欠損ノイズが再現可能な Sim2Real による LiDAR セグメンテーション

Sim2Real LiDAR Segmentation with Synthetic Raydrop Noise

○ 学 宮脇智也 (九州大)　中嶋一斗 (九州大)　劉瀟文 (九州大)　岩下友美 (JPL)　正 倉爪亮 (九州大)

Tomoya MIYAWAKI, Kyushu University, miyawaki.tomoya.696@s.kyushu-u.ac.jp
Kazuto NAKASHIMA, Kyushu University
Xiaowen LIU, Kyushu University
Yumi IWASHITA, NASA/Caltech Jet Propulsion Laboratory
Ryo KURAZUME, Kyushu University

In 3D scene understanding tasks using LiDAR data, constructing training data poses a challenge due to its high annotation cost. To this end, annotation-free simulator-based training has recently been gaining attention, while the domain gap between simulators and real environments often leads to decreased generalization performance. This paper introduces a Sim2Real domain adaptation method that mitigates the domain gap by reproducing realistic raydrop noise onto labeled simulation data using deep generative models, enhancing its applicability to real-world scenarios. We demonstrate the effectiveness of our approach in multiple segmentation tasks.

Key Words: Deep Learning, LiDAR, Sim2Real

1　緒言

3D LiDAR センサは，レーザ光に基づく距離センサの一種であり，周囲環境の物体の位置や形状を点群データとして計測することができる．最も一般的な計測方式では，複数の仰俯角・方位角に対してパルスレーザ光を照射し，反射光を計測するまでの時間を距離に換算する．自律移動ロボットや自動運転車に広く利用されており，高精度な自己位置同定や障害物検出といった環境認識に不可欠である．特に，LiDAR センサの点群に基づく物体検出やセグメンテーション [1, 2] は，ロボティクス・コンピュータビジョン分野の中心的タスクとして取り組まれてきた．これらの環境認識タスクにおける解法の多くは，教師付きデータを用いた深層学習に基づく多層ニューラルネットワークを利用している．しかし，学習に必要な大量のラベル付き点群を作成するために，膨大な時間とリソースを要することが大きな課題となっている．代表的な大規模ベンチマークデータセットである SemanticKITTI [3] は，ラベル付け作業に 1700 時間以上費やされたと報告されている．

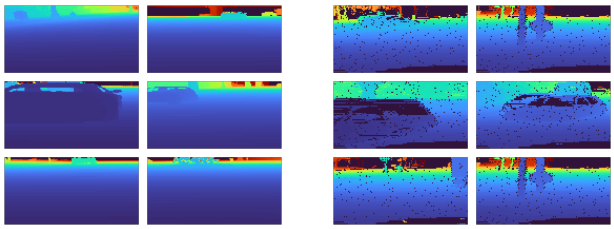
この問題に対する解決策の一つとして Sim2Real が注目されている．これはシミュレータ上で入力データと正解ラベルの対を自動的に合成し，合成データで学習した認識モデルを実環境に適応させる手法であり，高品質かつ大量のラベル付けが可能となる．しかし，一般に学習に利用する合成データとテストに利用する実データとの間のドメインギャップにより，実環境での汎化性能が大幅に低下することが多い．

学習データとテストデータのドメインギャップを解消する手法群はドメイン適応と呼ばれる．LiDAR データを用いた Sim2Real に関しては，これまでに，一般のドメイン適応タスクと同様に特徴分布を校正するアプローチ [2, 4] や実データの特徴を合成データに再現するアプローチ [1, 2, 4–6] が提案されている．本稿では，特に後者においてレーザ計測に伴う実データ特有の欠損ノイズを合成データに再現する手法に着目する．実験では，LiDAR 点群に対して点ごとの物体クラスを推定するセマンティックセグメンテーションを対象とする．特に，データ表現の異なる複数の実験設定において，欠損ノイズ再現による Sim2Real 効果を検証する．

2　関連研究

2.1　LiDAR セグメンテーション

LiDAR データを用いたセマンティックセグメンテーションにおける深層学習手法は，主に距離画像ベースの手法と点群ベースの手法の 2 つに分類される．距離画像ベースの手法では，LiDAR データを 2 次元の距離画像に変換し，これを入力として用いる．



