歩行者が交差点に存在するシーンにおける物体検出モデルの精度評価

鈴木　陽太郎[1)]　板谷　英典[1)]　平川　翼[1)]　山下　隆義[1)]　藤吉　弘亘[1)]

Evaluating the accuracy of object detection models in scenes where pedestrians are present at intersections

Yotaro Suzuki　Hidenori Itaya　Tsubasa Hirakawa　Takayoshi Yamashita　Hironobu Fujiyoshi

Evaluation of object detection models for automated vehicles requires a large amount of evaluation data. However, collecting evaluation data in a real environment is extremely costly. Therefore, it is expected that a variety of evaluation data can be created by using a computer graphics environment. In this study, an evaluation scene is created using a DIVP simulator to evaluate the detection accuracy of object detection models.

KEY WORDS: **Safety, Pedestrian detection/protection,** image processing/information processing(C1)

1. まえがき

安全かつ効率的なモビリティの実現に向け，自動運転システムの開発が加速しており，自動運自動車の実用化が近づきつつある．自動運転システムは，認知/判断/制御で構成され，それらの構成要素には人工知能（AI）の活用が期待されている．AI を用いた代表的な認知技術には，自車両周辺の物体を検出する物体検出技術が挙げられる．この技術により，自車両周辺の物体を高精度に検出することで，交通事故を未然に防ぐための適切な判断や制御へ繋げることが可能となる．

物体検出技術は，カメラ等のセンサにより取得した情報から特定の物体に対する位置や種類を推定する．この技術は様々な分野で活用され，数多くの物体検出モデルが提案されている．これら物体検出モデルを活用する際には，学習済みの物体検出モデルが正しく対象物体を検出できているか評価することが重要となるが，網羅的な評価データが必要となり，評価データの収集方法が課題である．

自動運転分野では，実環境において車載カメラや LiDAR などから収集されたデータセットが多く公開されている[1,2,3,4,5,6]．しかし，これらのデータセットは実環境で収集されているため，データ収集のコストが非常に高く，また事故に繋がるようなエッジケースはほとんど含まれていない．そこで，シミュレータなどの CG 環境を活用することで，データ収集のコストを抑え，エッジケースを含むシナリオでの評価が期待されている．　また CG 環境では，車両や歩行者等のアセット変更，天候や時刻の変更が容易であり，同一シーンにおける様々な環境条件での評価データを収集することができる．

本研究では，自動運転分野における車載カメラからの歩行者検出を対象とし，DIVP（Driving Intelligence Validation Platform）シミュレータ［7］を用いて収集した多様な評価データを用いて，物体検出モデルの性能を網羅的に評価する．本研究で評価するシナリオは，横断歩道手前に停車車両が存在し，歩行者が横断歩道を渡るシーンとする．歩行者，停車車両，天候，時刻の環境条件を変更し，複数の物体検出モデルにおける歩行者検出精度を比較する．

本研究の貢献は以下の通りである．

- 交通事故に繋がるシナリオにおいて，DIVP シミュレータを用いて多様な環境条件での評価データを収集した．この評価データを用いることで，自動運転分野における物体検出モデルの正確な評価が可能となる．
- 様々な物体検出モデルに対して，網羅的な評価データによる歩行者検出精度を比較することで，検出傾向の差異を明らかにする．これにより，物体検出モデルごとに歩行者検出が困難なシーンを明確にすることが可能となる．

2. 関連研究

本節では，物体検出モデルおよび，自動運転における評価シーンについて述べる．

2.1. 物体検出モデル

物体検出は画像や動画，センサ情報から，特定の物体を識別し，位置を特定する技術である．対象物体のカテゴリや位置，個数まで検出できることから，自動車の安全支援だけでなく，医療や建設，製造業など様々な分野で応用されている．これら物体検出モデルの学習には機械学習や深層学習が活用され，高精度な物体検出を実現している．以下では，物体検出の中でも代表的な画像ベースの物体検出モデルに着目して詳細を述べる．

画像ベースの物体検出モデルは，アンカーベースとアンカーフリー，Transformer ベースの 3 つに分類することができる[8]．アンカーベースモデルは，入力画像をグリッド分割し，各セルの中心を基準点としてアンカーを配置することで，ア

