

確率的欠損表現を用いた LiDAR データの生成モデリング

中嶋 一斗^{1,a)} 倉爪 亮^{1,b)}

概要

3D LiDAR データの生成モデルを構築するための敵対的生成ネットワークを提案する。提案手法は、逆深度マップとして仰俯角・方位角ごとの距離値を生成し、3D 点群に変換する。特に、逆深度マップの生成プロセスでは、確率的に可逆な欠損処理を導入することで、欠損あり学習データのみから潜在的な表面形状と欠損確率をモデリングする。実験では、生成データの品質・多様性および実データの再現性に関して提案手法の有効性を検証する。

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。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。近年では、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。一方で、用途によっては消費電力やレイテンシなどの制約により、高精度・高解像度な 3D LiDAR センサを利用できない場合がある。また、レーザ測距の原理上、鏡面体や物体境界においては反射光を集光できずに欠損が生じやすい。本研究では、このような低品質な LiDAR データを復元・補完推定するために、データの生成規則を統計的に学習した生成モデルの構築を目的とする。

2. 関連研究

2.1 LiDAR データの合成

近年、ニューラルネットワークに基づく生成モデルについて活発に議論されており、点群データへの応用も進んでいる。しかし、その多くは家具や車の CAD モデルから均一にサンプリングしたデータを対象にしている。一方で、3D LiDAR センサから得られる点群データは比較的点数が多く、放射状の距離計測に基づくため点密度も不均一である。物体検出などの認識タスクでは、各計測点を仰俯

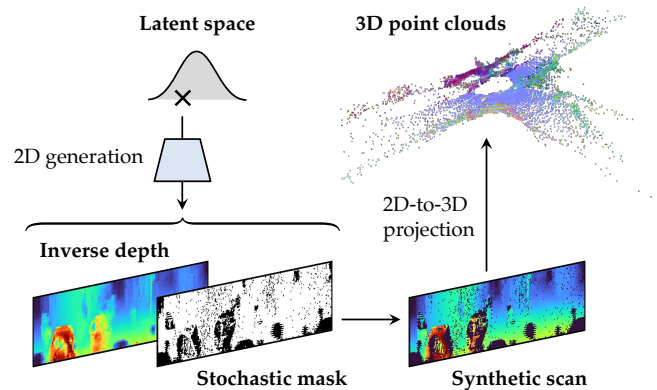


図 1 提案手法による 3D LiDAR データの生成フロー

角・方位角の 2D グリッドに投影した画像表現を用いることで、精度と計算量の改善に成功している。生成モデルの例としては、Caccia ら [2] が代表的な深層生成モデルである変分オートエンコーダ (VAE) と敵対的生成ネットワーク (GAN) を画像表現した LiDAR データに応用している。VAE を用いた実データの再構成評価により、従来の点集合表現に比べて高品質に生成できることを報告している。

2.2 ノイズを考慮した画像生成モデル

画像表現した LiDAR データは、3D データ空間の不均質性や計算量の問題を改善できる一方で、計測の物理的要因による点欠損が乗法性二値ノイズとして顕著に現れる。ニューラルネットワークに基づく生成モデルは離散分布を直接学習することができないため、これらのノイズが実データ分布とのギャップやモデル学習の不安定性を引き起こす場合がある。前述の Caccia ら [2] は、予めすべてのデータの点欠損を近傍点で補完することでこの問題に対処しているが、真の形状に即していない可能性が高い。

一方、自然画像のノイズ除去や生成モデルの分野では、ノイズのある学習データのみからノイズのないデータ分布を学習する手法がこれまでに提案されている。乗法性二値ノイズを扱った例として、Noise2Noise [8] ではノイズ除去モデル、AmbientGAN [1] では GAN の学習を実現している。これらの手法は、確率的なノイズ分布としてベルヌーイ分布を仮定することで、背後にあるデータ分布を推定している。ただし、LiDAR データのように画像内容に応じた

¹ 九州大学

^{a)} k.nakashima@irvs.ait.kyushu-u.ac.jp

^{b)} kurazume@ait.kyushu-u.ac.jp

二値ノイズを取り扱うことができない. NR-GAN [5] は, GAN の学習に際して, 画像内容に応じたノイズ分布そのものを推定することができるが, 乗法性二値ノイズに対応していない.