Synthetic　　Real

Fig.1: Synthetic and real LiDAR range images （Black areas indicate ray-drop noises）

距離画像とは，レーザの仰俯角 ϕ・方位角 θ をそれぞれ縦軸・横軸とした 2D グリッドに対し，LiDAR センサから得られた距離値を球面投影して格納したものである．このアプローチの代表的なモデルとして SqueezeSegV2 [2] がある．SqueezeSegV2 は，畳み込みニューラルネットワーク (CNN) を用いた SquuzeSeg [1] の改良モデルである．新たにドメイン適応パイプラインを導入することで LiDAR データにおける欠損ノイズの影響を軽減する．これにより，SqueezeSeg に比べて実データに対する適応性能が向上しており，また，Sim2Real におけるドメイン適応にも有効であることが実験的に示されている．本研究では，後述の距離画像を用いたセグメンテーションタスクにおけるアーキテクチャとして SqueezeSegV2 を採用する．欠損復元の効果を測るため，欠損復元以外のドメイン適応手法は利用しない．一方，点群ベースの手法では，LiDAR データをそのままの 3 次元点群の形式で扱う．点群は 3 次元空間内の点の集合であり，各点は位置情報と属性情報を持つ．点群ベースの手法の代表的なモデルとして RandLA-Net [7] がある．RandLA-Net は，大規模な点群データを効率的に処理するために設計されている．ランダムサンプリングと局所的な特徴集約を組み合わせることで，計算コストを大幅に削減しながらも高いセグメンテーション精度を達成している．

2.2　欠損ノイズのモデル化

LiDAR データにおける欠損ノイズは，照射したレーザ光が物体表面で拡散・減衰することで，反射光の検知に必要な受光強度が十分に得られず発生する場合が多い．しかし，照射される物体の材質や入射角によって複雑に変化するため，物理パラメータを同定しシミュレータ上で再現するのは難しい．図 1 に距離画像表現された合成データと実データの例を示す．

関連研究の多くは，距離画像表現された合成データに対して欠損ノイズを再現する．特に関連研究 [1, 2, 4, 8] では共通のアプローチを採用しており，距離画像を隠蔽する乗法性二値マスク $m_i \in \{0,1\}$ がレーザ照射角 i ごとの生起確率 $p_i \in [0,1]$ のベルヌーイ分布 $m_i \sim \mathrm{Bernoulli}(p_i)$ に従って生じるものと仮定し，二値マスクを欠損ノイズとして利用する場合の確率 p_i のモデル化を検討している．本研究では，そのうちの 1 つである，Nakashima ら [8] が提案した距離画像表現に基づく LiDAR データの敵対的生成ネットワーク (GAN) によって欠損ノイズを再現する．

3 GAN を用いた欠損ノイズの再現

3.1 LiDAR 距離画像の GAN

本研究で使用する GAN [8] は，一般的な GAN と同様に，潜在変数 $z \sim N(0, I)$ から画像 x_z を生成する生成器と，生成データ x_z と実データ x_{real} を識別する識別器から構成される．一方，生成器はデータ x_z を直接生成するのではなく，欠損なし距離画像 r_z と欠損確率マップ p_z を分離生成する．次に，欠損確率マップ p_z に従ってサンプリングされる乗法性二値ノイズ m_z によって距離値をマスクすることで欠損あり距離画像 $x_z = m_z \odot r_z$ を表現する．本研究では，INR-GAN [9] に基づく生成器を KITTI Raw データセット [10] で学習する．

3.2 距離画像復元に基づくシーン潜在変数の推定

学習された GAN は，潜在変数 z を探索することで所与のデータを再構成することができ，一般に GAN inversion と呼ばれる．ここでは，前述の欠損なし距離値出力 r_z を合成データ $\hat{x}$ に近づけるように以下のマスク付き相対誤差を最小化する潜在変数 $\hat{z} = \arg\min_z \mathcal{L}_{\mathrm{rec}}$ を求める．

$$\mathcal{L}_{\mathrm{rec}} = \frac{\|\hat{m} \odot (1 - r_z/\hat{x})\|_1}{\|\hat{m}\|_1}, \tag{1}$$

