

画像認識システムを使用した簡易型交通調査に関する基礎的研究

菅原宏明¹・小篠耕平²・藤井純一郎³・大久保順一⁴・小早川悟⁵・和田英之⁶

¹正会員 八千代エンジニアリング株式会社 技術開発研究所 (〒111-8648 東京都台東区浅草橋5-20-8)
E-mail:sugawara@yachiyo-eng.co.jp

²正会員 八千代エンジニアリング株式会社 技術開発研究所 (〒111-8648 東京都台東区浅草橋5-20-8)
E-mail:kh-ozasa@yachiyo-eng.co.jp

³正会員 八千代エンジニアリング株式会社 技術開発研究所 (〒111-8648 東京都台東区浅草橋5-20-8)
E-mail:jn-fujii@yachiyo-eng.co.jp

⁴非会員 八千代エンジニアリング株式会社 技術開発研究所 (〒111-8648 東京都台東区浅草橋5-20-8)
E-mail:jn-ookubo@yachiyo-eng.co.jp

⁵正会員 日本大学理工学部 交通システム工学科 (〒274-8501 千葉県船橋市習志野台七丁目24-1)
E-mail: kobayakawa.satoru@nihon-u.ac.jp

⁶非会員 日本大学理工学部 交通システム工学科 (〒274-8501 千葉県船橋市習志野台七丁目24-1)
E-mail: cshi16143@g.nihon-u.ac.jp

道路交通量を把握する方法には、道路管理者が設置している常時観測機器から得られるデータを活用する方法の他に、設置済みの監視用カメラや別途取り付けカメラの画像から交通量を読み取るシステム等が開発されている。しかし、これらのシステムでは高所から撮影した画像を想定しているものが多く、撮影機材の設置の制約条件により任意の場所や多地点での調査が難しい。

そこで本研究では、家庭用ビデオカメラと市販の三脚という入手および設置が容易な既存の機材で撮影した動画から深層学習を用いて自動で交通量をカウントするための手法を提案し、その中で用いるデータセットの違いによる性能の変化を検証した。

Key Words : traffic survey, traffic volume, deep learning, image recognition, annotation

1. 背景と目的

道路の交通量は、道路の計画、維持修繕や管理などについての基礎資料を得る際に必要なデータである。任意の場所における交通量調査の多くは、調査員がカウンタを用いて目視で計測する方法が一般的である。調査員を使った交通量調査は、長時間の作業による調査員の負担や膨大な人件費などが課題となっている。

一方、交通量の観測には道路管理者が設置している交通量常時観測装置の結果から集計する方法、簡易型トラカン（路面設置型、路側設置型など）を設置して交通量を観測する方法など、観測調査員を必要としない方法も存在する。しかし、交通量の観測装置は高所 6m 程度に設置している場合が多く、超音波や光ビーコンにより計測するためには機器設置の工事を必要とする。さらに、現在実用化されている交通量調査向けの画像解析システムにおいても高所から撮影した画像の使用を想定しているものが多い。

このように調査員を必要とせず、カメラで撮影した動画を用いた画像認識システムによる調査方法や解析手法などが検討されているが、動画を撮影できる場所が限られてしまうという課題が残る。さらに、生活道路におけ

る安全対策等を検討するための交通調査の実施を踏まえると、細街路などにおいて多地点で一斉に調査が可能な画像認識システムの必要性は高いと考える。

そこで本研究では、家庭用ビデオカメラと市販の三脚という入手および設置が容易な既存の装置で撮影した動画から深層学習を用いて自動で交通量をカウントするための手法を提案し、車種を正確に読み取れる教師データと映像の高さの組み合わせを明らかにすると共に、その中で用いるデータセットの違いによる性能の変化を検証する。

2. 既往研究と本研究の位置づけ

交通流調査におけるAIの応用については、例えば国土交通省ではETC2.0やAIカメラの活用について検討している。

また、星野¹⁾らはETC車載器の製造時に割り当てられる固有の情報であるワイヤレスコールナンバー（以下、WCN）を用いた生活道路調査の適用可能性についての研究を行っている。WCN路側観測器を地上から高さ6mに設置した場合のWCNの取得率は61.7%、高さ2mに設置

