

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—

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 †

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

† : 名古屋大学, {yamamotot, shinmuraf}@murase.is.i.nagoya-u.ac.jp,
ddeguchi@nagoya-u.jp, {kawanishi, ide, murase}@i.nagoya-u.ac.jp

‡ : 株式会社デンソー, kazuki_katou@denso.co.jp

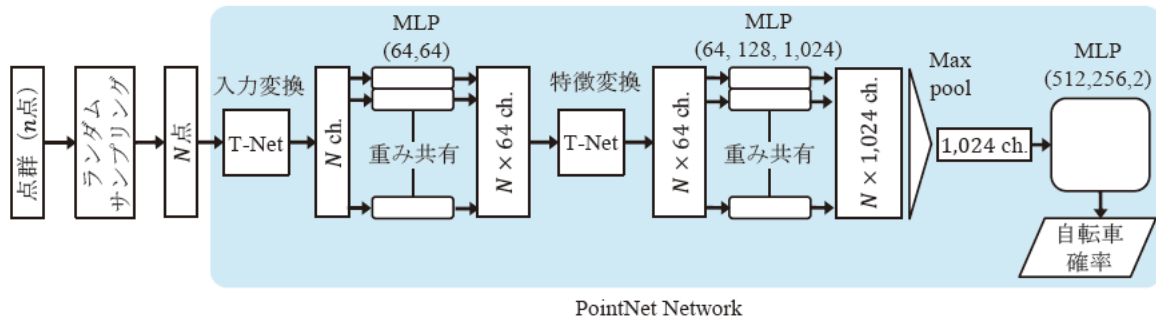
<要約> 近年, 自動運転の実用化を目指して, 距離センサである LIDAR を用いた周辺環境認識に関する研究が活発に行われている. また, 3D 点群データを直接扱う DNN 手法として PointNet が提案されており, 高い認識精度が得られることが示されている. しかし, PointNet は入力点数が固定であるという制約があり, 認識対象物体の点数が大きく変化する場合にそのまま適用することは難しい. そこで本発表では, LIDAR により計測された自転車の認識精度の向上を目的とし, 点群の形状を保持したダウンサンプリングと LIDAR の特性を考慮したアップサンプリングを行うサンプリングモジュールを提案し, この提案サンプリングモジュールの導入により, PointNet の入力の可変長化を行う. KITTI データセットを用いて評価実験を行い, 提案手法によって認識精度が向上することを確認した.

<キーワード> 自転車認識, PointNet, サンプリング, LIDAR, ITS

1 はじめに

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. 特にここ数年, 周辺環境認識のための車載センサとして LIDAR が注目されており, 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車両や歩行者の認識や検出が中心であるが, 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な技術である (以降, 車両本体と運転者を合わせて自転車と呼ぶ). また, LIDAR により計測された 3D 点群データから物体認識する手法が多数提案されている [1, 2, 3, 4, 5]. Kidono ら [1] は, 歩行者の大まかな形状を表現する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. Tatebe ら [2] は, 時系列情報を考慮することで点群の密度を向上させた特徴抽出手法と, 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観測され

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きており, 点群に対する物体認識手法も研究されている. Maturana ら [3] は, 点群の 3 次元位置の分布からボクセル表現を生成し, 3DCNN (3 Dimensional Convolutional Neural Network) 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, 点群をボクセル表現に一度変換する必要があり, 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. 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 (Deep Neural Network) 手法として提案された PointNet [4] が注目を集めている. 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 PointNet を応用した手法もいくつか提案されている. Zhou ら [6] は, 大規模な点群を領域分



PointNet Network

図 1 PointNet の処理手順. T-Net は入力された特徴を変換する DNN である. また, MLP は “Multi Layer Perceptron” を表し, 丸括弧内は各層のチャンネル数を表す.