ただし，$\hat{m}$ は合成データ $\hat{x}$ の欠損マスク，$\odot$ は要素積，$\|\cdot\|_1$ は L_1 ノルムである．本処理は Sim2Real タスクを学習する前にオフラインで実行する．

3.3 欠損確率マップに基づくノイズサンプリング

前項の最適化によって得られる $\hat{z}$ を用いて欠損確率マップ $p_{\hat{z}}$ を生成し，これを合成データ $\hat{x}$ の欠損ノイズ再現に利用する．具体的には，$p_{\hat{z}}$ を生起確率として，データごとの欠損マスクを $m_{\hat{z}} \sim \mathrm{Bernoulli}(p_{\hat{z}})$ によってサンプリングする．本サンプリングの計算コストは低いため，Sim2Real の対象タスクを反復学習する際にオンライン実行し，欠損の確率的な振る舞いを再現する．

4 実験

本章では，欠損ノイズを再現した合成データを用いてセマンティックセグメンテーションモデルを学習し，実データに対する適応性能を検証した結果を報告する．

4.1 実験設定

本実験で用いるデータセットを表 1 に示す．入力データ表現に応じて，2 種類の実験を設定する．1 つ目は，水平 360°・解像度 64×1024 の距離画像に対して前景・背景の 19 クラスを識別するタスクで，合成データの学習には SynLiDAR データセット [5]，実データの評価には SemanticKITTI データセット [11] を用いる．2 つ目は，1 つ目と同様の実験設定において，入力データ表現を距離画像から点群に変更する．具体的には，欠損ノイズが再現された合成データの距離画像を 3 次元空間上に再度マッピングし，得られた点群を入力に用いる．セマンティックセグメンテーションを行うモデルは，距離画像を入力とする実験 1 では SqueezeSegV2 [2]，点群を入力とする実験 2 では RandLA-Net [7] を用いる．結果の評価には，推定された領域と真値の領域の重畳度を示す intersection-over-union (IoU,%) を算出する．

4.2 比較手法

本稿では，欠損確率の再現方法について，4 種類の方法を比較する．(1) **再現なし**：欠損ノイズを重畳せずに合成データをそのままモデルに入力する．(2) **画素共通の頻度 (Global frequency)**：実データの学習セットの全画素からスカラーの頻度値を算出し，全画素に共通の欠損確率を格納した $H \times W$ の欠損確率マップを算出する．自動運転シミュレータの LiDAR モデル [12] に導入されている欠損モデルに類似する．(3) **画素ごとの頻度 (Pixel-wise frequency)** [2]：実データの学習セットから画素位置ごとの頻度値を算出し，データ共通の $H \times W$ の欠損確率マップを算出する．(4) **GAN 推論 (GAN inversion)**：第 3 章で紹介した GAN inversion を介してデータごとの欠損確率マップを生成する．

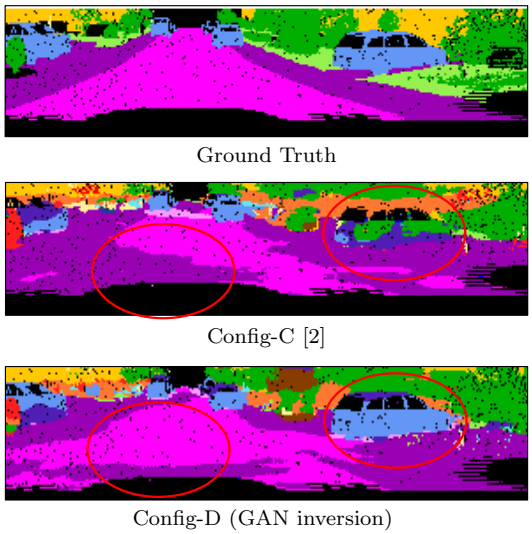
Table 1: Dataset

Dataset	Domain	Number of classes	Number of data
SynLiDAR [5]	Simulation	19†	198,396
SemanticKITTI [11]	Real	19†	43,552