1)　中部大学(487-0027 愛知県春日井市松本町 1200)
2)　*）講演者

ンカーを基準としたバウンディングボックスを推定する． アンカーベースモデルの代表例として，Faster R-CNN [9]や YOLOv5 が挙げられる．

Faster R-CNN は，物体検出モデルとして初めて入力から出力までを，単一モデルで学習可能としたモデルである．入力画像から物体候補領域を抽出する Region Proposal Network (RPN) と，RPN で抽出した物体領域が具体的に何かを識別する Fast R-CNN [10] で構成されている．これにより，検出速度と精度の向上を実現している．しかし，アンカーベースモデルは，アンカーの設定がハイパーパラメータであるため，スケールの異なる物体が検出困難な傾向がある．

YOLO [11] は，物体検出タスクにおいて入力画像から分類と物体位置の推定を同時に行う非常に高速なモデルである．YOLO には複数のバージョンがあり，学習データに対するデータ拡張などの学習テクニックや，検出精度と速度の向上が成されている． YOLOv5 は，モザイクデータ拡張を用いた学習により頑健性の向上を実現している．YOLOv5 には，バックボーンモデルのサイズが異なる YOLOv5n，YOLOv5s，YOLOv5m，YOLOv5l, YOLOv5x がある．YOLOv5n から YOLOv5x にかけてモデルサイズが大きくなり，大きいほど検出精度は高くや処理速度は遅くなる．

アンカーフリーモデルは， 入力画像から物体の上下左右の端をヒートマップとして予測することで，アンカーを必要としないアプローチである．一方で，アンカーベースモデルのアプローチと比較すると，汎化性能は低下することがある．アンカーフリーモデルの代表例として，CenterNet [12] や YOLOv8 等が挙げられる．

YOLOv8 は，バックボーンモデルや損失関数に着目した改良が施されており，YOLOv8 以前に提案された YOLO 手法と比較し，処理速度と検出精度ともに改善している．YOLOv5 と同様にバックボーンモデルのサイズが異なるモデルが公開されている．

自然言語や画像分類などにおいて高い性能を発揮している Transformer [13] を導入したアプローチが提案されている．Transformer をベースとしたモデルは，CNN をベースとしたモデルと異なり大域的な特徴抽出が可能なモデルである． しかし，アンカーベースモデルと比較すると Average Precision (AP) の精度は向上しているが，画像上の小さい物体に対する検出精度は低いことがある．Transformer ベースのモデルの代表例として，Detection Transformer (DETR) [14] や ViTDet [15] 等がある．DETR は， Transformer Encoder-Decoder 構造を採用し，Decoder 側で object query との Cross-attention を導入することで，物体ごとの効率的な分類と物体位置の推定を実現している．

2.2. 自動運転における評価シーン

自動運転システムの認知機能において，歩行者を正確に検出することが実環境における重大な事故の回避には求められる．そこで，交通環境における歩行者に着目し，歩行者の横断行動を解析した研究がいくつか報告されている [16, 17]．

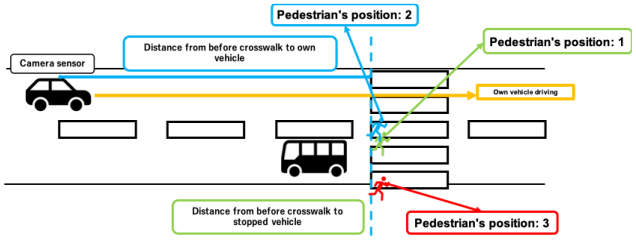


Fig.1 Contents of scenario creation

Zhang らは，歩行者がどのような状況で道路を横断行動するか，歩行者の横断行動に関する要因を調査している．双方向一車線道路の CG 環境を構築し，車道を走行する複数の車両間の車間距離に着目して評価している．また，どのタイミングで歩行者が道路を横断するかを走行中の車両間の距離から比較するため，横断行動時における車間距離と，歩行者が横断歩道を使用するかを機械学習により予測することで調査している．調査結果から，歩行者は車両を見逃すとリスクの高い行動を選択する傾向があり，歩行速度が速い歩行者ほど横断時に短い間隔を選ぶ傾向があることが確認されている．

