

# 全方位 LiDAR を用いた

## 1D Self-Attention Network によるセマンティックセグメンテーション

鈴木 貴大<sup>1)</sup> 平川 翼<sup>1)</sup> 山下 隆義<sup>1)</sup> 藤吉 弘亘<sup>1)</sup>

### 1D Self-Attention Network for Point Cloud Semantic Segmentation using Omnidirectional LiDAR

Takahiro Suzuki Tsubasa Hirakawa Takayoshi Yamashita Hironobu Fujiyoshi

Understanding the environment around a vehicle is essential for automated driving technology. For this purpose, omnidirectional LiDAR is used for obtaining surrounding information, and point cloud-based semantic segmentation methods have been developed. However, these methods require time to acquire point cloud data and to process the point cloud, which causes a significant positional shift of objects in practical application scenarios. In this paper, we propose a 1D self-attention network (1D-SAN) for LiDAR-based point-cloud semantic segmentation, which is based on a 1D-CNN for real-time pedestrian detection of omnidirectional LiDAR data. Because the proposed method can sequentially process segmentation during data acquisition with omnidirectional LiDAR, we can reduce the processing time and suppress positional shift. Moreover, for improving segmentation accuracy, we use the intensity as input data and introduce a self-attention mechanism into the method. The intensity enables us to consider object texture. The self-attention mechanism can consider the relationship between point clouds. Experimental results with the SemanticKITTI dataset show that the intensity input and the self-attention mechanism in the proposed method improve accuracy. In particular, the self-attention mechanism contributes to improving the accuracy for small objects. Also, we show that the processing time of the proposed method is faster than the other point-cloud segmentation methods.

**KEY WORDS:** Safety, Active safety, Road environment recognition, LiDAR, Semantic Segmentation (C1)

#### 1. まえがき

自動運転技術では、車両周辺の環境を把握することが必要不可欠である。物体検出手法は自動運転における基本的な手法である。これは、車載カメラの RGB 画像を用いる手法 (4~6)、Light Detection and Ranging (LiDAR) で取得した 3 次元点群を用いる手法 (7~10) に分類される。LiDAR は赤外線レーザを照射し、物体に反射して返ってくるまでの時間を計測することで、3 次元情報や反射強度を取得する技術である。LiDAR にはいくつかの種類があるが、全方位 LiDAR は、回転しながらレーザを周囲に照射し、LiDAR を原点とした 360 度全方位の 3 次元情報を取得する。全方位 LiDAR を用いたセマンティックセグメンテーション手法は数多く提案されている (10, 17~22)。しかし、全方位 LiDAR を用いる場合、データの取得に時間がかかり、処理完了までに物体の位置ずれが生じる可能性がある。

本研究では、全方位 LiDAR に基づくセマンティックセグメンテーション手法である 1dimensional self-attention network (1D-SAN) を提案する。提案手法では、360 度の点群データ全体の中の一部の点群を逐次処理する。本手法は、全方位 LiDAR の歩行者検出手法である 1 次元畳み込みニューラルネットワーク (1D-CNN) (1) に基づいている。1D-CNN は、LiDAR

から得られる距離値を 1 次元波形データとみなし、ネットワークへの入力としている。これにより、データ取得時に逐次処理を行うことが可能になり、物体の位置ずれを抑制することができる。これをセマンティックセグメンテーションタスクに応用し、複数のクラスに対応する。また、全方位 LiDAR により取得した反射強度も入力データとして用いる。反射強度は物体の材質により異なる性質があるため、物体の質感を考慮した識別が可能となる。さらに、self-attention block (2) に基づいた 1dimensional self-attention block (1D-SAB) を提案する。1D-SAB を導入することで、点群間の関係性を考慮することができる。反射強度と 1D-SAB により、全方位 LiDAR データの逐次処理を維持しつつ、セグメンテーション精度の向上を図る。