3. 提案手法

本研究では, 画像表現した LiDAR データの欠損の分布として, 前章で紹介した文献 [1], [8] のようにベルヌーイ分布に従う乗法性二値ノイズを仮定する. ただし, 各画素の欠損の生起確率を適応的に生成し, サンプリングした欠損を用いて LiDAR データを表現する (図 1). 潜在的なシーンの形状と計測に伴う欠損を分離表現することで, 生成モデルの表現力向上を図る. 以降では, LiDAR データの画像表現方法および敵対的生成学習の枠組み, ベルヌーイ分布の生起確率を学習するための勾配近似法について述べる. 提案手法の概要図を図 2 に示す.

3.1 データ表現

垂直 H ラインおよび水平 W 点のスキャンを行う 3D LiDAR センサでは $H \times W$ 組の距離値が得られる. 同時に, これらの計測点を縦軸を仰俯角, 横軸を方位角とした 2D グリッドに投影すると, 距離値を各画素値に持つ $H \times W$ の深度マップで表現できる. ただし, この深度マップ表現は, ダイナミックレンジのほとんどを少数画素の遠点が占めているため, 深度マップを直接学習した予備実験では安定に学習することができなかった. そこで, 本研究では各画素値の深度を逆数に変換した逆深度マップを採用する. 欠損領域は最大距離を設定し, 逆深度の値域は変換後に $[-1, 1]$ に正規化する.

3.2 逆深度マップの敵対的生成

敵対的生成ネットワークは一般的に生成器 G_θ と識別機 D_ϕ の 2 つのネットワークから構成される. ただし, (θ, ϕ) は学習パラメータである. 画像生成タスクの場合, 生成器 G_θ は潜在変数 $z \sim N(0, I)$ から画像 x_z を生成し, 識別機 D_ϕ は生成画像 x_z と実画像 x_{real} を識別する. 2 つのネットワークは, 識別機 D_ϕ の出力で与えられる目的関数の min-max ゲームで最適化が行われる. 本研究では, 以下の R_1 正則化付き non-saturating loss [7] を採用した. ただし, γ は係数である.

$$\mathcal{L}_\theta = -\mathbb{E}_x[\log D_\phi(x)] - \mathbb{E}_z[\log(1 - D_\phi(G_\theta(z)))] + \frac{\gamma}{2} \mathbb{E}_x[\|\nabla_x D_\phi(x)\|^2] \quad (1)$$

$$\mathcal{L}_\phi = -\mathbb{E}_z[\log D_\phi(G_\theta(z))]. \quad (2)$$

3.3 Gumbel-Sigmoid 分布による確率的欠損表現

本研究では, 上述の画像 x_z の生成過程に確率分布に

基づく欠損処理を導入する. まず, 密な逆深度マップ $\tilde{x}_z \in \mathbb{R}^{H \times W}$ と測距レーザの反射しやすさを表す信頼度マップ $\pi_z \in \mathbb{R}^{H \times W}$ を仮定する. 両者は, 潜在変数 z で条件付けされている点に注意されたい. 欠損を表す二値マスク $m_z \in \{0, 1\}^{H \times W}$ は, π_z を生起確率としたベルヌーイ分布に従ってサンプリングされる.

$$m_z \sim \text{Bernoulli}(\pi_z) \quad (3)$$

識別器で評価する最終的な逆深度マップ x_z は, 密な逆深度マップ \tilde{x}_z に二値マスク m_z を適用することで得られる.

$$x_z = m_z \odot \tilde{x}_z + (1 - m_z) \odot \alpha, \quad (4)$$

\odot は要素積を表し, α は欠損を表す定数である.

ただし, 式 (3) のサンプリングは可微分な操作ではないため, 勾配を伝播して信頼度マップ π_z を修正することができない. そこで, 本研究では straight-through (ST) Gumbel-Sigmoid [4] により二値マスク m_z を再パラメータ化することで, 勾配を近似的に推定する. まず, 連続緩和した次のソフトマスク \tilde{m}_z を導入する.

$$\tilde{m}_z = \text{sigmoid}\left(\frac{e + g_1 - g_2}{\tau}\right), \quad e = \ln\left(\frac{\pi_z}{1 - \pi_z}\right). \quad (5)$$

ここで, $g_1, g_2 \in \mathbb{R}^{H \times W}$ は $\text{Gumbel}(0, I)$ に従って独立にサンプリングされ, π_z の対数オッズ e に摂動を与える. τ は $\text{sigmoid}(\cdot)$ の勾配を制御するハイパーパラメータで, τ が小さいほどソフトマスク \tilde{m}_z は二値マスクに近づく.

