

1. はじめに

VoxNet[1] は, LIDAR 等の距離センサで得られたポイントクラウドを占有グリッドに変換し, 3D Convolutional Neural Network(3D-CNN) に入力することで 3 次元物体を認識する手法である. しかし, 占有グリッドに変換する際, 物体の大きさ等の情報が欠落する. 本研究では VoxNet の高精度化を目的とし, 対象となる物体の補助情報を付与した 3D-CNN を提案する.

2. VoxNet

VoxNet[1] は, 大規模なポイントクラウドを直接 CNN で処理するために, ポイントクラウドデータを占有グリッドに変換し, 3D CNN に入力する. 占有グリッドは, 畳み込み処理が可能となるように 3 次元のグリッド特徴に変換したものであり, 形状に依存せず 3 次元物体認識が可能となる. また, データサイズを小さくできるため, 高速な処理が可能となる. 3D CNN は, 分類タスクに有効な 3 次元の局所空間フィルタを学習できる. そして, 複数の層を積み重ねることによりそれらの共起性を捉えることができる. しかし, 占有グリッドにすることで物体の大きさ等の情報が利用されず, 情報量が減ってしまう.

3. 提案手法

提案手法では, 3D CNN の全結合層に物体の補助情報を付与することで認識精度の向上を図る. 補助情報は LIDAR 中心からの物体方向 (角度) や大きさ (高さ, 幅, 奥行き) とする. 提案手法は以下の 4 つのステップからなり, Step1 ~3 までの流れを図 1 に示す.

Step1 ポイントクラウドのクラスタリングを行い, 物体候補領域を抽出する. 取得した際に物体の位置から角度を算出する.

Step2 物体候補領域から物体の幅と高さ, 奥行き, 角度を求める.

Step3 物体候補領域から占有グリッドを生成する. 占有グリッドは各軸方向に対し, 32 分割した局所グリッド領域を設定し, 局所グリッド内にポイントクラウドデータがある場合は 1, ない場合は 0 とする.

Step4 占有グリッドを 3D-CNN に入力する. 識別には Softmax 関数を使用, 各クラスの確率を算出する.

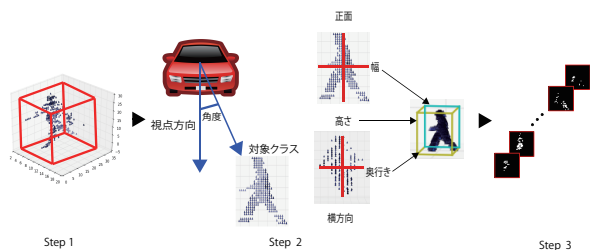


図 1: Step1 から Step3 までの処理の流れ

本手法で用いる 3D CNN の構造を図 2 に示す. 畳み込み層 3 層, 全結合層 2 層から構成されている. 付与する 4 つの補助情報は全結合層 1 層目に入力する.

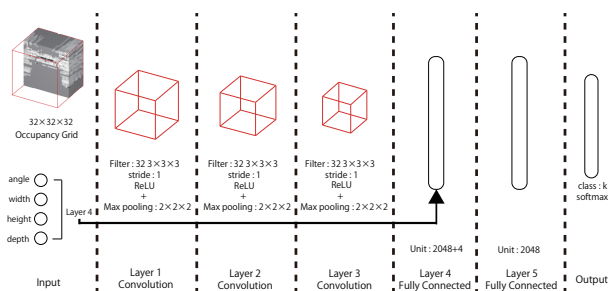


図 2: 3D-CNN の処理の流れ

4. 評価実験

3D-CNN に物体の補助情報を付与することの効果の評価実験により示す.

4.1. シミュレーションデータによる実験

本実験では, シミュレータにより生成したポイントクラウドデータを学習と評価に用いる. 認識クラスは, Car, Othervehicle, Bus, Bicycle, Pedestrian, Child, Stationary, UtilityPole, Curb, Guardrail, Wall の 12 クラスである. サンプルはデータ拡張により, 学習に 855,409 枚, 評価に 131,876 枚を用いる. バッチサイズは 60, 学習回数は 10epoch である.

実験結果を図 3 に示す. 対象物体の角度および大きさ情報を付与することで, 平均認識率が約 10% 向上した. 特に Child の認識率が大幅に向上し, 補助情報を付与することで大人と子供を区別ができることが確認できた. Bus は OtherVehicle と Wall に誤認識された. これは, OtherVehicle は大型車を含み, 大きさも類似しているため, 補助情報を付与しても識別が困難であると考えられる.

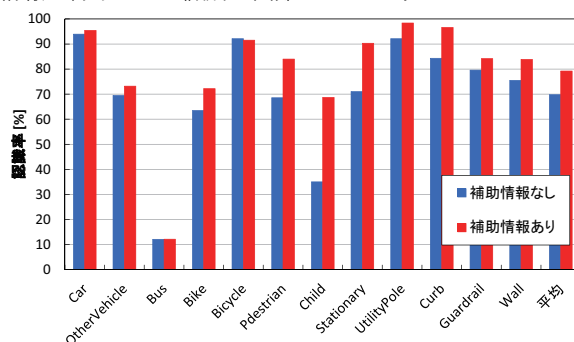


図 3: シミュレーションデータの実験結果

4.2. 実環境データによる実験

本実験では, KITTI Vision Benchmark Suite(KITTI) データセットを使用する. 認識クラスは, Car, Van, Truck, Tram, Pedestrian, Person_sitting, Cyclist, Misc の 8 クラスである. サンプルは学習に 13,550 枚, 評価サンプルに 4,518 枚使用する.

実験結果を図 4 に示す. 補助情報を付与することで認識精度が約 6% 向上し, 実環境下でも効果があることが確認できた. 特に Person_sitting の精度が大幅に向上した. 一方, 車種が多く含まれている Misc は Car 等に類似しているものが多く, 精度が低下した.

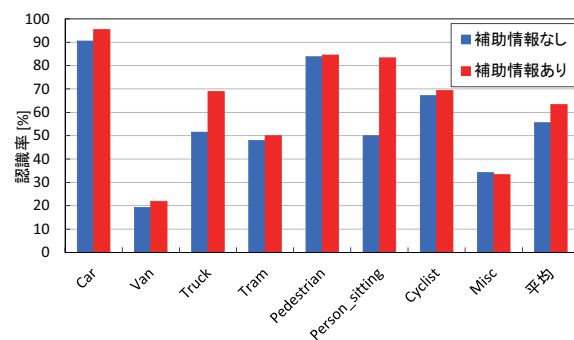


図 4: 実環境 (KITTI) の実験結果

5. おわりに

物体の補助情報を付与することで 3D-CNN の精度を約 10% 向上させることができた. そして, 実環境では約 6% 向上した. 今後は, 効果的な占有グリッドの作成方法について検討をする.

参考文献

[1] D. Maturana, et al., "VoxNet: A 3D Convolutional Neural Network for Real-Time Object Recognition", In IROS. 2015.