

入力可変長 PointNet による自転車認識 *

山本大貴** 新村文郷** 出口大輔** 川西康友**
井手一郎** 加藤一樹*** 村瀬洋**

Cyclist Recognition via Size-Adaptable PointNet

Taiki YAMAMOTO, Fumito SHINMURA, Daisuke DEGUCHI, Yasutomo KAWANISHI,
Ichiro IDE, Kazuki KATO and Hiroshi MURASE

In recent years, many researches on surrounding environment recognition using a LIDAR have been actively conducted in order to develop autonomous driving. PointNet, which is a DNN-based object recognition method that directly processes 3D point clouds, has shown good performance. However, it has a limitation that since the number of input points is fixed, it cannot be applied to point-clouds containing varying numbers of points. Therefore, this paper proposes a novel sampling module that allows the input of an arbitrary number of points into the PointNet. The proposed module features two functions: down-sampling that can maintain the shape of target objects, and up-sampling based on LIDAR characteristics. To evaluate the effectiveness of the proposed method, an experiment is conducted on point clouds provided by the KITTI Vision Benchmark Suite. The experimental result shows that the recognition method using PointNet with the proposed sampling module outperforms the conventional methods.

Key words: cyclist recognition, 3D point cloud, LIDAR, size-adaptable PointNet

1. 緒 言

近年、自動運転の実用化を目指して、自車両の周辺環境認識に関する研究が活発に行われている。特に最近では、車両周辺の環境認識のための車載距離センサとして注目されている LIDAR (Light Detection and Ranging) を用いた周辺環境認識技術の実用化を目指して広く研究が進められている。これらの研究の多くは自動車や歩行者の認識や検出を目的としているが、一般道における自動運転を実用化するためには、自転車の認識や検出も重要である。自転車は歩行者に比べて移動速度が速いため、自動車との衝突の危険性が高い。衝突事故を防ぐためには、迅速かつ正確な自転車の検出が求められる(以降、自転車の車体とその搭乗者を含めて自転車と呼ぶ)。そこで本論文では、LIDAR を用いた高精度な自転車認識の実現を目的とする。

LIDAR は、計測範囲に LASER 光を照射し、その反射光が計測されるまでの時間や反射光の反射強度を解析することで、車両周辺に存在する対象物体までの距離やその反射特性を 3次元点群データ(以降、点群と呼ぶ)として取得できるセンサである。自転車や歩行者から計測された点群を可視化した例を図 1 に示す。点群に含まれる点の位置の分布を見ることで、物体の大まかな輪郭形状を把握することができ、その情報は物体を認識する上で非常に重要だと考えられる。本論文では、点群に含まれるこのような点の分布の形状を、点群の分布形状と呼ぶ。

LIDAR を用いた物体認識は、LIDAR により計測された点群から立体物を抽出する物体候補抽出と、物体候補の点群のクラスを認識する物体認識の処理から構成される。自転車の候補点群の抽出については、これまで多くの物体候補抽出手法が提

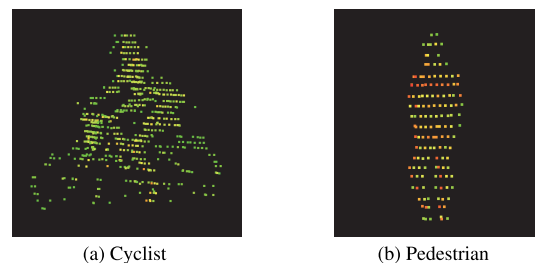


Fig. 1 Samples of point cloud

案されており¹⁾²⁾、物体候補としては自転車点群のほかに、自動車や歩行者、電柱から計測された点群が抽出される。本論文では、それらの物体候補として得られた点群から自転車を認識する自転車認識手法に焦点を当てる。これまでに、点群から物体を認識する手法が多数提案されている³⁾⁻⁷⁾。その中でも、深層ニューラルネットワーク (Deep Neural Network; DNN) に点群を直接入力して認識する PointNet⁶⁾ が注目されている。PointNet は、入力された点群から点の順序や回転に不変な特徴を学習することで高い認識性能を得ている。しかし、点群の入力点数が固定という制約があり、認識のためには点群から一定の点数を適切にサンプリングしなければ性能が大きく低下することが知られている。従来は認識対象の点群が高密度に計測されていると仮定し、単純なランダムサンプリングによって一定の入力点数に調整したうえで認識処理を行っている。しかし、対象との距離が離れるに従い計測される点群の密度は低下し、特に自転車の車体のように細い部分から計測できる点数は少なくなる。よって、従来手法で用いられている単純なランダムサンプリングでは車体部分の点を選択されず、元の点群の分布形状から大きく変化する可能性がある。特に自転車認識においては、自転車の車体部分を含めた形状が重要なため、点群の分布形状の保持は重要と考えられる。また、計測される点数が入力

* 原稿受付 令和元年 5 月 7 日

掲載決定 令和元年 9 月 4 日

** 名古屋大学大学院 (愛知県名古屋市千種区不老町)

*** 株式会社デンソー (愛知県刈谷市昭和町 1-1)

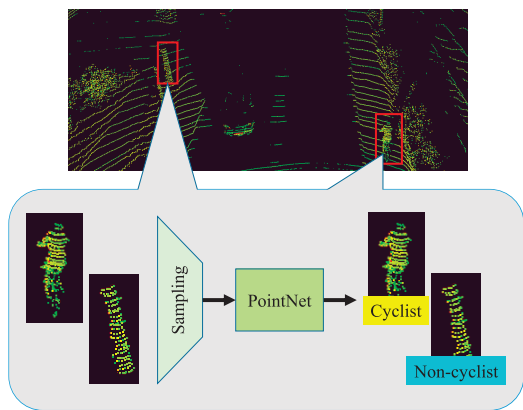


Fig. 2 Outline of the proposed size-adaptable cyclist recognition method

点数より少ない場合は PointNet を適用できない。単純に重複を許してアップサンプリングする方法も考えられるが、同じ点の情報を用いることになるため、認識精度の向上は見込めない。これらの問題に対処するためには、PointNet に入力する計測点を適切に選択する必要がある。

そこで本論文では、LIDAR により計測された自転車の認識精度向上を目的とし、任意の入力点数に対して適用を可能にした入力可変長 PointNet を提案する (図 2)。具体的には、点群の分布形状を保持したダウンサンプリングと LIDAR の測定精度を考慮したアップサンプリングを行うサンプリングモジュールを提案する。ダウンサンプリングでは、低密度部分の計測点が保存されるようにサンプリングを行うことで、自転車の点群の分布形状を保持したサンプリングを可能とする。アップサンプリングでは、LIDAR の測定精度を考慮し、対象までの距離に応じて点群に摂動を加えて点群の密度の増加を図る。そして、この提案サンプリングモジュールを導入することにより、PointNet の入力の可変長化を可能とし、認識精度が向上することを示す。

本論文の貢献は以下の通りである。

- LIDAR から得られる点群からの自転車認識手法の提案。
- 自転車の形状および LIDAR の測定精度を考慮したサンプリングモジュールの提案。
- PointNet の入力の可変長化による自転車の認識精度の向上。

なお、本論文は文献⁸⁾を発展させてまとめたものである。

本論文の構成は、まず 2. で本論文に関連する研究について述べる。次に、3. で従来の PointNet の入力を可変長にする提案手法の概要について述べ、4. で提案手法の実装方法について述べる。5. では、提案手法の有効性を確認するために行った実験について述べ、6. でその考察を述べる。最後に、7. で本論文のまとめと今後の課題について述べる。

