig. 11 shows qualitative results of the cross-dataset generalization experiments from the paper for the task of lane detection. The baseline model trained on 100% real CULane [29] data (first column in Fig. 11) results in lots of false negatives (highlighted in red squares) and false positives (highlighted in yellow squares) when tested on the TuSimple[5] dataset. Among the models trained with a mix of real and simulated data from Unreal Engine (A + S), the best model is the one trained on 70% real and 30% sim data (second column in Fig. 11) and results in fewer false negatives as compared to the baseline, but the number of false positives goes up. Overall, the best results are obtained with a model trained on 40% real and 60% sim2real data (A + G) with a significantly reduced number of false positives and negatives (last column in Fig. 11). Fig. 12 shows qualitative results of the cross-data generalization experiments for the single-image depth task. The best result is achieved by the model trained on a mix of 40% real and 60% sim data (A + S). For sim2real data augmentation, quantitatively, the best result is achieved by the model trained on a mix of 50% real and 50% sim2real data (A + G). However, for the sake of fair comparison, results in Fig. 12 are shown with the A + G model with 60% synthetic data. As highlighted by the zoomed-in section within each depth map, it can be clearly seen that training on a mix of real and simulated data improves the quality of depth map, especially around periphery of the vehicle silhouettes (Row 3). Moreover, adding sim2real data to the real dataset improves the quality of the predicted depth maps even further (Row 4).

Section A.5 provides additional quantitative results for the three tasks. In particular, Table. 7 provides additional quantitative insights into the role of synthetic data augmentation in improving the number of true positives and false positives for the slot detection task. Table 6 shows how increasing dropout regularization does not help improve generalization performance of models trained on 100% real data. Fig. 14 and Table 8 together provide a summary of cross-dataset testing results for lane detection models trained on TuSimple and tested on CULane. Consistent with the results in the main paper from models trained on CULane and tested on TuSimple, synthetic data augmentation helps deflate inherent bias in the TuSimple dataset and improve cross dataset generalization performance. Fig. 16 provides additional quantitative results from the cross-dataset generalization experiments for the task of depth estimation.

[4] https://www.unrealengine.com/en-US/
[5] https://github.com/TuSimple/tusimple-benchmark/tree/master/doc/lane_detection

A.1. Noise Factor Distribution of Simulated Data

The simulated data for both the parking slot detection and lane detection tasks was generated using an in-house Unreal Engine-based pipeline. Table 4 lists the noise factors along with their range of variation that were used to generate the simulated data for the lane detection task. For each noise factor, values were randomly sampled from a uniform distribution within the specified range. The road spline was randomly generated as well but with some checks to ensure the spline was smooth and did not break or loop back on itself. A total of 300 scenarios were thus created with 300 frames in each leading to a total of 90000 simulated images. Table 5 lists the noise factors varied for the parking slot detection task to generate 7 simulated scenarios resulting in a total of 15565 images for training.

Table 4: List of noise factors along with their range of variation used for generating the simulated data for the lane detection task. All factors except for sun intensity and cloud density are integer values. Factors for which no units are specified are unitless by design.

Noise Factor	Sun Intensity	Cloud Density	Sun Angle Pitch + Yaw (deg)	Traffic Density	Traffic Speed Std. Dev.	No. of Lanes	Speed Limit (mph)
Range	[0, 3]	[0, 2.5]	[0, 180]	[5,20]	[0, 30]	[1, 4]	[50, 90]

