

深層学習に基づく RGB 画像を用いた低解像 LiDAR データの高解像化

山川 賢人[†] 坂上 文彦[†] 佐藤 淳[†]

名古屋工業大学大学院[†]

1. はじめに

近年、自動車の自動運転や運転支援の実現が進み、自動車に様々なセンサーを搭載するのが一般的となっている。特に自動運転においては、RGB カメラに加えて LiDAR などのセンサーを搭載することが想定されている。RGB カメラは、低コストで高精細な画像を取得することができるが、深度画像を直接取得することができない。深層学習を用いて RGB 画像から深度画像を推定する方法が多く提案されているが、精度が低くドメインシフト問題が課題となっている。

一方 LiDAR は、深度画像を直接取得できるという利点がある。しかし、垂直方向の解像度が低く、価格も高価である。このため、高解像な深度画像を低コストに取得する技術の開発が課題となっている。

そこで本稿では、高解像 RGB 画像と低解像深度画像を組み合わせることで、高解像な深度画像を取得する手法を提案する。提案法では、低解像な深度画像を高解像な欠損した深度画像として捉え、欠損した高解像深度画像の画像インペインティングを敵対的学習 (GAN) [1] により実現する。

2. 深層学習に基づく深度画像の高解像化

本研究では図 1 に示す敵対的生成ネットワーク (GAN) を用いる。低解像深度画像と高解像 RGB 画像を入力とし高解像距離画像を出力する Generator と、真値の深度画像と生成された深度画像を判別する Discriminator より成る。

Generator は、多くの画像生成タスクで使用されている U-net をベースとしている。本研究では、高解像 RGB 画像と低解像深度画像を Generator に組み合わせる際、2 つの手法を提案する。提案法 1 のネットワーク構造を図 2 に示す。この構造は、従来から複数の情報を結合するために用いられ

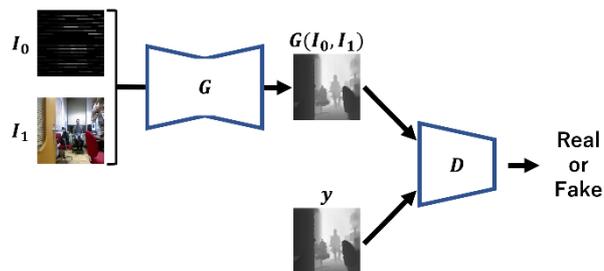


図 1 ネットワーク構造

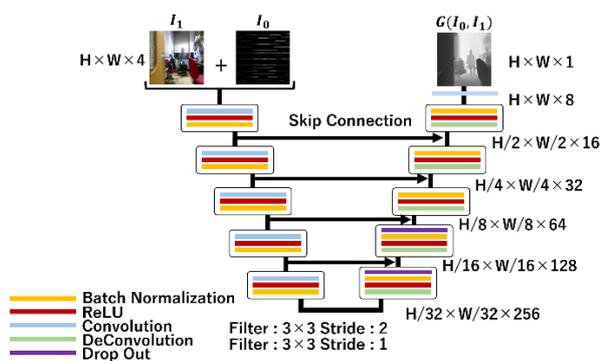


図 2 提案法 1

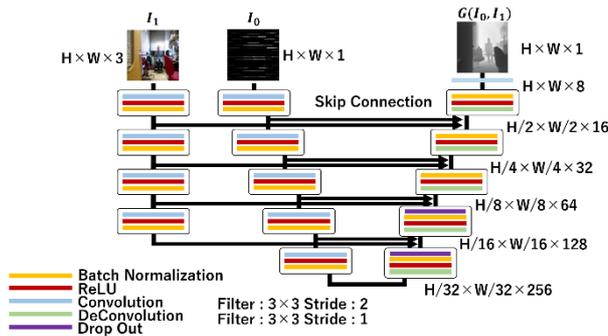


図 3 提案法 2

ている手法である。一方、提案法 2 のネットワーク構造を図 3 に示す。この手法では RGB 画像と深度画像に対して別々に畳み込みを行い、各層の画像特徴量を Skip Connection によって結合する。この構造では、RGB 画像の高解像情報を保持して次の層の画像特徴量へと畳み込むことができる。そのため、U-net のデコーダ部分で画像を生成する際、より高解像な深度画像の生成をすることができる。

Generating High Resolution Depth Image from Low Resolution LiDAR Data using RGB Image Based on Deep Learning

[†]Kento Yamakawa, [†]Fumihiko Sakaue, [†]Jun Sato

[†]Nagoya Institute of Technology

GAN をトレーニングして得られた Generator を G^* とする. G^* は次式に基づいて学習する.