した場合の取得率は35.0%であり、設置高さによって精度に大きく差が表れている。生活道路におけるWCNを用いた調査は、機器の設置高さの条件によって設置可能箇所が限られるという課題が明らかになった。

岩崎²⁾らの研究では、ネットワークカメラを6mの高さに設置し画像解析した結果、車両を囲む矩形の面積から大型車両、小型車両の区別が可能としている。しかし、大型車両の誤認や車両の影の影響を受けるといった課題が明らかになった。

鈴木³⁾の研究では多機能ビデオ画像処理システム Traffic Analyzerを開発し、交通量計測が95%以上という実用上には十分な精度で通過車両を検出することを可能にした。しかし、撮影アングルは陸橋や歩道橋など、車両から垂直な場所が必要であるため、任意の場所での調査が容易に実施できないという課題が残る。

そこで、本研究では第一段階として生活道路等の細街路を通過する自動車交通量の実態を動画から自動的に把握することが可能な画像認識システムの実現を目指し、市販の三脚で撮影可能な60cmから150cmまでの高さにカメラを設置し、それぞれの撮影した動画の認識率などを検証し、精度を最も向上させることができる撮影高さについて明らかにする。

3. 動画撮影調査の概要

(1) 調査場所と調査方法

調査場所は千葉県船橋市習志野台8丁目を調査地点とした。調査日時は令和元年8月21日の7時15分から9時15分の2時間である。調査地点を図-1に示す。路側から車道にカメラを向け、図-1で示した矢印の方向に走行する車両を撮影した。



図-1 調査地点図

車種区分を精度良く読み取る高さを検証するため、表-1のように撮影方向と高さを設定した。カメラの高さは60cm, 110cm, 150cmの3つの高さで、道路に対して前方斜め30度に3台のカメラを全4地点に、また、地点1と地

点4には反対側車線の斜め30度（反対側車線を走行する車両の後方）から撮影するカメラを1台ずつ追加し、計14台のカメラで撮影した。地点1と地点4は約600m離れているが、沿道の建物立地状況についてみると図-2に示すとおり大きくは変わらない状況である。なお、動画ファイルはMP4形式で記録した。

表-1 撮影の方向と高さ

撮影方向	車道路面からの高さ
前方斜め30度	60cm
	110cm
	150cm
後方斜め30度	110cm



図-2 地点1および地点4の背景

(2) 調査結果

各地点の動画から人手により集計した地点別の調査時間2時間あたりの車種別通過台数を表-2に示す。自動車を小型乗用車(P)、小型貨物車(t)、大型乗用車(B)、普通貨物車(T)の4車種で集計を行ない、本研究ではこの通過車両台数を真値として取り扱うこととした。

表-2 通過車両台数（台/2時間）

車種	P	t	B	T	合計
地点1	980台	215台	16台	129台	1340台
地点2	112台	21台	2台	1台	136台
地点3	143台	16台	1台	1台	161台
地点4	955台	198台	13台	129台	1295台

4. アノテーションと画像解析

(1) データセットの提案

本研究では車種別交通量を検出するために Weiliu⁴⁾を参考に深層学習の画像認識手法の一つである Single Shot MultiBox Detector（以下、SSD）を適用した。SSDは1秒間に10フレーム以上認識が可能であること、一般的なオブジェクト分類モデルとして多くの実績があること、また近年のグラフ共通化やグラフコンパイルの手法に

よりハードウェアに依存せず、エッジ機器での稼働が可能ことから、図-3 に示す手順で深層学習により車両認識および車種判別、交通量のカウンルの検証を行っている。また検出されたオブジェクト・トラッキング・アルゴリズムとしては J.C.Nascimento⁵⁾ による Centroid Tracking を用いた。