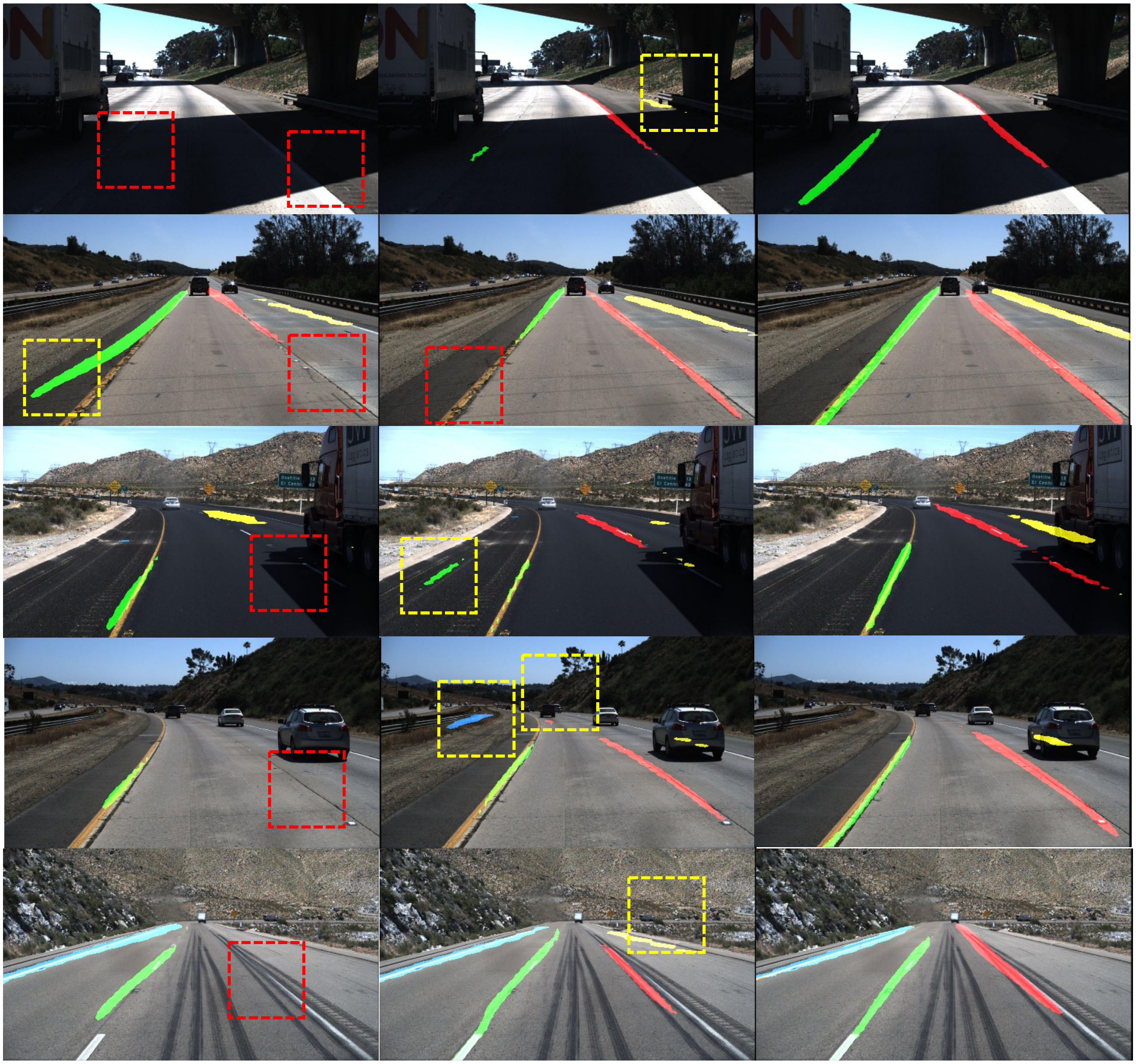
Table 5: List of noise factors varied for generating simulated data for the slot detection task. The header S.No. stands for scenarios numbers, which indicate the 7 different scenarios simulated based on the noise factors descriptions listed against them.

S.No.	**Weather**	**Parking Density**	**Line Color**	**Line Damage**	**Line Thickness**	**Time of Day**	**Sun Angle**	**Sun Intensity**	**Ground Material**	**True Negatives**	**Cloud Opacity**	**No. of Frames**
1	clear	heavy	yellow	1	0.05	10 am	(-60, 45)	5	cracked	trees, signs	1	2084
2	clear	medium	white	0.1	0.12	8 am	(-30, 45)	3	asphalt	trees, signs	1	2071
3	clear	medium	yellow	0.1	0.15	10 am	(-60, 90)	10	cracked	side walk	1	2427
4	clear	light	white	0.5	0.15	12 pm	(-90, 90)	8	cracked	side walk	3	2415
5	clear	light	yellow	0.2	0.1	2 pm	(-120, 60)	10	asphalt	side walk	1	2133
6	overcast	heavy	white	0.5	0.1	4 pm	(-160, 75)	1	cracked	side walk	3	1632
7	overcast	light	yellow	0.5	0.05	8 am	(-30, 60)	1	asphalt	grass	1	2803

A.2. Parking Slot Detection: Qualitative Results

Figure 10: Qualitative comparison of slot detection results on the held-out Parking B test set. Each row shows results on one example test image. Here, black boxes denote ground truth and green boxes are model predictions. The three columns from left to right show results from models trained on real Parking A dataset only, models trained on a mix of real and simulated data and models trained on a mix of real and sim2real data translated to look like Parking A real data. Note the number of true positives (TPs) and confidence scores increases from left to right. The first row shows how the number of TPs increases and the last row shows how confidence score goes from undetected to detected with a confidence improvement from 31% to 93% confidence for the same image. Confidence scores are best viewed by zooming into the relevant figure.