2. 関連研究

LIDAR から得られる点群を用いた物体認識の研究がこれまでも報告されている^{3) 4) 9)–11)}。Kidono ら³⁾は、高解像度 LIDAR を用いて遠方の歩行者を検出するため、スライス特徴量と反射強度の分布に関する特徴量を提案している。スライス特徴とは、点群を一定の高さ毎に分割し、分割されたスライス状の点群の分布形状の組み合わせにより全身の形状を表現する

特徴量である。Tatebe ら⁴⁾は、複数の時刻に取得された同一物体の点群を統合することで点群の密度を増加させる特徴抽出手法と、点群の時間変化を表現する特徴抽出手法を提案し、低密度な点群として計測された歩行者の認識精度の向上を実現している。これらは人手により設計された特徴量であり、これらの組み合わせにより、遠方で低密度な点群として計測された歩行者の認識精度の向上を実現している。

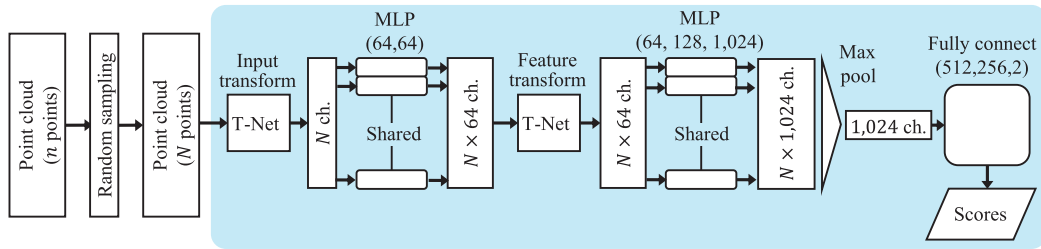
一方、画像認識分野を中心に深層学習手法が発展しており、点群に対する物体認識手法も提案されている^{5)–7) 12)}。Maturana ら⁵⁾は、点群の 3 次元位置の分布からボクセル表現を生成し、このボクセル表現を 3 次元畳み込みニューラルネットワーク (3 Dimensional Convolutional Neural Network; 3DCNN) に入力する深層学習手法を提案している。このボクセル表現の生成により点群の 3 次元位置の分布を表現可能であるが、個々の点の位置情報等の情報が失われる。また、点群の規模が大きくなるとボクセル表現のデータ量が莫大になり、認識処理に必要な時間が多くなる。Beltrán ら¹²⁾は、大規模な点群から鳥瞰図画像を生成し、生成された鳥瞰図画像を Faster-RCNN¹³⁾ を応用した CNN 構造に入力することにより、鳥瞰図画像上での歩行者や自転車、車両を検出する手法を提案している。鳥瞰図から元の点群へ検出位置を投影することにより、3 次元空間中でも歩行者等を検出できることを示している。この鳥瞰図表現は、観測対象物体から一定以上の密度の点群が得られていれば点群の密度等に影響を受けることなく検出が可能であるが、対象との距離に応じて精度が低下することが報告されている。遠方にある物体ほど計測される点群の密度は低下するため、対象との距離に応じて精度が低下することは多くの物体認識手法が共有する課題である。そのため、高精度な物体認識のためには、距離による精度低下を抑えることが重要である。

この問題に対して、点群を他形式に変換することなく直接認識する DNN 手法として提案された PointNet⁶⁾ が注目されている。PointNet は点群の各点に対して個別に畳み込みを行い、Max pool 層により点の順序に不変な特徴抽出を可能としている。この抽出された特徴は、3 次元物体の形状間の差異を表現するのに有効な Hausdorff 距離の近似になっていることが示されている。これにより点群の分布形状を学習することが可能となり、高い認識性能を得ている。また、この PointNet を応用した手法も提案されており、Zhou ら⁷⁾は、大規模な点群を領域分割した後、各領域から PointNet による特徴抽出と物体候補領域の抽出を行う Region Proposal Network¹³⁾ を組み合わせることにより、自車両の周辺環境認識を行っている。しかし、PointNet は入力点数が固定であることから、認識対象物体から計測される点数が大きく変化する場合は、点群から一定の点数を適切にサンプリングしなければ性能が大きく低下するという問題がある。

3. 提案手法の概要

PointNet⁶⁾ の処理手順を図 3 に示す。前述したように、PointNet は点群を直接入力して認識できるが、入力点数 N が固定という制約がある。

入力点数が N 点以上の点群に対しては、ダウンサンプリングにより一定の入力点数 N に調整を行う。しかし、搭乗者の身体のような太い部分と自転車の車体のような細い部分とで点群の密度に差がある。単純なランダムサンプリングを用いた場合、図 4 に示すように、点の密度が高い搭乗者の身体部分の



PointNet architecture

Fig. 3 Outline of the original PointNet framework

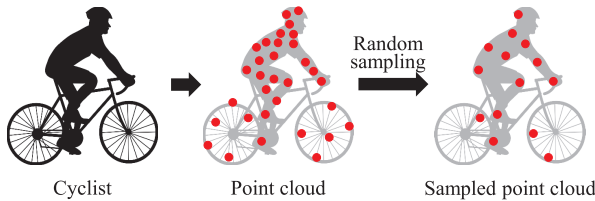


Fig. 4 Example of random sampling

計測点が多く選択され、密度が低い自転車の車体のような細い部分の計測点を選択されず、元の点群の分布形状から大きく変化してしまう。そこで、低密度部分の点が保存されるよう計測点を選択することで車体のような細い部分の計測点を確実に残し、元の点群の分布形状を極力保持しつつ、入力点数を N 点に調整することを考えた。

また、入力点数が N 点未満の点群に対しては PointNet を適用できない。PointNet を適用するためにはアップサンプリングが必要であるが、重複を許してアップサンプリングした場合、同じ点の情報を使うため認識性能の向上は見込めない。一方で、無作為に点を増やせば、元の点群の分布形状を大きく変化させてしまう可能性がある。そこで、LIDAR の測定精度を考慮して計測点の存在する位置に関する確率分布を生成し、それをもとにサンプリングを行うことで、入力点数を N 点に調整する。

本論文では上記を実現するため、下記の 2 種類の機能を持つ新しいサンプリングモジュールを提案する。

- 機能 1
計測点数 N 点以上の入力点群に対し、点群の分布形状を保持したダウンサンプリング。
- 機能 2
計測点数 N 点未満の入力点群に対し、LIDAR の測定精度を考慮して点群の密度を増加させるアップサンプリング。

提案サンプリングモジュールを導入することにより、任意の入力点数に対して PointNet の適用を可能にし、自転車に対する認識精度の向上を狙う。図 5 に提案手法の処理手順を示す。なお本論文では、LIDAR により計測された全ての点群から自転車候補点群の抽出が行われた後の処理を想定する。

3.1 点群の分布形状を保持するダウンサンプリング手法

自転車から計測された点群の分布形状を保持するため、低密度部分の計測点が残されるようサンプリングを行う。まず、自転車候補点群を水平方向に X 分割、奥行方向に Y 分割、鉛直方向に Z 分割する。自転車候補点群に含まれる計測点の総