Rasouli らは，歩行者の横断行動予測タスクにおいて，タスクを 3 つに分けて再定義し，歩行者の行動予測モデルの評価をしている．ここでタスクの再定義は，歩行者が行動する際の意図，歩行者の行動に起因するイベント，自動車に対する歩行者の行動によるリスクの 3 つである． この評価指標を様々な歩行者の行動予測モデルを評価することで，モデルごとの傾向を網羅的に解析するとともに，歩行者の行動に重要な要素を明確にしている．

これらの研究では，歩行者に着目した研究がなされているが，歩行者の検出から物体検出モデルの評価を行う研究はされていない．そこで，本研究では DIVP シミュレータを用いて道路を横断する歩行者を対象とした評価シナリオを作成した．評価シナリオは，歩行者，停車車両，天候，時刻に着目することで，様々な環境条件下における評価を可能とする．また，これらのデータを用いた評価を行うことで，複数の物体検出モデルを様々な条件で網羅的に評価する．

3. 評価シナリオの作成

本研究では，横断歩道を渡る歩行者に着目し，様々な物体検出モデルを評価する．DIVP シミュレータを用いて様々な環境条件下での評価データ作成し，物体検出モデルごとの歩行者の検出精度から比較する．ここで，評価シナリオは全て同一のシナリオであり，天候等の環境条件のみを変更して評価データを作成する．本節では，評価シナリオの作成概要について述べる．

Table.1 Environment conditions of evaluation scenario

Conditions	Contents
Weather	Sunny, Cloudy, Rainy
Time	14:00, 16:00
Pedestrian	adult male, elderly, child, adult male (umbrella), adult female (stroller)
Clothing color of pedestrians (adult male)	black, blue, gray
Stopped vehicles	Prius, Hiace

Fig.2 Example of a scene created in a CG environment.

3.1. 評価シナリオの作成概要

本研究では，評価データ作成にDIVPシミュレータを採用する．DIVPシミュレータは，実環境内で起こる物理現象を高品質に再現可能であり，天候やオブジェクト，時刻等のみを同一シーン下で変更可能なシミュレータである．本研究に用いる評価シナリオの例をFig.1に示す．

評価シナリオでは，横断歩道上に歩行者が存在する双方向一車線道路とし，道路を通行する車両を自車両とする．ここで，自車両に搭載した車載カメラ画像を物体検出モデルの入力画像とする．本評価シナリオでは，横断歩道手前に停車車両を配置することで，停車車両による歩行者の遮蔽に対する評価データを収集する．ここで歩行者の配置は，横断歩道上である歩行者が完全に隠れる位置/歩行者が部分的に隠れる位置，歩道上の3種類である．これにより，歩行者と停車車両間の遮蔽率に対する評価を可能とする．また自車両のスタート位置は，停車車両から50m離れた位置とし，自車両を10km/hで走行させる．これにより，自車両と歩行者間の距離に対する評価を可能とする．

3.2. 評価シナリオの作成条件

同一シナリオにおいて，環境条件のみ変更することで歩行者の検出精度を評価する．評価シナリオの作成にあたり，シナリオの各環境条件をTable.1に示す．変更する環境条件は，天候，時間帯，停車車両の車種，歩行者の4属性である．天候は，晴天/曇天/雨天の3種類とし，各天候での時間帯を14時/16時の2種類として評価シナリオを作成した．停車車両の車種により歩行者との遮蔽率が変わるため，停車車両はプリウスとハイエースの2種類とした．検出対象である歩行者の属性ごとでの評価を可能とするため，歩行者の属性は成人男性，子供，高齢者，傘をさした成人男性，ベビーカーを押す成人女性の5種類とした．ここで，成人男性は，服色が黒と青，灰色の3種類である．

評価シナリオの条件は，上記の組み合わせによる84通りであり，評価データの総数は24,480枚である．評価シナリオの例をFig.2に示す．

Fig.3 Training data collected in Tokyo coastal area.

