

深層学習ベースの物体検出手法を用いて車載カメラ画像から交通環境内の二輪車を検出する際の学習モデルの構築方法による検出性能の変化

中島 健介¹⁾ 佐久間 尚紀²⁾ 中安 淳也²⁾ 木谷 友哉³⁾

本研究では、自動車に車載された可視光カメラにて撮影した映像から、交通環境内の二輪車を深層学習による物体検出手法によって検出するとき、学習モデルの構築方法が検出精度にどのように寄与するかを明らかにする。物体検出手法はYOLO (v4-tiny)を用いた。教師データに用いる二輪車画像は、物体検出でよく使われるCOCOデータセット中の二輪車画像3085枚に加え、動画共有サイト上の動画や著者らの車両のドライブレコーダー映像の二輪車が映ったフレームからさらに5000枚程度用意した。これらの二輪車画像をもとに、次のポリシーに基づいて教師データを作成した。まず、車載カメラ映像中の二輪車を見つけることを想定し、実際の交通環境内に無い二輪車の画像を取り除いた。そして画像をアノテーションする際のクラスについて、1つのクラスとした場合、外観の異なる4車種にクラス分けをした場合、4つの撮影方向別にクラス分けをした場合を評価するためにそれぞれのデータセットを作成した。これらのデータセットを学習したモデルが、どの程度の距離から二輪車を検出できるかを評価した。その結果、実際の交通環境に近い二輪車の画像を学習用データとして利用することで検出距離が5~20m程度から10~25m程度に向上することがわかった。一方で、クラスを車種や撮影方向ごとに増やしても検出距離に変化はみられず、15~30m程度であった。また検出対象の動画をトリミングすることで、45m程度の検出距離となった。

キーワード: エレクトロニクス・制御, イメージ認識システム, 高度交通システム, 二輪車, 被視認性

1. はじめに

財団法人交通事故総合分析センター⁽¹⁾によると、2015~2019年の5年間で日本での二輪車搭乗者の死傷者数累計は約25万人である。そのうち対四輪車での事故によるものは90%以上である。そのような事故の種類は出会い頭、右折事故の順で多い。出会い頭では車両が急に出現することが多いと考えられるため、イメージングセンサで事前に対応可能なのは右直事故である。右直事故の91%は四輪ドライバーが右折、二輪車が直進で四輪車の過失が大きい。このような事故は表1に示すように、四輪車側の見落としが大きな原因である。以上より、二輪車と自動車との事故を減らすためには、自動車の運転者からの二輪車の視認性を上げることが重要であると考えられる。

一方で近年、機械を用いることで人の目では視認しにくいものを検出し、運転者のサポートを行う先進運転支援システム: ADAS (Advanced Driver-Assistance Systems) の普及が進んでいる。しかし、現状として歩行者や自転車などの検出を行う研究は多いが二輪車を対象としたものは少ない。そこで私たちは、二輪車の検出を通じて、どのようにすれば二輪車と自動車の交通事故を減らすことができるかを研究した。本研究では、可視光カメラ映像中の二輪車を検出する。その際、物体検出モデルの学習に用いる画像に対して独自のアノテーションを行い、検出性能が向上するかを検証する。

1) 静岡大学大学院総合科学技術研究科情報学専攻 (432-8011, 静岡県浜松市中区城北3-5-1)

2) 静岡大学情報学部

3) 静岡大学大学院情報学領域 (E-mail:t-kitani@kitanilab.org)

Table. 1 二輪車の右折時事故の人的要因⁽¹⁾

	右折時
発見の遅れ	87%
判断の誤り	12%

2. 関連研究

2.1. 対向車の運転者からの二輪車の視認性に関する関連研究

文献⁽²⁾では、運転者が事故に関連する対象物を交通環境内で検知できたかを主に夜間の事故状況を再現して多数ケーススタディしている。それらの中には相手が二輪車を想定した研究が行なわれており、接近する二輪車のヘッドライトが電柱などで遮られるときにハロを起こしたり1つのヘッドライトが2つに見えたりするなどの状況が報告されている。文献⁽³⁾では、二輪車と四輪車のヘッドライトの被視認性について6台の二輪車を用いた検証を行なっている。四輪車と二輪車のヘッドライトの被視認性を定量化した結果、二輪車の方が著しく低いことが示されている。また、二輪車に昼間走行灯(デイライト)を装着することで知覚距離を有意に伸ばせることも示している。文献⁽⁴⁾では、二輪車の多くのヘッドライトは1灯式で四輪車に比べて幅が狭く被視認性が悪いため、9台の二輪車のヘッドライトのビームパターンを測定して、どの程度の範囲を照らせるのかを検証している。文献⁽⁵⁾では、夜間の衝突事故を想定した時の運転者の応答の評価方法について述べられている。この研究結果によると、四輪車と二輪車を比較するとヘッドライトの照度とドライバーの応答距離が著しく低いことがわかる。

2.2. YOLOを用いた道路上の物体検出に関する研究

文献⁽⁶⁾では、交差点を撮影した動画からYOLOを用いて交通量と車両走行軌跡の計測を行っている。交差点付近の電柱

上部に設置されたカメラで撮影した動画画像を利用し、1 フレームごとに YOLO での処理及びトラッキングを行うアルゴリズムを構築することで交通量と車両走行軌跡を計測できることを確認した。車種別に計測された車両に二輪車も含まれているが、二輪車だけに特化したチューニングはされていない。