数を n 、各分割領域 $V_{x,y,z}$ に含まれる計測点の数を $|V_{x,y,z}|$ とする。全計測点から無作為に点を選択するとき、各分割領域 $V_{x,y,z}$ の計測点を選択される確率 $P(x,y,z)$ は式 (1) により算出される。

$$P(x,y,z) = \frac{|V_{x,y,z}|}{n} \quad (1)$$

点群が高密度の分割領域では $|V_{x,y,z}|$ が大きいため、分割領域内の計測点を選択される確率 $P(x,y,z)$ が高い。一方で、低密度の分割領域では $|V_{x,y,z}|$ が小さいため、 $P(x,y,z)$ が低くなり分割領域内の計測点を選択されにくい。低密度部分の計測点を保存するためには、低密度の分割領域内の計測点を選択される確率を高くすることが重要である。そこで、低密度の分割領域内から計測点を選択されやすくなるよう、各計測点に対するサンプリング時の重み ω を式 (2) により算出する。

$$\omega(x,y,z) = \frac{1}{|V_{x,y,z}|} \quad (2)$$

算出した重み $\omega(x,y,z)$ に基づいて逆関数サンプリングを行うことで含まれる計測点の数が少ない分割領域 $V_{x,y,z}$ からもサンプリングを行うことを可能とする。図 6 に提案ダウンサンプリング手法の概要を示す。

3.2 LIDAR の測定精度に基づくアップサンプリング手法

LIDAR は LASER 光を照射し、その反射光を解析して距離を測るため、観測対象までの距離に応じて計測された値に誤差が生じる。そのため、同じ位置を計測した場合でも、計測点の位置が誤差の分だけ異なって計測される可能性があり、計測点は LIDAR の測定精度に応じた空間的な位置の曖昧性を持つ。そこで本論文では、各計測点を距離に応じた確率分布で表現し、点群の存在尤度マップ \mathcal{M} を生成する。そして、生成した存在尤度マップ \mathcal{M} から N 点をサンプリングする。図 7 に提案アップサンプリング手法の概要を示す。

存在尤度マップ \mathcal{M} は、各計測点 \mathbf{p}_i を用いて式 (3) により求められる。

$$\mathcal{M} = \sum_i g(\mathbf{p}_i, \Sigma_i) \quad (3)$$

$g(\cdot)$ はカーネル関数であり、ここでは 3 次元正規分布を用いる。これにより、計測点を中心とし、計測点毎に異なる分散共分散行列 Σ_i をもつ 3 次元混合正規分布 $g(\mathbf{p}_i, \Sigma_i)$ として存在尤度マップを得る。分散共分散行列 Σ_i は各計測点までの距離 d_i に依存し、式 (4) のように表される。

$$\Sigma_i = \begin{pmatrix} \sigma_i^2 & 0 & 0 \\ 0 & \sigma_i^2 & 0 \\ 0 & 0 & \sigma_i^2 \end{pmatrix} = \sigma_i^2 \mathbf{I} \quad (4)$$

ここで、 σ_i^2 は d_i と LIDAR の測定精度に基づいて計算される。本論文では、LIDAR の測定精度が全方向に対して一様とみな

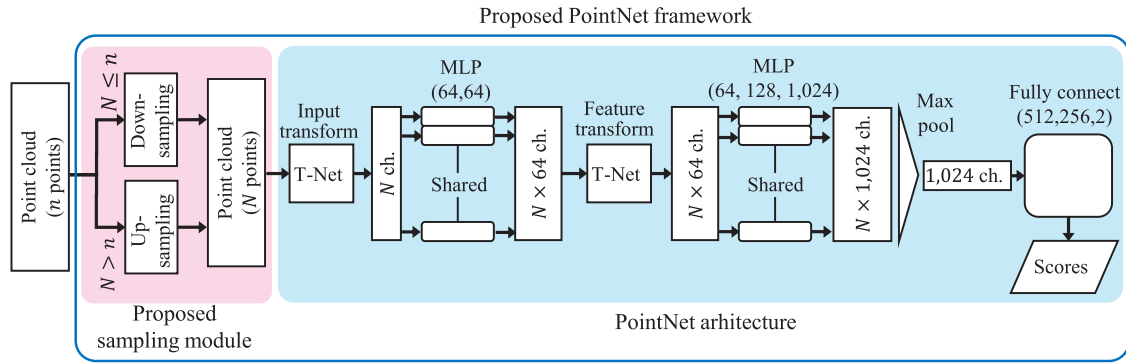


Fig. 5 Outline of the proposed PointNet framework

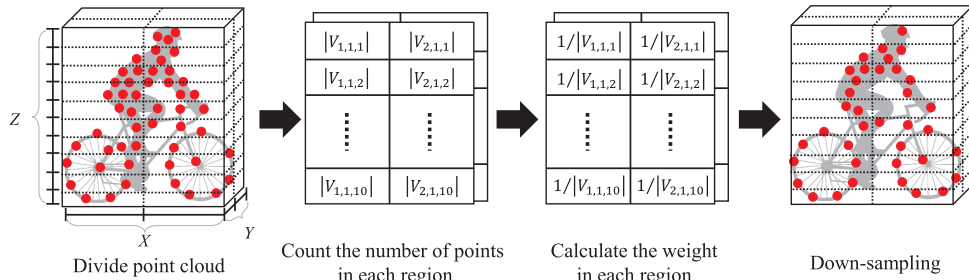


Fig. 6 Outline of the proposed down-sampling module ((X, Y, Z) = (2, 2, 10))

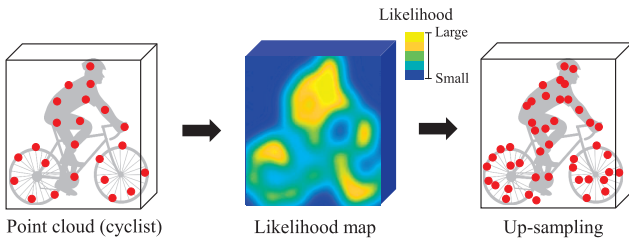


Fig. 7 Outline of the proposed up-sampling module

し、全ての方向で共通の σ_i を用いる。建部らの手法¹⁴⁾を参考に、 σ_i は計測点までの距離の2乗に比例するとすると、 σ_i は式(5)のように表される。

$$\sigma_i = \alpha d_i^2 \quad (5)$$

ここで、LIDARからの距離が D の位置の物体を計測した際に、計測された値に含まれる空間的な誤差が Ω とすると、距離 D の計測点に対して $\sigma_i = \lambda \Omega$ となるよう、 α は式(6)のように算出される。

$$\alpha = \frac{\lambda \Omega}{D^2} \quad (6)$$

ここで、 λ は摂動の大きさを制御するために用いる変数である。したがって、式(4)中の標準偏差 σ_i は、原点から計測点 \mathbf{p}_i までの距離 d_i を用いて、式(7)により求める。

$$\sigma_i = \frac{\lambda \Omega}{D^2} d_i^2 \quad (7)$$

生成された存在尤度マップ \mathcal{M} から N 点を抽出することにより、アップサンプリングを行う。

4. 提案サンプリングモジュールの実装

提案サンプリングモジュールを以下のようにして実装する。

4.1 ダウンサンプリング手法

3.1節の考え方に基づき、サンプリング時の乱数の偏りを極力抑えつつ、低密度部分の計測点が保存されるようにダウンサンプリングする。具体的には、以下の手順により、同じ計測点が複数回選択されることを防ぎ、低密度部分を含む各領域から計測点が確実にサンプリングされるよう処理を行う。