4.3 実験結果

表 2 にそれぞれ距離画像，点群を入力とした SynLiDAR → SemanticKITTI の 19 クラスセグメンテーションの結果を示す．表には，実データで学習した場合の結果も付記している．

表 2 より，距離画像・点群のいずれのデータ表現を用いた場合においても，欠損ノイズ再現手法によって平均 IoU が向上していることがわかる．なかでも，GAN 推論手法によってデータごとの欠損確率を推定した手法がそれぞれ 17.6%，22.4%と最も高い．特に一部物体クラスの汎化性能は大きく向上しており，距離画像・点群のどちらの入力手法においても car クラスの IoU は大きく向上している．距離画像と点群を入力に用いた場合のセグメンテーション結果の一例をそれぞれ図 2，3 に示す．次点の画素ごとの頻度を用いた手法では分類精度の悪い領域が，GAN 推論手法では高精度に分類できていることが確認できる．この結果から，GAN を用いたデータごとの欠損ノイズの再現は，多クラスのセグメンテーションのような複雑なタスク，そして複数のデータ表現に対して汎用的に有効であると言える．GAN 推論の汎化性能が他の Sim2Real 手法よりも優れていた要因として，実データ特有の物体ごとの欠損の再現精度が高いことが考えられる．図 4 に再現された欠損ノイズと実データの例を示す．実データにおける car クラスは車窓や車体部分で欠損しやすいという特性があるが，この特性を最も再現できているのは GAN 推論手法である．一方で，実データで学習した場合の IoU とは大きく離れている．この要因として，第 3.2 章で紹介した GAN inversion のシーン復元精度が挙げられる．GAN 推論手法はデータごとに欠損ノイズを再現できるという点で他手法よりも優れているが，GAN inversion によるシーン復元精度が低い場合，シーンに対応して得られる欠損ノイズの再現精度も低くなる．実際に，図 4 に示し



Ground Truth

Config-C [2]

Config-D (GAN inversion)

Fig.2: Example of segmentation results (Range image based method) : Red circles indicate the areas that should be identified as a car or a road.

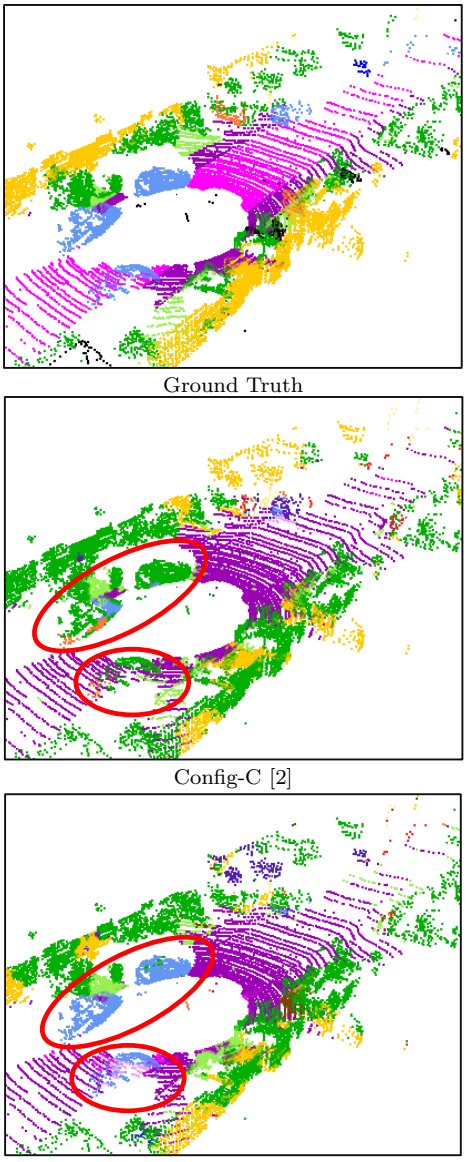


Fig.3: Example of segmentation results (Point cloud based method) : Red circles indicate the areas that should be identified as a car.