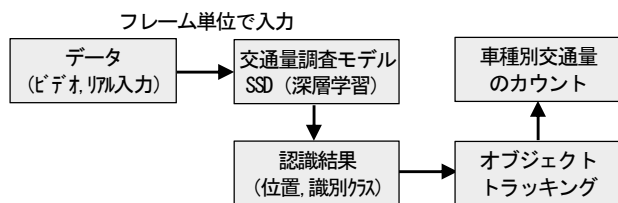


図-3 深層学習による交通量観測手法の手順

(2) 教師画像の作成

図-4 は車種別交通量検出のシステムのために新たに作成したデータセットの作成手順を示したものである。

VoTT を用いたアノテーション画像を COCO 形式に変換するプログラムを新たに作成した。

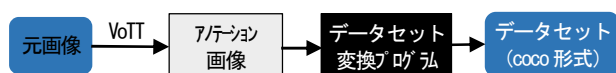


図-4 COCO 形式データの作成手順

データセットの作成では、道路行政関連調査（道路交通センサ調査等）で使われる車種区分を考慮し、小型乗用車、大型乗用車（バス）、小型貨物車、普通貨物車、自転車、その他二輪車、歩行者の 7 クラスとした。また、様々な用途に合わせて将来的に拡張が可能であること、また、短時間での学習が可能であることを目標とした。

COCO 2014 のデータセットには学習用データが 82,783 枚、クラス総数は 80 であり、交通関係として車、バス、トラック、バイク、自転車、人が含まれている。本研究の目標を踏まえて、教師画像のセットにおいては COCO の画像は使わず、データセットの形式を用いて新たに 12,000 枚、クラス数 7 のセットを作成した。

5. 3つの教師データによる解析

各高さにつき 1 種類の教師データを作成するため、教師データは 3 種類となる。読み取りと教師データの一覧を表-3 に示す。

表-3 読み取りと教師データの一覧

撮影高さ	地点 1	地点 4
60cm	読み取り	教師データ
110cm	教師データ	読み取り
150cm	教師データ	読み取り

教師データと同一地点かつ高さの動画は解析対象とはしないため、解析したのは地点 1 の高さ 60cm、地点 4 の高さ 110cm、地点 4 の高さ 150cm である。図-5 はこれら 3 動画の解析結果を実台数比で示したものである。ここで実台数比とは、画像認証システムが動画から車両認識と車種判別まで認識した台数の実台数に対する比率であり、車両認識率と車種判別率の積で示される（表-4）。

表-4 評価指標の定義

名称	定義
車両認識率	動画に映っている車両を車種区分に関わらず認識した比率
車種判別率	認識した車両の中で車種区分まで正しく判別した比率
実台数比	車両認識率と車種判別率の積で、実台数に対して車両の認識から車種区分まで正しく認識した比率

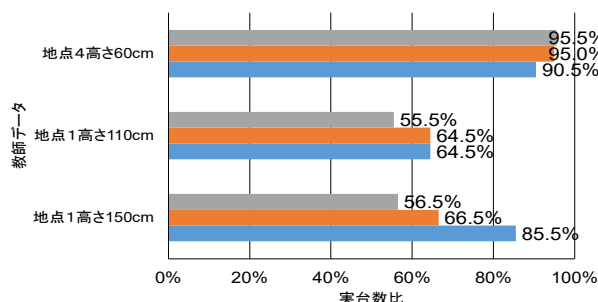


図-5 教師データで共通した映像の実台数比の比較

全ての動画において地点 4 の高さ 60cm の動画を教師データとした場合の実台数比が高く、地点 1 の高さ 110cm の動画を教師データとした場合は、解析した動画全てが他の教師データより低く、地点 1 の高さ 150cm の動画を教師データとした場合には、撮影高さが高くなるにつれて実台数比が向上するという傾向を示した。

図-6 は、車両認識率と車種判別率の関係を教師データの長さ別に示したものである。実台数比を向上させる（車種別に交通量をカウンルする）ためには、車両認識率、車種判別率両者を向上させる必要がある。

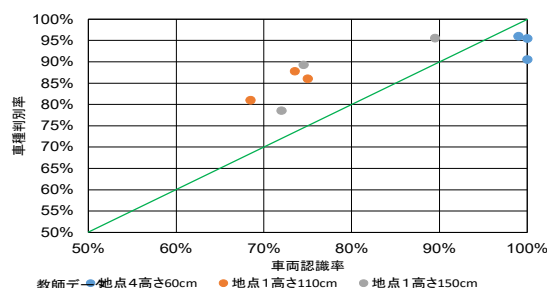


図-6 車両認識率と車種判別率の関係

本研究での条件下では車両認識率、車種判別率、実台数比の全てにおいて、教師データの撮影高さが 60cm、150cm、110cm の順で高く、さらに、読み取る動画と教師データと高さが同じであればより精度が高いという傾向を示した。

