条件付きフローマッチングによる LiDAR データ生成モデルのサンプリング高速化

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

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基づく走行環境の認識に不可欠である．一方，ビーム数の少ない安価なセンサによって点群密度が低下したり，鏡面物体や悪天候環境において計測データが欠損する場合には，性能低下を引き起こす可能性がある．

この問題に対しては，あらかじめ実データを学習した深層生成モデルを計測環境の事前情報として用いることで，アップサンプリングやデータ修復に応用するアプローチが発表されている．特に近年では，潜在変数分布とデータ分布間の反復的な確率過程によってデータ生成を表現する拡散モデル（diffusion models）が，有望な生成モデルとして注目されている．拡散モデルは，それまでに広く利用されてきた VAE (variational autoencoders) や GAN (generative adversarial networks) に比べて，学習安定性および生成品質において優れており，自然画像をはじめとした多くのドメインで有効性が示されている．一方，高品質な生成データを得るためには，ニューラルネットワークを用いた再帰的な状態遷移のステップ数を十分大きくする必要があり，計算コストの高さが実ロボット応用における課題となる．

そこで本研究では，拡散モデルと同様の学習安定性と生成品質の高さを持ちながら，少ないステップ数で高品質なデータ生成を実現する条件付きフローマッチング [6] に着目する．条件付きフローマッチングによる生成モデルは，拡散モデルと同じように確率密度分布の遷移を考える確率フロー（probability flow）の一種であるが，等速直線軌道に基づく常微分方程式によって定式化しており，理想状態ではステップ数によらず高品質なデータが得られる．本稿では，条件付きフローマッチングを LiDAR データ生成に応用するための手法とモデルアーキテクチャを提案し，KITTI-360 データセット [4] を用いた評価実験によって有効性を示す．

2. 関連研究

LiDAR データの生成モデルは，球面投影による画像表現に基づいて VAE を学習する手法 [1] や GAN を学習する手法 [1, 7, 8] がこれまでに提案されている．さらに近年では，画像表現に基づく拡散モデルを用いた手法 [9, 10, 15] が大幅な品質向上を実現している．例えば，Zyrianov ら [15] は，画素空間で拡散モデルを学習する離散時間 SMLD (score matching with Langevin dynamics) を用いて，LiDAR 距離・反射強度画像を生成する LiDARGen を提案した．Nakashima ら [9] は，画素空間で拡散モデルを学習する連続時間 DDPM (denoising diffusion probabilistic models) を用いて，LiDAR 距離・反射強度画像を生成する R2DM を提案した．Ran ら [10] は，学習済みオートエンコーダの低解像度潜在空間上で DDPM を学習する LDM (latent diffusion models) [11] を用いて，LiDAR 距離画像を生成する LiDM を提案した．

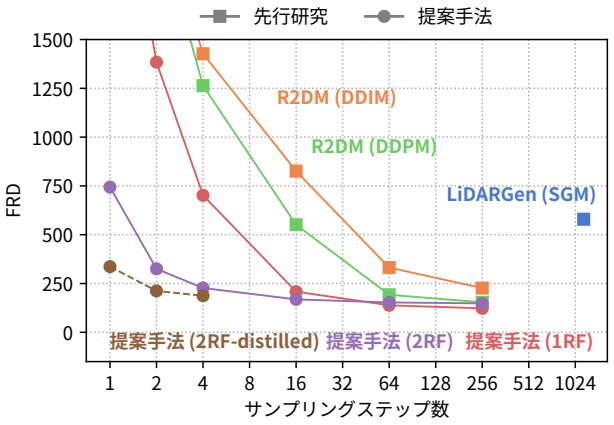

図 1 確率フローに基づく生成モデルを用いた条件なしサンプリングにおける，ステップ数と生成画像品質指標 FRD [15] のトレードオフ

画素空間で拡散過程を学習する場合 [9, 15]，反復的補正によって画素値を精度高く表現できる一方で，各ステップにおける関数評価の計算コストが高い．一方，潜在空間で拡散過程を学習する場合 [10]，空間解像度の圧縮によって計算コストが低下するものの，画素値の精度はオートエンコーダ (1 ステップ生成) に依存するため高周波成分を損ないやすい．後者の欠点は，知覚的品質を評価する自然画像ドメインにおいて問題になりにくいが，点群再構成のために画素値の精度が要求される距離画像においては課題となる．以上の理由から，本研究では画素空間上のデータ生成を採用しつつ，等速直線軌道に基づく条件付きフローマッチング [6] によって求解に要するステップ数を削減することで，高品質かつ高速な LiDAR データ生成モデルを構築する．

3. 提案手法

3.1 データ表現

先行研究 [1, 7–10, 15] と同様に，球面投影による正距円筒画像表現を用いる．正距円筒画像表現では，レーザの仰俯角・方位角をそれぞれ縦軸・横軸とした 2D グ

リッドに対して，各計測点に含まれる距離値・反射強度値を投影する．ただし，画像を多く占める近距離画素のダイナミックレンジを改善するために，距離値を対数スケールに変換する [9, 15].