A.3. Traffic Lane Detection: Qualitative Results



(a) Inferences from model trained with 100% real data

(b) Inferences from model trained with 70% real + 30% sim

(c) Inferences from model trained with 40% real + 60% sim2real

Figure 11: Qualitative results from the cross-dataset generalization experiments (train on CULane, test on TuSimple) in the paper for the task of lane detection. Columns (a), (b) and (c) show the results from models trained on 100% real data, 70% real + 30% sim data and 40% real + 60% sim2real data respectively. The red dashed squares highlight false negatives and the yellow dashed squares highlight false positives.

A.4. Depth Estimation: Qualitative Results

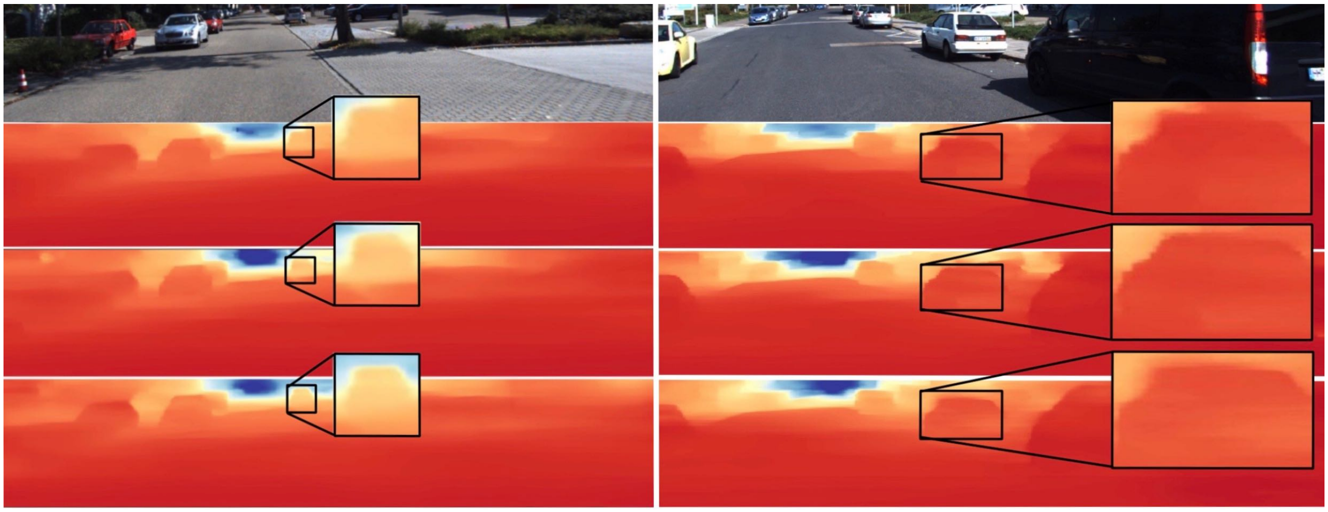


Figure 12: Qualitative comparison of results from single-image depth estimation models trained on the KITTI Odometry Sequence 00 + Virtual KITTI datasets and tested on the KITTI Tracking dataset. Row 1 shows two different test images from KITTI tracking dataset, Row 2 shows the corresponding depth estimation results from a model trained on 100% real data, Row 3 shows the depth estimation results from a model trained on 40% real + 60% sim and Row 4 shows the depth estimation results from a model trained on 40% real + 60% sim2real data.The addition of sim and sim2real data in the training mix (rows 3 and 4) improves the crispness of the depth estimation along vehicle boundaries as shown by the expanded insets.

A.5. Case Studies: Additional Quantitative Results

Table 6: Summary of MobileNetV2 SSD based parking slot detection cross-dataset testing results (train on Parking A, test on Parking B) for varying dropout percentages. F-Measure was 0% for all models.

Dropout	0%	0.5%	1%	10%	90%
False Positives (↓)	252	58	**3**	91	920

Table 7: Summary of cross-dataset testing results for parking slot detection. Here, TP and FP denote the number of true positives and false positives respectively.

Train	**Test**	**TP** (↑)	**FP** (↓)
A	B	0	252
A + S (40%)	B	89	**35**
A + G (50%)	B	**303**	149

(a) Real TuSimple Frame (b) Simulated (c) GAN Translated

Figure 13: Example real, simulated and sim2real translated images used for training models in Fig. 14 and Table. 8

Table 8: Summary of results in Fig. 14. Here, A and B denote the CULane and TuSimple datasets respectively. S denotes simulated images and G denotes the sim-to-real GAN translated equivalent of S. Example B, G and D images are shown in Fig. 13. For all these experiments, SCNN was trained on 512×288 images. For cross-dataset testing, CULane images were downsized and then padded (along height) to match the training resolution of 512×288 while simultaneously maintaining the original aspect ratio. For synthetic data augmentation rows, results are shown for the best model in terms of F-measure on cross-dataset testing in green for A + S and in blue for A + G.