まず、自転車候補点群を囲む最小の3次元矩形を水平方向に X 分割、奥行方向に Y 分割、鉛直方向に Z 分割する。式(2)の各計測点に対するサンプリング時の重み ω により、各分割領域 $V_{x,y,z}$ から計測点が1点選択される確率は等しいため、各分割領域内から無作為に1点ずつ計測点を抽出し、この抽出処理を反復する。反復回数 β は式(8)により算出される。

$$\beta = \left\lfloor \frac{N}{XYZ} \right\rfloor \quad (8)$$

XYZ は分割領域の総数を表す。ただし、同じ計測点が複数回抽出されることを防ぐため、各反復で1度選択された計測点は抽出しないものとする。分割領域内の全ての計測点を抽出した場合は、以降の反復ではその領域内からは抽出しない。これにより計測点の抽出処理が行われなかった回数を ϕ とする。最後に、不足分の $N - (\beta + \phi)$ 点について、選択されていない全ての計測点から重複なしに無作為に抽出する。以上の処理により、式(2)の重み ω に基づく逆関数サンプリングを模擬する。

4.2 アップサンプリング手法

3.2節の考え方に基づき、LIDARの測定精度を考慮してアップサンプリングする。

自転車候補点群が $n (< N)$ 点で構成されているとする。まず、各計測点 \mathbf{p}_i に対して、原点からの距離 d_i に応じた位置の摂動を加えた点 $\hat{\mathbf{p}}_i$ を生成する。この摂動の大きさは、平均0、標準偏差 σ_i の正規分布に従う乱数により求め、標準偏差 σ_i

は 3.2 節の説明に基づき LIDAR の測定精度に基づいて定める。これにより、元の計測点と位置情報が異なる点が n 点追加される。この生成処理を γ 回反復する。反復回数 γ は式 (9) により算出される。

$$\gamma = \left\lceil \frac{N}{n} \right\rceil \quad (9)$$

最後に、不足分の $N - \gamma n$ 点について、4.1 節と同様の方法により元の n 点で構成されている自転車候補点群から $N - \gamma n$ 点の計測点を選択し、摂動を加えて新たに追加する。以上の処理により、式 (3) の存在尤度マップ \mathcal{M} からのサンプリングを模擬する。

4.3 PointNet の学習と認識

学習では、学習用に用意した自転車候補点群に対して提案サンプリングモジュールを適用して N 点の点群を生成し、PointNet で学習することにより、自転車認識器を構築する。従来の PointNet では、最初に 1 度だけ N 点へのサンプリングを行って入力点群を生成し、学習に用いる。そのため、最初のサンプリングで選択されなかった計測点は学習に使用されない。このままでは、サンプリング前の点群の多様性が失われる可能性がある。そこで本手法では、PointNet の学習を繰り返すたびに提案サンプリングモジュールを適用して N 点の点群を生成する。これにより学習のたびに使用される計測点を変化させ、PointNet の汎化性を向上をねらう。

認識時は、認識対象の自転車候補点群に対して提案サンプリングモジュールを適用して生成した N 点の点群を入力とし、構築した自転車認識器を用いて、自転車認識を行う。

5. 評価実験

提案手法の有効性を確認するために行った実験について述べる。

本実験では、KITTI Vision Benchmark Suite の評価基準^{*1}を考慮し、車両前方 30 m までの範囲内に存在する自転車の認識を想定した。自動車が時速 40 km で走行する状況を考えると、自転車が車両前方 30 m の位置に接近した時点から認識できれば、自動ブレーキ等により衝突前に停止できると考えられる。また、30 m より近い位置に自転車が突然現れて衝突を回避できない場合でも、衝突前に極力速度を低下させて被害を軽減するため、30 m 以内の位置の自転車の認識が重要である。自動車が 30 m で停止できる理由は、車速が 40 km/h のときの一般的な停止距離が約 20 m であること¹⁵⁾、自転車認識処理に 1 秒掛かる場合の空走距離が約 10 m であることから、30 m あれば自動車が停止できるためである。なお、本実験では自転車認識処理に掛かった計算時間は 1 秒以内であった。なお、30 m 以上遠い位置に存在する自転車の認識については、6.5 節の比較実験にて、30 m 以上遠い位置に存在する自転車を含むデータを用いて精度評価した結果を示す。

5.1 データセット

本実験では、KITTI データセット¹⁶⁾ から抽出した自転車点群、非自転車点群を用いて実験用データセットを構築した。具体的には、KITTI データセットに付与されている自転車のアノテーションを用いて、以下の条件を満たす自転車点群を抽出

^{*1} KITTI データセット内の画像中で縦 40 画素以上の大きさの物体の認識を評価基準の 1 つとしており、この基準は KITTI データセット内の点群中の前方約 30 m 以内の位置に存在する物体に相当する。

Table 1 Number of cyclist point clouds and non-cyclist point clouds in the dataset

	Number of point clouds
Cyclist	805
Non-cyclist	2,735

Table 2 Number of point clouds per number of points

Number of points	1-64	65-128	129-256	257-
Cyclist	165	248	167	225
Non-cyclist	625	859	684	567

Table 3 Number of point clouds per distance

Distance	0-10 m	10-20 m	20-30 m
Cyclist	135	365	305
Non-cyclist	340	1,110	1,285

した。

- 車両前方 30 m 以内の位置に存在する自転車点群
- 遮蔽なし、または遮蔽が一部の自転車点群

また、Point Cloud Library¹⁷⁾ の EuclideanClusterExtraction 関数によるクラスタリングを用いて、以下の条件を満たすクラスタを非自転車点群として抽出した。

- 車両前方 30 m 以内の位置に存在
- クラスタに含まれる計測点が 30 点以上
- 以下の式 (10), (11), (12) の全てを満たす。ただし、クラスタの水平方向の左右幅を w 、奥行方向の奥行幅を l 、鉛直方向の高さを h とする。

$$w \leq 2.5 \text{ m かつ } l \leq 2.5 \text{ m} \quad (10)$$

$$w \geq 1.0 \text{ m または } l \geq 1.0 \text{ m} \quad (11)$$

$$0.7 \text{ m} \leq h \leq 2.0 \text{ m} \quad (12)$$

表 1 にそれぞれの点群のデータ数を示す。また、表 2 に点群内に含まれる点数毎の点群のデータ数を、表 3 に距離毎の点群のデータ数をそれぞれ示す。自転車点群および非自転車点群に対し、5 分割交差検証によって認識精度を評価した。

5.2 評価方法

本実験では、提案サンプリングモジュールを導入した PointNet の有効性を確認するため、以下に示す提案手法、比較手法 1、比較手法 2、比較手法 3、比較手法 4 の 5 種類の手法を比較した。

- 提案手法：Size-Adaptable PointNet (SA PointNet)
提案サンプリングモジュールを導入した PointNet により認識を行う手法。
- 比較手法 1：SVM
人手で設計したハンドメイド特徴量^{3) 4)} をカーネル SVM により認識を行う手法。
- 比較手法 2：VoxNet
3DCNN である VoxNet⁵⁾ により認識を行う手法。
- 比較手法 3：PointNet
従来の PointNet⁶⁾ により認識を行う手法。入力点数を N にするためランダムダウンサンプリングと重複を許すアッ

プサンプリングを用いる。

● 比較手法 4 : PointNet + Sampling

4.1 節の提案ダウンサンプリング手法と 4.2 節の提案アップサンプリング手法を導入した PointNet により認識を行う手法。

評価指標として、ROC (Receiver Operating Characteristic) 曲線と p-AUC (partial Area Under the Curve¹⁸⁾) を用いた。p-AUC とは、ROC 曲線のある区間のみから求めた AUC (Area Under the Curve) であり、最大値が 1 になるように正規化されている。本論文では、p-AUC を ROC 曲線の誤認識率が 5% 以下の区間で求めることで、自転車認識にとって重要となる誤認識率が低い区間に着目して、認識精度を評価する。