#### 2. 関連研究

LiDAR から取得したもののみならず、点群を用いた手法は多数提案されている (7~27)。これらは、点群をボクセルとして扱う手法、点群を画像上に投影し画像として扱う手法、3 次元点群をそのまま用いる手法の 3 つのアプローチに分類される。また、上記のアプローチに加えて、1D-CNN による歩行者検出がある。本章では、点群を用いた各アプローチおよび代表的な手法を簡潔に説明する。

1) 中部大学 (487-8501 愛知県春日井市松本町 1200)

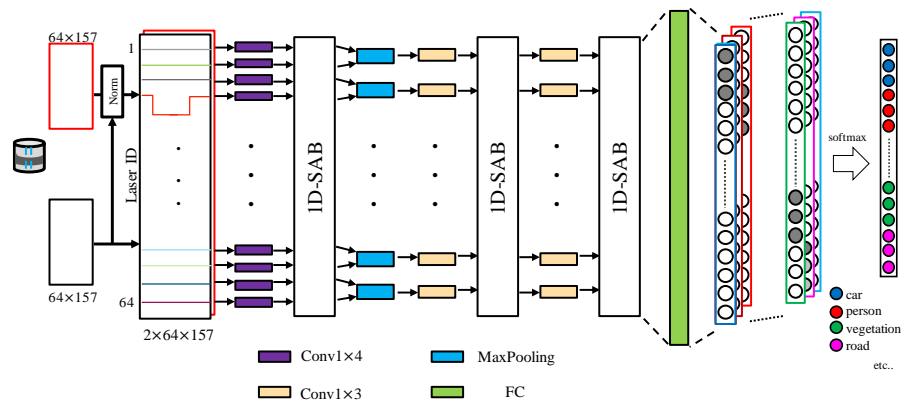


Fig.1 Network structure of 1D-SAN

### 2.1. ボクセルとして扱う手法

点群をボクセルとして扱う手法では、まず 3 次元点群をボクセルに変換する。その後、ボクセル化された点群情報に対し、3 次元畳み込み処理などを行い、各タスクに対応する (14, 15)。

点群をボクセル化することで、3 次元点群本来の情報を保持しやすく、3 次元畳み込みによる円滑な特徴抽出が可能である。また、3 次元点群の疎であるという欠点を改善し、ボクセルごとにまとまっており各タスクへ扱いやすくなっている。一方で、ボクセルは律法的に形成されるため、計算コストが高く処理が遅くなりやすいという欠点がある。

### 2.2. 点群を画像として扱う手法

点群を画像として扱う手法では、まず 3 次元点群情報を 2 次元画像のように変換する。変換された 2 次元画像に対し、通常の画像への処理と同様に、2 次元畳み込み処理などを行う (9, 10, 13, 16~18, 20~22)。

3 次元点群を 2 次元画像のように変換することで、通常のような畳み込み処理が可能となり、処理の高速化を実現している。一方で、変換された画像上において、ある画素とその近傍の画素は実際の近傍点ではないなど、3 次元点群本来の情報が欠落する可能性がある。

### 2.3. 3 次元点群を直接入力する手法

3 次元点群情報を直接処理する手法では、点群をボクセル化や 2 次元画像へ変換したりせず、そのままネットワークへ入力し処理を行う (7, 8, 11, 12, 19, 23~27)。入力には、3 次元座標情報や物体の反射強度値などを用いる。

### 2.4. 1D-CNN による歩行者検出

従来手法の問題点として、検出処理に時間がかかるため、自動車など高速で移動するものに搭載して検出する場合、検出した位置と実際の位置にずれが生じる場合がある。この問題に着目し、位置ずれを抑えることを可能とした 1D-CNN による歩行者検出 (1) がある。

1D-CNN による歩行者検出では、全方位 LiDAR の回転と共に逐次的に処理を行う。まず、LiDAR から取得できる距離値をレーザ ID ごとに 1 次元波形データとみなし、逐次的に 1D-CNN

に入力することで LiDAR の回転と共に歩行者検出を可能としている。実際の車両の走行中、LiDAR の回転とともに歩行者分の点群数が確保できたら逐次ネットワークへ入力し、歩行者か背景かの識別を行う。さらに、歩行者として識別した点群に対してクラスタリングを行うことで、高精度化を実現している。