2.3. 学習クラスの細分化に関する研究

文献⁽⁷⁾では、サッカーにおけるリアルタイム選手識別のため、YOLO を用いてサッカー選手の背番号を認識する手法が提案されている。認識するクラスは1番から11番までの11クラスとし、画像枚数は各クラス約400枚、合計約4000枚で学習を行い、1番から11番まで合計363枚の画像で評価を行った。予測の信頼度の閾値を0.3にすることで画像1枚あたり約0.29秒、F値0.781の精度で識別が可能であり、識別クラスの細分化による有効性を示した。

3. 二輪車検知精度のための学習データのクラス分け設計手法

3.1. 物体検出手法 YOLO

YOLO とは You Only Look Once の略で、深層学習を利用した物体検出手法である。YOLO は深層学習フレームワーク darknet を使用して実装されており、これらは GitHub* 上でオープンソースとして公開されている。また同サイト上では COCO データセット[†]で学習済みの物体検出モデルが提供されている。YOLO の特徴は処理が高速な点で、「検出」と「識別」を同時に行っている。以下、YOLO の検出の仕組みについて⁽⁸⁾より引用する。図1のように、YOLO はまず入力画像を $S \times S$ のグリッドに分割する(図中左)。次に、それぞれのグリッドに対し B 個のバウンディングボックスと信頼度を推測する(図中中央上)。推測したバウンディングボックスに物体が含まれている確率を $Pr(Object)$ 、推測したバウンディングボックスと正解データとの重なり具合を IOU_{pred}^{truth} とすると、信頼度は式(1)で表される。

$$Pr(Object) \times IOU_{pred}^{truth} \quad (1)$$

同時に、それぞれのグリッドで C 個のクラスについて条件付きクラス確率を推測する(図中中央下)。条件付きクラス確率は式(2)で表される。

$$Pr(Class_i | Object) \quad (2)$$

最後に、式(1)と式(2)を掛け合わせることで、最終的な信頼スコアを得る(図中右)。

$$Pr(Class_i | Object) \times Pr(Object) \times IOU_{pred}^{truth} \quad (3)$$

$$= Pr(Class_i) \times IOU_{pred}^{truth} \quad (4)$$

この信頼スコアは、各バウンディングボックスのクラスの予測確率とバウンディングボックスがどれだけ正確に物体を囲っているかを表す。

YOLO は各クラスの平均精度 (Average Precision) の全クラスでの平均である mAP (mean Average Precision) や処理速

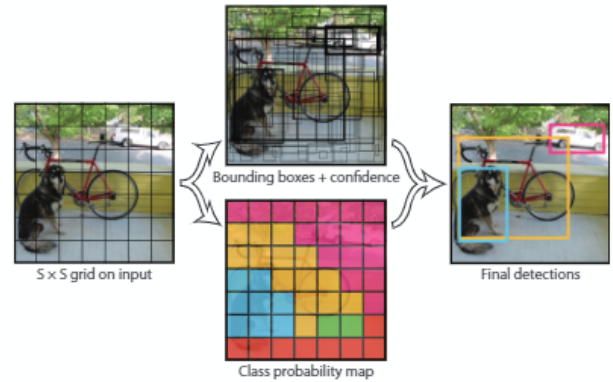


Fig.1 YOLO における検出の仕組み (文献⁽⁸⁾より引用)

度が優れている。改良版である YOLOv4⁽⁹⁾ は Tesla V100 を用いた COCO データセットによる評価で IOU 閾値を 0.5 としたとき、mAP は 62.8、FPS は 96 である。これは近い時期に発表された EfficientDet⁽¹⁰⁾ の mAP=52.2、FPS=62.5 や HardNet⁽¹¹⁾ の mAP=54.3、FPS=41.5 などと比べ、高精度で高速な処理が可能な手法である。本研究の最終目的は、独自に学習したモデルを用いて車載カメラで撮影された道路上の映像をリアルタイムに処理することであるので、物体検出アルゴリズムとして YOLO を採用するのが適していると考えた。

3.2. 学習データのクラス分けの提案

本研究では、独自に作成した二輪車画像データセットを用いて YOLO のアルゴリズムによる物体検出モデルを学習する手法を提案する。物体検出手法 YOLO では、Microsoft 社など数社が提供する COCO データセットという機械学習用に画像内の物体に種類ごとのラベルを付けたデータセットによって学習されており、80 種の物体、つまり 80 クラスを検出するように学習されている。そこには“motorcycle”のクラスも含まれているが、実際に画像を確認すると、図2に示すような建物の屋上から俯瞰で二輪車を撮影した画像や二輪車のミラーのみを撮影した画像など、交通において他車両から見られる二輪車とはいえない画像も含まれていた。このように写る二輪車の姿は、車両に搭載される可視光カメラ映像に対して物体検出することを考慮すると、対象のモデルでは学習する必要はないと考えた。

