

深層学習による動画像解析を用いた 交通量調査技術の開発

齊藤 隆仁¹・池田 大造²

¹正会員 (株)NTTドコモ サービスイノベーション部 (〒100-6150 東京都千代田区永田町2丁目11番地
1号 山王パークタワー)

E-mail:takato.saitou.bu@nttdocomo.com

²正会員 (株)NTTドコモ サービスイノベーション部 (〒100-6150 東京都千代田区永田町2丁目11番地
1号 山王パークタワー)

E-mail:ikedad@nttdocomo.com

カメラを活用した交通量調査技術の開発が盛んに行われており、道路交通センサスなどの公的統計への活用が期待されている。これらの技術を活用し、店舗の往来把握や駐車場の利用状況把握のように時間や空間的に疎である環境への適用検討も進められている。高所に設置されたCCTVカメラと異なり、民間企業や個人では接地位置の撮影の方が撮影コストが低い。しかし、接地位置の映像は対象物に遮蔽が発生し易い傾向がある。そこで本研究では、自治体での活用にとどまらず民間企業や個人が利用できる交通量調査システムの開発をめざし、遮蔽の影響を低減する物体認識および物体追跡技術を実装した。複数の交差点で可搬カメラによる交通量観測を実施し、動画像解析の観測結果と人手観測により精度検証を行い、画角の違いによる観測精度への影響を考察する。

Key Words : road traffic census, video-based car detection, vehicle type estimation, occlusion

1. 背景

自治体による都市計画や道路建設の効果測定をはじめ、交通量調査の観測結果が広く利用され、民間企業では店舗の往来把握や出店計画への利用を目的として交通量調査を実施する機会も増えている。交通量調査には、人手観測と機械観測の大きくふたつの手法がある。人手観測の精度を高く維持するためには、経験豊富な観測員の確保が重要と言えるが、近年では人口減少等の影響により観測員の確保が難しくなっている。機械観測は一定の精度が期待でき、観測手段として超音波やミリ波を利用した常設トラフィックカウンターが活用されている。一方で、設置コストが高いことから主要幹線道路等への設置で留まっている。そこで設置コストが低いCCTVカメラや可搬式なトラフィックカウンターによる観測が広く使用されている。映像を解析し交通量を調査する手法は車種を推定できる可能性や交通状況を目視で確認できるため、活用に向けた検討が進められている。その一方で暗所や天候条件の影響を受けやすく、車が密に重なり合うことで遮蔽が発生し認識精度が低下するといった課題があり、CCTVカメラのように高所にカメラを設置する際

には、対応策が取られることが多い。しかし、一時的にカメラ観測を行うためには、ポールカメラ等を利用して高所に設置することになるが、手間がかかる上に安全面への懸念が残る。

そこで汎用的な三脚にカメラを固定設置した状況で撮影し、動画像を解析することで交通量観測が可能であれば、自治体による活用にとどまらず民間企業や個人が利用しやすくなると考え、深層学習を利用した動画像解析技術を交通量調査へ応用した交通量調査システムを開発した。

2. 動画像解析による交通量調査技術

(1) 関連研究と課題

国交省や自治体を実施する全国道路・街路交通情勢調査(道路交通センサス)の対象は車両であるが、民間企業では車両の交通量にとどまらず、人を対象とした通行量観測に対しても期待がある。そこで本稿では、車両と人を対象として動画像を用いた観測手法を提案する。動画像を利用して車両や人を検知することは、スマートシティ分野の重要なテーマであり、CCTVカメラの映像を

解析する技術が多く開発されている。

動画による交通量調査技術は大きくわけて、(1) 動画からフレームごとに車両を検出し、車種を推定する、(2) フレーム間の同一車両を識別し車両を追跡する、(3) 車両の移動方向から通過を判定し集計する、といった処理過程をとることが多い。例えば、Buiらの研究¹⁾では、1台のCCTVカメラで撮影した動画画像を利用してConvolutional Neural Networkにより構築されたYolov3²⁾を使い、car/bus/truck/bikeの4種類を対象とした車両検知と車種推定を行い、DeepSort³⁾を使って同一車両の識別と車両追跡を行っている。Yoloをはじめとした検出手法は、検出対象物の周辺に膨大なバウンディングボックス(BBox)を算出し、この中からひとつを選択することで最終的に対象物を検出する。BBoxを選択する方法としてNon-Maximum Suppression (NMS) が広く用いられている。NMSはBBox間の重なり度合いを示す指標であるIoU (Intersection over Union) を用い、閾値処理でBBoxを絞っていく。閾値には0.3~0.5の値が用いられることが多いが、Huangらの研究⁴⁾では、真値と全く同じ領域のBBoxを設定した理想的な状況で閾値0.5のNMSを施すと約10%のBBoxがNMSによって除外されてしまうことを指摘していることから、交通量調査に用いる方法としてはNMSによらない検出手法が必要であると考えられる。車両通過判定と集計を行う方法としては、動画画像上に仮想的な検知線を設定する方法が広く利用されるが、Buiらは交差点領域に矩形を複数設定する方法を使い、対象物の遮蔽による観測数の誤差低減を図っている。しかし矩形を緻密に設計する必要があり、少数の交差点映像であれば活用できるものの、数多くの交差点映像を対象とする場合や、駐車場または店舗敷地のような環境下では矩形の設計が難しい可能性がある。また遮蔽による観測誤差への対策は、処理過程の後段ではなく、前段で講じたほうが効果的であるといえる。交差点の映像は検出した対象物が密に重なり合うため、車両検出の段階で遮蔽に強い技術を応用することが適切であると考えられる。