$$G^* = \arg \min_G \max_D \mathcal{L}_{GAN}(G, D) + \lambda \mathcal{L}_{L1}(G) \quad (1)$$

ここで, \mathcal{L}_{GAN} は次式に示す敵対的 Loss を表す.

$$\mathcal{L}_{GAN} = E_{y \sim p_{data}(y)} [\log D(y)] + E_{I_0, I_1 \sim p_{data}(I_0, I_1)} [\log(1 - D(G(I_0, I_1)))]$$

また, \mathcal{L}_{L1} は次式に示す L1 Loss を表す.

$$\mathcal{L}_{L1}(G) = E_{y, I_0, I_1 \sim p_{data}(y, I_0, I_1)} [|y - G(I_0, I_1)|]$$

y は高解像深度画像の真値, I_0 は低解像深度画像, I_1 は高解像 RGB 画像を表す. 式(1)に従ってネットワークを学習することで低解像深度画像から高解像深度画像を生成する Generator が得られる.

3. データセット

次に, 本研究で使用したデータセットについて述べる. 提案法の学習では, 深度画像と RGB 画像のペアが必要である. そこで, NYU Depth Dataset [2] を用いて学習データセットを構築した. NYU Depth は室内画像のデータセットであり, 2284 組の深度画像と RGB 画像のペアで構成されている. このデータセットから, 深度画像と RGB 画像を 256×256 にリサイズし, 2184 組を訓練に, 100 組をテストに使用した. 本研究では, このようにして作成したテストデータによる合成画像実験と, 実際の LiDAR から低解像深度画像を取得してテストを行う実画像実験の 2 つの実験を行った.

4. 実験

Generator と Discriminator は 5000 エポック学習を行い, バッチサイズは 32, 学習の最適化には学習率 0.001 の Adam を用いた.

提案した手法を用いて, 低解像深度画像と RGB 画像から高解像深度画像を生成した合成画像実験の定量評価の結果を表 1 に示す. 比較のため, 低解像深度画像のみから高解像深度画像を生成したものを従来法とした.

表 1 は解像深度画像の精度を RMSE, PSNR で表したものである. この結果より, RGB 画像を組み合わせた提案法では, どちらの指標においても従来法に比べ精度が向上していることが確認できる. また, 高解像情報を保持した提案法 2 が最も良い結果となることが確認できた.

表 1 提案法の定量評価

	従来法	提案法 1	提案法 2
RMSE	6.7396	5.9306	5.8986
PSNR	32.0736	33.1247	33.1436

次に, 実際の LiDAR の低解像深度画像を用いた実画像実験の結果を示す. 合成画像実験と同様に, NYU Depth Dataset を用いて学習を行い, LiDAR (Velodyne VLP-16) から得られた低解像深度画像を学習したネットワークに入力して提案手法の性能を評価した. RGB カメラと LiDAR の間のキャリブレーションは, 事前に射影変換を用いて行った.

実画像実験による高解像深度画像の復元結果を図 4 に示す. 赤枠部分や緑枠部分に注目すると, 従来法に比べ提案法 1, 2 では, 滑らかに復元されていることが確認できる.

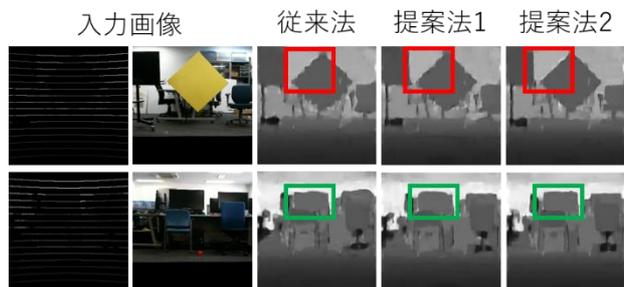


図 4 実画像実験の結果

5. まとめ

本論文では, LiDAR から得られた低解像深度データを入力として, 高解像深度画像を取得する手法を提案した. 高解像 RGB 画像と低解像深度画像を組み合わせる GAN ベースのネットワークを提案した.

提案したネットワークを用いて高解像深度画像を生成するために, 合成画像と実画像の実験を行った. 合成画像実験と実画像実験のどちらにおいても, 提案法がより正確な高解像深度画像を生成できることを示した.

参考文献

- [1] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, Generative Adversarial Networks, Advances in Neural Information Processing Systems, 2014.
- [2] N. Silberman and R. Fergus, Indoor Scene Segmentation using a Structured Light Sensor, International Conference on Computer Vision, 2011.