Train	**Test**	**Precision** (↑)	**Recall** (↑)	**F-Measure** (↑)
B	B	80.2%	91.7%	85.6%
B + S (40%)	B	80.1%	91.5%	85.4%
B + G (20%)	B	79.1%	90.1%	84.2%
B	A	2.8%	3.7%	3.2%
B + S (40%)	A	**5.3%**	**6.6%**	**5.9%**
B + G (20%)	A	**5.9%**	**7.8%**	**6.8%**

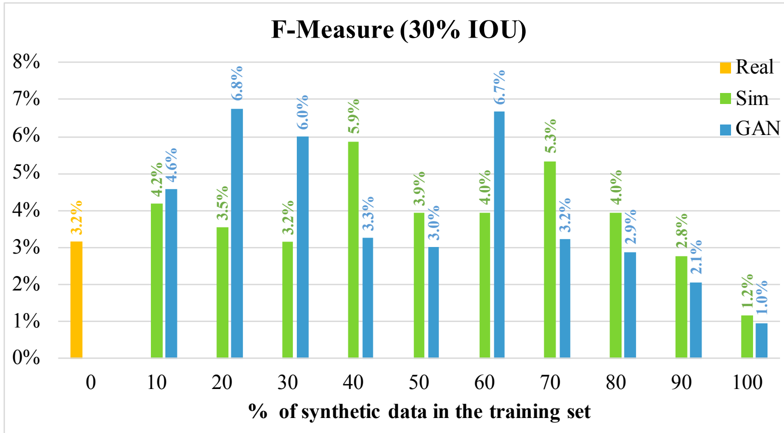


Figure 14: Plot of F-measure for cross-dataset testing of lane detection models trained on a mix of real TuSimple (dataset B) and synthetic images (either simulated or sim2real translated) and tested on real CULane (dataset A) images. As you move from left to right, the ratio of synthetic data in the training set increases.

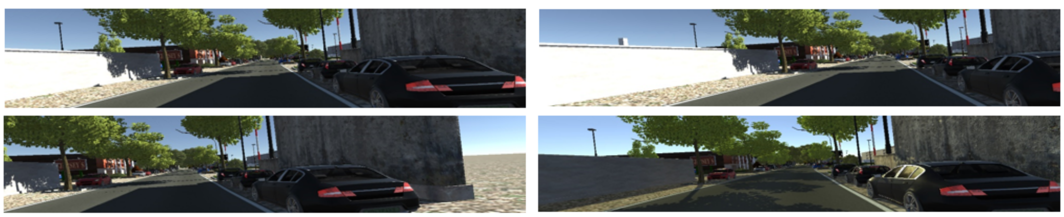


Figure 15: From top to bottom: Images from vKITTI Clone, 15L, 15R and Morning subsets used as simulation data for the single image depth task.

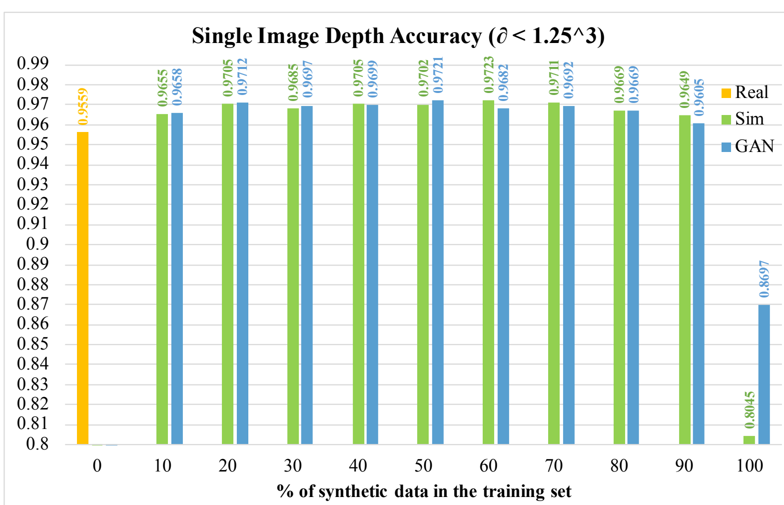


Figure 16: Accuracy results for the single image depth task (higher is better).