3.2 条件付きフローマッチング

潜在変数 $\boldsymbol{x}_0 \sim p_0$ とデータ $\boldsymbol{x}_1 \sim p_1$ の確率フローを，以下の常微分方程式（ordinary differential equations, ODE）によって表現する．

$$\frac{d\boldsymbol{x}_t}{dt} = v_\theta(\boldsymbol{x}_t, t) \tag{1}$$

ただし，$v_\theta(\boldsymbol{x}_t, t)$ は，時刻 $t \in [0, 1)$ における中間状態 $\boldsymbol{x}_t$ に対して速度場を表現するニューラルネットワークである．以降に，$v_\theta(\boldsymbol{x}_t, t)$ の学習方法と，再学習による経路最適化，サンプリングについて述べる．

初期フローの学習　本研究では Rectified Flow [6] を採用し，$\boldsymbol{x}_0$ と $\boldsymbol{x}_1$ の最適輸送を与える等速直線な条件付きフロー $\boldsymbol{x}_1 - \boldsymbol{x}_0$ によって速度場を教示する．すなわち，以下の損失関数を最小化する．

$$\mathcal{L}_{\mathrm{CFM}} = \| (\boldsymbol{x}_1 - \boldsymbol{x}_0) - v_\theta(\boldsymbol{x}_t, t) \|_2^2 \tag{2}$$

ただし，$\boldsymbol{x}_1 \sim p_1$，$\boldsymbol{x}_0 \sim p_0$，$t \sim \mathrm{Uniform}(0, 1)$ であり，速度場モデルの入力 $\boldsymbol{x}_t$ は内分点 $t\boldsymbol{x}_1 + (1-t)\boldsymbol{x}_0$ によって与える．以降は，初期フローを 1-RF と呼ぶ．

積分経路の直線化　上記の初期フローは，独立な分布に従う $\boldsymbol{x}_0$ と $\boldsymbol{x}_1$ によって学習されるため，ODE の直線的な積分経路を得ることは難しい．一方，潜在変数 $\boldsymbol{x}_0 \sim p_0$ と初期フローによって生成される $\tilde{\boldsymbol{x}}_1 = \mathrm{ODE}(\boldsymbol{x}_0)$ によって，速度場を再学習することで潜在変数分布からデータ分布への輸送コストが減少することが知られており，この手続きを Reflow [6] と呼ぶ．本研究では，再学習に用いる時刻のサンプリングとして，$t = 0$ および $t = 1$ 付近の頻度を高くする U 型分布 [3] を採用する．また，以下の Pseudo-Huber 損失 [3] を最小化する．

$$\mathcal{L}_{\mathrm{RF}} = \sqrt{\| (x_1 - x_0) - v(x_t, t) \|_2^2 + c^2} - c, \tag{3}$$

ただし，定数 c は関連研究にしたがって，データの次元数 D に対して $c = 0.00054\sqrt{D}$ と定める．一度の Reflow を経たモデルを 2-RF とする．

特定時刻によるモデル蒸留　潜在変数と生成データの組を用いて，さらに特定の時刻のみを再学習することで，速度場モデルの推定を固定ステップの生成に特化させることができ，この手続きを蒸留と呼ぶ．k-RF の蒸留によって得られるモデルを，k-RF-distilled とする．

サンプリング　学習したフローを用いたサンプリングは，汎用数値積分ソルバを用いて式 1 の ODE を求解することで実行できる．本稿では，ODE の数値解法として，Reflow および蒸留における学習データ生成には適応的ステップサイズの Dormand-Prince 法 (`dopri5`)，すべての手法評価には固定ステップサイズの標準的な Euler 法 (`euler`) を用いる．