最後にソフトマスク \tilde{m}_z は次の閾値関数で画素位置 (i, j) ごとに二値化され, 二値マスク m_z を得る.

$$m_z^{i,j} = \begin{cases} 1 & \tilde{m}_z^{i,j} \geq 0.5 \\ 0 & \tilde{m}_z^{i,j} < 0.5 \end{cases} \quad (6)$$

確率的な二値マスク m_z を順伝播時に用いるが, 逆伝播時は式 (6) を恒等関数とみなして, 二値マスク m_z の勾配をソフトマスク \tilde{m}_z で近似する. 以上より, 欠損を制御する信頼度マップ π_z は, 3.2 章で導入した敵対的生成学習を通して, 逆深度マップ \tilde{x}_z と同時にモデリングされる.

4. 実験

4.1 データセット

KITTI Odometry データセット [3] を用いて提案手法の有効性を検証する. KITTI Odometry データセットは, 自動車上部に搭載した垂直 64 ラインの 3D LiDAR センサ (Velodyne HDL-64E) で 22 箇所の市街地や住宅地を連続的に計測したものである. 学習セットは 19,130 スキャン, 検証セットは 4,071 スキャン, テストセットは 20,351 スキャンの点群から構成される. 本研究では, 直交座標の座標集合で与えられている公開データを仰俯角に応じて分割した後, それぞれを 256 点に間引くことで 64×256 の逆深度マップを生成した.

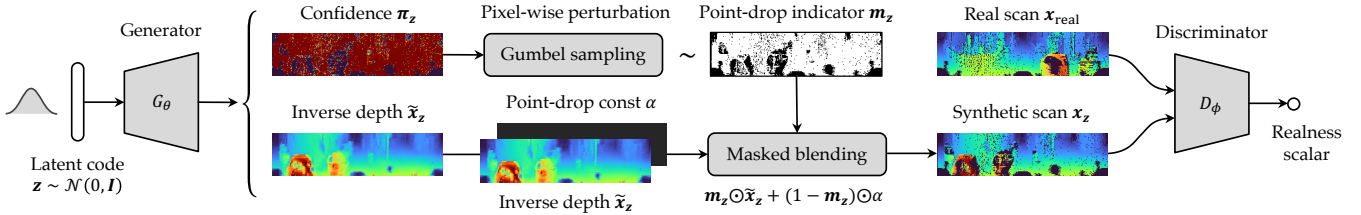


図 2 提案する敵対的生成フレームワークの概要図

4.2 モデル構造と学習設定

本研究のベースラインとして、先行研究 [2] を参考に、5つの転置畳込み層から構成される生成器と、5つの畳込み層から構成される識別器を設定した。ただし、出力サイズは 64×256 に変更、活性化関数は Leaky ReLU に統一、バッチ正規化は排除する。すべてのパラメータは $\mathcal{N}(0, 1)$ で初期化し、実行時に He の初期化定数でスケールする [6]。離散ノイズの影響を緩和するために、識別器に入力する前にアンチエイリアシング [5] を適用する。すべてのカーネル操作は、左右両端のパディングを互いに循環させ、逆深度マップの円筒構造を保持したまま処理する。

提案手法は、ベースラインの生成器について出力チャンネル数のみを変更し、逆深度マップと信頼度マップを生成する。両者は式 (4) に従って、最終的な識別器入力を得る。

すべての実験は、ミニバッチサイズを 32、学習係数は $G_\theta \cdot D_\phi$ ともに 0.002、反復回数を 25×10^6 回、 R_1 正則化の係数 γ は 1 に固定し、Adam を用いてパラメータを更新した。また、DiffAugment [9] を用いて識別器入力に対してデータ拡張を適用した。