割した後, 各領域からの PointNet による特徴抽出と物体候補領域の抽出を行う Region Proposal Network [7] 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サンプリングする必要があり, 適切なサンプリングを行わなければ性能が大きく低下することが知られている. そこで本発表では, LIDAR により計測された自転車の認識精度向上を目的とし, 点群の形状を保持したダウンサンプリングと LIDAR の特性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. 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 の入力の可変長化が可能となり, 認識精度も向上することを示す.

2 アイデア

図 1 に PointNet の処理手順を示す. PointNet は点群を他形式に変換することなく直接認識できるが, 入力点数 N が固定であるという制約がある. そのため認識対象物体が高密度に計測されていると仮定し, 単純なランダムダウンサンプリングによって, 一定の入力点数 N に調整したうえで認識処理を行っている. しかし, 遠方に存在する自転車からは高密度な点群を得ることは困難であり, 自転車の車体部分のような細い部分から計測できる点数も少なくなる. このことから, 図 2 に示すように, 単純なランダムダウンサンプリングでは, 車体部分等の計測点が認識に利用されず, 元の形状から大きく変化してしまう可能性がある. また, 認識対象物体から得られる点数が PointNet の入力点数 N 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



図 2 自転車から計測される点群とランダムサンプリング後の点群.

が, 同じ計測点の情報を使うことになるため, 性能向上は見込めない. これらの問題に対して, 自転車の形状を保持可能な新しいサンプリングモジュールを提案し, これを PointNet に導入することで解決を図る. 提案サンプリングモジュールの機能を以下に示す.

- 機能 1
計測点数 N 以上の入力点群に対し, 点群の形状を保持したダウンサンプリング
- 機能 2
計測点数 N に満たない入力点群に対し, LIDAR の測距精度に基づき摂動を加えるアップサンプリング
- 機能 3
学習時にバッチ単位でサンプリング処理を行うデータ拡張

上記の機能を持つ提案サンプリングモジュールを導入することにより, PointNet の入力を可変長化し, 自転車に対する認識精度の向上を狙う. 図 3 に提案手法の処理手順を示す.

2.1 点群の形状を保持するダウンサンプリング手法

自転車の車体部分は人の身体部分と比較すると, 車体部分は細いため, 計測できる点数が少なくなる. そのため, 自転車は人の身体部分と自転車の車体部分から計測される点の数に大きな差があり, ランダムサン

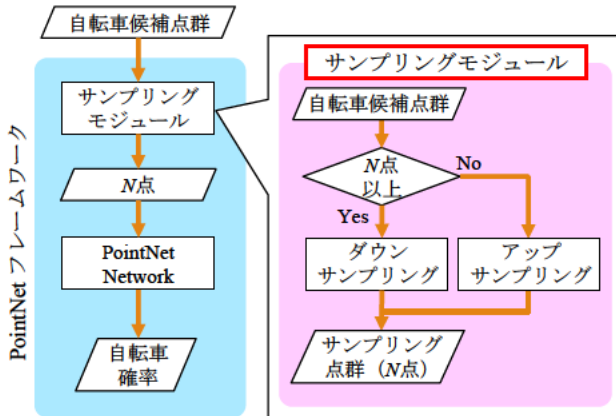


図3 提案手法の処理手順.

プリング手法を行うと、点の密度が大きい身体部分の点が多く抽出される。これにより、計測点の密度が小さい車体部分からは計測点が抽出されず、元の形状から大きく変化してしまう可能性がある。そこで、点群全体の密度が均一となるようにサンプリングを行うことで、車体部分からの確実なサンプリングを可能とする。図4にダウンサンプリング手法の概要を示す。

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。各分割領域 $V_{x,y,z}$ に含まれる計測点の数を $\alpha_{x,y,z}$ とすると、領域 $V_{x,y,z}$ に含まれる計測点 p_i の重み w_i を式 (1) により算出する。

$$w_i = \frac{1}{\alpha_{x,y,z}} \quad (1)$$

算出した重み w_i に基づいて inverse transform sampling [8] を行うことで、 $\alpha_{x,y,z}$ の値が小さい分割領域 $V_{x,y,z}$ からもサンプリングを行うことを可能とする。