た例では物体レベルの特徴的な欠損が再現できているが，図 5 に示す例では車窓などの物体レベルの特徴的な欠損が再現できていない．今後の課題として，データごとの欠損ノイズ再現精度のばらつきを解消することが挙げられる．

5 まとめと今後の展望

本研究では，LiDAR 計測に伴う欠損ノイズの生起確率を学習済み GAN によって再現することで Sim2Real を行う手法を紹介した．LiDAR 距離画像・点群に基づくセマンティックセグメンテーションを対象とした Sim2Real 実験を行った結果，提案手法の有効性が示された．データ表現を点群に変更した場合も欠損ノイズ再現手法が有効であったことから，３次元物体検出など，点群を入力とする他のタスクへの適応も期待できる．一方，欠損ノイズの再現精度は，GAN を用いた距離画像の復元精度によって大きく左右されており，実データで学習した場合の IoU との大きな差がみられた．今後はより欠損ノイズ再現精度の高い手法を構築し，GAN 推論手法を上回る認識精度を達成することを目指す．

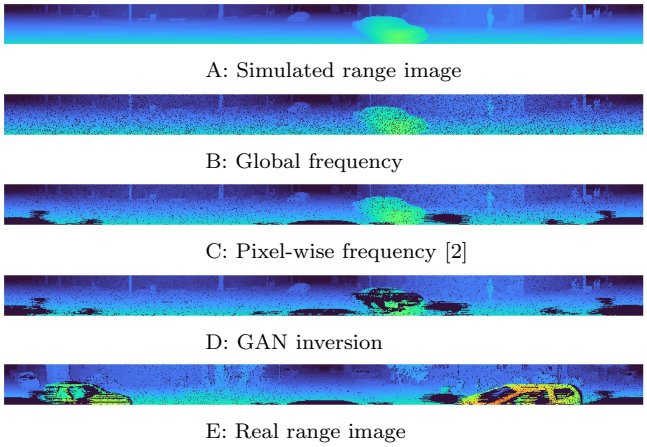


Fig.4: Examples of synthetic and realistic ray-drop noises

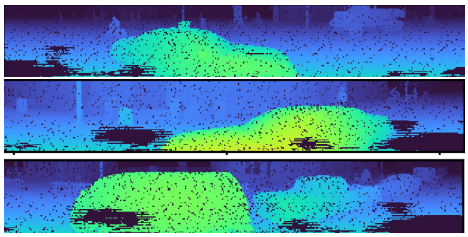


Fig.5: Examples of inaccurate synthetic ray-drop noises

謝辞

本研究の一部は JSPS 科研費 JP23K16974，JSPS 科研費 JP20H00230 の助成を受けたものである．

参考文献

[1] B. Wu, A. Wan, X. Yue, and K. Keutzer, “SqueezeSeg: Convolutional neural nets with recurrent CRF for real-time road-object segmentation from 3D LiDAR point cloud,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1887–1893, 2018.

[2] B. Wu, X. Zhou, S. Zhao, X. Yue, and K. Keutzer, “SqueezeSegV2: Improved model structure and unsupervised domain adaptation for road-object segmentation from a LiDAR point cloud,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4376–4382, 2019.

[3] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, J. Gall, and C. Stachniss, “Towards 3D LiDAR-based semantic scene understanding of 3D point cloud sequences: The SemanticKITTI dataset,” *The International Journal on Robotics Research (IJRR)*, vol. 40, no. 8-9, pp. 959–967, 2021.

[4] S. Zhao, Y. Wang, B. Li, B. Wu, Y. Gao, P. Xu, T. Darrell, and K. Keutzer, “ePointDA: An end-to-end simulation-to-real domain adaptation framework for LiDAR point cloud segmentation,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, pp. 3500–3509, 2021.

[5] A. Xiao, J. Huang, D. Guan, F. Zhan, and S. Lu, “Transfer learning from synthetic to real LiDAR point cloud for semantic segmentation,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, pp. 2795–2803, 2022.

[6] S. Manivasagam, S. Wang, K. Wong, W. Zeng, M. Sazanovich, S. Tan, B. Yang, W.-C. Ma, and R. Urtasun, “LiDARsim: Realistic LiDAR simulation by leveraging the real world,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11167–11176, 2020.

