

1. はじめに

自動車に搭載する認識アルゴリズムを評価する際、同一シーンにおいて天候等の異なる状況下のデータを収集することは難しい。そこで、CG を用いて実環境を再現することで、異なる天候下の同一シーンデータの生成が可能となる。一方で、実環境の画像と CG の生成画像には、アピアランスの違いから発生するドメイン差があると認識アルゴリズムの検出精度に影響を与えることになる。そこで本研究では、実環境と CG で作成したデータに対して、深層学習モデルによる物体検出結果を比較して両データの差異について明らかにする。

2. DIVP シミュレータによる CG データの作成

本研究では、実世界の東京臨海部を撮影したシーンを、Driving Intelligence Validation Platform (DIVP) のシミュレータを用いて再現する。DIVP は、自動運転システムの評価を行う環境を構築する内閣府のプロジェクトで開発されたシミュレータである。RAW データからの現像処理をするため、DIVP シミュレータは実環境との偽色の再現性が一般公開データセットに比べて高い。JPEG 画像は色合い等の調整後に画質が劣化してしまうが、RAW データであれば調整後も画質が劣化することがない。テストデータ数は 6 つあり、各データは実環境と CG データを含む。CG データには、曇り、雨、13 時、16 時のバリエーションも含まれ、検出対象車両ではない車両は描写されていない。

3. YOLOv4

YOLOv4 [1] は、YOLO を拡張して検出精度と計算速度を向上させたモデルであり、backbone と neck, head から構成される。backbone は CSPDarknet53 により画像の特徴を抽出する。neck は backbone から受けとった特徴マップを Feature Pyramid Network(FPN) と Path Aggregation Network (PANet) により受容野の広い特徴量に変換する。head はクラス分類の予測や bounding box の位置を予測する。

4. 評価実験

実画像と CG 画像に対して YOLOv4 を用いて物体検出を行い、遠方車両の検出精度を比較する。YOLOv4 の学習には東京臨海部で収集した物体検出用データセットを用いる。本データセットの学習枚数は 19,261 枚である。

4.1. 定量的評価

本実験では、2 つのシーンを用いて比較実験を行う。各データをシーン 1、シーン 2 とし、各シーンに対して検出対象の車両との距離ごとの検出精度を比較する。図 1 に示すようにシーン 1 は対象車両に日が当たっており、シーン 2 は対象車両に日が当たっていない状態を対象とする。シーン 1 の検出精度を図 2、シーン 2 の検出精度を図 3 に示す。横軸は検出対象の車両までの距離、縦軸は車両の検出率を表す。また、青線は実環境上の検出精度、橙線は CG の検出精度を表す。シーン 1 では、100m までは実環境と CG で同程度の検出精度だが、100m 以上遠くなると CG 内での検出精度は実環境よりも低下する。シーン 2 では 20~60m の距離では実環境と CG で同程度の検出精度だが、60m 以上遠くなると CG 内での車両の検出は不可能であった。



図 1: 評価対象のシーン (CG データ)

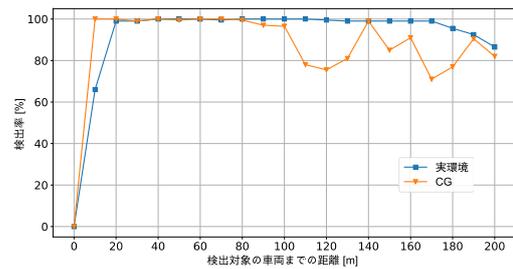


図 2: シーン 1 の検出率

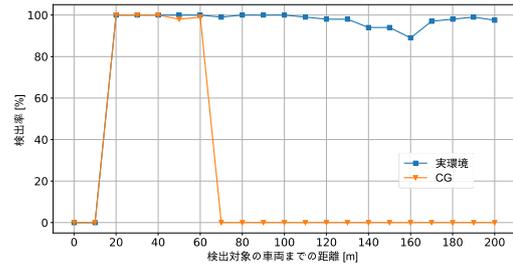


図 3: シーン 2 の検出率

4.2. 定性的評価

各シーンで実環境と CG を比較した例を図 4 に示す。図 4 の (a) と (b) はシーン 1、図 4 の (c) と (d) はシーン 2 を示す。シーン 1 では、実環境と CG で遠方車両に対する検出が同程度であった。シーン 2 では、CG 内の検出対象の車両から遠くなると検出が不可能であったが、近くなると実環境と同程度の検出が可能であった。図 1 に示すように、シーン 1 とシーン 2 では CG 内の対象車両に対する影の影響が異なる。シーン 1 は影の影響が小さいデータ、シーン 2 は影の影響が大きいデータであり、評価実験からシーン 2 はシーン 1 に比べ検出精度が低下することから、実環境と CG 間で影がドメイン差の要因ではないかと考える。

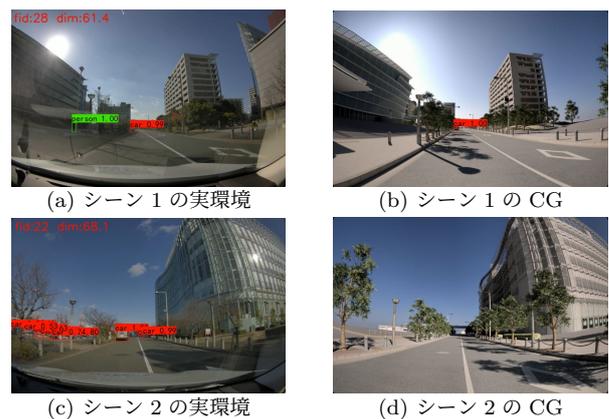


図 4: 実環境と CG データに対する検出例

5. おわりに

本実験では物体検出の精度から実画像と CG のドメイン差について調査した。CG 画像では、検出対象に対し影の影響が大きいことが判明した。今後は実環境と CG のドメイン差を低減するために影領域に対する CG の改良へとつなげる。

参考文献

[1] A. Bochkovskiy, et al., "YOLOv4: Optimal Speed and Accuracy of Object Detection", CVPR, 2021.