2.2 LIDAR の測距精度に基づく距離に応じた擾動を加えるアップサンプリング手法

遠方に存在する自転車からは高密度な点群を得られないため、PointNet により認識を行う場合は点群の密度を増加させる必要がある。そこで建部ら [5] の考え方を参考にし、LIDAR の測距精度を考慮して計測点を確率分布で表現し、点群の存在尤度マップ M を生成する。具体的には、LIDAR から照射される LASER 光は幅 (太さ) を持ち、その幅は観測される物体までの距離が遠くなるほど大きくなる。そのため、LIDAR の測距精度と距離に応じて、計測点の計測位置に関する確率分布が変化する。そして、生成した存在尤度マッ

プ M からサンプリングを行うことで N 点を抽出する。図5にアップサンプリング手法の概要を示す。

まず、存在尤度マップ M は、各計測点 p_i を用いて式 (2) により求められる。

$$M = \sum_i g(p_i, \Sigma_i) \quad (2)$$

$g(\cdot)$ はカーネル関数であり、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 $g(p_i, \Sigma_i)$ 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 (3) のように表される。

$$\Sigma_i = \begin{pmatrix} \sigma_i & 0 & 0 \\ 0 & \sigma_i & 0 \\ 0 & 0 & \sigma_i \end{pmatrix} = \sigma_i I \quad (3)$$

ここで、 σ_i は d_i と LIDAR の測距精度に基づいて計算される。生成された存在尤度マップ 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。

3 提案サンプリングモジュールの実装

2節の考え方を実現するため、以下の処理を用いて提案サンプリングモジュールを実装する。

3.1 ダウンサンプリング手法

2.1節の考え方に基づき、点群全体の密度が均一となるようにサンプリングする手法を述べる。遠方に存在する自転車から計測できる点群全体の密度は低く、サンプリング時の乱数の偏りを無視できない可能性がある。そこで、以下の手順により、計測点の密度の低い領域の欠損を防ぐ。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。次に、各分割領域内に計測点が存在する場合、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。この抽出処理の反復により、2.1節の式 (1) の重み w_i に基づく inverse transform sampling を模擬する。 j 回目の抽出処理により抽出される点数を S_j とすると、 S_j は抽出されていない計測点が存在する領域数を表し、また反復回数 β は以下の条件を満たす最大の整数となる。

