# 拡散モデルの条件付き生成を用いた LiDAR データの Sim2Real ドメイン変換

宮脇 智也<sup>1,a)</sup> 中嶋 一斗<sup>1,b)</sup> 劉 瀟文 $^{1,c}$ ) 岩下 友美 $^{2,d}$ ) 倉爪 亮 $^{1,e}$ )

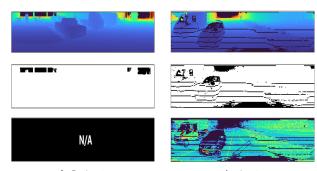
# 概要

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である. この 問題に対して、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. しかし、合成データと実データの間の ドメインギャップにより, 実環境での汎化性能が低下する ことが多い. 本稿では、拡散モデルベースの LiDAR 生成 モデルを応用し、シミュレータの合成データに対して実 データの特徴を付与することで、ドメインギャップを解消 する Sim2Real ドメイン変換手法を提案する.

# 1. はじめに

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. 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重要である. 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。しかし、これらのタスクでは、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.

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. Sim2Real 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



合成データ

実データ

**図 1** LiDAR 距離画像の合成データ [2] と実データ [9] (一部の方位 角を切り出して表示, 上から距離値, 欠損ノイズ, 反射強度)

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る. LiDAR データの Sim2Real に関しては、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, 3] や実デー タの特徴を合成データに再現するアプローチ [1-5] が提案 されている. 本研究は、実データの特徴を再現する後者の アプローチによって、Sim2Real 適応性能を向上させるこ とを目的とする.

データのドメイン変換や編集に関するタスクでは, 生成 モデルを応用する手法が台頭している。特に、生成モデル のひとつである拡散モデル [6] は、サンプル品質と学習安 定性が高く、多くの分野で主流フレームワークとなりつつ ある. また、拡散モデルの生成過程は学習後に条件を付与 することが可能であり、超解像や欠損補完などの線形逆問 題 [7,8] への応用が注目されている. 本稿では、LiDAR データを学習した拡散モデルの生成を合成データで条件付 けることで、実データ特有の特徴である欠損ノイズ、およ び反射パルスレーザの強度である反射強度を高い品質で再 現する手法を提案する.

#### 九州大学

# 2. 関連研究

#### 2.1 欠損ノイズと反射強度のモデル化

LiDAR データにおける欠損ノイズとは、照射したレーザ

ジェット推進研究所

miyawaki@irvs.ait.kyushu-u.ac.jp

 $<sup>^{\</sup>rm b)}$  k\_nakashima@ait.kyushu-u.ac.jp

c) liu@irvs.ait.kyushu-u.ac.jp

yumi.iwashita@jpl.nasa.gov

kurazume@ait.kyushu-u.ac.jp

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データ欠損を指す.また,反射強度とは,LiDAR が計測する反射パルスレーザの強度を示す.これらの特徴は,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.合成データには欠損ノイズが見られず,また反射強度値は存在しないことがわかる.

これまでに, 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手法が提案されている. Zhaoら [3] は、 教師なし画像変換手法である CycleGAN を用いることで、 合成データに対して欠損画像を再現する手法を提案した. Nakashima ら [10] は LiDAR 距離画像の距離値と欠損確率 pの共起関係を学習した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手法を提案している. Wu ら[2]は、実データの距離画像から反射強度画像を推定す る教師あり学習によって, 合成データの反射強度再現を提 案している. Xiao [4] らは, 実データ・合成データの敵対的 学習に基づく点群変換と欠損画像生成、実データの教師あ り学習による反射強度生成による 3 段階の Sim2Real 手法 を提案している. 一方, これらの手法によって欠損ノイズ や反射強度などの特徴が付与された合成データは、実デー タと比べて未だギャップがあり、高精度な特徴再現手法が 必要である.

# 3. 提案手法

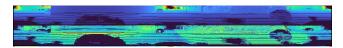
本章では、まず学習済み拡散モデルを利用した条件付き 生成アルゴリズムを紹介した後、本研究で使用する LiDAR データの拡散モデルについて説明する.最後に、LiDAR データの Sim2Real ドメイン変換を実現する提案手法につ いて述べる.

#### 3.1 前提知識:条件付き生成アルゴリズム

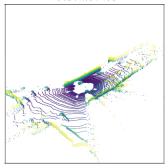
データの条件なし生成を学習した拡散モデル (unconditional diffusion models) は、複数ステップの確率的な状態 遷移によってランダムにデータを生成することができるが、その生成過程を特定の参照データや指標にしたがって制御することで、再学習することなく条件付き生成モデルとして応用する研究が盛んに行われている。特に、生成過程を制御し、画像のノイズ除去やドメイン変換などの画像編集に応用する研究は近年急速に発展している [7, 11–13].

#### 3.1.1 SDEdit

Meng ら [13] は,ユーザの提供するガイド情報に基づいて生成過程を制御することで教師なしドメイン変換を行う SDEdit(Stochastic Differential Editing)を提案した. SDEdit では,まず拡散過程によってユーザが提供したガイド画像  $x_{\rm guide}$  を時刻  $t_1$  のノイズレベルに遷移させる.