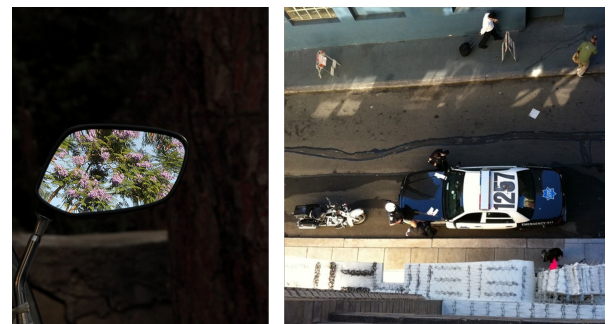


Fig.2 除外する画像の例 (COCO データセットより抜粋)

そこで提案手法では、まず COCO データセット中の画像で二輪車のラベル“motorcycle”が付与された画像に対し、前述の

*<https://github.com/AlexeyAB/darknet>

†<https://cocodataset.org/>

学習に利用しない姿で二輪車が写る画像を排除したデータセットを作成した。さらにそこから“motorcycle”クラスを二輪車の車種及び撮影方向ごとに区分してクラス分けしたデータセットを作成した。これらのデータセットを用いて物体検出モデルの学習を行なうことで、認識精度を向上させられるとの仮説を立て検証を行なう。なお、本研究においては二輪車に対する検出のみを扱うこととし、データセットは自転車や車のような二輪車でない物体が写る画像を含まない。

3.3. 提案手法の検出精度に対する有効性

3.2. 節で述べた手法で作成したデータセットを学習することで、二輪車検出精度が向上するかを検証した。提案手法により、COCO データセットから精査画像データセットと7車種別データセットを作成した。後述する3.4. 節の車種クラスとは異なり、fullcowl, naked, american, offroad, big scooter, scooter, part of motorcycle の7クラスを設定した。

“motorcycle”のクラスに対する評価の結果を表2に示す。ただし、7車種別データセットは二輪車が複数のクラスに分けられているため、mAPの値となっている。画像を精査したデータセットはCOCOデータセットに比べ、適合率は67%から74%に、再現率は53%から65%に向上した。また、そこから7車種別に分類したデータセットを学習したモデルは、再現率は65%から64%へ下がったものの、適合率は74%から79%に向上した。以上より、学習データセットについて、詳細

Table. 2 モデル評価結果

学習したデータセット	適合率 [%]	再現率 [%]	F 値	AP [%]
COCO データセット	67	53	0.59	55
精査画像データセット	74	65	0.69	72
7車種別データセット	79	64	0.70	81*

(*)この値はmAP

なクラス分けをすることで二輪車の検出精度を高められることが確認できた。本研究では、これらの車種別のクラス分けに加え、学習データセットのクラスを二輪車の方向別に分けたデータセットを作成して、検出距離を大きくすることができるか検証する。

3.4. 提案手法により作成したデータセットについて

YOLOにおける学習画像数は一般に1クラス当たり2000枚程度が目安[‡]とされている。そこで、各クラスの学習用画像枚数として2000枚程度を確保できるように努めた。収集した画像に対するアノテーションは、全て手作業で実施した。本研究では以下の6種類のデータセットを作成した。

motorcycle-coco データセット (2842 枚)：COCO データセットから“motorcycle”を含む画像を抜き出し、前述の学習に利用しない姿で二輪車が写る画像を排除したデータセット

motorcycle-5000 データセット (5810 枚)：YouTube上の車載動画や自前で撮影した動画において、二輪車が映るフレームを学習用画像にしたデータセット

motorcycle-3000 データセット (2842 枚)：後の評価のため、motorcycle-5000の画像枚数をmotorcycle-cocoと合わせたデータセット

motorcycle-8000 データセット (8652 枚)：motorcycle-cocoとmotorcycle-5000の画像を合わせたデータセット

車種別データセット (8652 枚)：motorcycle-8000に対し、図3上段で示す4車種ごとへ二輪車をクラス分けしたデータセット

方向別データセット (8652 枚)：motorcycle-8000に対し、図3下段で示す4方向ごとへ二輪車をクラス分けしたデータセット

また車種別・方向別データセットで設定したクラスと、画像に付与される各クラスのラベル数は以下の通りである。

- 車種別データセット
 - moto_cowl (2186): カウルで車体が覆われた車両
 - moto_adv (1086): アドベンチャーやオフロードタイプの車両
 - moto_naked (4293): カウルで車体が覆われていない車両
 - moto_scooter (5115): 原付スクーターの車両
- 方向別データセット
 - moto_Front (2920): 二輪車の前面が写っている画像
 - moto_Right (2799): 二輪車の右側が写っている画像
 - moto_Back (3561): 二輪車の背面が写っている画像
 - moto_Light (3400): 二輪車の左側が写っている画像

また各データセットに含まれる画像を学習・評価用に8:2の割合で分割した。車種別及び方向別のデータセットについて、どのような画像が該当するのかを図3に例示する。

4. 提案手法における二輪車検出距離評価実験

4.1. 実験の目的

本実験の目的は、一般的な交通状況下を想定した環境で、車載可視光カメラ映像に対する物体認識技術を用いた二輪車検出の検出距離や適合率・再現率を定量的に評価することである。

