

点群データに対するセマンティックセグメンテーションにおける CG データによる事前学習の有効性検証

中部大学	正会員	○中村智也	中部大学	非会員	山下隆義
中部大学	非会員	藤吉弘亘	中部大学	非会員	平川 翼
清水建設（株）	正会員	大貫奈々美	清水建設（株）	正会員	山本真哉

1. はじめに

月面環境において、自律型の建設機械を用いて無人の建設活動を行うためには、建設機械周辺の環境を把握することが必要不可欠である。周辺環境を把握するための手法として、建設機械に搭載したカメラで撮影した RGB 画像を用いる手法、または Light Detection and Ranging(LiDAR)で取得した 3 次元点群を用いる手法がある。LiDAR は、夜間などの RGB 画像では撮影しにくい条件でも 3 次元点群を取得できる。そのため、本稿では LiDAR を入力デバイスとして用いる。周辺環境の認識対象には、他の建設機械や人のような移動物体だけでなく、盛り土のような作業対象も含まれる。盛り土の場合、高さや広がり方などが多様である。そのため、一般的に剛体の物体を認識対象とする物体検出では検知できないことも考えられる。そこで、本稿では 3 次元点群の各点に対してクラス分類を行うセマンティックセグメンテーションを物体検知技術として用いる。セマンティックセグメンテーションは、画像内の全画素や点群の各点に対してラベルやカテゴリを関連付けるタスクである。学習には大量の実環境データが必要であり、不十分な場合は精度の低下を招く。そこで、CG により大量のデータを生成して学習に用いることで、実環境データを補填することが期待される。本稿では、点群データに対するセマンティックセグメンテーションにおいて、CG データによる事前学習が有効であるか検証する。

2. セマンティックセグメンテーション

点群データに対するセマンティックセグメンテーションの代表的な手法である SalsaNext[1]は、3 次元点群を擬似画像へ変換し処理を行う。SalsaNext は Context Module, Pixel-Shuffle Layer, Dilated Convolution を導入し、識別精度の向上とリアルタイム性を実現している。SalsaNext のネットワーク構造を図 1 に示す。

3. 提案手法

本研究では、CG データセットを用いて事前学習モデルを作

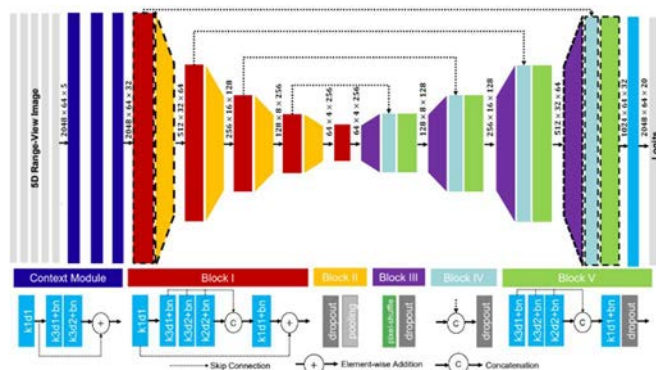


図 1: SalsaNext のネットワーク構造

成し、実環境データセットでファインチューニングすることで事前学習の有効性を検証する。

3.1 CARLA を用いた CG データセット作成

セマンティックセグメンテーションの学習において、データセットが小規模であると、過学習などにより汎化能力が向上しない。そこで、実環境データセットに類似したデータを CG で作成する。CG データセットを用いて事前学習することで汎用的なモデルを獲得する。そして、実環境データセットでファインチューニングすることで実環境のドメインに適応させる。本稿では、CARLA[2]を用いて工事現場を模したマップを作成する。作成したマップ例を図 2(a)に示す。マップ上では、ランダムに盛り土や建機、人を配置して点群を取得する。図 2(a)のマップで取得した点群データを図 2(b)に示す。生成したマップ数は 7、フレーム数は 3458 である。クラスは Background, Sandpile, Vehicle, Human である。