距離画像表現



鳥瞰図

**図 2** KITTI Raw データセットで学習した R2DM [14] による生成 サンプル

$$x(t_1) = \alpha(t_1)x_{\text{guide}} + \sigma(t_1)z, \quad z \sim N(0, I) \tag{1}$$

ただし、x(t) は時刻 t における潜在変数であり、t=1 でノイズ  $x(1)\sim N(0,I)$ 、t=0 でデータ x(0) を表す。 $\alpha(t_1)$  と  $\sigma(t_1)$  はそれぞれ時間に依存するスカラー関数であり、 $\alpha(t_1)$  はデータの大きさ、 $\sigma(t_0)$  はノイズの大きさを表す。次に時間  $t=t_1$  から t=0 まで逆拡散過程によって徐々にノイズを除去することで、ガイド画像に基づいた画像を生成する.

#### $3.1.2 \quad \Pi GDM$

拡散モデルは,逆問題を解く有効なアプローチとして近年注目されている.ここで,逆問題とは,劣化のある測定値yから元の信号 $x_0$ を推測する問題であり,超解像,インペインティング,JPEG復元などが例として挙げられる.逆問題は以下の式で定式化される.

$$y = Hx_0 + z \tag{2}$$

ここで,H は既知の測定行列,z はガウスノイズである. Song [7] らは,ノイズの多い,非線形な,さらには微分不可能な測定を含む逆問題を解決する擬似逆行列ガイダンス付き拡散モデル  $\Pi$ GDM(Pseudoinverse-guided Diffusion Models)を提案した. $\Pi$ GDM ではベイズ則を用いて,以下のように条件付きスコア関数を定義している.

$$\nabla_{x_t} \log p_t(x_t|y) = \nabla_{x_t} \log p_t(x_t) + \nabla_{x_t} \log p_t(y|x_t)$$
(3)

ここで、最初の項は条件なし生成を学習した通常の拡散 モデルによって計算でき、2番目の項は以下の近似を用い て求める

$$\nabla_{x_t} \log p_t(y|x_t) \approx r_t^{-2} (\underbrace{H^{\dagger} y - H^{\dagger} H \hat{x}_0}_{\text{ig} \pm \bar{q}})^{\top} \frac{\partial \hat{x}_0}{\partial x_t}$$
(4)

 $H^{\dagger}$  は Moore-Penrose 擬似逆行列,  $\hat{x}_0$  は各ステップにおけ

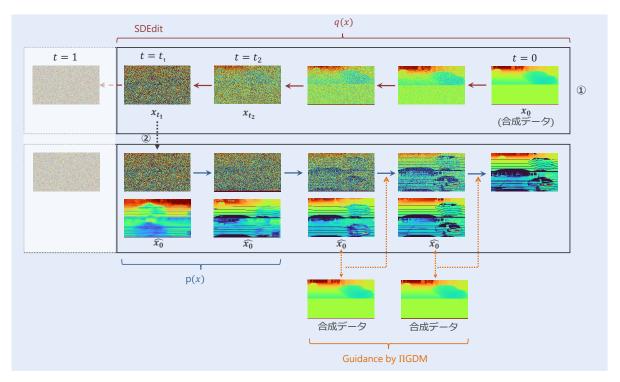


図 3 提案手法

る元の信号  $x_0$  の推定値, $r_t$  は時刻 t におけるデータ依存の標準偏差である。 $\Pi GDM$  は,式 3,式 4 で定式化される条件付きスコア関数に従って逆拡散過程を実行することにより,測定値 y に基づいた生成制御を行うことができる.

#### 3.2 LiDAR データの拡散モデル

本研究では、条件付き生成に用いる拡散モデルとして、Nakashima ら [14] の R2DM を採用する。R2DM は、Li-DAR のビーム角度に関する帰納バイアスを導入した連続時間 DDPM によって、距離・反射強度画像を学習する手法であり、学習済みモデルの条件付き生成を用いた欠損補完を実証している。本研究では、31M パラメータの公開実装を 285M パラメータに増強し、KITTI Raw データセット [9] を用いて学習した。生成サンプルの例を図 2 に示す。

# 3.3 Sim2Real ドメイン変換への応用

本研究では、実データによって事前学習された拡散モデルを用いて、合成データに基づく条件付け生成を行うことで、Sim2Realドメイン変換を行う.ただし、合成データは欠損を含まない距離画像と全画素欠損した反射強度画像の組、実データは欠損を含む距離画像と反射強度の組から構成される(図 1 参照).本章では、前節で述べた SDEditと  $\Pi GDM$  を組み合わせた条件付き生成アルゴリズムを提案する.提案手法の概要を図 3 に示す.