4.2. 実験方法

一般財団法人日本自動車研究所 (JARI) の特異環境試験場[§]にて、自動車に車載された可視光カメラによる二輪車撮影実験を行った。

(1) 使用機材

実験で使用した機材を表3に示す。車載カメラは図4に示すようにフロントバンパーとフロントガラス上に設置し、フロントバンパーには地面から約85cmでフロント先端部に、フロントガラス上には地面から約150cmの位置に取り付けた。検証の結果、フロントガラスのカメラの方が多少良い性能を示したが全体の傾向に大きな差がなかったことと、小型の自動車に車載することも考慮して、以降はフロントバンパーのカメラで撮影された結果について取り扱う。

[‡]<https://github.com/AlexeyAB/darknet>

[§]<https://www.jari.or.jp/test-courses/jtown/test-path/special-environment-area/>

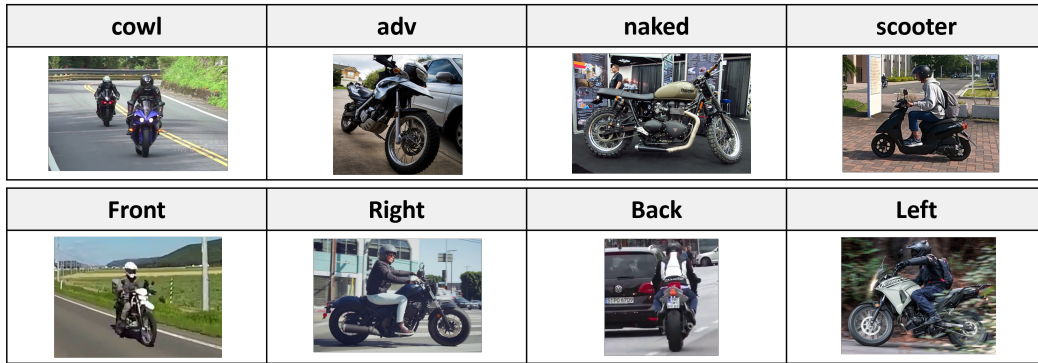


Fig.3 各クラスに属する二輪車画像の例

Table. 3 実験で使った機材

機材の種類	用途	製品名	備考
可視光カメラ	二輪車の撮影	GoPro HERO9 Black	60FPS, 1080p
可視光カメラ	メジャーの撮影	insta360 ONE X2	50FPS, 1440p
自動車	カメラを搭載	TOYOTA VOXY	-
二輪車	撮影対象	SUZUKI V-Strom250	-



Fig.4 自動車上のカメラ設置位置と二輪車

(2) 実験条件

実験環境として雨、霧、日照等の環境条件を再現可能な屋内施設の「特異環境試験場」、様々な交差点形状を再現可能な屋外施設の「多目的市街路」にて実験を行った。環境条件として以下の4つの条件を設定した。

- 屋内: 特異環境試験場内において、天井照明を点灯した環境
- 霧: 特異環境試験場内において、天井照明を点灯し、さらに霧を発生させた環境で視程は約30m
- 日照: 特異環境試験場内において、天井照明を点灯し、さらに撮影車両と同一車線上を日照装置で照らした環境
- 屋外: 多目的市外路において、天気は晴天の環境

このような環境において、四輪車に搭載された可視光カメラでそれぞれ4つの構図における二輪車を撮影した。撮影した構図は図5に示すように、二輪車が四輪車と同じ車線の前方で停止している構図「追従」、二輪車が四輪車の対向車線で停止している構図「対向」、交差点において四輪車の右手側で二輪車が停止している構図「右から」、交差点において四輪車の左手側で二輪車が停止している構図「左から」の4つである。

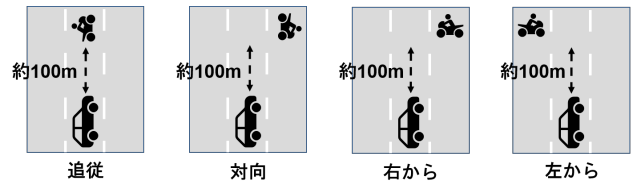


Fig.5 撮影構図

(3) 実験手順

撮影手順の概要を示す。自動車から100m離れたところにライダーが乗った状態で静止した二輪車に時速10km弱で近づいていき、追従・対向の構図では安全性を考慮し二輪車と自動車の距離が10mとなった時点で停止、右・左からの構図では二輪車の横を通過し二輪車が映らなくなったところで停止する。そこからバックによって元の地点に戻る。これと同時に二輪車と自動車の距離を計測するため、自動車運転席側の扉付近に設置したカメラによって、地面に敷かれたメジャーを撮影した。また、撮影した動画の各フレームに後処理で時刻を付与するため、各構図の開始、終了時に可視光カメラには時計の時刻を映した。二輪車のヘッドライトは常に点灯し、自動車のヘッドライトは消灯させた。カメラで撮影した動画の各フレームの時刻付与は次のように行った。動画の時刻付与では、まず可視光カメラで撮影した動画の開始、終了時に映り込ませた時計の時刻から経過時間を算出した。そして動画のフレームレート (fps: frames per second) でその経過時間を除算、1フレームごとの経過時間を計算し開始フレームから終了フレームまでのすべてのフレームに時刻の付与を行った。自動車側方の道路沿いのメジャーを撮影した動画からの距離算出は次のように行った。まず、メジャー動画から目視で5m, 10m, ..., と5mおきにフレーム番号を確認し、メジャーの距離をそれぞれの区間のフレーム数で除算して1フレームあたりに自動車が進んだ距離を

