

深層学習を用いた道路標識検出のための 3 次元都市モデルに基づく学習用画像の生成 Generating Training Images based on a 3D City Model for Road Sign Detection using Deep Learning

加藤 立登[†] 西口 敏司[‡] 水谷 泰治[‡] 橋本 渉[‡]

Ryuto Kato Satoshi Nishiguchi Yasuharu Mizutani Wataru Hashimoto

1. はじめに

自動運転技術に関する研究開発が盛んに実施されており [1], 車載カメラで撮影した映像から周囲の状況を推定するための手法の一つとして画像認識技術が用いられている。一方, 高精度な認識が可能な手法として深層学習が活用されるようになってきた。しかしながら, 深層学習を用いた識別には, 大量の正解ラベル付き学習用画像を用意する必要があり, 実空間における道路標識を様々な位置・方向から撮影した画像を用意することは非常に困難な作業となる。

そこで本研究では, 3DCG 環境下の 3 次元都市モデルに道路標識モデルを配置し, 環境の明るさや, 標識を撮影するカメラの位置や向きを変えて撮影することで生成した道路標識を含む学習用画像が, 深層学習を用いた道路標識検出にどの程度有効か検討する。

2. 道路標識検出のための学習用画像の生成手法

深層学習では, 多数の正解ラベル付き学習データを用意することが必要である。画像中に映った物体の認識を目的とした深層学習では, データ拡張と呼ばれる手法が利用される。この手法は, 画像を幾何的に変形したり, 回転したり, 色成分を人工的に変化させたりすることで, 疑似的に学習用画像を増加させる手法である。しかしながら, このような変形の結果得られる画像上の物体の様子は, 見た目, 角度, 色, 物体の周囲の背景が現実の様子と大きく異なるという問題がある。そこで本研究では, 3 次元の都市モデルと道路標識モデルを用いて, より現実に近い見た目を持つ, 道路標識を含んだ学習用画像を多数生成する手法を提案する。

学習用画像を生成するにあたっては, 3 次元都市モデル内に配置した, 太陽の位置と色味, および, 標識を撮影するカメラの標識に対する位置と向きをパラメータとし, 撮影毎にこれらのパラメータを変化させることで, 道路標識を含む多様な学習用画像の生成を行う。

2.1 太陽光の変化の再現

3 次元都市モデルにおいて, 太陽光は平行光源としてモデル上に配置する。現実空間での影の変化を再現するために, 平行光源の位置を撮影ごとに変化させる。また, 現実の標識や背景は, 環境の影響で様々な色の変化があるため, 昼光色を基準とした平行光源の RGB 値を変化させる。具体的には, 0~1 に正規化された RGB 値に対して, R 値については-0.4~+0.4 の値を, G 値については, -0.1~+0.1 の値を, B 値については-0.2~+0.2 の値を元の RGB 値それぞれにランダムに加える。なお, 変化の程度は経験的に定めた。

[†] 大阪工業大学 大学院情報科学研究科

[‡] 大阪工業大学 情報科学部

2.2 道路標識に対するカメラの位置と傾きの変化

本研究では, 3 次元都市モデルに配置したカメラで撮影した様子を学習用画像として生成する。ここで生成される画像内の標識の大きさと位置, 傾きの変化を再現するために, カメラの道路標識に対する位置と向きを変化させる。

まず, 道路標識を視野の中心にとらえつつ撮影可能な基準となる位置を予め 10 ヶ所設定しておく。これを元に, 道路標識との距離を 5 段階で設定し, 各段階において, 前後に一定の範囲でランダムに移動させる。さらに, 左右に 3 段階の位置を設定し, 各段階において, 左右に一定の範囲でランダムに移動させる。また, カメラの向きは, 道路標識の中心を見た方向を中心とし, この中心を含めた, 上, 下, 左, 右, 左上, 左下, 右上, 右下の 9 段階の向きをあらかじめ設定し, 各向きにおけるカメラのローカル座標系でのオイラー角の x 軸と y 軸の値に対し, x 軸は 6.5~7.7, y 軸は 6.5~7.3 の範囲でランダムな倍率をかける。以上から, 一つの道路標識に対して, $5 \times 3 \times 9 \times 10 = 1,350$ 枚の画像を生成することができる。

2.3 道路標識の矩形領域の取得

これまでの方法で道路標識を含む現実のシーンを再現した画像が生成できるが, これらの画像を検出器の学習用画像として与える際には, 画像内の標識を含む矩形領域が必要となる。この矩形領域を画像ごとに手で付与するのは大変な労力が必要となるため, 本研究では画像上の道路標識を含む矩形領域を自動的に抽出する。

このために, まず, カメラの姿勢を, 学習用画像を生成した状態で保持したまま, 道路標識モデルの色を黒, それ以外のモデルの色を全て白として撮影する。これにより, 道路標識部分が黒色のマスク領域としたマスク画像が得られる。次に, このマスク領域に外接し, x 軸方向と y 軸方向に平行な辺を持つ矩形領域を求める。得られた矩形領域を, 学習データの正解領域として使用する。これらの処理の様子を図 1 に示す。