4.3 生成データの品質・多様性評価

テストセットの逆深度マップとサンプリングした逆深度マップに対して、2D レベルと 3D レベルの評価を行った。2D レベル評価としては、逆深度マップに対して、切り出したパッチ集合間で画像類似度を計算する sliced Wasserstein distance (SWD) を計算した。3D レベル評価としては、逆深度マップから変換した点群に対して、Jensen-Shannon divergence (JSD), coverage (COV), minimum matching distance (MMD), 1-nearest neighbor accuracy (1-NNA) の 4 つの集合間類似度を計算した。ただし、3D レベル評価は COV · MMD · 1-NNA における距離行列の計算時間を削減するために、furthest point sampling により点群を 2048 点にダウンサンプリングした。点群間の距離尺度には、chamfer distance (CD) を用いた。ベースラインの逆深度マップを点群に変換する際は、欠損値 α を判定する最適な相対許容誤差 $\in [0.001, 0.1]$ を検証セットを用いて探索した。評価結果を表 1 に示す。すべての指標において提案手法が良い結果を示した。提案手法による生成データの例を図 3 に示す。形状に応じた欠損確率を学習し、潜在的なシーンの形状と欠損を分離生成できていることが分かる。

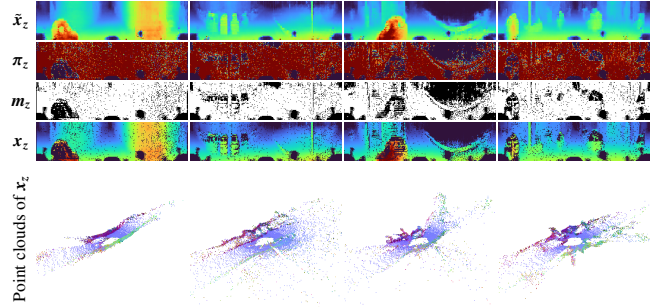


図 3 提案手法による生成データの例

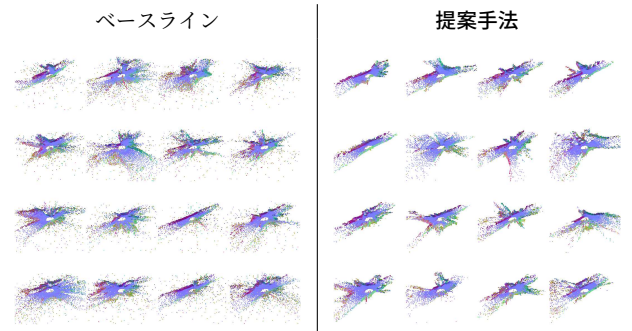


図 4 サンプリングした点群の比較

また、点群の生成例を図 4 に示す。提案手法は欠損に伴うジャンプエッジを表現できるため、ベースラインに比べてノイズ点が少ないことが分かる。

4.4 実データの再現性評価

テストセットの未知データ x_{target} を再現するように潜在変数 z を最適化し、生成モデルの表現力を比較する。具体的には、以下の L1 距離をもとに z を求める。

$$\hat{z} = \arg \min_z \frac{\sum_{i,j} \mathbb{1}[x_{target} \neq \alpha]^{i,j} \|x_{target}^{i,j} - \tilde{x}_z^{i,j}\|_1}{\sum_{i,j} \mathbb{1}[x_{target} \neq \alpha]^{i,j}}, \quad (7)$$

ただし、 $\mathbb{1}[x_{target} \neq \alpha] \in \{0, 1\}^{H \times W}$ は目標画像の欠損を表す二値マスクである。生成画像と目標画像の一致度は、単眼深度推定分野で一般的に用いられている relative absolute error (Abs Rel), relative squared error (Seq Rel), root mean squared error (RMSE), 対数スケールにおける RMSE (RMSE log), および閾値 t に基づく精度 ($\delta < t$) を計算した。また、変換した点群間で chamfer distance (CD) を計算した。評価結果を表 2 に示す。すべての指標において提案手法が良い結果を示した。

表 1 KITTI Odometry データセットを用いた品質・多様性評価

	3D				2D
	JSD $\times 10^2$	COV (%)	MMD $\times 10^3$	1-NNA (%)	SWD $\times 10$
ベースライン	6.41	5.91	2.42	100.0	1.62
提案手法	2.38	38.2	1.07	92.3	1.56
学習セット	0.93	35.0	0.87	96.7	1.85