算出する。それをもとに 5m の区間では自動車は等速であると仮定し線形補間することで距離を算出した。5m ごとのフレームの確認は目視で行ったので有効数字は 1m とする。撮影した動画に対し、算出した各フレームの時刻、二輪車と自動車の距離とその動画の撮影フレームレートをプログラムを用いて書き込んだ。

4.3. 評価方法

(1) 検出距離の判定基準

今回の評価においては、各物体検出モデルが出力する信頼スコアが 0.25 以上であれば二輪車が検出されたとみなし、検出距離については誤検出を抑制するため最初に 9 フレーム連続で二輪車が検出されたときの 1 フレーム目における二輪車と四輪車の距離を検出距離とする。ドライブレコーダでは LED 式の信号と同期して信号が映らなくなることを防止するために、商用電源の周波数である 50Hz や 60Hz と同期しないように 27.5fps のフレームが使われることが多い。9 フレームはここでは約 0.3 秒に相当し、時速 50km では約 4.5m 進む距離である。

(2) 使用した物体検出モデル

物体検出手法 YOLOv4 は計算量が大きく、一般に高性能なグラフィックボードを搭載した PC で評価されることが多い。今回の評価においては後に車載されることを想定し、軽量版である YOLOv4-tiny を利用して 3.4 節において作成した 6 種類のデータセットに加え、COCO データセットを合わせた合計 7 種類のデータセットをそれぞれ学習したモデルを評価した。

4.4. 実験結果

二輪車に対する検出結果について、フロントバンパーのカメラ映像に検出をかけた結果を表 4 に示す。数値が記入されていない箇所については、9 フレーム連続での二輪車検出が無かったことを意味する。

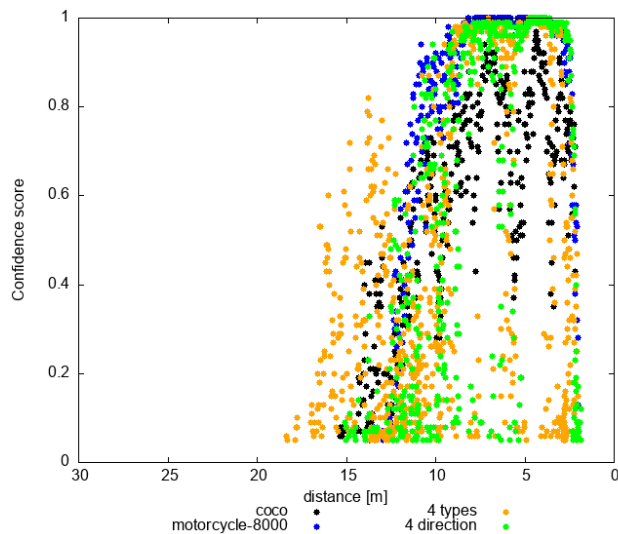


Fig.6 対向の構図における学習モデル別の二輪車検出-1

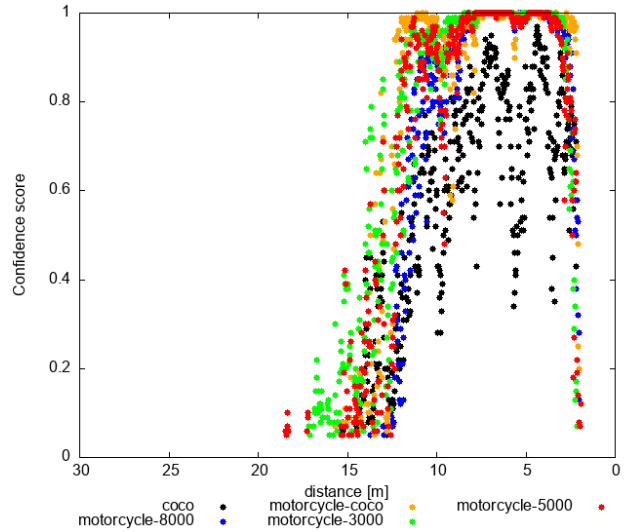


Fig.7 対向の構図における学習モデル別の二輪車検出-2

また、二輪車との距離ごとに各モデルが出力した信頼スコアについて、屋内における対向の構図での結果を、図 6, 7 に示す。なお両図において二輪車との距離が 2m 以下になったときに急激に信頼スコアが低下している理由は、自動車と二輪車の距離が近くなりすぎて隣の車線の対向車である二輪車がカメラの画角から外れたためである。