$$N \geq \sum_{j=1}^{\beta} S_j \quad (4)$$

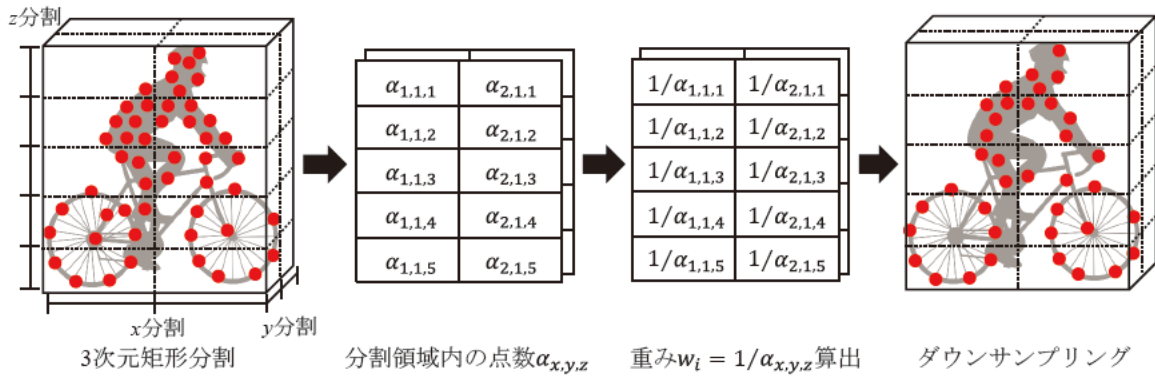


図 4 ダウンサンプリング手法の概要 ((X, Y, Z) = (2, 2, 5) における例)。

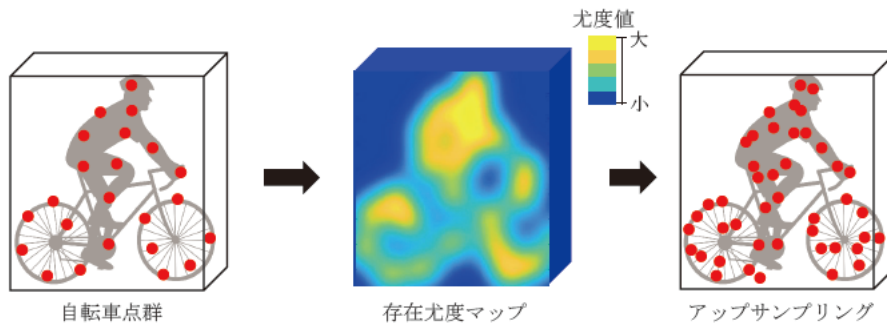


図 5 アップサンプリング手法の概要。

ただし、各反復を通して同じ計測点は選択せず、また分割領域内のすべての計測点を抽出した場合は、以後その領域内からは抽出しない。最後に、不足分の $N - \sum_{j=1}^{\beta} S_j$ 点について、すべての分割領域内から重複なしで無作為に抽出する。

3.2 アップサンプリング手法

2.2 節の考えに基づき、LIDAR の測距精度に基づき距離に応じた摂動を加えてアップサンプリングする手法を述べる。自転車候補点群が n ($< N$) 点で構成されているとする。まず各計測点に対して、LIDAR の測距精度に基づく原点からの距離 d_i に応じた位置の摂動をその計測点の位置情報に加え、新たに $(\gamma - 1)n$ 点を追加する。ただし、 γ は $N > \gamma n$ を満たす最大の整数とする。原点からの距離 d_i に応じた摂動は、平均 0、標準偏差 σ_i の正規分布に従う乱数により求める。最後に、不足分の $N - \gamma n$ 点について、3.1 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。この処理により、2.2 節の式 (2) の存在尤度マップ M からのサンプリングを模擬する。

3.3 学習時のバッチ単位でのサンプリング処理

従来の PointNet は n 点で構成される点群に対して、 N 点へのランダムサンプリングを行うことで入力点群を生成した後、バッチ単位での学習を行う。そのため、PointNet の学習に用いられる計測点は入力点群の生成時に固定され、バッチ単位で常に同じ点群が用いられる。そのため、サンプリング前の点群の多様性が失われる問題がある。そこでサンプリングモジュールにより、バッチ毎に機能 1 (3.1 節) のダウンサンプリング及び機能 2 (3.2 節) のアップサンプリングを行うことで抽出される計測点に変化し、点群の多様性を増加することが可能となる。

3.4 PointNet の学習と認識

図 6 に、3.1 節、3.2 節、3.3 節で述べた提案サンプリングモジュールを導入した PointNet の処理手順を示す。学習毎に提案サンプリングモジュールを通して自転車候補点群から 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。認識時は、提案サンプリングモジュールを通して自転車候補点群から生成した 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。