4. 評価実験

本章では，3節で述べたDIVPシミュレータを用いて作成した評価データにもとづき，歩行者の検出精度から様々な物体検出モデルの傾向について述べる．本研究で対象とする物体検出モデルは，YOLOv5，YOLOv8，DETR，Faster R-CNNの4つである．物体検出モデルの評価指標としては以下の2通りである．

・距離と遮蔽率に着目した歩行者検出精度の比較
・様々な環境条件下での歩行者検出精度の比較

4.1. 物体検出モデルの学習データ

自動運転車に搭載する物体検出モデルは，実運用を見据えて学習には実環境データを用いることが一般的である．そこで本実験では，実環境データで学習された物体検出モデルを，DIVPシミュレータで作成したデータを用いて物体検出モデルを評価する．上記の理由に従い，本実験で用いる物体検出モデルの学習には，実世界の東京臨海部にて撮影された実環境データを用いる．学習に用いる実環境データは学習データが19,261枚，評価データが2,140枚である．学習に用いる実環境データの例をFig.3に示す．

4.2. 距離と遮蔽率に着目した歩行者検出精度の比較

本節では，遮蔽率または距離に着目し，物体検出モデルごとの歩行者に対する検出精度を用いて比較する．

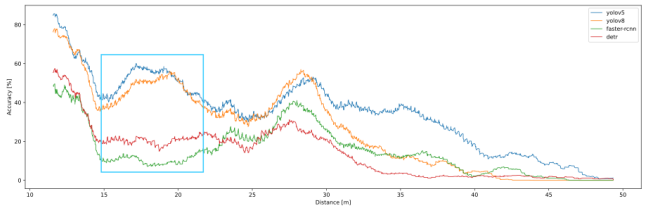


Fig.4　Comparison of Detection Accuracy by Object Detection Model by Distance

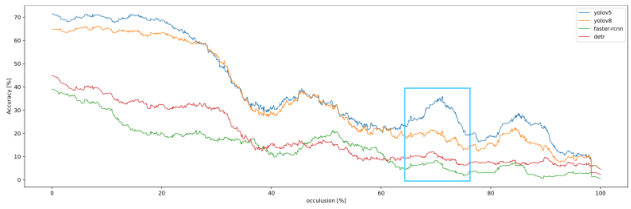
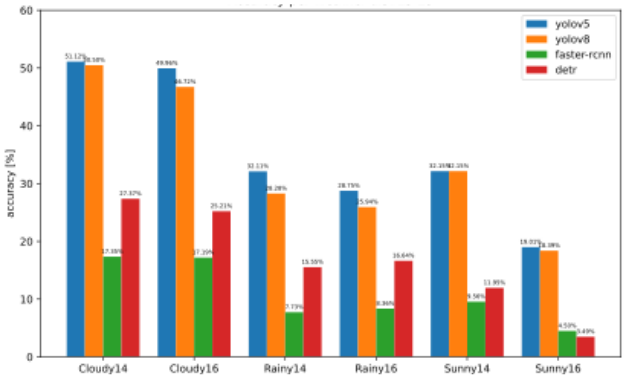
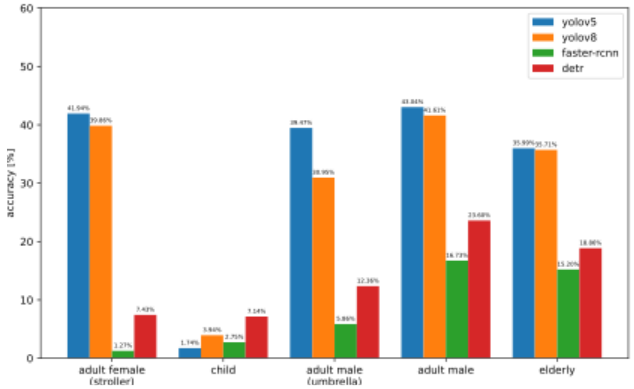


Fig.5　Comparison of Detection Accuracy by Object Detection Model by Occulusion Rate