しかし、この手法では検出対象を歩行者のみとしている。実際の自動運転では、歩行者以外の物体を検出することも重要である。また、入力データを距離値のみとしているため、物体の形状だけで判断してしまう。そのため、同形で異なる物体が存在した場合、誤識別をしてしまう可能性がある。

### 3. 提案手法

本章では、提案手法の詳細な説明を行う。1-dimensional self-attention network (1D-SAN) の構造を図 1 に示す。まず、LiDAR から距離値と反射強度値を取得し、外れ値へ値を補間する。次に、これらのデータを 1 次元波形データと仮定し、チャンネル方向へ結合する。このデータを 1D-SAN へ入力し波形データのレーザ ID ごとのクラス確率を出力する。この処理により、1D-CNN の利点を維持しつつ多クラス分類を実現する。また、提案したネットワークには 1-dimensional self-attention block (1D-SAB) を導入している。1D-SAB により、点群間の相対的な位置関係を重み決定に反映させることが可能となる。以上により、各物体の質感や点群間の重要な関係性を考慮した識別ができる。

#### 3.1. 点群データの前処理

点群データは、全方位 LiDAR から照射したレーザが物体に反射し返ってきた点の集合であり、その時間から距離値と反射強度値を取得することができる。本研究にて用いる LiDAR は Velodyne 社製の HDL-64E である。HDL-64E は 64 本のレーザを搭載しており、0.175 度の間隔で全方位に照射することから、1 フレームにつき点群は縦 64、横 2048 の合計 131,072 点となる。しかし、全方位 LiDAR には空や鏡面体など反射光が取得困難なシーンも含まれる。反射光が取得できなかった場合、その地点では値が取得できず外れ値となるため、外れ値部分の

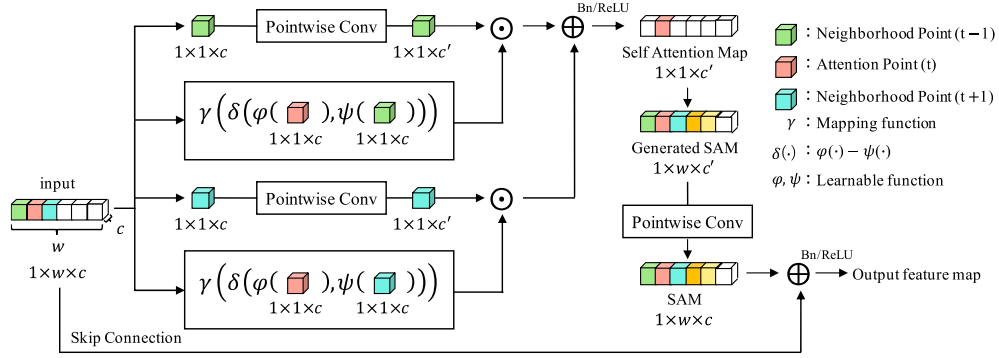


Fig.2 Detailed structure of 1D-SAB

値の補間が必要である。本研究では、距離値の外れ値には、照射角度が $0^\circ$ 以上のときは空と仮定しレーザの最大照射距離値120、照射角度が $0^\circ$ 未満のときはレーザが地面に当たったと仮定し対応した値を補間する。同様に、反射強度の外れ値にも、照射角度が $0^\circ$ 以上のとき0.0、照射角度が $0^\circ$ 未満のときは道路や歩道などの地面クラスの反射強度の平均値0.29を補間する。補間後の距離値と反射強度をレーザID毎に結合することで、レーザID毎の波形データを作成する。