本手法では、まず合成データに対して、[1,0] で定義された連続時間 R2DM において  $t=t_1$  までノイズを加える (q(x)). その後は、 $t=t_1$  から R2DM による生成を行う. 生

成過程では、 $t=t_2$  までは条件付けをせずに逆拡散 (p(x)) を実行する.  $t=t_2$  以降は、 $\Pi$ GDM によって合成データ(ガイド画像)による条件付け生成を行う.この際,式 4 にしたがって各タイムステップにおける元の信号  $x_0$  の推定値 $\hat{x}_0$  と合成データとの間の誤差を計算するが,全ての画素で誤差を計算した場合,最終的に $\hat{x}_0$  に含まれる欠損ノイズが除去される可能性がある.そのため, $\hat{x}_0 \in [-1,1]$  に含まれる閾値  $\eta$  の距離値については, $\Pi$ GDM の誤差計算から除外することで,再現された欠損ノイズを保持する.本研究では, $t_1=0.7,\ t_2=0.57,\ \eta=-0.5$  と定めた.

# 4. 実験

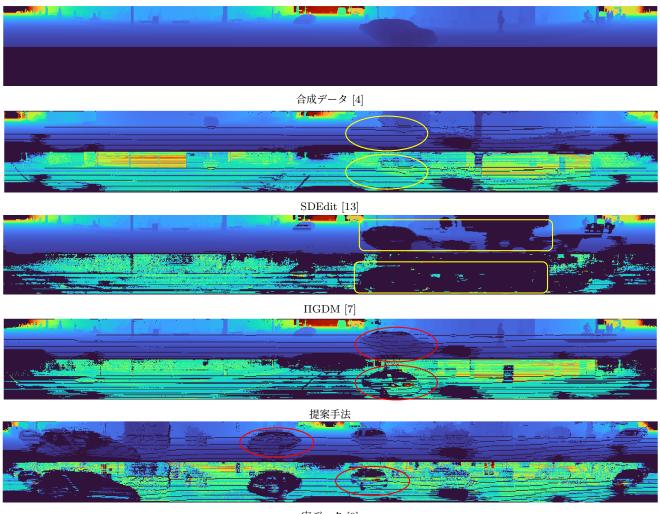
本章では、LiDAR の合成データセットである SynLiDAR データセット [4] 内の合成データに対し、提案手法および 比較手法を用いて欠損ノイズと反射強度を再現した結果を報告する.

# 4.1 比較手法

欠損ノイズと反射強度の再現方法について, (1) **SDEdit**, (2) **IIGDM**, (3) **提案手法**の 3 種類の方法を比較する.

# 4.2 実験結果

図4に実験結果を示す.最上部に示すデータが欠損ノイズおよび反射強度の再現対象となる合成データ,最下部に示すデータが実データの一例である.SDEdit のみの場合,実データのような欠損ノイズと反射強度を生成できているものの、黄色の枠で示した箇所のように、合成データ



実データ [9]

図 4 欠損ノイズと反射強度の再現結果

のシーン内容 (車両や街並み) を反映できていない. また, IIGDM のみの場合, 合成データのシーン内容を保持できているものの, 黄色枠で示した箇所のように, 不自然な欠損ノイズや反射強度が再現されており, 実データと比べるとギャップがある. 一方, 提案手法による再現は, 赤枠で示した箇所に見られるように, 合成データのシーン内容を忠実に保持しつつ, 実データに近い欠損ノイズと反射強度を再現することができている.

# 5. まとめと今後の展望

本稿では、LiDAR データを学習した拡散モデルの生成を合成データで条件付けることで、実データ特有の特徴である欠損ノイズと反射強度を高い品質で再現する手法を提案した.今後は、提案手法で実データの特徴を再現した合成データの品質を定量的に評価し、Sim2Real セマンティックセグメンテーションなどのタスクを実施することで提案手法の有効性を検証する予定である.

# 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 realtime road-object segmentation from 3D LiDAR point cloud," in *Proceedings of the IEEE International Con*ference 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] 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 Li-DAR point cloud segmentation," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, pp. 3500–3509, 2021.

- [4] 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.
- [5] 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.
- [6] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," 2020.
- [7] J. Song, A. Vahdat, M. Mardani, and J. Kautz, "Pseudoinverse-guided diffusion models for inverse problems," in *International Conference on Learning Repre*sentations, 2023.
- [8] B. Kawar, M. Elad, S. Ermon, and J. Song, "Denoising diffusion restoration models," in *Proceedings of* the Advances in neural information processing systems (NeurIPS), 2022.
- [9] 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.
- [10] K. Nakashima, Y. Iwashita, and R. Kurazume, "Generative range imaging for learning scene priors of 3D LiDAR data," in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 1256–1266, 2023.
- [11] A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, and L. Van Gool, "Repaint: Inpainting using denoising diffusion probabilistic models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pat*tern Recognition (CVPR), pp. 11461–11471, June 2022.
- [12] S. Xie, Z. Zhang, Z. Lin, T. Hinz, and K. Zhang, "Smart-brush: Text and shape guided object inpainting with diffusion model," in 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 22428–22437, 2023.
- [13] C. Meng, Y. He, Y. Song, J. Song, J. Wu, J.-Y. Zhu, and S. Ermon, "SDEdit: Guided image synthesis and editing with stochastic differential equations," in *International Conference on Learning Representations*, 2022.
- [14] K. Nakashima and R. Kurazume, "Lidar data synthesis with denoising diffusion probabilistic models," in *IEEE International Conference on Robotics and Automation* (ICRA 2024), pp. -, 2024.