これらの処理により, 検出器の学習に必要な学習用画像, 正解ラベル, および, 道路標識の正解領域が得られる。

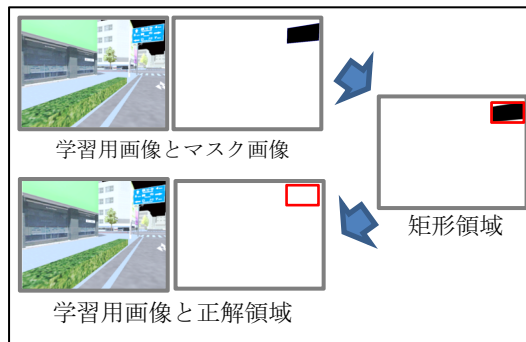


図 1 学習用画像と正解領域の生成

3. 実験

3 次元都市モデルおよび道路標識モデルを扱うために、Unity を使用した。道路標識としては、225 種類取り上げ、標識オブジェクトの表面にそれぞれのテクスチャを張り付けた。2.3 節で説明した道路標識の矩形領域の抽出には、Unity の Layer 機能、Renderer 機能、および、OpenCV の外接矩形検出関数を利用した。また、本研究で用いた計算機の性能は、CPU は Intel Core i7-7700K 4.2GHz、メインメモリは 32GB、GPU は GeForce GTX 1080Ti である。

3.1 学習用画像と正解データの生成にかかる時間

デジタルカメラ等で道路標識を含んだ学習用画像を収集する方法と比べ、提案手法がどの程度時間短縮となるか、推計した。学習用画像を 303,750 枚、正解領域を計算するためのマスク画像を 303,750 枚生成した。これらの画像の生成には、約 9 時間 40 分を要した。また、事前に行った、標識オブジェクトに張り付ける標識画像の作成には、約 13 時間 45 分、マスク画像からの標識の正解領域と正解ラベルの生成には約 20 分を要した。従って、303,750 枚の学習用画像と正解データの生成に合計 23 時間 45 分を要し、1 枚当たり約 0.28 秒で学習用画像と正解領域を生成できた。一方、デジタルカメラ等を用いて手で画像を収集する場合は、1 枚の撮影に約 3 秒、手で正解領域を付与する場に 1 枚当たり約 5 秒かかるとし、さらに、撮影の移動時間も考慮すると 1 枚当たり要する時間と労力はけた違いに大きくなる。以上から、学習用画像を準備する労力の観点からは、提案した画像生成手法が有用であるといえる。

3.2 生成画像を用いて学習した検出器の識別精度

デジタルカメラを用いて撮影した画像、および、Google ストリートビューの画面から切り取った画像のうち、30 枚の画像を評価用として、検出器の識別精度を調べた。検出器としては、YOLO version2 を用いた。なお、YOLO の学習とテストに用いた画像の解像度はいずれも幅 640 画素、高さ 480 画素である。また、検出対象とする道路標識は国土交通省が一覧で定めている 2016 年 9 月時点の分類[2]を参考にして、225 クラスにしたものを利用した。

識別精度の評価値として F 値を用いた。算出手順としては、検出の出力結果をその画像の中に対象の標識 225 種が存在するかどうかという識別の問題に置き換えて F 値を算出した。

実験では、30 枚の全評価用画像のうち、出現した標識のクラス数は、対象全 225 クラスのうちの 31 クラスであった。さらに、この 31 クラスのうち、テスト画像中にそのクラスが含まれている画像枚数が 4 枚以上であった 3 クラスについて、これら 3 クラスの精度 (Precision)、再現率 (Recall)、F 値を表 1 に示す。表 1 からわかるように、標識番号 128 以外は F 値がすべて 40%以上、かつ、精度が 100%であった。一方、標識番号 128 は F 値が 0%であった。また、標識番号 322 の F 値が 75.0%で最も高かった。なお、本稿の実験で使ったテスト画像に対しては、全体的に精度より再現率が低いという傾向を示す結果であった。図 2 に、これらの 3 クラスの標識の例を示す。

実験結果から、標識 322 に関しては、提案手法により F 値 75.0%の高い精度で検出が可能であることがわかった。また、F 値が 40%以上の標識 322、328 に関しては、精度が

100%という値を示していることから、テスト画像の中に対象の標識が含まれていると予測した結果で、間違っている確率は低いと考えられる。一方、F 値が 0%であった標識 128 に関しては、本稿の手法では適切に検出器の学習が行えない可能性が高いと思われる。

表 1 で評価した 3 クラスの標識に関して、実験結果と図 2 に示した標識の特徴を照らし合わせると、F 値が 40%以上であった標識に関しては、標識内の矢印や線、円などの図形の数や文字の数が少ないことがわかる。このため、比較的単純な図形や少ない文字で構成された標識に関しては、本手法は検出器の学習に一定の有効性があると考えられる。一方、F 値が 0%の標識 128 については、矢印の上に道路番号の図形が付与され、さらに文字も多く含まれている。これより、複雑な図形が含まれている標識や様々な図形・文字で構成されている標識に対しては、本手法では十分な検出器の学習が行えないと考えられる。

これらのことから、本研究の手法において、より精度の高い検出器を得るためには、文字の内容の変化や標識内の様々な図形の変化を考慮した学習用画像が重要であると思われる。

表 2 検出器の実験結果

標識番号	精度(Precision)	再現率(Recall)	F値
128	0.0%	0.0%	0.0%
322	100.0%	60.0%	75.0%
328	100.0%	25.0%	40.0%



図 2 実験で評価対象となった標識の例

4. おわりに

本稿では、道路標識検出に深層学習を適用するための多様な学習用画像を生成するために、3 次元都市モデルを用いる手法を提案した。実験結果から、3DCG を用いて生成した道路標識を含む画像を学習用画像として用いても、道路標識の検出が行える検出器を生成できる可能性があることがわかった。

今後の課題としては、道路標識を含むより忠実なシーンの再現や、現実の道路標識画像も組み込んだ学習用画像データセットの生成などが挙げられる。

参考文献

- [1] 保坂明夫, 青木啓二, 津川定之, “自動運転—システム構成と要素技術—,” pp.106 - 111, 森北出版, 2015.
- [2] 国土交通省, “道路：道路標識等 道路標識の概要等 道路標識一覧,” <http://www.mlit.go.jp/road/sign/sign/douro/ichiran.pdf>, 2016 - 09 - 27 閲覧.