5.3 実験環境

入力点数を $N = 128$ 点、4.1 節の提案ダウンサンプリング手法における 3 次元矩形の分割数を $(X, Y, Z) = (2, 2, 10)$ とした。計測点が低密度な部分から点を多くサンプリングするためには分割数を大きくすることが望ましいが、自転車の車体部分は横幅が小さいため、幅方向の分割数を大きくすると自転車の計測点が存在しない領域が多く発生する。自転車の方向が未知であるため、本実験では自転車がどの方向を向った場合でも対応できるように、水平方向と奥行方向の分割数を 2 とした。鉛直方向の分割数は、Kidono らの提案したスライス特徴量³⁾に倣い、10 とした。KITTI データセットで用いられている Velodyne LiDAR 社の HDL-64E の測定精度¹⁹⁾によると、25 m の距離にある物体を測定したときに計測値に生じる誤差が ± 2 cm であるため、 $\Omega = 0.02$ m、 $D = 25$ m とした。また、アップサンプリング時の摂動の値を大きくするため、 $\lambda = 3$ とした。

PointNet の実装には、Qi が提供しているプログラム²⁰⁾を用いた。入力には、各計測点の 3 次元位置情報 (x, y, z) (座標) に加え、反射強度を合わせた 4 次元の情報を用いた。学習におけるエポック数は 600、バッチサイズは 32 に設定した。エポック数は、学習時の損失関数の推移を見つづ、過学習が起きないように実験的に求めた。バッチサイズは文献⁶⁾と同じ値を用いた。

VoxNet の実装には、Maturana が提供しているプログラム²¹⁾を参考に実装した。ボクセルの大きさは $20 \times 20 \times 20$ 次元とした。学習におけるエポック数は 100、バッチサイズは 32 に設定した。PointNet と同様に、エポック数は実験的に求め、バッチサイズは文献⁵⁾と同じ値を用いた。

5.4 実験結果

実験により得られた各手法の ROC 曲線を図 8 に示す。

提案手法は他手法と比べて、ROC 曲線が左上に位置していることが確認でき、自転車に対する認識精度の向上が確認できる。また、表 4 に実験により得られた各手法の p-AUC を示す。PointNet を用いる手法 (提案手法、比較手法 3 および比較手法 4) が高い p-AUC を得られており、その中でも提案サンプリングモジュールを利用する提案手法がより高い p-AUC を得られていることが確認できる。そのため、提案サンプリングモジュールの導入により、自転車に重要な誤認識率が低い場合における認識精度の向上を確認した。

Table 4 p-AUC of the results

Method	p-AUC
Proposed: SA PointNet	0.974
SVM	0.906
VoxNet	0.935
PointNet	0.963
PointNet + Sampling	0.966

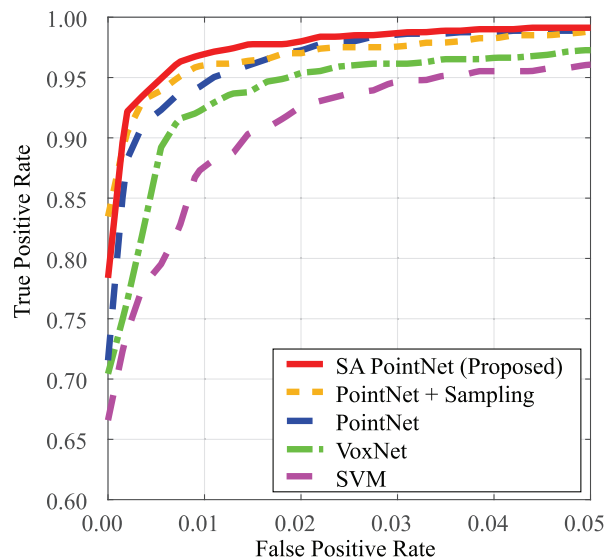


Fig. 8 ROC curves of the results ($N = 128$)

6. 考 察

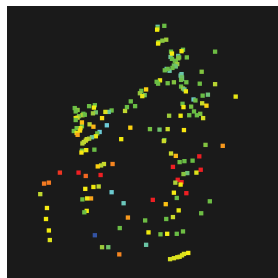
実験結果をもとに、提案サンプリングモジュールの効果や入力点数 N を変化させたときの認識性能、自転車までの距離までの認識性能について考察する。

6.1 提案サンプリングモジュールの効果

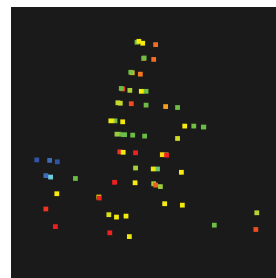
提案手法と比較手法 3 (PointNet) に着目すると、提案手法の認識精度が上回っていることが確認できる。提案手法により、自転車の車体部分等の計測点の密度が低い部分から確実な計測点の抽出を行い、元の点群の分布形状からの大きな変化を防ぐことで、認識精度が向上したと考えられる。また、原点からの距離に応じて各計測点に摂動を加え、点群の密度を擬似的に大きくすることが、認識精度の向上に貢献したと考えられる。

実際に提案サンプリングモジュールを適用し、自転車点群をダウンサンプリングした結果を図 9 に示す。図 9(a) は 128 点以上の計測点で構成される自転車点群のため、ダウンサンプリングが適用される。ここで、提案サンプリングモジュールによりダウンサンプリングされた結果が図 9(b) であり、ランダムサンプリングされた結果が図 9(c) である。図中の白い破線の楕円で示した部分に注目すると、ランダムサンプリングでは自転車のフレームやタイヤといった車体の細い部分の計測点が失われたのに対し、提案サンプリングモジュールでは計測点をサンプリングに成功していることが分かる。

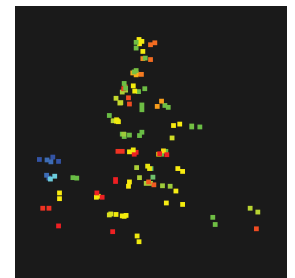
また、非自転車点群である歩行者点群をダウンサンプリングした結果を図 10 に示す。この歩行者点群は、提案手法により非自転車として認識された点群の一例である。サンプリング前の点群が図 10(a) であり、提案サンプリングモジュールによりダウンサンプリングされた結果が図 10(b) である。ダウンサン



(a) Original point cloud of a cyclist

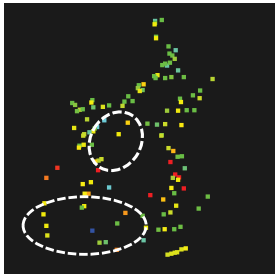


(a) Original point cloud of a cyclist

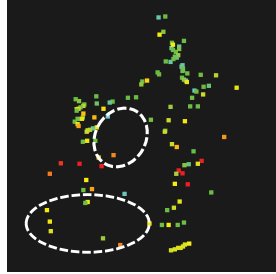


(b) After up-sampling by the proposed module

Fig. 11 Example of up-sampling results

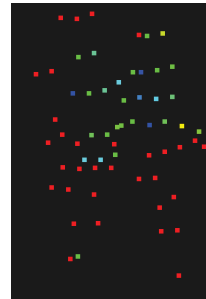


(b) After down-sampling by the proposed module

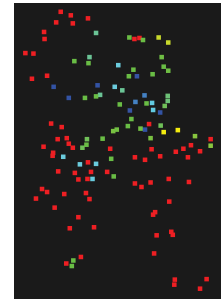


(c) After down-sampling by random sampling

Fig. 9 Example of down-sampling results

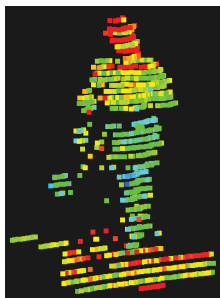


(a) Original point cloud of two pedestrians

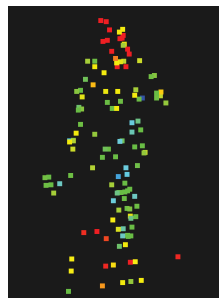


(b) After up-sampling by the proposed module

Fig. 12 Example of up-sampling results for non-cyclist



(a) Original point cloud of a pedestrian



(b) After down-sampling by the proposed module

Fig. 10 Example of down-sampling results for non-cyclist

プリング後も歩行者点群の元の点群の分布形状が保持されており、自転車点群の分布形状とは異なっていることが分かる。そのため、非自転車として正しく認識できたと考えられる。