表 2 KITTI Odometry データセットを用いた実データの再現性評価

	Abs Rel	Seq Rel	RMSE	RMSE log	$\delta < 1.25$ (%)	$\delta < 1.25^2$ (%)	$\delta < 1.25^3$ (%)	CD $\times 10^3$
ベースライン	0.0875	0.722	3.94	0.175	90.3	96.2	98.4	1.06
提案手法	0.0705	0.479	3.34	0.149	92.7	97.2	98.8	0.61

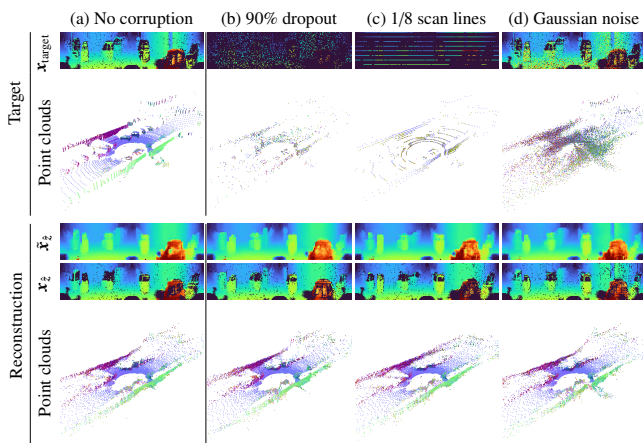


図 5 提案手法を用いた欠損データの復元

図 5 には、テストセットの目標データに対して、(a) そのまま潜在変数を最適化させた場合、(b) 90%の点をランダムに欠損させた場合、(c) 垂直ライン数を 1/8 に減らした場合、(d) 距離値に対してガウシアンノイズ $N(0, 0.01)$ を付加した場合の提案手法による生成結果を示す。細部の再現性に課題があるものの、欠損した未知データの部分観測に対して、密な逆深度マップを推定し、元データに類似した欠損を再現できていることが分かる。

5. まとめと今後の予定

本稿では、3D LiDAR データの生成モデルを構築するためのアプローチとして、確率的欠損表現を導入した逆深度マップの敵対的生成ネットワークを提案した。生成データの品質・多様性および実データの再現性に関する評価では、確率的欠損表現を用いないベースラインに対して大きい改善が見られた。実際の生成例からは、欠損のある学習データのみから、潜在的なシーン形状と欠損確率を分離生成できることを確認した。また、人工的ないくつかの欠損に対しても、密な点群を復元できることを示した。今後は、モデルの改良および 3D LiDAR データを用いた認識タスクへの応用に取り組む。

謝辞

本研究は JSPS 特別研究員奨励費 19J12159 および JSPS 科研費 JP20H00230 の助成を受けたものである。

参考文献

- [1] Bora, A., Price, E. and Dimakis, A. G.: AmbientGAN: Generative models from lossy measurements, *Proceedings of the International Conference on Learning Representations (ICLR)* (2018).
- [2] Caccia, L., van Hoof, H., Courville, A. and Pineau, J.: Deep Generative Modeling of LiDAR Data, *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5034–5040 (2019).
- [3] Geiger, A., Lenz, P., Stiller, C. and Urtasun, R.: Vision meets robotics: The kitti dataset, *The International Journal of Robotics Research (IJRR)*, Vol. 32, No. 11, pp. 1231–1237 (2013).
- [4] Jang, E., Gu, S. and Poole, B.: Categorical Reparameterization with Gumbel-Softmax, *Proceedings of the International Conference on Learning Representations (ICLR)* (2017).
- [5] Kaneko, T. and Harada, T.: Noise robust generative adversarial networks, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8404–8414 (2020).
- [6] Karras, T., Aila, T., Laine, S. and Lehtinen, J.: Progressive Growing of GANs for Improved Quality, Stability, and Variation, *Proceedings of the International Conference on Learning Representations (ICLR)* (2018).
- [7] Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J. and Aila, T.: Analyzing and improving the image quality of stylegan, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8110–8119 (2020).
- [8] Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M. and Aila, T.: Noise2Noise: Learning Image Restoration without Clean Data, *Proceedings of Machine Learning Research*, Vol. 80, Stockholm, Sweden, PMLR, pp. 2965–2974 (2018).
- [9] Zhao, S., Liu, Z., Lin, J., Zhu, J.-Y. and Han, S.: Differentiable Augmentation for Data-Efficient GAN Training, *Advances in Neural Information Processing Systems (NeurIPS)* (2020).