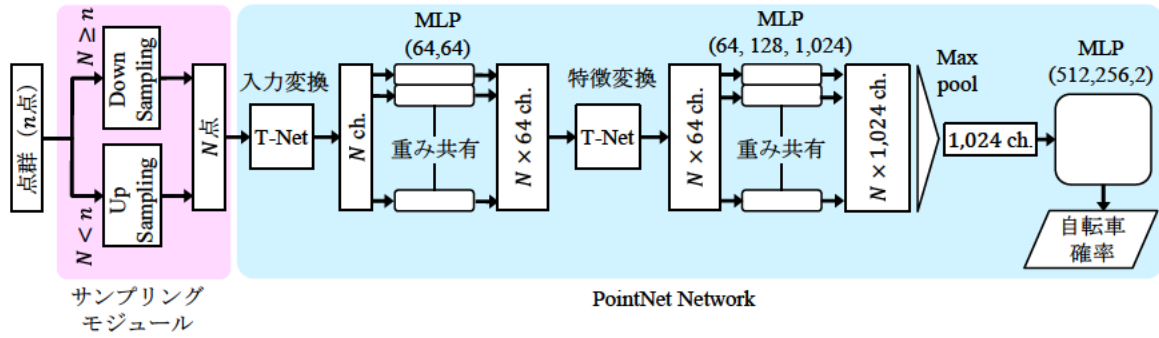


図 6 提案サンプリングモジュールを導入した PointNet の処理手順。

4 評価実験

本節では、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る。

4.1 データセット

本実験では、KITTI データセット [9] 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した。

- 車両前方 30 m 以内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 Library [10] 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 以内に存在
- 最小点数：30 点
- 以下の式 (5), (6), (7) の全てを満たす。なお、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。

$$w \leq 2.5 \text{ m かつ } l \leq 2.5 \text{ m} \quad (5)$$

$$w \geq 1.0 \text{ m または } l \geq 1.0 \text{ m} \quad (6)$$

$$0.7 \text{ m} \leq h \leq 2.0 \text{ m} \quad (7)$$

表 1 にそれぞれの点群のデータ数を示す。自転車点群及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。

4.2 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

表 1 自転車点群と非自転車点群のデータ数

点群の種類	データ数
自転車点群	805 個
非自転車点群	2,735 個

手法 1、提案手法 2、従来手法 1、従来手法 2、従来手法 3 の 5 つの手法を比較した。

- 提案手法 1
従来 PointNet [4] に、提案サンプリングモジュールにおける 3.3 節の機能 3 のデータ拡張のみを導入した手法。ただし、3.1 節と 3.2 節のサンプリング手法を、ランダムダウンサンプリング手法とランダムアップサンプリング手法で代用する。
- 提案手法 2
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。ただし、提案サンプリングモジュールは 3.1 節、3.2 節、3.3 節に示す全機能を使用する。
- 従来手法 1
人手で設計したハンドメイド特徴量 [1, 2] をカーネル SVM により認識を行う手法。
- 従来手法 2
3DCNN である VoxNet [3] により認識を行う手法。
- 従来手法 3
従来 PointNet [4] により認識を行う手法。各点群データをランダムサンプリング処理により N 点にすることで学習用及び評価用データセットを生成した後、認識を行う。

評価指標として、ROC (Receiver Operating Characteristic) 曲線と p-AUC (partial Area Under the Curve) [11] を用いた。p-AUC とは、ROC 曲線のある

表 2 各手法の p-AUC

手法	p-AUC
提案手法 1	0.963
提案手法 2	0.974
従来手法 1	0.904
従来手法 2	0.935
従来手法 3	0.963

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。本発表では、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。

4.3 従来手法 (PointNet 及び VoxNet) の実装

PointNet の実装には、Qi が提供しているプログラム [12] を用いた。入力には、各計測点の 3 次元位置情報 (x 座標, y 座標, 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。入力点数を $N = 128$ 点、3.1 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。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。式 (3) の標準偏差 σ_i は、原点から計測点 p_i までの距離 d_i を用いて、式 (8) により求めた。

$$\sigma_i = \frac{3\Omega}{D^2} d_i^2 \quad (8)$$

ここで、KITTI データセットで用いられている Velodyne LiDAR 社の HDL-64E の測距精度に基づき、 $\Omega = 0.02$ m, $D = 25$ m とした。

次に、VoxNet の実装には、Maturana が提供しているプログラム [13] 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。

4.4 実験結果

図 7 に実験により得られた各手法の ROC 曲線を示す。提案手法 2 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 2 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 1, 提案手法 2, 従来手法 3) が高い p-AUC を得られており、特に提案サンプリングモジュールの全機能を利用する提案手法 2 が最も高い p-AUC を得