次に、提案サンプリングモジュールを適用し、自転車点群をアップサンプリングした結果を図 11 に示す。図 11(a) は 128 点未満の計測点で構成される自転車点群のため、本実験の設定ではアップサンプリングが必要な点群である。提案サンプリングモジュールによりアップサンプリングされた結果が図 11(b) である。各計測点に LIDAR の測定精度に基づく摂動を加えることで、自転車の点群の分布形状を大きく変化させることなく、点群の密度の増加に成功していることが分かる。

また、非自転車点群である歩行者点群をアップサンプリングした結果を図 12 に示す。この歩行者点群は、並んだ 2 人の歩行者から計測された点群であり、提案手法により非自転車として認識された点群の一例である。サンプリング前の点群が図 12(a) であり、提案サンプリングモジュールによりアップサンプリングされた結果が図 12(b) である。アップサンプリング前後の歩行者点群で、両者とも点群の分布形状が自転車点群の分布形状とは異なっていることが分かる。

これらの結果から分かる通り、提案サンプリングモジュールは自転車点群に対して分布形状を保持するようなサンプリング

を実現しており、識別精度の向上に有効であることが確認できた。

また、提案手法と比較手法 3、比較手法 4 の 3 種類に着目すると、従来の PointNet に対して 4.1 節の提案ダウンサンプリング手法と 4.2 節の提案アップサンプリング手法を導入することで精度が向上し、さらに 4.3 節の学習手法を導入することで精度が向上することが確認できた。

6.2 入力点数 N を変化させたときの認識性能

PointNet の入力点数 N を 64 および 256 に変化させたときの ROC 曲線を図 13 と図 14 にそれぞれ示す。なお、比較手法 1 (SVM) および比較手法 2 (VoxNet) は、入力点群の数に関わらず全ての点群を用いて特徴量へ変換するため、サンプリングは行わず全ての入力点群を用いている。そのため認識精度に変化はなく、図 8 と同じである。また、表 5 に提案手法と従来の PointNet の入力点数 N を変化させたときの p-AUC を示す。

まず、 $N = 64$ の認識精度に注目する。 N がより小さくなったことで、ダウンサンプリングが適用される入力点群の割合が増加し、ダウンサンプリング手法の貢献が大きくなる実験設定である。提案手法が従来の PointNet よりも高い精度を示していることから、提案ダウンサンプリング手法が有効に機能していることが確認できる。 $N = 128$ の場合と比較すると、従来の PointNet の認識精度が低下しているのに対し、提案手法の認識精度はわずかに低下しただけに留まっていることが分かる。これは、認識に用いられる計測点が減り、従来の PointNet で用いられるランダムダウンサンプリングでは自転車点群の分布形状を保持できず認識精度が低下したのに対し、提案ダウンサンプリング手法では自転車点群の分布形状を保持できたため認識精度の低下を最小限に抑制できたと考えられる。

次に、 $N = 256$ の認識精度に注目する。 N がより大きくなったことで、アップサンプリングが適用される入力点群の割

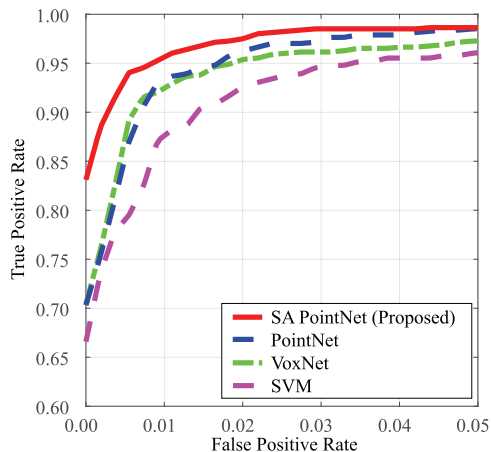


Fig. 13 ROC curves of the results ($N = 64$)

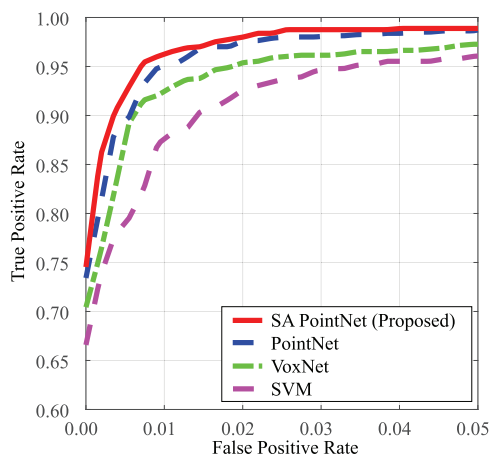


Fig. 14 ROC curves of the results ($N = 256$)

合が増加し、アップサンプリング手法の貢献が大きくなる実験設定である。先と同様に、提案手法が従来の PointNet よりも高い精度を示していることから、提案アップサンプリング手法が有効に機能していることが確認できる。 $N = 128$ の場合と比較すると、提案手法と従来の PointNet の両手法ともに認識精度がわずかに低下した。基本的に、自転車点群を構成する計測点が多いほど自転車の詳細な形状が分かり、少ないほど自転車の大きな形状しか分からない。 PointNet は入力点数 N の点群の分布形状をもとに学習するので、 N を大きくすれば自転車の詳細な形状まで学習することになる。しかし、提案アップサンプリング手法により点群の密度を疑似的に高くしているが、入力点数 N が大きくなるほど、元々の計測点の数が少ない自転車点群から詳細な形状を復元することは難しい。ゆえに、認識精度がわずかに低下したと考えられる。

提案手法と従来の PointNet の両手法において、 $N = 128$ のときに認識精度が最も高い結果が得られた。前述した通り、 PointNet の入力点数 N を小さくすると、認識に用いられる計測点が減るため認識対象の形状を捉えることが難しくなる。一方で、入力点数 N を大きくすると、認識対象の形状を捉えやすくなるが、アップサンプリングでの形状の復元に限界があるため、計測点の少ない自転車点群に対する認識精度が低下する。これらはトレードオフの関係にあるため、認識対象に応じて最適な入力点数を選択することは、今後の取組が必要な課題

Table 5 p-AUC of the results in case of various N

Method	p-AUC
SA PointNet ($N = 64$)	0.968
SA PointNet ($N = 128$)	0.974
SA PointNet ($N = 256$)	0.968
PointNet ($N = 64$)	0.943
PointNet ($N = 128$)	0.963
PointNet ($N = 256$)	0.958