3.3 速度場モデルのアーキテクチャ

画素空間の確率フローを学習するために，HDiT (hourglass diffusion transformers) [2] を採用する．HDiT は，Transformer ベースの拡散モデルとして提案されたものであり，窓走査に基づく軽量な Neighborhood Attention によって，大幅な計算コスト削減のもと画素空間の学習を可能にしている．本研究では，以下の変更点を加えた．(1) Neighborhood Attention の窓走査は左右境界を循環させる．(2) LiDAR アップサンプリングタスク [14] を参考に，窓サイズを正方形から (1,4) に変更する．(3) 相対位置表現として，ビーム角度によって 2D Axial RoPE を条件付ける．(4) 絶対位置表現として，トークンごとに学習可能なバイアスを加算する．

4. 評価実験

4.1 実験設定

先行研究 [9, 15] に従い，KITTI-360 データセット [4] を用いる．KITTI-360 データセットは，自動車上部に設置した 2 種類の LiDAR センサ（Velodyne HDL-64E, SICK LMS200）および魚眼カメラによってドイツの市街地を計測した 9 シーケンスで構成される．本研究では，Velodyne HDL-64E によって計測された 81,106 スキャンデータを用い，各スキャンデータに含まれる点群は，球面投影によって解像度 64×1024 の距離・反射強度画像に変換する．

4.2 比較手法

本実験では，離散時間 SMLD の LiDARGen [15]，連続時間 DDPM の R2DM [9]，LDM ベースの LiDM [10] と比較する．ただし，LiDM はモデル内に絶対位置バイアスがなく，生成データはランダムな水平回転を含むため，学習可能なバイアス項を入力各画素に加算するモデル LiDM+APE を新たに学習した．また，LiDARGen と R2DM は距離画像と反射強度画像を生成するのに対して，LiDM は距離画像のみを生成する．

4.3 評価指標

実データ集合（データセット）と生成データ集合の分布類似度を定義する以下の指標群によって，生成モデルの品質・多様性を評価する．本研究ではまず，特徴表現の Fréchet 距離に基づく FRD [15] と FPD [13] を計算する．また，点群の鳥瞰図表現（BEV）から作成したヒストグラムに基づく MMD [15] と JSD [15] を計算する．

4.4 結果

表 1 に，評価結果を示す．上段のステップ数（NFE）が多い場合，提案手法は比較手法と同等以上の評価スコアを示した．また，下段のステップ数を 1 にした場合，すべての先行研究は著しく評価スコアが低下するのに対し，提案手法は評価スコアの低下を抑えることができている．図 2 に，1-RF と 2-RF について，同じ潜在変数に対する異なるステップ数の生成結果を示す．ステップ数が 256 の場合，両者の生成データに大きな違いは見られないが，1-RF の生成結果はステップ数が減少するにつれて点群の形状が歪んでいる．一方，2-RF

