欠損確率の再現による LiDAR Sim2Real の検討

○宮脇 智也（九州大学）　中嶋 一斗（九州大学）　Xiaowen Liu（九州大学）
岩下 友美（ジェット推進研究所）　倉爪 亮（九州大学）

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などの 3D シーン理解タスクでは，大量の学習データ作成に要するアノテーションコストが課題となっている．この問題に対しては，シミュレータによって自動的に合成したラベル付きデータを学習し，実環境に適応させる Sim2Real が注目されている．本稿では，シミュレータの合成データに対して欠損ノイズを再現することで実環境に適応させる Sim2Real 手法を紹介する．距離画像表現に基づく代表的なセグメンテーションタスクにおいて提案手法の有効性を示す．

1. はじめに

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．これらのシーン理解タスクにおける解法のほとんどは，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おり，SemanticKITTI [3] や nuScenes [4] に代表される大規模ベンチマークデータセットを用いて多くの成果が報告されている．一方で，データセットを構成する大量のラベル付き点群を作成するには，膨大な時間とリソースを要することが課題となっている．この問題に対して，シミュレータ上でラベル付き学習データを自動的に合成し，学習したモデルを実環境に転用する Sim2Real が注目されている．

学習データとテストデータの分布の不一致を解消するための手法群はドメイン適応と呼ばれ，Sim2Real もそのうちの一種である．これまでに，一般のドメイン適応タスク同様に特徴分布を校正するアプローチ [2, 5] 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, 5–7] が提案されている．本稿では，特に後者に着目し，レーザ計測に伴う実データ特有の欠損ノイズを再現する方法について議論する．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タスクを対象として，欠損ノイズ再現による Sim2Real の効果について報告する．

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．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．これらの欠損ノイズの分布の違いは，LiDAR データの Sim2Real において性能低下を引き起こすことが知られており [2]，欠損ノイズの正確な復元が重要となる．本章ではまず，LiDAR データの距離画像表現と欠損ノイズのモデル化について述べる．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 グリッドに各距離値を投影すると，各画素に距離値 r を持つ $H \times W$ サイズの距離画像が得られる．欠損ノイズのモデル化に関する関連研究の多くは，この距離画像表現に基づいており，本研究もこれに準ずる．図 1 に，代表的な LiDAR データセットにおける合成データと実データの距離画像の例を示す．

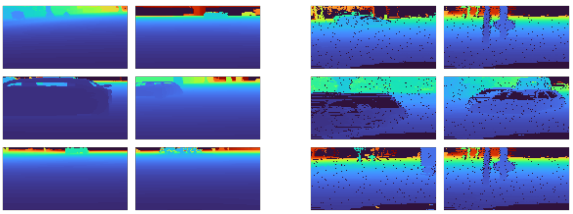

図 1 LiDAR 距離画像の合成データ [2] と実データ [8]
（一部の方位角を切り出して表示，黒色領域が欠損）

本稿では，関連研究 [1, 2, 5, 9] と同様に，これらの距離値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 のモデル化について検討する．

例えば，Wu ら [1, 2] は，距離画像の画素位置ごとの欠損頻度を現実のデータセットから予め算出し，学習時は欠損頻度からサンプリングした欠損ノイズを合成データに付与する方法を提案している．しかし，この手法は全ての合成データで同じ欠損頻度 p_i を用いるため，データに含まれる物体ごとの特徴を表現することはできない．また，Zhao ら [5] は，対応のない合成データと実データの集合を用いて，CycleGAN に基づく欠損推定ネットワークを学習している．Manivasagam ら [7] は，欠損マスクの推定を二値分類問題として定式化し，欠損推定ネットワークを学習している．これらの手法は二値分類に伴うソフトマックス出力を確率値 p_i として，欠損ノイズを擬似的にサンプリングしている．しかし，分類学習に基づく多層ニューラルネットワークのソフトマックス出力確率は，多くの場合学習データの実際の分布に則さないことが知られている [10]．一方，Nakashima ら [9] は，LiDAR 距離画像に対する敵