図 6 では、COCO データセットのオリジナルの "motorcycle" クラス、単一クラスの motorcycle-8000、その約 8000 枚を 4 車種に分けた学習モデル、同じく 4 方向に分けた学習モデルの評価結果を示している。特徴的であるのは、4 types の検出が早いこと、motorcycle-8000 以外の 3 つのモデルが 5m 程度の距離の時に一度信頼性スコアが下がっていることが上げられる。COCO データセットでは交通環境内の二輪車以外の構図の写真が一部混ざってしまっていたことが信頼性スコアが安定しないことにつながったと考えられる。4 types や 4 directions では、それぞれ 4 つのクラスを持つため、途中で当初認識したクラスと異なるクラスと判定したために、クラスの変り目において一時的に信頼スコアが下がっている。とくに 4 directions では、当初カメラは対向二輪車の正面を捉えていたものが、二輪車に近づくに連れて左側車線から二輪車の右側方を見る位置関係に変化するため、その影響が出たものと考えられる。ここから分かることは、学習に使う二輪車画像は対象環境で想定している構図に合ったものを選ぶのがよいこと、適切にクラス分けをすると検出距離を長くできる可能性があるが、クラスの切り替わり時に一時的に信頼性スコアの低下を招くと言うことである。後者については、クラスを特定できなくとも広い意味で二輪車に関連するクラスとして認識していれば検出と判断するようなシステムのロジックが必要となる。

図 7 では、COCO データセットのオリジナルの "motorcycle" クラス、その "motorcycle" から不適切な画像を取り除いたもの、それらに YouTube から切り出した交通環境内の二輪車の画像を増やしたものでの評価結果である。いずれも二輪車の単一クラスである。特徴的であるのは、COCO データセットが一時的に信頼性スコアの低下を起しているのに対し、他の二輪車

Table. 4 検出距離 [m] (四輪車下部の可視光カメラ映像にて二輪車を対象)

構図	環境条件	COCO	motorcycle -coco	motorcycle -3000	motorcycle -5000	motorcycle -8000	4車種 (4 types)	4方向 (4 directions)
追従	屋内	-	-	16	16	-	16	-
	霧	-	14	20	20	19	16	17
	日照	-	-	-	-	-	-	-
	屋外	-	13	18	16	16	16	16
対向	屋内	13	12	14	12	11	16	10
	霧	7	8	9	8	5	1	5
	日照	11	12	12	13	12	16	11
	屋外	16	16	18	20	16	16	16
右から	屋内	16	21	25	28	24	28	24
	霧	-	-	-	-	-	-	-
	日照	6	21	24	28	24	28	25
	屋外	24	24	28	30	26	26	23
左から	屋内	18	26	24	25	23	25	25
	霧	-	12	11	10	12	10	11
	日照	7	14	12	12	14	12	12
	屋外	18	18	34	31	31	31	18

単一クラスのモデルでは一度信頼性スコアが高くなると落ちていないことである。仮説ではクラスを細分化することが検出精度の向上につながると考え、特に相対する方向（撮影方向）によるタイヤ数の見え方が大きく影響すると予想した。しかし、学習用データ枚数を十分用意し、移動体として対象の見え方や撮影角度が変化するような場合で評価実験を行うと、適切な二輪車の写真を使って二輪車は単一クラスにまとめた方が検出が安定するという結果が得られた。

4.5. 考察

(1) 学習に使用したデータセットによる検出結果の比較

COCO と motorcycle-coco：表4より、前者のデータセットを学習したモデルに対し後者のデータセットを学習したモデルの方が検出距離が長い傾向があることがわかる。また前者では検出距離の判定基準を満たせなかった条件でも、後者では基準を満たせられるようになっている。さらに図7のように、前者よりも後者の方が高い信頼スコアで二輪車が検出できており、3.2.節で述べた二輪車画像の精査が二輪車検出距離の向上に寄与したと考えられる。

motorcycle-coco と motorcycle-3000：データセットに含まれる二輪車画像について、画像枚数は等しいが、前者はランダム、後者は交通環境内の画像を集中的に集めている。表4より、前者のデータセットを学習したモデルに対し後者のデータセットを学習したモデルの方が検出距離が長い傾向があることから、二輪車検出に用いる物体検出モデルの学習には実際の交通環境内の二輪車画像を用いる方が二輪車検出距離の向上に寄与すると考えられる。

motorcycle-3000, motorcycle-5000, motorcycle-8000：表4からは、データセット中の二輪車画像の増加に伴って検出距離が向上したとはいえない。また図7においても距離

ごとの信頼スコアの推移には大きな変化は見られないことから、画像枚数の増加による二輪車検出距離の向上には限度があると考えられる。

motorcycle-8000, 車種別, 方向別：motorcycle-8000 は“motorcycle”の単一クラスのみからなるデータセットである。これと車種別データセット及び方向別データセットを比較すると、図6に示すように、一部の条件では車種別データセットを学習したモデルが他のモデルより遠い距離から二輪車を検出することはあるものの、表4からわかるように車種別・方向別のモデルで検出距離は全体的に向上しておらず、詳細なクラス分けが二輪車検出距離の向上に寄与したとはいえない。

(2) 9フレーム連続で二輪車検出ができなかった実験環境の条件

表4の結果から、日照の環境条件下の追従の構図及び霧の環境条件下の右からの構図において、9フレーム連続での二輪車検出ができていないことが確認できる。前者については、図8に示すように、日照装置と二輪車が同一車線にあることで二輪車の姿が見えづらくなったことが原因だと考えられる。