### 3.2. ネットワーク構造

本節では、1D-SANの構造について詳細な説明を行う。1D-SANは図1に示すように3層の畳み込み層、3層の1D-SAB、1層の全結合層から構成されている。すべての畳み込み層では、水平方向にのみ畳み込みが行われる。入力データの縦方向のサイズおよび出力ユニット数は、LiDARのレーザID数に依存する。本研究にて用いるLiDARは64本のレーザを照射するため64である。横方向のサイズは、セマンティックセグメンテーションに最適である30度分の点群数157とする。30度分が最適であると判断した理由は、入力角度を変更した比較実験において総合的な精度が最も高かったためである。入力チャンネル数は、距離値と反射強度の2チャンネルである。入力データの1次元畳み込みは、レーザIDごとに行う。各畳み込み層の後、1D-SABに入力することで点群間の関係性を考慮した処理を行う。その後、全結合層にて特徴を結合し、softmax関数にて入力データの中央の各レーザIDに対するクラス確率を算出し、識別結果を出力する。これを逐次的に行うことで、全方位に対するセマンティックセグメンテーションが可能となる。

### 3.3. 反射強度の追加

セマンティックセグメンテーションの識別精度を向上させるため、入力データに物体の反射強度を追加する。反射強度には、金属など反射率の高い物体では光が拡散するため弱くなり、布など反射率の低い物体では光が正確に返ってくるため強くなるという性質がある。したがって、反射強度をネットワークへの入力に追加することで、物体の質感を考慮した識別が期待できる。また、反射強度は物体への距離が遠くなるほど減衰し、返ってくる値が小さくなる。そこで、LiDARからの距

離値を用いて正規化することで値を補正し、距離による減衰の影響を抑制する。反射強度の正規化には、光の減衰の法則を用いる。光の減衰の法則は、光の強さは光源から物体までの距離の2乗に反比例するというものである。距離値による反射強度の正規化の式を式(1)に示す。

$$I' = I * (2 * d)^2 \quad (1)$$

ここで、 $I'$ は正規化後の反射強度、 $I$ は正規化前の反射強度、 $d$ はLiDARから得られた距離値である。式(1)を適用することで、距離により減衰した値を復元することが可能となり、反射強度を入力に用いることができる。

### 3.4. 1-Dimensional Self-Attention Block (1D-SAB)

本節では、1D-SABの詳細な説明を行う。1D-SABの構造を図2に示す。作成した1次元波形データを、レーザIDごとに1D-SABに入力される。入力したデータは1点ずつ処理され、対応する点のself-attentionを算出する。図2の赤色の値を処理の注目点としたとき、緑色が近傍点1、青色が近傍点2となる。各近傍点に対して、pointwise conv処理を行う。また、注目点と近傍点を学習可能な関数 $\varphi, \psi$ へ入力する。そして、 $\varphi, \psi$ を用いて関係関数 $\delta$ を求める。関係関数 $\delta$ の定義式を式(2)に示す。 $x$ が注目点、 $x_t$ が近傍点を示す。

$$\delta(\varphi(x), \psi(x_t)) = \varphi(x) - \psi(x_t) \quad (2)$$

その後、マッピング関数 $\gamma$ によってチャンネル数を1つ目の出力と合わせる。そして、上述の特徴量との要素積を算出する。この処理を近傍点分行い、それらを総和することでself-attention map (SAM)を生成する。生成したSAMは、pointwise conv処理により入力チャンネル数と同じチャンネル数にする。この出力に、スキップ機構として入力データを加算し、最終出力とする。1D-SABを用いることで、点群間の重要な位置に大きな重みを与えることができ、点群間の関係性を考慮することができる。

### 4. 評価実験

本章では、提案手法の有効性を検証する。評価実験では、セマンティックセグメンテーションへの反射強度の有効性および1D-SABの有効性を検証する。また、他の手法との精度比較も行う。