対的生成ネットワーク (GAN) の学習を通して，距離値と欠損確率 p の共起関係を学習する手法を提案している．

3. GAN inversion による欠損確率推定

第 2 章で述べた通り，学習データの頻度 [2] や二値分類に伴うソフトマックス出力確率 [7] に基づくモデル化では，サンプリングされる欠損ノイズと実際の欠損ノイズの分布が乖離する可能性がある．本稿では，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) [9] を利用した欠損確率の再現手法を紹介する．

3.1 LiDAR 距離画像の GAN

本研究で使用する GAN [9] 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 [11] に基づく生成器を KITTI Raw データセット [8] 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 $\hat{z} = \arg\min_z \mathcal{L}_{\text{rec}}$ を求める．

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

ただし，$\hat{m}$ は合成データ $\hat{x}$ の欠損マスク，$\|\cdot\|_1$ 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

前章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 $m_{\hat{z}} \sim \text{Bernoulli}(p_{\hat{z}})$ 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

本稿では，LiDAR 距離画像の画素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し，欠損ノイズ再現による Sim2Real 効果について報告する．

4.1 実験設定

表 1 に本実験で用いるデータセットを示す．対象クラス数に応じて，2 種類の実験を設定する．1 つ目は，水平 90°・解像度 64×512 の距離画像に対して車と歩行者の 2 クラスを識別するタスクで，合成データの学習には GTA-LiDAR データセット [2]，実データの評価には同じクラスでアノテーションされた KITTI Raw データセット [8] のサブセット [2] を用いる．2 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 [6]，実データの評価には SemanticKITTI データセット [12] を用いる．セマンティックセグメンテーションを行うモデルには，SqueezeSegV2 [2] 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．

表 1 使用するデータセット．† SynLiDAR [6] と SemanticKITTI [12] は共有する 19 クラスのみを利用.

データセット	ドメイン	クラス数	データ数
GTA-LiDAR [2]	Simulation	2	121,087
KITTI Raw [2]	Real	2	10,848
SynLiDAR [6]	Simulation	19†	198,396
SemanticKITTI [12]	Real	19†	43,552

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**：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，欠損確率とする．自動運転シミュレータの LiDAR モデル [13] 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** [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**：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．

4.3 実験結果

表 2 に GTA-LiDAR → KITTI Raw の実験結果，表 4 に SynLiDAR → SemanticKITTI の実験結果を示す．それぞれの表には，先行研究の結果 [5, 6] および実データで学習した場合の結果も付記している．

表 2（GTA-LiDAR → KITTI Raw）では，再現なしのクラス平均 IoU が 1.7%であるのに対して，欠損ノイズを再現した手法はいずれも 40%を超えており，特に GAN 推論によってデータごとの欠損確率を推定した手法が 46.3%と最も高い．さらに，複数のドメイン適応手法を組み合わせた先行研究 ePointDA [5] や実データで学習した場合の IoU を超えており，シミュレーション合成による学習データのスケールアップと提案手法による Sim2Real の有効性が定量的に示された．図 2 に，再現された欠損ノイズを示す．GAN 推論による欠損ノイズは物体に応じた欠損が表現できている．

また，表 3 では，第 3.3 章で紹介した GAN 推論に基づく欠損ノイズ生成において，反復学習時に毎回サンプリングしたものを確率的ノイズ，初回のサンプリング結果で固定したものを確定的ノイズとし，比較する．実験結果から，クラスごとの IoU，クラス平均の IoU

表 2 欠損確率の再現方法による比較（GTA-LiDAR [2] → KITTI Raw [8]）

手法	IoU (%) ↑		
	■ car	■ pedestrian	クラス平均
再現なし	1.1	2.4	1.7
画素共通の頻度	55.2	25.1	40.2
画素ごとの頻度 [2]	59.0	22.5	40.7
GAN 推論	67.3	25.2	46.3
ePointDA [5]	66.2	24.8	45.5
実データで学習	70.1	16.5	43.3