また後者については図9で示すように、右からの構図は他の構図と比較すると1車線分カメラから遠い位置に二輪車が存在するため、カメラで撮影する際により強く霧の影響を受けることが考えられる。

4.6. 対象動画のトリミングによる検出距離の向上

表4から、二輪車の検出距離は高々30m程度であることがわかる。時速50kmで走行する車の停止距離が33mであることを考慮すると、より長い距離で二輪車を検出することが望ましい。そこで、検出距離向上のための追加検証を実施した。

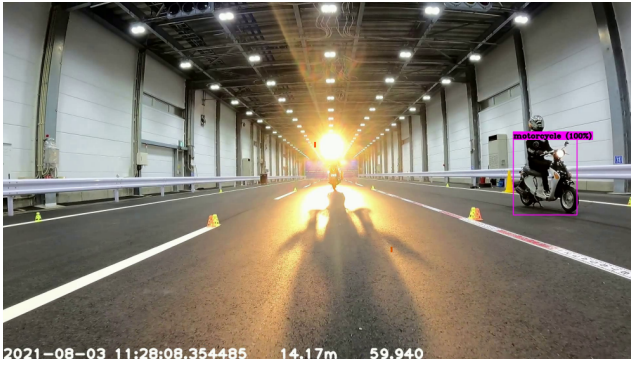


Fig.8 日照装置によるカメラ映像への影響

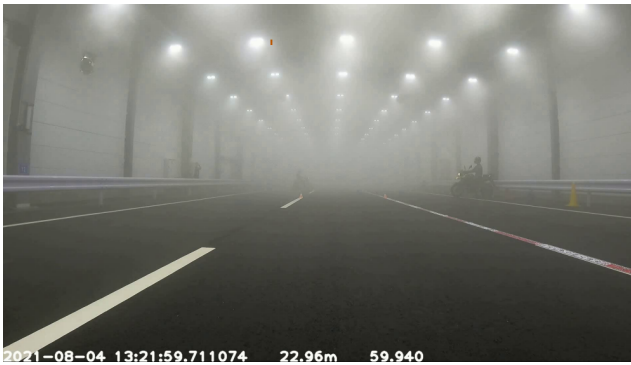


Fig.9 霧によるカメラ映像への影響

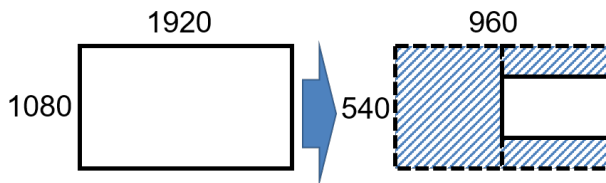


Fig.10 動画に対するトリミング (右からの構図の場合)

物体検出手法 YOLO では映像全体が一定の大きさにリサイズされて処理される。今回撮影した 1920×1080 ピクセルの映像も YOLO 内部では一旦 416×416 ピクセルの映像に変換された後に物体認識処理が行われる。そのため、映像内での対象物の大きさが相対的に小さすぎる場合は、検出対象動画をトリミングして、対象物を拡大することで検出距離を伸ばせる可能性があると考えた。動画のトリミングは図 10 に示すように 1920×1080 のカメラ映像に対して、検出対象の二輪車が動画に残るように縦横それぞれ半分の大きさにトリミングを実施した。

トリミングした動画に対する二輪車検出を実施した際の結果を表 5 に示す。表 4 と比較すると、日照での追従と霧での右からを除く全ての条件で検出距離が向上したことがわかる。また単一クラスのデータセットと車種別・方向別のデータセットをそれぞれ比較すると、ともに検出距離が向上している傾向がある。提案手法による詳細なクラス分けは、それ単体では効果的ではないが、実験動画のトリミングのような別の検出距離向上をもたらす手法と併せて適用することで有効な手法となると考えられる。

Table. 5 トリミング前後における検出距離 (前 / 後) [m] (四輪車下部の可視光カメラ映像にて二輪車を対象)

構図	環境条件	motorcycle -8000	4 車種 (4 types)	4 方向 (4 directions)
追従	屋内	- / 25	16 / 31	- / 23
	霧	19 / 33	16 / 26	17 / 32
	日照	- / -	- / -	- / -
	屋外	16 / 34	16 / 35	16 / 29
対向	屋内	11 / 18	16 / 27	10 / 20
	霧	5 / 17	1 / 15	5 / 14
	日照	12 / 19	16 / 27	11 / 21
	屋外	16 / 31	16 / 31	16 / 31
右から	屋内	24 / 50	28 / 54	24 / 43
	霧	- / -	- / -	- / -
	日照	24 / 43	28 / 54	25 / 44
	屋外	26 / 55	26 / 57	23 / 56
左から	屋内	23 / 44	25 / 44	25 / 57
	霧	12 / 18	10 / 17	11 / 18
	日照	14 / 33	12 / 40	12 / 48
	屋外	31 / 49	31 / 49	18 / 49