Table 1 Evaluation of effectiveness of proposed method

Approach	intensity	car	bicycle	motorcycle	truck	othe-vehicle	person	bicyclist	motorcyclist	road	parking	sidewalk	other-ground	building	fence	vegetation	trunk	terrain	pole	traffic sign	mean-IoU
1D-CNN(1)		54.7	0.0	0.0	0.0	0.3	0.0	0.0	0.0	69.6	3.9	39.3	0.0	59.9	20.6	57.9	4.2	40.2	15.0	0.0	19.4
Ours (w/o SAB)	✓	57.1	0.0	0.0	0.3	0.1	0.0	0.0	0.0	<b>72.6</b>	10.7	<b>43.0</b>	0.0	61.3	17.6	59.9	3.7	42.0	13.9	2.4	20.2
Ours (w SAB)	✓	<b>58.1</b>	<b>3.3</b>	<b>2.2</b>	<b>4.8</b>	<b>3.4</b>	<b>6.0</b>	<b>10.7</b>	<b>0.4</b>	<b>72.6</b>	<b>16.8</b>	38.7	<b>1.2</b>	<b>66.9</b>	<b>28.4</b>	<b>70.7</b>	<b>5.5</b>	<b>54.7</b>	<b>29.7</b>	<b>32.2</b>	<b>26.6</b>

Table 2 Quantitative comparison with other segmentation methods

Approach	car	bicycle	motorcycle	truck	othe-vehicle	person	bicyclist	motorcyclist	road	parking	sidewalk	other-ground	building	fence	vegetation	trunk	terrain	pole	traffic sign	mean-IoU
PointNet(7)	46.3	1.3	0.3	0.1	0.8	0.2	0.2	0.0	61.6	15.8	35.7	1.4	41.4	12.9	31.0	4.6	17.7	2.4	3.7	14.6
PointNet++(8)	53.7	1.9	0.2	0.9	0.2	0.9	1.0	0.0	72.0	<b>18.7</b>	41.8	<b>5.6</b>	62.3	16.9	46.5	13.8	30.0	6.0	8.9	20.1
SPGraph(19)	<b>68.3</b>	0.9	<b>4.5</b>	0.9	0.8	1.0	6.0	0.0	49.5	1.7	24.2	0.3	68.2	22.5	59.2	27.2	17.0	18.3	10.5	20.0
SPLATNet(25)	66.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	70.4	0.8	41.5	0.0	<b>68.7</b>	27.8	<b>72.3</b>	<b>35.9</b>	35.8	13.8	0.0	22.8
Ours (w/o SAB)	57.1	0.0	0.0	0.3	0.1	0.0	0.0	0.0	<b>72.6</b>	10.7	<b>43.0</b>	0.0	61.3	17.6	59.9	3.7	42.0	13.9	2.4	20.2
Ours (w SAB)	58.1	<b>3.3</b>	2.2	<b>4.8</b>	<b>3.4</b>	<b>6.0</b>	<b>10.7</b>	<b>0.4</b>	<b>72.6</b>	16.8	38.7	1.2	66.9	<b>28.4</b>	70.7	5.5	<b>54.7</b>	<b>29.7</b>	<b>32.2</b>	<b>26.6</b>

#### 4.1. 実験概要

##### 4.1.1. データセット

データセットには, SemanticKITTI (3) を用いる. これは, 自律走行を目的とした KITTI データセット (28) に基づいて作成された実環境のデータセットである. 3次元点群データは HDL-64E で取得されている. SemanticKITTI は 22 シーンで構成されており, フレーム数は 43,000 である. このうち, シーン 00 から 10 の 21,000 フレームを学習用, シーン 11 から 21 までの 22,000 フレームを評価用とする. 学習用シーンのうち, シーン 08 を検証用データとして学習を行う.

SemanticKITTI はこの KITTI データセットすべての点群データに対してアノテーションが施されており, 人や車, 建物など計 22 クラスが定義されている. 人や車は動きの有無でもクラス分けがされている. 本研究では, 動的物体を静的物体と同様に扱うため, 識別対象は 19 クラスである.

##### 4.1.2. 比較手法

比較手法を以下に示す. 反射強度と 1D-SAB の有効性と処理速度を評価するために, 以下の手法の精度を比較した.

- 1D-CNN : 1D-CNN (1) の出力を 19 クラスの確率を算出するように拡張. 反射強度は入力に用いない.
- Ours(w/o SAB) : 反射強度を入力とし, 提案手法から 1D-SAB を取り除いたモデル.
- Ours(w SAB) : 反射強度の入力, 1D-SAB を両方を使用.