である。

6.3 自転車の位置までの距離毎の認識性能

車両前方 0–10 m の位置に存在する自転車点群に対し、提案手法と比較手法 2 (VoxNet)、比較手法 3 (PointNet) を適用して自転車認識を行ったときの ROC 曲線を図 15 に示す。同様に、10–20 m、20–30 m の位置に存在する自転車点群の ROC 曲線を図 16、図 17 にそれぞれ示す。なお、PointNet の入力点数 $N = 128$ とした。また、表 6 に各実験により得られた各手法の p-AUC を示す。

まず、距離が 0–10 m のときの認識精度に注目する。近距離では点群の密度が高くなるので、多くの点群の入力点数が 128 点以上の計測点で構成される。よって、多くの点群に対しダウンサンプリングが適用される。提案手法が従来の PointNet よりも高い精度を示していることから、提案ダウンサンプリング手法が有効に機能していることが確認できる。また、高密度に点群が計測されるため、VoxNet が高い精度を示していることが分かる。

次に、距離が 10–20 m のときの認識精度に注目する。この距離でも多くの点群の入力点数は 128 点以上であるが、距離が 0–10 m の点群と比べると計測点が少ないため、ダウンサンプリングの前後で点群の変化は小さい。そのため、提案手法と従来の PointNet が比較的近い精度を示した。

最後に、距離が 20–30 m のときの認識精度に注目する。遠距離では点群の密度が低くなるので、多くの点群の入力点数が 128 点以下の計測点で構成される。よって、多くの点群に対しアップサンプリングが適用される。誤認識が小さい区間にて提案手法が従来の PointNet よりも高い精度を示しており、単純に重複を許してアップサンプリングする PointNet と比較して、提案アップサンプリング手法が有効に機能していることが確認できる。また、距離が離れるにしたがって VoxNet の認識性能が大きく低下しているのに対し、提案手法では認識性能の低下を抑制できていることが分かる。

6.4 アップサンプリング時の摂動の大きさ

本論文では、LIDAR の測定精度が全方向に対して一様とみなし、アップサンプリング時に計測点に対し摂動を加えた。しかし、実際には水平方向や鉛直方向と奥行方向の測定精度は異なると考えられる。建部ら¹⁴⁾は、水平方向について、LIDAR のビーム幅と水平方向の測定誤差の 2 つのセンサ特性を用いて計測点の存在する位置に関する確率分布を生成している。LIDAR から照射される LASER 光は幅を持ち、その幅は計測される物体までの距離が遠くなるほど大きくなる。その幅の範囲内に物体が存在すれば点として計測されるため、計測点は LASER 光の幅分の位置の曖昧性を持つことを考慮して確率分布を生成している。一方、奥行方向については、奥行方向の測定誤差を用いて確率分布を生成している。このような建部らの

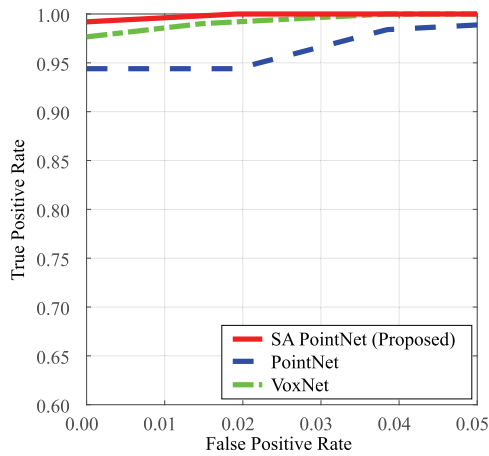


Fig. 15 ROC curves of the results (0–10 m)

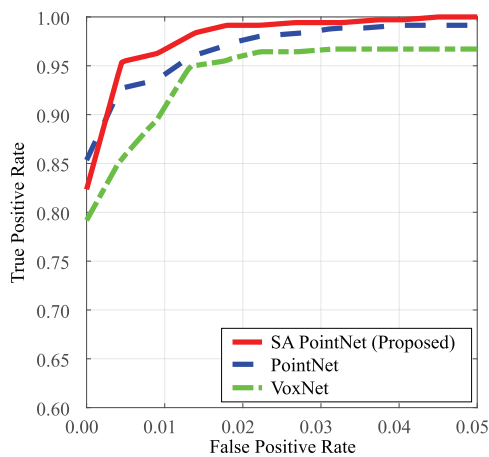


Fig. 16 ROC curves of the results (10–20 m)

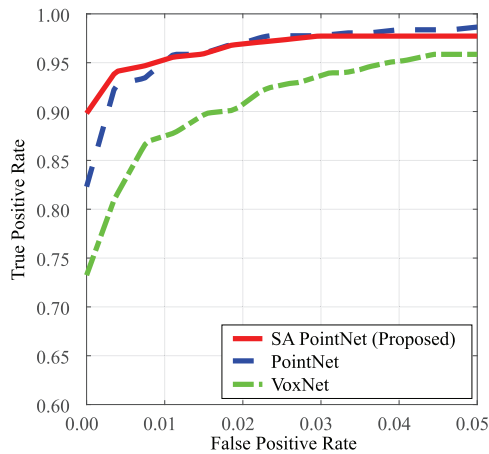


Fig. 17 ROC curves of the results (20–30 m)

モデル化の利用については、今後の取組が必要な課題である。

6.5 Frustum PointNets との組み合わせによる自転車検出精度の評価実験

RGB-D データからの物体検出手法である Frustum PointNets²⁾ を自転車候補抽出として利用し、点群からの物体認識部分に提案手法を導入することで、提案手法の有効性を検証した。

Table 6 p-AUC of the results in case of various distances

Distance	Method	p-AUC
0–10 m	SA PointNet	0.998
	PointNet	0.954
	VoxNet	0.991
10–20 m	SA PointNet	0.979
	PointNet	0.967
	VoxNet	0.939
20–30 m	SA PointNet	0.965
	PointNet	0.962
	VoxNet	0.906

Frustum PointNets は、まず画像から物体検出を行い 2 次元空間上での物体候補の矩形の位置を求める。2 次元空間上での矩形は、3 次元空間上では 2 次元空間上での矩形を上底とした四角錐台 (Frustum) として定義される。そこで、2 次元空間上での矩形位置をもとに、点群を四角錐台状に切り出す。切り出された点群から物体に含まれる計測点を抽出し、その物体の 3 次元位置を求める手法である。本実験では、Frustum PointNets が四角錐台状に切り出された点群から自転車候補の計測点を抽出した際に、その点群が自転車か否かを提案手法により認識することで、自転車の認識精度が向上するか否かを評価した。

Frustum PointNets の実装には、Qi が提供しているプログラム²²⁾ を用いた。Frustum PointNets の各パラメータは、提供されているプログラムの標準設定値を用いた。提案手法の各パラメータは先の実験と同様に、入力点数は $N = 128$ 点、3 次元矩形の分割数を $(X, Y, Z) = (2, 2, 10)$ 、 $\Omega = 0.02$ m、 $D = 25$ m、 $\lambda = 3$ とした。データセットは、KITTI データセット¹⁶⁾ に含まれる学習用に用意されたデータを 2 分割し、3,712 枚を学習に使用し、残りの 3,769 枚を評価に用いた。評価方法は、KITTI Vision Benchmark Suite における 3D Object Detection Evaluation の評価基準¹⁶⁾ を用いた。

実験により得られた提案手法を導入した Frustum PointNets と従来の Frustum PointNets の Precision-Recall 曲線を図 18 に示す。曲線は、Easy 難易度²⁾ および Moderate 難易度³⁾ のときの各手法の精度を示している。曲線が右上に位置するほど高精度であり、提案手法を導入することで精度が向上することが確認できる。Precision が向上していることから、Frustum PointNets が誤って自転車と認識していた非自転車点群が、提案手法によって正しく非自転車点群として認識されたと考えられる。したがって、提案手法を組み合わせることで、従来の Frustum PointNets よりも高精度な自転車検出が可能であることを確認した。