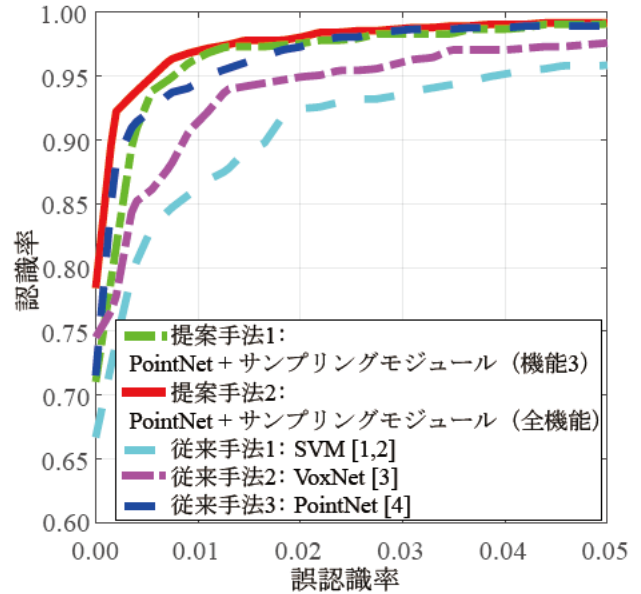


図 7 各手法の ROC 曲線 (認識率 60% から 100%, 誤認識率 0% から 5% までの範囲を描画)

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、自転車に重要な低誤認識率時における認識精度の向上を確認した。

4.5 考察

まず提案手法 1 と従来手法 3 に着目すると、提案手法 1 の認識精度が従来手法 3 を上回っていることが確認できる。これは、提案手法 1 は学習毎に自転車候補点群からサンプリングされる計測点が変わることによって、学習データの多様性が大きくなっているためであると考えられる。次に、提案手法 2 と提案手法 1 に着目すると、提案手法 2 の認識精度が提案手法 1 を上回っていることが確認できる。よって、提案手法 2 により、自転車の車体部分等の計測点の密度が小さい部分から確実な計測点の抽出を行い、元の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。

5 むすび

本発表では、点群の形状を保持したダウンサンプリングと LIDAR の特性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、この導入により PointNet の入力を可変長化する手法を提案した。具体的には、提案ダウ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，点群全体の密度が均一になるようにサンプリングを行うことで，車体部分等の点群の密度が小さ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 の測距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 手法や他の 3DCNN 手法，SVM 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及び性能評価が挙げられる．

謝辞

本研究の一部は，科学研究費補助金による．

参考文献

- [1] K. Kidono, T. Miyasaka, A. Watanabe and J. Miura, “Pedestrian recognition using high-definition LIDAR,” Proc. 2011 IEEE Intelligent Vehicles Symposium, pp.405–410, June 2011.
- [2] 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, pp.157–164, Feb. 2017.
- [3] 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, pp.922–928, Sept. 2015.
- [4] 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, pp.652–660, July 2017.
- [5] 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)，pp.1017–1024，Dec. 2018.
- [6] 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, pp.4490–4499, June 2018.
- [7] S. Ren, K. He, R. Girshick and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” Proc. 29th Conf. on Neural Information Processing Systems, pp.91–99, Dec. 2015.
- [8] L. Devroye, “Non-uniform random variate generation”, Springer, 1986.
- [9] 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, pp.3354–3361, June 2012.
- [10] R. Rusu and S. Cousins, “3D is here: Point Cloud Library (PCL),” Proc. 2011 IEEE Conf. on Robotics and Automation, pp. 1–4, May 2011.
- [11] S. Walter, “The partial area under the summary ROC curve,” Statistics in Medicine, 24(13), pp.2025–2040, May 2005.
- [12] C.R. Qi, “PointNet: Deep learning on point sets for 3D classification and segmentation,” <https://github.com/charlesq34/pointnet/>. (2019/1/6 参照)
- [13] D. Maturana, “3D/Volumetric convolutional neural networks with Theano+Lasagne,” <https://github.com/dimatura/voxnet/>. (2019/1/6 参照)