学習設定は, 20 epoch, バッチサイズは 24 とする. 損失関数には, クロスエントロピー誤差, 最適化手法には MomentumSGD を使用し, 初期学習率は 0.01 とする. 学習時は 1 epoch ごとに 1/2 倍に減衰させる. さらに, 点群に対するセマンティックセグメンテーション手法である PointNet (7), PointNet++ (8), SPGraph (19), SPLATNet (25) と比較を行う.

##### 4.1.3. 評価方法

評価には, Intersection over Union (IoU) を用いる. IoU は, 各点群の正解ラベルとセグメンテーション結果がどの程度合致しているかを示す指標である. 正解した点群数を True Positive (TP), 正解クラスを別のクラスと予測した点群数を False Positive (FP), 別のクラスを正解クラスと予測した点群数を False Negative (FN) とすると, IoU は式 (3) のように表すことができる.

$$\text{IoU} = \frac{TP}{TP + FP + FN} \quad (3)$$

また, 総合的な評価指標として Mean IoU を用いる. Mean IoU はクラスごとの IoU の平均をとることで求められるため, 式 (3) のように表すことができる.  $C$  はクラス数を示す.

$$\text{mIoU} = \frac{1}{C} \sum_{i=1}^C \frac{TP}{TP + FP + FN} \quad (4)$$

#### 4.2. 反射強度の有効性の評価

まず, 反射強度の有効性を評価する. 従来手法と提案手法の精度比較を表 1 に示す.

1D-CNN と Ours(w/o SAB) を比較すると, 反射強度を導入することで, mIoU が 0.8pt 向上していることがわかる. また, 反射強度を導入した場合, 19 クラス中 9 クラスの IoU が向上している. しかし, fence や trunk など反射強度を考慮しても IoU が低下しているクラスもある. このことから, 物体によっては反射強度を考慮しても精度が向上しない場合もあると考えられる. 一方で, 自動運転において重要であると考えられる car や road, traffic-sign に対する IoU が向上していることが確認できた. また, IoU が低下しているクラスに関しては, 従来手法とほとんど同等の精度を発揮している.

Table 3 Comparison of processing speed

	speed(msec)
PointNet(7)	208.333
1D-CNN(1)	0.0814
Ours(w/o SAB)	0.0888
Ours(w SAB)	0.3488

これらにより、反射強度の導入はセマンティックセグメンテーションにおいて有効であると言える。

#### 4.3. 1D-SAB の有効性の評価

次に、1D-SAB の有効性を評価する。表 1 より、1D-SAB を導入したモデルの mIoU が 26.6% となっており、1D-SAB を導入しない場合と比較して 6.4pt 向上している。これにより、self-attention 機構により点群間の関係を考慮することは、セマンティックセグメンテーションの精度向上に有効であると言える。クラスごとの IoU では、1D-SAB を導入することで 19 クラスのうち 17 クラスの IoU が最も高い値となっている。特に、bicyclist や pole, traffic-sign といった小物体の精度が大きく向上した。このことから、SAB を用いて重みの決定をすることは、特に小物体など点群数が少ないクラスに関する精度向上に貢献すると言える。

#### 4.4. 従来手法との精度比較

従来手法との精度比較を表 2 に示す。1D-SAB を導入した提案手法は mIoU が 26.6% となっており、最も高精度となった。クラスごとの IoU を見ると、1D-SAB を導入した提案手法は、19 クラスのうち 11 クラスが最も高精度となった。特に、fence や pole, traffic-sign といった小物体の精度が他の方と比較し大きく向上した。また、残りの 8 クラスも他の手法とほとんど同程度の精度を発揮していることがわかる。一方で、1D-SAB を導入していない場合も、距離値と反射強度の入力のみで SPLATNet 以外の手法と比較し、mIoU が向上した。