本分析だけで教師データの撮影高さにより生じた差の要因を明らかにすることは難しいが、考えられる要因として以下の点が考えられる。図-7に示すように高さ150cmで撮影した動画は高さ60cmに比べて車両を見下ろすことになり、車両の特徴を見出しにくいこと、高さ60cmの場合には背景の建築物が車両の陰に隠れ、車両の認識がしやすい条件になったことなどが想定される。今後、条件が異なる地点での分析を行い検証する必要があると考える。



図-7 撮影高さによる車両の見え方と背景の位置関係

6. データセットの改善効果

(1) データセットの作成

データセット作成のために4断面のビデオ撮影を2時間行ったが、調査ではビデオの設置高を3パターンとし合計12セットの録画を行い、合計24時間分の動画を録画している。データセットはこれら12セットの動画の中から作業時間を考慮し6セット、各1時間2,000枚を選定し、合計12,000枚の画像についてアノテーション作業を行った。

本稿では最初に作成した本データセットを実交通データセットと呼ぶ。実交通データセットを用いた検出結果については、「4.アノテーションと画像解析」で示したとおり検出率は地点・高さに影響されることがわかっている。

実交通データセットでは実際に存在していた車種の割合に依存してデータセットが作成されるため、必然的に実交通量の多い車種区分の割合が高くなる。

本研究では、この車種別台数の割合が偏っていることが検出率に影響を及ぼしているのではないかと仮説を立て、これを検証した。検証では6つの実交通データセットから割合の少ない車種のデータを抜き出し、車種別のアノテーション枚数を増やす方法でデータセットを作成し、これらの検出率について調べた。本論文ではこのデータセットを混合車種データセットと呼ぶ。

(2) 深層学習モデルの作成

本研究では、先述のようにデータセットに含まれる車種区分の比率を変えたものを複数作成し、検出結果がどのように変化するかについて検証した。図-8は実交通データセットと準備した中で最も精度が高かった混合車種データセットについて示したものである。

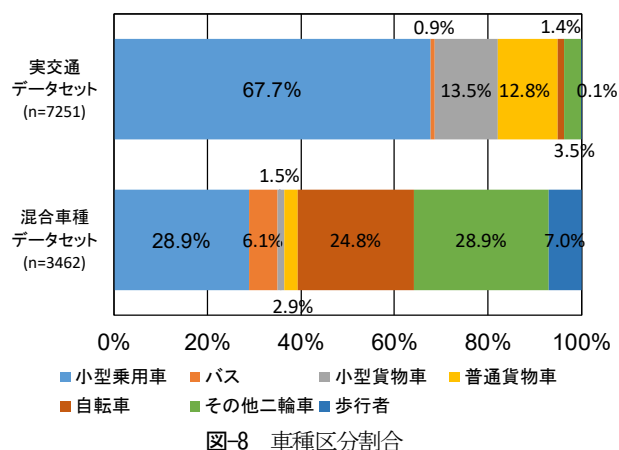


図-8 車種区分割合

(3) 学習結果

モデル学習は16万エポックまで行い、4万エポック毎での性能を比較した。図-9は学習過程を示したものであり、学習回数を重ねると損失が減少していることが分かる。

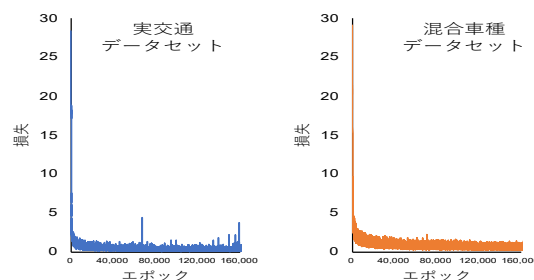


図-9 学習過程

(4) 精度の評価

モデルの精度評価ではIoU (Intersection over Union) が0.5以上のものを正解とした。図-10に4万エポック毎の車両認識率のみと実台数比（車両認識率と車種判別率の積）を示す。車両認識率については、混合車種データセットで90%以上、実交通データセットでも80%前後と高い認識率を示した。