7. 結 言

本論文では、LIDAR により計測された点群から自転車を認識する手法を提案した。具体的には、PointNet の入力点数が固定であるという制約に対し、点群の分布形状を保持したダウンサンプリングと LIDAR の測定精度を考慮したアップサンプリングを行うサンプリングモジュールを提案し、これらの導入により PointNet の入力を可変長化した。具体的には、提案ダウ

^{*2} KITTI データセット内の画像中で縦 40 画素以上の大きさの物体の認識に対応する。
^{*3} KITTI データセット内の画像中で縦 25 画素以上の大きさの物体の認識に対応する。

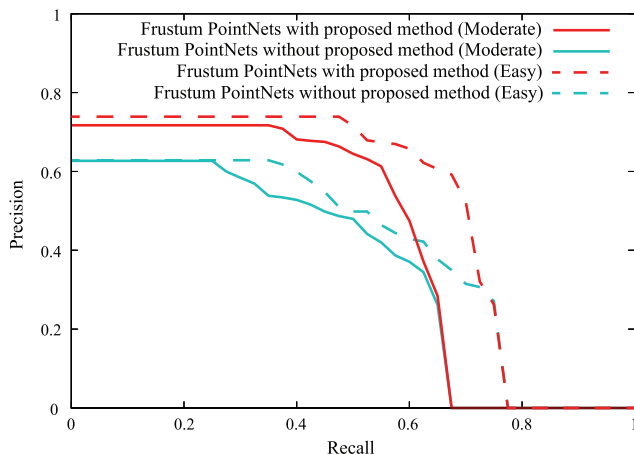


Fig. 18 Precision-Recall curves of the results

ンサンプリング手法によって、点群を囲む 3 次元矩形を分割後、各分割領域内の計測点の数から密度を求め、点群の低密度部分が保存されるよう計測点の選択確率に重み付けしてサンプリングを行うことで、自転車の車体部分等の点群の密度が低い部分からの確実なサンプリングを可能とした。また、提案アップサンプリング手法によって、計測点毎に LIDAR の測定精度に基づく距離に応じた摂動を加えることで点群の密度を向上した。KITTI データセットから抽出した自転車点群、非自転車点群を用いた評価実験により、提案手法は従来の PointNet や他の 3DCNN を用いた手法、SVM を用いた手法と比べて自転車の認識精度が向上することを確認した。

今後の課題として、他物体に対する認識精度の評価や、より認識が難しいデータセットに対する認識精度の評価、他の PointNet を応用した手法に対する提案手法の実装および性能評価が挙げられる。

謝 辞

本研究の一部は、科学研究費補助金 (17H00745) による。

参 考 文 献

- 1) A. Kanazaki and T. Harada: 3D selective search for obtaining object candidates, Proc. 2015 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (2015) 82.
- 2) C. R. Qi, W. Liu, C. Wu, H. Su and L. J. Guibas: Frustum PointNets for 3D object detection from RGB-D data, Proc. 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (2018) 918.
- 3) K. Kidono, T. Miyasaka, A. Watanabe, T. Naito and J. Miura: Pedestrian recognition using high-definition LIDAR, Proc. 2011 IEEE Intelligent Vehicles Symposium (2011) 405.

- 4) Y. Tatebe, D. Deguchi, Y. Kawanishi, I. Ide, H. Murase and U. Sakai: Can we detect pedestrians using low-resolution LIDAR? —Integration of multi-frame point-clouds—, Proc. 12th Int. Conf. on Computer Vision Theory and Applications, **5** (2017) 157.
- 5) D. Maturana and S. Scherer: VoxNet: A 3D convolutional neural network for real-time object recognition, Proc. 2015 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (2015) 922.
- 6) C. R. Qi, H. Su, K. Mo and L. J. Guibas: PointNet: Deep learning on point sets for 3D classification and segmentation, Proc. 2017 IEEE Conf. on Computer Vision and Pattern Recognition (2017) 652.
- 7) Y. Zhou and T. Oncel: VoxelNet: End-to-end learning for point cloud based 3D object detection, Proc. 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (2018) 4490.
- 8) 山本大貴, 新村文郷, 出口大輔, 川西康友, 井手一郎, 加藤一樹, 村瀬洋: LIDAR から得られる 3D 点群を用いた自転車認識 ~可変長 PointNet の検討~, 動的画像処理実用化ワークショップ 2019 講演論文集 (2019) 289.
- 9) K. O. Arras, Ó. M. Mozos and W. Burgard: Using boosted features for the detection of people in 2D range data, Proc. 2007 IEEE Int. Conf. on Robotics and Automation (2007) 3402.
- 10) L. Spinello, K. O. Arras, R. Triebel and R. Siegwart: A layered approach to people detection in 3D range data, Proc. 24th AAAI Conf. on Artificial Intelligence (2010) 1625.
- 11) 高田 智, 榎田修一: LIDAR を用いた歩行者検出のための Real AdaBoost に基づく特徴選択, 自動車技術会論文集, **49**, 4 (2018) 799.
- 12) J. Beltrán C. Guindel, F. M. Moreno, D. Cruzado, F. P. Garcia and A. De La Escalera: BirdNet: A 3D object detection framework from LiDAR information, Proc. 21st IEEE Int. Conf. on Intelligent Transportation Systems (2018) 3517.
- 13) S. Ren, K. He, R. Girshick and J. Sun: Faster R-CNN: Towards real-time object detection with region proposal networks, IEEE Trans. on Pattern Analysis and Machine Intelligence, **39**, 6 (2017) 1137.
- 14) 建部好輝, 出口大輔, 川西康友, 井手一郎, 村瀬洋: 低密度 LIDAR 点群からの歩行者検出, 精密工学会誌, **84**, 12 (2018) 1017.
- 15) Global Road Safety Partnership, Speed management: A road safety manual for decision-makers and practitioners (2008) http://whqlibdoc.who.int/publications/2008/9782940395040_eng.pdf?ua=1 2019.7.26 アクセス.
- 16) A. Geiger, P. Lenz and R. Urtasun: Are we ready for autonomous driving? The KITTI vision benchmark suite, Proc. 2012 IEEE Conf. on Computer Vision and Pattern Recognition (2012) 3354.
- 17) R. Rusu and S. Cousins: 3D is here: Point Cloud Library (PCL), Proc. 2011 IEEE Conf. on Robotics and Automation (2011) 1.
- 18) S. Walter: The partial area under the summary ROC curve, Statistics in Medicine, **24**, 13 (2005) 2025.
- 19) Velodyne LiDAR, HDL-64E S2 datasheet.2010_lowres, https://velodynelidar.com/lidar/products/brochure/HDL-64E%20S2%20datasheet.2010_lowres.pdf 2019.1.6 アクセス.
- 20) C. R. Qi: PointNet: Deep learning on point sets for 3D classification and segmentation, <https://github.com/charlesq34/pointnet/> 2019.1.6 アクセス.
- 21) D. Maturana: 3D/Volumetric convolutional neural networks with Theano+Lasagne, <https://github.com/dimatura/voxnet/> 2019.1.6 アクセス.
- 22) C. R. Qi, W. Liu, C. Wu, H. Su and L. J. Guibas: Frustum PointNets for 3D object detection from RGB-D data, <https://github.com/charlesq34/frustum-pointnets/> 2019.7.1 アクセス.