#### 4.5. 定性的評価

図 3 に SemanticKITTI へのセグメンテーション結果の可視化例を示す。図 3(b) のように、1D-CNN では car を vegetation, road を terrain と誤認識している部分がある。それに対し、Ours(w/o SAB) は 1D-CNN と比較して car や road を正しく認識できている。さらに、Ours(w SAB) では、より ground truth に近い認識ができており、特に pole や trunk といった小物体への認識精度が向上していることが確認できる。

#### 4.6. 処理速度の比較

処理速度の比較を表 3 に示す。PointNet の速度は LiDAR の 1 回転分のデータに対する処理時間、1D-CNN および提案手法の速度は逐次処理を行うデータ範囲に対する処理時間として算出した。処理速度の計測は、すべて NVIDIA Quadro RTX 8000 を用いて行っている。表 3 から、1D-CNN が 0.0814msec となり、最も高速であることがわかった。これは、距離値のみを 1 次元畳み込み処理をすることから、最も軽量であるためだと

考えられる。反射強度を入力データとして用いたモデルは、1D-CNN と比較して 0.0074msec 速度が低下した。また、精度比較の結果が最も良かった 1D-SAB を導入したモデルも、1D-CNN や Ours(w/o SAB) と比較して大幅に速度が低下し 0.3488msec となった。しかし、全方位 LiDAR は 5Hz~20Hz で稼働するため、200msec 以下の速度で処理できている提案手法は、リアルタイム性を維持できていると言える。

#### 5. まとめ

本研究では、全方位 LiDAR を用いた 1-dimensional self-attention network (1D-SAN) によるセマンティックセグメンテーション手法を提案した。提案手法は、LiDAR データの一部を処理し、セグメンテーション結果を逐次出力することで、処理時間を削減することができる。SemanticKITTI を用いた評価実験では、反射強度と 1D-SAB を用いることで、処理時間を抑えつつセグメンテーション精度を向上させることを可能とした。特に、小物体の精度が向上することがわかった。今後の課題としては、実用化に向けてより高い精度を維持しつつ、より低い計算コストを実現することが挙げられる。

#### 謝辞

本研究は、総合科学技術・イノベーション会議の戦略的イノベーション創造プログラム (SIP) 第 2 期/自動運転 (システムとサービスの拡張) 「自動運転技術 (レベル 3, 4) に必要な認識技術等に関する研究」 (管理法人: NEDO) によって実施されました。

#### 参考文献

- (1) Y. Kunisada, et al., “Pedestrian Detection Method Based on 1D-CNN During LiDAR Rotation”, in ITSC, 2018.
- (2) H. Zhao, J. Jia, and V. Koltun, “Exploring self-attention for image recognition”, in CVPR, 2020.
- (3) J. Behley, et al., “SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences”, in ICCV, 2019.
- (4) N. Dalal, and B. Triggs, “Histograms of oriented gradients for human detection”, in CVPR, 2005.
- (5) C. Papageorgiou, and T. Poggio, “A trainable system for object detection”, International journal of computer vision, vol. 38, pp.15-33. 2000.
- (6) Y. LeCun, et al., “Backpropagation applied to handwritten zip code recognition”, Neural computation, pp.541-551, 1989.
- (7) C. R. Qi, et al., “PointNet: Deep learning on point sets for 3d classification and segmentation”, in CVPR, 2017.
- (8) C. R. Qi, L. Yi, H. Su, and L. J. Guibas, “PointNet++: Deep hierarchical feature learning on point sets in metric space”, in NeurIPS, 2017.
- (9) X. Chen, et al., “Multi-view 3d object detection