これに対し車種判別率も考慮した実台数比をみると混合車種データセットの方が70%前後と高く、検出率について優位であることが示されている。

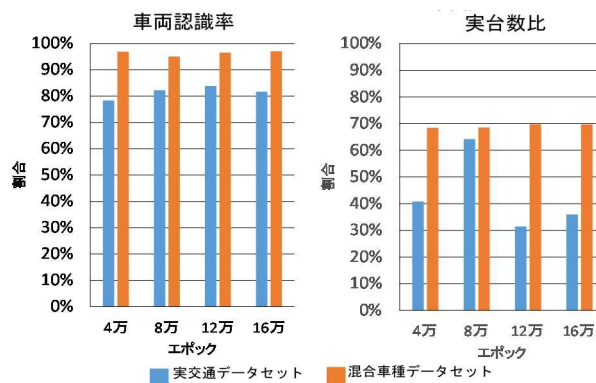


図-10 作成モデルの精度

また、車種判別のエラーについて確認したところ、例えば12万エポックにおいて小型貨物で約8%、小型乗用で約6%とエラーの割合が高かった。これは、図-11に示すように小型貨物と小型乗用で同じ形状（同一車種）であるにもかかわらず車種区分が異なるケースが発生することが要因であると考えられる。



図-11 認識エラー事例（小型貨物車を小型乗用車と認識）

7. まとめと今後の課題

本研究では、簡易な手法で撮影した動画を用いて深層学習による車種別の交通量カウントのためのデータセットの違いによる性能の変化について検証した。

その結果、画像の撮影高さ3パターン別に作成した教師データで検出精度の検証を行ったところ、教師データと解析対象となる映像の撮影高さが同じである場合に最も認識精度が高いという結果ではあるが、その中でも撮影高さ60cmの教師データによる解析が最も精度が高いという結果になった。

さらに、教師データは実交通データセットよりも混合データセットにすることで検出精度の向上や学習時間短縮の可能性を見出すことができた。しかし、この傾向が他地点でも共通なのかを明らかにするために、他地点での同様な調査、検証が必要である。

なお、実交通データセットと混合データセットの両者において小型貨物と小型乗用の判別エラーが多いが、これは形状が同一な車両であるにもかかわらず車種区分が異なることが要因であると想定される。

今後はナンバープレート情報を読み取り、複合識別を行うことで車種判別の精度を高めるための手法について研究する予定である。

参考文献

- 1) 星野一輝, 小嶋文, 市本哲也, 鈴木達也, 上田透, 片山賢治, 久保田尚: WCNを用いた生活道路調査の適用可能性に関する研究, 土木学会論文集 D3, Vol.74, No.5, 1817-1826, 2018.
- 2) 岩崎洋一郎, 永村幸大, 中宮俊幸, 岩本祥二郎, 宮田俊彦, 倉本俊昌, 北島俊孝, 瀬戸口恵: 交通量調査自動化のためのネットワークカメラを用いた交通流計測手法, 電気関係学会九州支部連合大会講演論文集, pp.19-20, 2015.
- 3) 鈴木一史, 中村英樹: 交通流解析のためのビデオ画像処理システム Traffic Analyzer の開発と性能検証, 土木学会論文集, Vol.62, pp.276-287, 2006.
- 4) Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg: SSD: Single Shot MultiBox Detector
- 5) J. C. Nascimento, An algorithm for centroid-based tracking of moving objects

(2020.3.7 受付)

A basic study on simplified traffic survey using image recognition system

Hiroaki SUGAWARA, Kohei OZASA, Junichirou FUJII, Junichi OOKUBO,
Satoru KOBAYAKAWA and Hideyuki WADA

In order to grasp the road traffic volume, in addition to the method of utilizing the data obtained from the constant observation equipment installed by the road manager, the traffic volume can be calculated from the images of the installed monitoring cameras and cameras attached separately. A reading system has been developed. However, many of these systems assume images taken from high places, and it is difficult to conduct surveys at arbitrary places or at multiple points due to restrictions on installation of photographing equipment.

In this study, we propose a method to automatically count the traffic volume using deep learning from videos taken with existing equipment that is easy to obtain and install, such as a home video camera and a commercially available tripod. The effect of performance on the data set used was verified.