(a)　weather/time

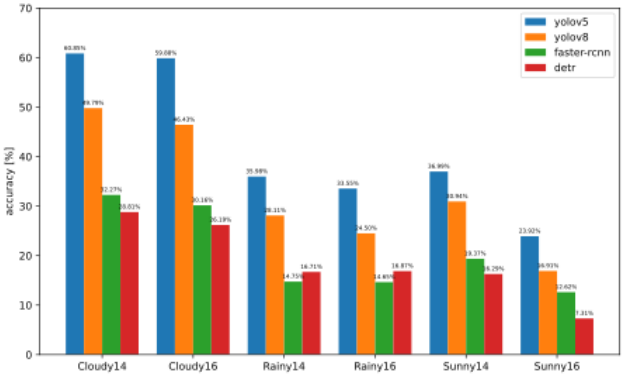


(b)　pedestrian

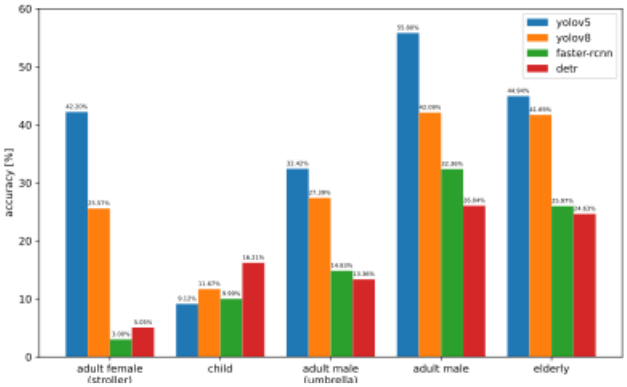
Fig.6　Accuracy comparison of object detection models (between distance 15 – 23 m)

4.2.1.　距離ごとの歩行者検出精度

自車両から歩行者までの距離に着目した物体検出モデルの検出精度を Fig.4 に示す．ここで，横軸は自車両から歩行者までの距離(m)，縦軸は歩行者に対する検出精度(%)である．Fig.4 から，距離が 15m から 23m 間では YOLOv5 と YOLOv8 の検出精度が向上しているのに対し，他モデルの検出精度は低下している．また YOLOv5 と YOLOv8 間では，距離が近いシーンではどちらも同様の検出精度であるが，距離が遠いシーンでは YOLOv5 が YOLOv8 より検出精度が高いことが確認できる．



(a)　weather/time



(b)　pedestrian

Fig.7　Accuracy comparison of object detection models (between occlusion rate 65 – 75%)

4.2.2.　遮蔽率ごとの歩行者検出精度

停車車両と歩行者の遮蔽率に着目し，物体検出モデルごとの歩行者に対する検出精度を Fig.5 に示す．横軸は停車車両と歩行者の遮蔽率(%)，縦軸は歩行者に対する検出精度(%)である．Fig.5 から，遮蔽率が約 70%のシーンにおいて YOLOv5 の検出精度が向上している一方で，他モデルは検出精度が低下していることがわかる．

4.3.　様々な環境条件下での歩行者検出精度の比較

本節では，3.2 節で述べた様々な環境条件に着目し，物体検出モデルごとの歩行者に対する検出精度用いて比較する．

4.3.1.　距離 15-23m 間における歩行者検出精度の比較

4.2.1 節で述べた結果に着目し，距離が 15m から 23m 間における様々な環境条件下での物体検出モデルごとの歩行者検出精度を Fig.6 に示す．ここで，Fig.6(a)は天候/時刻ごと，Fig.6(b)は歩行者属性ごとでの歩行者検出精度である．

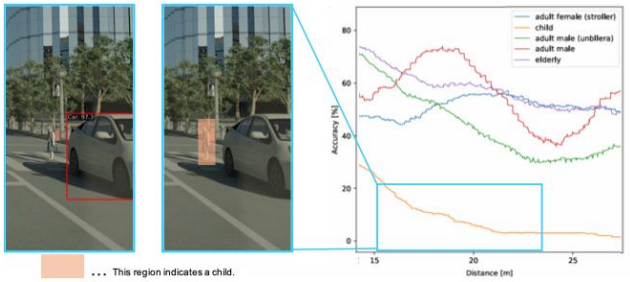


Fig.8　Example of detection results for child between distances 15 -23 m in YOLOv5