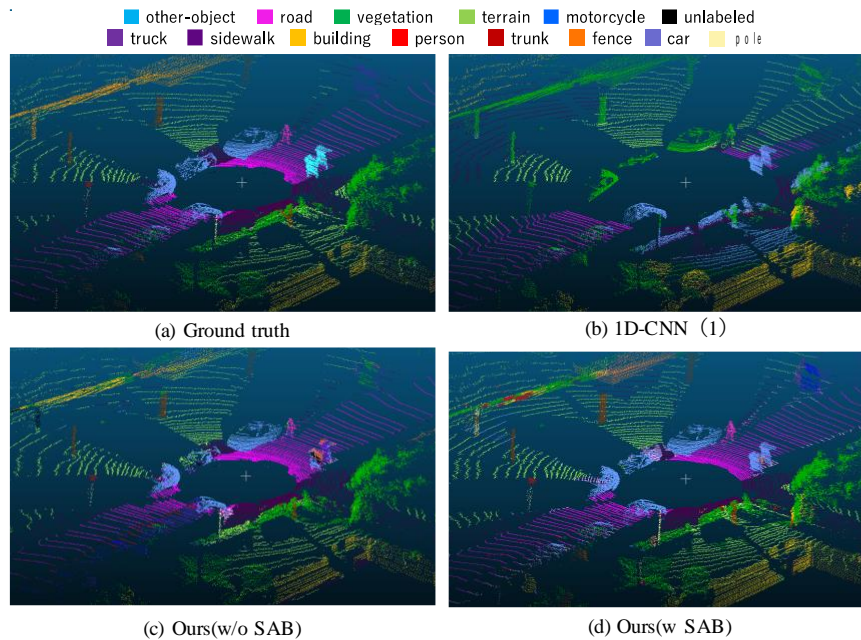


Fig.3 Visualization results on SemanticKITTI test set

network for autonomous driving”, in CVPR, 2017.

(10) B.Wu, et al., “SqueezeSeg : Convolutional neural nets with recurrent crf for real-time road-object segmentation from 3d lidar point cloud”, in ICRA, 2018.

(11) L.Spinello, et al., “Tracking people in 3d using a bottom-up top-down detector”, in RAS, 2011.

(12) C.R.Qi, et al., “Frustum pointnets for 3d object detection from rgb-d data”, in CVPR, 2018.

(13) M.Simon, et al., “Complex-yolo : Real-time 3d object detection on point clouds”, arXiv, 2018.

(14) D.Maturana, and S.Scherer, “Voxnet: A 3d convolutional neural network for real-time object recognition”, in IROS, 2015.

(15) Y.Zhou, and O.Tuzel, “Voxelnet:End-to-end learning for point cloud based 3d object detection”, in CVPR, 2018.

(16) I.Alonso, et al., “3d-mininet: Learning a 2d representation from point clouds for fast and efficient 3d lidar semantic segmentation”, in CVPR, 2020.

(17) G.P.Meyer, et al., “LaserNet: An efficient probabilistic 3d object detector for autonomous driving”, in CVPR, 2019.

(18) T.Cortinhal, et al., “Salsanext: Fast, uncertainty-aware semantic segmentation of lidar point clouds for autonomous driving”, arXiv, 2020.

(19) L.Landrieu, and M.Simonovsky, “Large-scale point cloud semantic segmentation with superpoint graphs”, arXiv, 2018.

(20) B.Wu, et al., “SqueezeSegv2: Improved model structure and unsupervised domain adaptation for road-object segmentation from a lidar point cloud”, in ICRA, 2019.

(21) C.Xu, et al., “SqueezeSegv3: Spatially-adaptive convolution for efficient point-cloud segmentation”, arXiv, 2020.

(22) A.Milioto, et al., “RangeNet++: Fast and Accurate LiDAR Semantic Segmentation”, in IROS, 2019.

(23) H.Thomas, et al., “KPConv: Flexible and deformable convolution for point clouds”, in CVPR, 2019.

(24) Q.Hu, et al., “Randla-net: Efficient semantic segmentation of large scale point clouds”, in CVPR, 2020.

(25) H.Su, et al., “SPLATNet: Sparse lattice networks for point cloud processing”, in CVPR, 2018.

(26) Y.Zhang, et al., “Polarnet: An improved grid representation for online lidar point clouds semantic segmentation”, in CVPR, 2020.

(27) M.Tatarchenko, et al., “Tangent convolutions for dense prediction in 3D”, in CVPR, 2018.

(28) A.Geiger, et al., “Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite”, in CVPR, 2012.