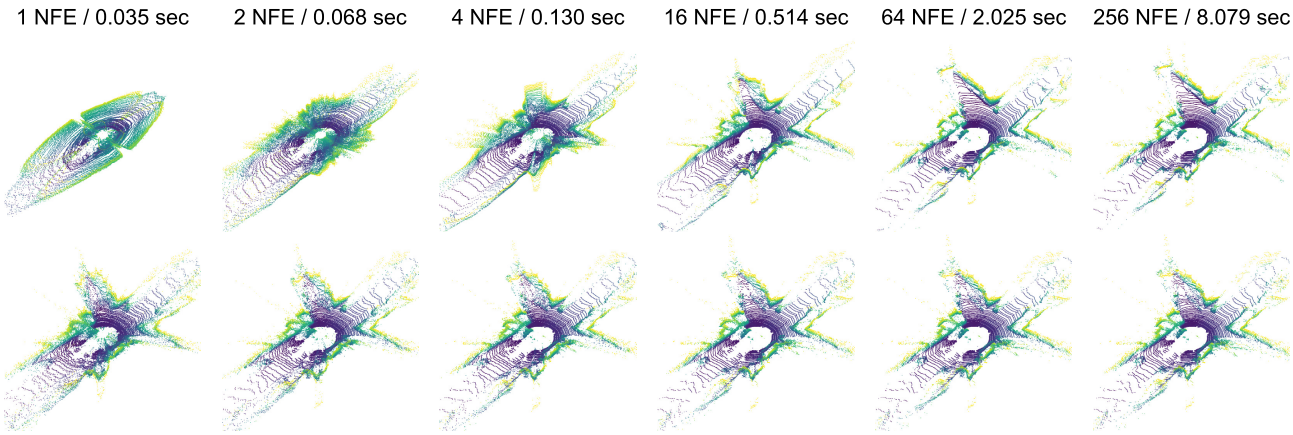

図 2　1-RF（上段）と 2-RF（下段）による提案手法の生成例

表 1　KITTI-360 データセットを用いた生成データの評価

手法	モデル構造（パラメータ数）	反射強度	フレームワーク	NFE	評価指標↓ FRD	FPD	MMD×10^4	JSD×10^2
LiDARGen [15]	RefineNet (30M) [5]	✓	SMLD	1,160	579.39	90.29	7.39	7.38
R2DM [9]	Efficient U-Net (31M) [12]	✓	DDPM	256	154.14	3.79	0.73	2.19
LiDM [10]	LDM (258M) [11] + APE (0.5M)	✗	DDIM	200	N/A	372.81	0.67	4.30
提案手法	HDiT (80M) [2]	✓	1-RF	256	122.82	9.32	0.30	2.16
			2-RF	256	148.10	11.06	0.29	2.24
R2DM [9]	Efficient U-Net (31M) [12]	✓	DDPM	1	2981.78	237.40	89.13	33.48
LiDM [10]	LDM (258M) [11] + APE (0.5M)	✗	DDIM	1	N/A	1240.25	188.41	36.14
提案手法	HDiT (80M) [2]	✓	1-RF	1	2724.66	3967.51	96.90	33.76
			2-RF	1	743.97	27.94	11.05	3.86
			2-RF-distilled	1	336.60	13.08	3.45	2.37
			2-RF-distilled	2	212.09	11.00	3.12	2.39
			2-RF-distilled	4	187.12	10.92	3.08	2.40

では微細な歪みがあるものの，全体の点群形状が保持できている．

5.　まとめと今後の予定

本研究では，条件付きフローマッチングを用いた LiDAR データの生成モデルを提案した．KITTI-360 データセットを用いた生成データの評価実験によって，ステップ数によらず提案手法が高品質な LiDAR データを生成できることを示した．今後は，Reflow および蒸留の損失関数を改良し，さらなる性能向上を目指す．また，学習済みフローを用いた応用手法を開発する．

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] Lucas Caccia, Herke van Hoof, Aaron Courville, and Joelle Pineau. Deep generative modeling of LiDAR data. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5034–5040, 2019.

[2] Katherine Crowson, Stefan Andreas Baumann, Alex Birch, Tanishq Mathew Abraham, Daniel Z Kaplan, and Enrico Shippole. Scalable high-resolution pixel-space image synthesis with hourglass diffusion transformers. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2024.

[3] Sangyun Lee, Zinan Lin, and Giulia Fanti. Improving the training of rectified flows. *arXiv preprint arXiv:2405.20320*, 2024.

[4] Yiyi Liao, Jun Xie, and Andreas Geiger. KITTI-360: A novel dataset and benchmarks for urban scene understanding in 2D and 3D. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol. 45, No. 3, pp. 3292–3310, 2022.

[5] Guosheng Lin, Anton Milan, Chunhua Shen, and Ian Reid. Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1925–1934, 2017.

[6] Xingchao Liu, Chengyue Gong, and qiang liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2023.

[7] Kazuto Nakashima, Yumi Iwashita, and Ryo 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.

[8] Kazuto Nakashima and Ryo 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.

[9] Kazuto Nakashima and Ryo Kurazume. LiDAR data synthesis with denoising diffusion probabilistic models. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 14724–14731, 2024.

[10] Haoxi Ran, Vitor Guizilini, and Yue Wang. Towards realistic scene generation with LiDAR diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14738–14748, 2024.

[11] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10684–10695, 2022.

[12] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, et al. Photorealistic text-to-image diffusion models with deep language understanding. In *Proceedings of the Advances in neural information processing systems (NeurIPS)*, Vol. 35, pp. 36479–36494, 2022.

[13] Dong Wook Shu, Sung Woo Park, and Junseok Kwon. 3D point cloud generative adversarial network based on tree structured graph convolutions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3859–3868, 2019.

[14] Bin Yang, Patrick Pfreundschuh, Roland Siegwart, Marco Hutter, Peyman Moghadam, and Vaishakh Patil. TULIP: Transformer for upsampling of LiDAR point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15354–15364, 2024.

[15] Vlas Zyrianov, Xiyue Zhu, and Shenlong Wang. Learning to generate realistic LiDAR point clouds. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 17–35, 2022.