4.7. 他手法との比較

二輪車を検出するその他の手法として、画像内の人物を骨格推定ライブラリの OpenPose⁽¹²⁾ を用いて認識することで、二輪車ライダーの検出ができるかについても、実映像に適用して評価を行った。二輪車ライダーの骨格を正確に推定可能な距離は 10~15m 程度であり、物体検出と比較すると検出距離が小さい。また計算量が大きいため、本研究で用いた映像をリアルタイム以上の速度で処理することができず、事故防止の観点で車載して利用することは現実的ではなかった。

5. まとめ

本論文では、二輪車画像から提案手法によって 6 種類のデータセットを作成し、これらのデータセットと COCO データセットを学習したモデルに対して各モデルの特異環境試験場および多目的市街地で撮影した動画を用いた検出距離の評価を行った。学習モデルの比較では、汎用的な性質を持つ COCO データセットベースのものよりも、実際の交通環境の二輪車画像を元にしたデータセットの方が良好な検出性能を示した。また、学習に用いる画像枚数の増加による検出性能の向上には限度があり、学習に用いる画像は 1 クラスあたり 3000 枚程度で十分であることが示された。一方で、車種別・方向別に二輪車のクラスを細分化したデータセットを学習したモデルは、単一クラスのデータセットを学習したモデルと比較すると検出性能が向上したとはいえない。現在市販されている機材やオープンソースの物体認識ソフトウェアを用いて二輪車を検出することができる距離は高々 30m 程度であるが、機材の性能や技術が今後向上していくことでより長い距離での検出が可能になると考えられる。トリミングされた動画における二輪車検出では車

種別・方向別データセットを学習したモデルの方がより良好な検出性能を示したことから、他の検出性能向上をもたらす手法と併せてクラスの細分化を適用するのが有効的であると考えられる。

References

- (1) 堤陽次郎, “夜間における四輪車と二輪車の右直事故の特徴分析,” (公財) 交通事故総合分析センター 第 23 回研究発表会, 2020.
- (2) E. Klein and G. Stephens: “Visibility Study - Methodologies and Reconstruction”, Future Transportation Technology Conference & Exposition, SAE International, pp. 1–19, Aug. 1992.
- (3) N. Haferkemper, J.H. Sprute, C. Schiller, and T.Q. Khanh, “The Effect of LED Daytime Running Lights on Motorcycle Conspicuity and Detection Distance Estimation - Latest Test Results”, SAE 2010 World Congress & Exhibition, pp. 1–9, Apr. 2010.
- (4) G. de Paula Brito, J.A.F. Borges, and G.F.G. Nogueira: “Analysis of Blind Areas in Different Categories of Vehicles Considering the Evolution of Projects and Legislation”, 23rd SAE Brasil International Congress and Display, pp. 1–8, Sep. 2014.
- (5) W.T. Neale, N. Mckelvey, D. Pentecost, and D. Koch: “Motorcycle Headlamp Distribution Comparison”, WCX World Congress Experience, SAE International, pp. 1–25, Apr. 2018.
- (6) 中島 浩平, 岩崎 洋一郎, 竹原 洋志, 宮田 俊彦, 倉本 俊昌, 北島 俊孝, 瀬戸口 恵: 物体検出アルゴリズム YOLOv3 を用いた交通量と車両走行軌跡の自動計測, 第 72 回電気・情報関係学会九州支部連合大会講演論文集, Vol. 2019, pp. 61–62.
- (7) 梶原 弘一郎, 大城 英裕, 行天 啓二, 高見 利也: 背番号認識への物体認識型深層学習の適用, 第 71 回電気・情報関係学会九州支部連合大会講演論文集, Vol. 2018, pp. 165–166.
- (8) Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, “You Only Look Once: Unified, Real-Time Object Detection”, arXiv:1506.02640 [cs.CV], pp. 1–10, May 2016.
- (9) Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection”, arXiv:2004.10934 [cs.CV], pp. 1–17, Apr. 2020.
- (10) Mingxing Tan, Ruoming Pang, Quoc V. Le, “Efficient-Det: Scalable and Efficient Object Detection”, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10781–10790, 2020.
- (11) Ping Chao, Chao-Yang Kao, Yu-Shan Ruan, Chien-Hsiang, Huang, and Youn-Long Lin, “HarDNet: A low memory traffic network”, Proceedings of the IEEE

International Conference on Computer Vision (ICCV), pp. 3552–3561, 2019.

- (12) Tomas Shimon, Shih-En Wei, Yaser Sheikh, “OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields”, IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 34, Issue 7, pp. 1444–1450, Dec. 2018.

参考情報等

利益相反

本研究は 2019~2021 年度一般財団法人日本自動車研究所 (JARI) 委託研究「ITS/AD 環境における二輪車課題の抽出と対応検討」による支援を受けたものである。

著者の貢献

- 中島健介 最終提案手法の立案・実装・性能評価, 解釈・考察, データ収集, 論文執筆
- 佐久間尚紀 初期提案手法の立案・実装・性能評価, データ収集, 解釈・考察
- 中安淳也 最終提案手法の立案・実装・性能評価, 解釈・考察, データ収集, 論文執筆
- 木谷友哉 研究の企画・構想, 手法の立案, 解釈・考察, データ収集, 論文校正