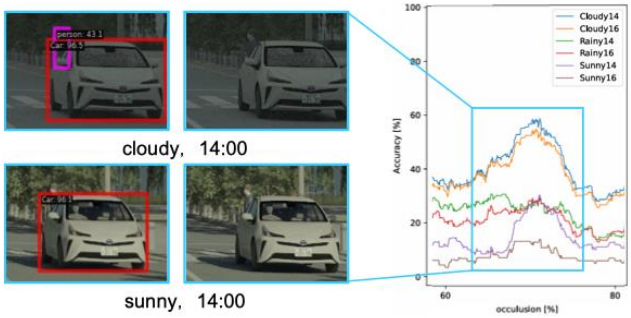


Fig.9　Example of detection results for cloudy, 14:00 between occlusion 65 -75 % in YOLOv5

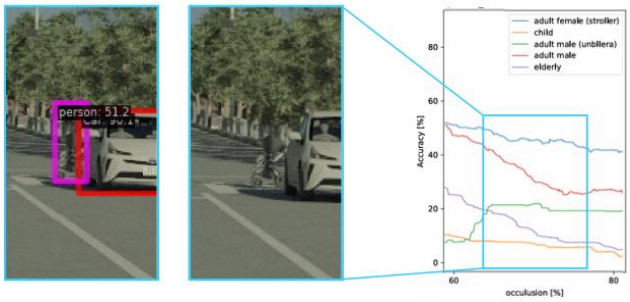


Fig.10　Example of detection results for adult female (stroller) between occlusion 65 -75 % in YOLOv5

Fig.6(a)から，各天候/時刻に対しては一貫して，YOLOv5 と YOLOv8 の検出精度が高く，他の物体検出モデルでは低いという 4.2.1 節で述べた傾向と同じであることが確認できる．一方で Fig.6(b)から，歩行者属性が子供の場合では，YOLOv5 の検出精度 1.74%と他モデルより低く，これは 4.2.1 節で述べた傾向とは異なる結果である．

そこで，距離が 15m から 23m 間における子供に対する YOLOv5 の検出結果例を Fig.8 に示す．Fig.8 のグラフから，子供の検出精度は他の歩行者属性に比べ 20%以下であり著しく低い．また可視化結果から，子供が停車車両に遮蔽していないシーンにおいても子供を検出できていないことが確認できる．これは，子供が他の歩行者属性に比べ画像上のスケールが非常に小さいため，アンカーベースの YOLOv5 では検出困難であったと考えられる．

4.3.2.　遮蔽率 65-75%間における歩行者検出精度の比較

4.2.2 節で述べた結果に着目し，遮蔽率が 65%から 75%間における様々な環境条件下での物体検出モデルごとの歩行者検出精度を Fig.7 に示す．ここで，Fig.7(a)は天候/時刻ごと，Fig.7(b)は歩行者の属性ごとでの歩行者検出精度である．Fig.7(a)に示す天候ラベルの 14 もしくは 16 は時刻を示す．

Fig.7(a)(b)から，歩行者属性が子供の場合を除き，YOLOv5 が最も検出精度が高く，これらは 4.2.2 節と同じ傾向である．ここで，歩行者属性が子供の場合において YOLOv5 の検出精度 9.12%と低い理由は，4.3.1 節で述べた理由と同様であると考えられる．また，天候が曇天のシーンや歩行者属性が成人男性とベビーカーを押す成人女性のシーンは，特に YOLOv5 の検出精度が他の物体検出モデルと比較し，著しく高いことが確認できる．

そこで，遮蔽率が 65%から 75%間における曇天の 14 時と晴天の 14 時に対する YOLOv5 の検出結果例を Fig.9 に示す．Fig.9 から，曇天の場合では歩行者と背景の色のエッジをとらえることが晴天の場合よりもわかりやすいため曇天のシーンの検出精度が向上したと考えられる．また，遮蔽率が 65%から 75%間におけるベビーカーを押す成人女性に対する YOLOv5 の検出結果例を Fig.10 に示す．Fig.10 のグラフから，ベビーカーを押す成人女性において検出精度が最も高い．また可視化結果から，停車車両がプリウスの場合，ベビーカーを押す成人女性は，成人女性の頭やベビーカーが遮蔽されてない．そのため，ベビーカーを押す成人女性は遮蔽率が高くても検出可能であったと考えられる．