表 3 GAN 推論における確率的ノイズと確定的ノイズの比較．P は精度，R は再現率．

サンプリング	IoU (%) ↑		
	■ car	■ pedestrian	クラス平均
確率的	64.2	24.2	44.2
確定的	67.3	25.2	46.3

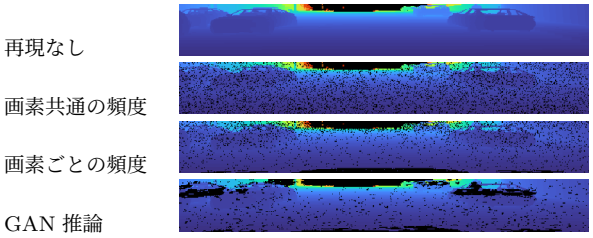

図 2　再現した欠損ノイズの比較（GTA-LiDAR）

ともに確率的ノイズを用いた学習方法が優れていることが確認できる．このことから，欠損ノイズ再現に基づく LiDAR Sim2Real の今後の指針の 1 つとして，欠損ノイズの直接推定ではなく，欠損確率の推定とサンプリングが適していると考えられる．

一方，表 4（SynLiDAR → SemanticKITTI）では，再現なしのクラス平均 IoU が 6.8%であるのに対して，欠損ノイズを再現した手法はいずれも 13%超と僅かな改善が見られた．また，表 2 と同様に GAN 推論を用いた手法が最も高い平均 IoU 14.7%を示した．一方で，実データを学習した場合の IoU とは大きく離れており，画素ごとの頻度と GAN 推論を用いた手法の差は小さい．考えられる要因として，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．例えば，図 3 に示す例では，画像全体に分散する欠損ノイズの分布を表現できているものの，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，GAN inversion におけるシーン潜在変数の推定精度を向上させることが挙げられる．

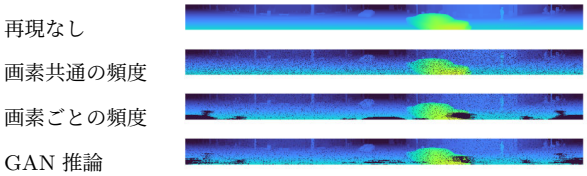

図 3　再現した欠損ノイズの比較（SynLiDAR）

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 2 種類の設定で Sim2Real 実験を行った結果，提案手法の有効性が示された．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，距離画像の解像度・物体クラス数の異なる 2 種類の実験設定で，欠損ノイズ復元の効果に差が見られた．今後は，欠損ノイズ生成の手がかりとなる GAN の性能改善，よりロバストな GAN inversion 手法の開発に取り組む．また，点群表現に基づくセグメンテーションや他の LiDAR データセットを用いた Sim2Real 実験を実施し，欠損ノイズ復元の効果についてより詳細に調査する予定である．

謝辞

本研究の一部は，JSPS 科研費 JP23K16974，JST【ムーンショット型研究開発事業】グラント番号【JPMJMS2032】の支援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] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, “nuScenes: A multimodal dataset for autonomous driving,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11621–11631, 2020.

[5] 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.

[6] 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.

[7] 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*

表 4　欠損確率の再現方法による比較（SynLiDAR [6] → SemanticKITTI [12]）

	IoU (%) ↑																			
手法	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
再現なし	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
画素共通の頻度	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
画素ごとの頻度 [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
GAN 推論	30.2	1.9	7.8	1.0	3.1	3.1	15.0	0.2	42.0	3.0	32.0	0.0	56.0	4.4	50.4	14.2	17.5	10.5	3.2	14.7
PCT [6]	56.0	7.0	17.1	2.8	9.9	23.7	43.7	5.6	55.3	0.8	22.9	0.0	50.1	8.4	65.3	23.1	43.5	28.8	7.5	24.8
実データで学習	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

the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11167–11176, 2020.

[8] 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.

[9] K. Nakashima and R. Kurazume, "Learning to drop points for LiDAR scan synthesis," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 222–229, 2021.

[10] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 1321–1330, 2017.

[11] 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.

[12] 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.

[13] 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.