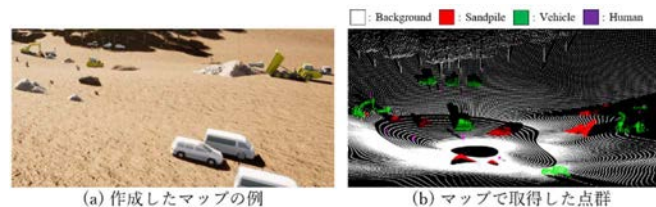


図 2: 取得した点群データの例

3.2 事前学習

事前学習モデルとは、大規模なデータセットを用いて学習

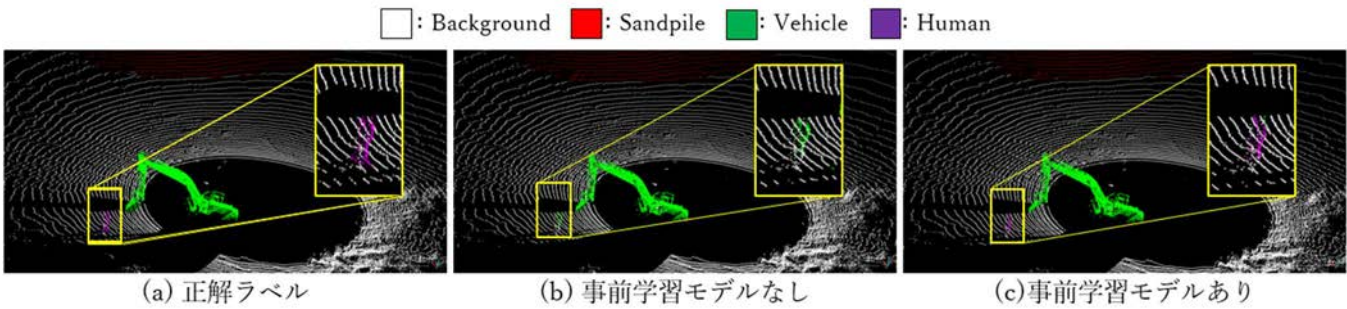


図 3：事前学習モデルの有無による検出結果の比較

したモデルである。事前学習モデルは、ファインチューニングに利用することが可能である。事前学習及びファインチューニングをセットで活用することで、学習データが小規模でも高精度な認識性能を達成することが可能である。本実験では、CG データセットを用いて学習を行い、事前学習モデルを作成する。

3.3 実環境データでの学習

CG データセットを用いて事前学習モデルを作成した後、実環境データセットを用いてファインチューニングを行い、実環境データに対する精度の向上を図る。ファインチューニングは事前学習モデルに、追加したデータセットでモデル全体を再学習する。実環境データは、工事現場で取得した点群データである。実環境データセット内のフレーム数は 383 である。

4. 評価実験

本実験では、事前学習モデルの有無による精度を比較する。

4.1 実験概要

事前学習モデルの学習にはCGデータセットの5シーン(2568フレーム)の内、データ使用率を100%, 50%, 33%に変更して精度を比較する。学習回数は100エポックである。学習時のネットワークにSalsaNext, 最適化手法に Adam, 初期学習率は 1.0×1.0^{-4} とする。事前学習モデルの評価には CG データセットの2シーン(890 フレーム)と実環境データ(33 フレーム)を用いる。ファインチューニングには実環境データ(350 フレーム)を用いる。学習回数は300エポックである。評価には実環境データ(33 フレーム)を用いる。評価指標にはIoUを用いる。IoUの式を式(1)に示す。ここで、TPは真陽性、FPは偽陽性、FNは偽陰性である。IoUは正解領域とモデルが予測した領域の重なり具合を表す指標である。正解領域と予測領域の重なりが大きいほど値が大きくなる。

$$IoU = \frac{TP}{TP+FP+FN} \tag{1}$$

4.2 実験結果

事前学習モデルを CG データと実環境データで評価した結

果を表 1 に示す。表 1 より、CG データで評価した時の学習データの利用率 100%は、50%, 33%と比較した場合、全クラスにおいて精度が向上している。また、実環境データで評価した場合、全クラスにおいて精度が低い。事前学習モデルの有無による評価結果を表 2 に示す。表 2 より、学習データの利用率 100%の事前学習モデルを使用する場合、Background は 0.4pt, Sandpile は 4.4pt, Vehicle は 1.6pt, Human は 24.6pt, Average は 7.8pt 精度が向上した。また、事前学習モデル使用した時の Human クラスにおいて、学習データの利用率 100%の場合に最も高い精度である。事前学習モデルの有無による検出結果を図 3 に示す。図 3 より、事前学習モデルを使用することで、Human の認識率が向上した。

表 1：事前学習モデルの定量的評価

学習	評価	学習データの利用率[%]	Background	Sandpile	Vehicle	Human	Average
CG	CG	100	96.7	59.5	31.8	61.4	62.3
		50	94.1	42.9	21.8	41.8	50.1
		33	93.6	38.6	18.7	14.5	41.3
CG	実環境	100	88.0	13.5	12.2	1.4	28.7

表 2：事前学習モデルの有無による定量的評価の比較

事前学習	学習データの利用率[%]	Background	Sandpile	Vehicle	Human	Average
✓		93.9	52.1	34.3	0.0	45.0
	100	94.3	56.5	35.9	24.6	52.8
	50	94.2	54.7	35.4	0.2	46.1
	33	94.1	54.0	35.1	0	45.9

5. おわりに

本稿では、点群データに対するセマンティックセグメンテーションにおける CG データによる事前学習の有効性を検証した。評価実験により、CG データを用いた事前学習は有効であることを確認した。今後は、CG データセットの増量とクラス間の点群数の不均衡問題について改善する方法を検討する。

謝辞

本研究の一部は「宇宙無人建設革新技術開発」の一環として国土交通省から研究開発を受託して行った。

参考文献

[1] Tiago Cortinhal, et al, “SalsaNext: Fast, Uncertainty-aware Semantic Segmentation of LiDAR Point Clouds for Autonomous Driving”, ISVC2020
[2] Dosovitskiy, Alexey, et al. “CARLA: An open urban driving simulator.” Conference on robot learning. PMLR2017.