[7] Q. Hu, B. Yang, L. Xie, S. Rosa, Y. Guo, Z. Wang, N. Trigoni, and A. Markham, “Randla-net: Efficient semantic segmentation of large-scale point clouds,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020.

[8] K. Nakashima and R. Kurazume, “Learning to drop points for LiDAR scan synthesis,” in *Proceedings of the IEEE/RSJ*

Table 2: Comparison by methods of simultating ray-drop probability (SynLiDAR [5] → SemanticKITTI [11])

(a) Range image based method

Config	Trainig domain+Ray-drop prior	IoU (%) ↑																			
		car	bicycle	motorcycle	truck	other-vehicle	person	bicyclist	motorcyclist	road	parking	sidewalk	other-ground	building	fence	vegetation	trunk	terrain	pole	traffic-sign	mIoU
A	Simulation	2.1	0.1	0.7	0.9	1.9	1.3	4.0	0.0	3.6	1.8	24.6	0.0	32.4	4.3	31.1	8.7	5.1	5.3	1.7	6.8
B	Simulation+Global frequency	20.7	1.9	3.3	1.2	2.4	3.6	10.0	0.0	19.0	3.2	28.1	0.0	56.2	6.2	51.4	12.9	20.3	14.0	4.8	13.6
C	Simulation+Pixel-wise frequency [2]	24.8	1.8	6.8	1.2	3.6	2.9	13.3	0.0	31.1	2.0	29.8	0.0	53.1	4.8	49.3	14.8	19.5	11.4	4.4	14.4
D	Simulation+GAN inversion	33.1	3.6	9.9	1.1	4.0	4.5	22.8	0.2	41.0	4.9	32.6	0.0	55.1	7.1	54.2	17.0	20.5	16.1	5.9	**17.6**
E	Real	88.0	0.0	12.8	22.4	15.0	2.6	12.4	0.0	92.8	36.9	77.5	0.2	80.6	35.7	77.2	28.2	69.5	20.0	0.7	35.4

(b) Point cloud based method

Config	Trainig domain+Ray-drop prior	IoU (%) ↑																			
		car	bicycle	motorcycle	truck	other-vehicle	person	bicyclist	motorcyclist	road	parking	sidewalk	other-ground	building	fence	vegetation	trunk	terrain	pole	traffic-sign	クラス平均
A	Simulation	16.8	1.4	2.0	0.7	2.8	7.3	8.3	0.0	1.0	1.0	23.4	0.1	14.1	5.5	51.3	4.4	29.8	20.5	4.1	10.2
B	Simulation+Global frequency	28.7	6.9	5.2	0.3	0.3	12.1	50.8	0.4	11.4	3.4	27.7	0.1	45.2	5.6	50.9	20.6	36.1	27.9	2.3	17.7
C	Simulation+Pixel-wise frequency [2]	30.7	6.8	8.5	1.4	1.1	15.6	51.5	0.4	2.8	3.8	28.2	0.0	47.9	5.6	60.4	25.3	43.7	29.4	8.6	19.5
D	Simulation+GAN inversion	56.1	17.0	5.3	2.9	1.7	29.4	56.7	0.0	11.7	7.7	30.0	0.0	39.5	7.8	64.0	18.3	42.2	28.1	6.4	**22.4**
E	Real	90.3	10.9	18.9	28.0	29.0	43.8	47.8	0.0	90.2	39.7	76.0	0.8	81.4	43.4	81.9	53.2	73.1	49.4	26.1	46.5

International Conference on Intelligent Robots and Systems (IROS), pp. 222–229, 2021.

[9] I. Skorokhodov, S. Ignatyev, and M. Elhoseiny, "Adversarial generation of continuous images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10753–10764, 2021.

[10] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *The International Journal of Robotics Research (IJRR)*, vol. 32, no. 11, pp. 1231–1237, 2013.

[11] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall, "SemanticKITTI: A dataset for semantic scene understanding of LiDAR sequences," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.

[12] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proceedings of the Annual Conference on Robot Learning (CoRL)*, pp. 1–16, 2017.