この結果から，YOLOv5 は他のモデルに比べて遮蔽率が高いシーンにおいても歩行者の頭やベビーカー等が露出している場合，検出精度が向上することがわかった．しかし，歩行者が子供の場合は画像上のスケールが非常に小さいため遮蔽率の低いシーンにおいて検出精度は低下することがわかった．

5. まとめ

本研究では，画像ベースな物体検出モデルの網羅的な評価に向けて，DIVP シミュレータを用いて様々な環境条件における評価データを作成し，複数の画像ベースな物体検出モデルの評価に取り組んだ．評価から，自車両と歩行者間の距離が 15m から 23m 間では YOLov5 と YOLOv8 の検出精度が向上しているが，YOLOv5 は子供の検出はスケールが小さいことから困難であった．停車車両による歩行者の遮蔽率が 65%から 75%間のシーンにおいて YOLOv5 の検出精度が向上しており，特に成人男性やベビーカーを押す成人女性の検出精度は著しく向上した．これは歩行者の頭やベビーカーが停車車両から出ていたためであった．また，DIVP シミュレーションを用いた評価シナリオの作成では，晴天 16 時のみで画像のコントラストが高くなる点や，歩行者と背景物体の色情報が近いことで，歩行者が同化する点などを明確にできた．

今後の展望として，評価対象である物体検出モデルの増加や，環境条件ごとの詳細な解析に取り組むことで物体検出モデルごとの特性を明確にすることが挙げられる．

謝辞

本研究の一部は経済産業省の受託研究プロジェクトである「無人自動運転等のCASE対応に向けた実証・支援事業（自動運転技術（レベル3，4）に必要な認識技術等の研究）」において実施されたものである．

参 考 文 献

(1) P. Sun, et al., “Scalability in Perception for Autonomous Driving: Waymo Open Dataset.”, CVPR, 2020.
(2) M.Liao, et al., “DVI: Depth Guided Video Inpainting for Autonomous Driving.”, ECCV, 2020.
(3) F. Yu, et al.,“A2D2: Audi Autonomous Driving Dataset.”, CVPR, 2020.
(4) F. Yu, et al.,“BDD100K: A Diverse Driving Video Database with Scalable Annotation Tooling.”, arXiv, 2018.
(5) M. Cardts, et al., “The cityscapes dataset for semantic urban scene understanding.”, CVPR, 2016.
(6) A. Geiger, et al., “Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite.”, CVPR, 2012.
(7) 内閣府 戦略的イノベーションプログラム(SIP), “Driving intelligence validation platform,” 2019, https://divp.net/.
(8) W. Chen, et al., “2D and 3D object detection algorithms from images.”, Array, vol.19, p.100305, 2023.
(9) S. Ren, et al., “Faster R-CNN: Towards real-time object detection with region proposal network.”, NeurIPS, 2015.
(10) R. Girshick, et al., “Fast R-CNN.”, ICCV, 2015.
(11) J. Redmon, et al., “You Only Look Once: Unified, Real-Time Object Detection.”, CVPR, 2016.
(12) X. Zhou, et al., “CenterNet Heatmap Propagation for Real-time Video Object Detection.”, ECCV, 2020.
(13) A. Vaswani, et al.,“Attention is all you need.”, NeurIPS, 2017.
(14) N. Carion, et al., “End-to-End Object Detection with Transformers.”, ECCV, 2020.
(15) Y. Li, et al., “Exploring Plain Vision Transformer Backbones for Object Detection”, ECCV, 2022.
(16) C.Zhang, et al., “Predicting and Analyzing Pedestrian Crossing Behavior at Unsignalized Crossings.”, IV, 2024.
(17) A.Rasouil, et al.,“Diving Deeper Into Pedestrian Behavior Understanding: Intention Estimation, Action Prediction, and Event Risk Assessment.”, IV, 2024.