(2) 提案する交通量調査システム

このような背景から、交通量調査技術の処理過程のうち車両検知と車種推定手法として、著者らはNMSを必要としない手法であるDetection Transformer⁵⁾ (DETR) を応用し、対象物体が密になり遮蔽が生じやすい状況に対しても堅牢な車両検出および車種推定の実現をはかる。フレーム間の同一車両の識別と車両追跡を行う手法には、照明変化に堅牢であり、フレーム間で一時的に対象物が遮蔽された場合でも個別識別子の変化を抑えることができるDeepSort³⁾を採用する。DeepSort³⁾は対象物の類似性を学習することで同一物体を認識し個別識別子を与える手法であり、類似性の指標としてはコサイン類似度が一般的

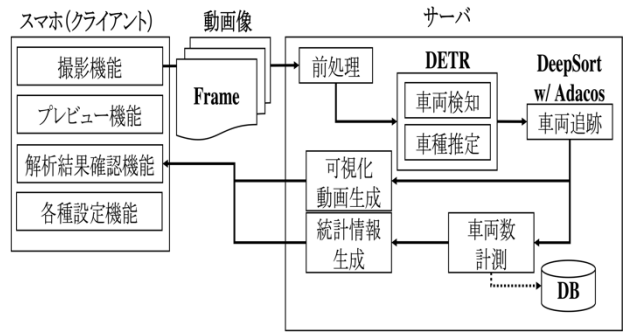


図-1 システム概要図

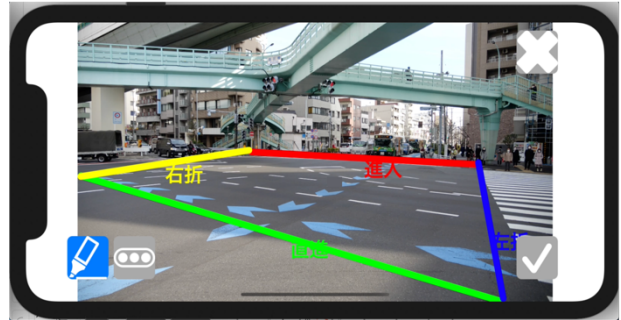
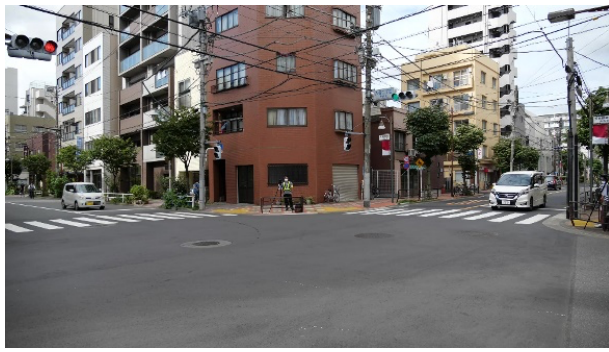


図-2 検知線の設定画面

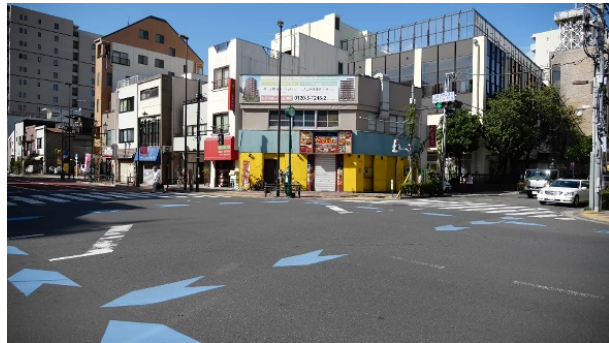


図-3 開発アプリのイメージ図

に用いられる。しかし、コサイン類似度は比較したい特徴量の次元が大きいときは類似度が低くなる傾向にある。そのため、車両のように形状の類似性が高い場合は追跡が失敗する可能性が高いと考えられる。そこで著者らが提案するシステムでは、Adacos⁶⁾によって同一物体と異なる物体との距離学習を行うことで同一物体識別の精度向上を図った。また、民間企業や個人のユーザが活用できる交通量調査技術をめざし、スマートフォンアプリとサーバにより構成されたシステムのプロトタイプを開発した。図-1にプロトタイプシステムの概要を示す。本システムは、スマートフォンのアプリケーション上で動作



(a) 交差点A (片道1車線)



(b) 交差点B (片道2車線)



(c) 交差点C (片道3車線)

図4 交差点撮影動画

するクライアントサーバから構成される。車両検知等の交通量調査に関わる技術は主にサーバ上に実装され、クライアントから撮影された動画フレームに基づき処理が行われ、解析結果はサーバからクライアントに転送された上で可視化される。車両の移動方向から通過を判定するために検知線を導入し、図-2のようにスマートフォンの画面上をタップすることで検知線（進入、右折、左折、直進）を容易に設定することができる。図-3はスマートフォン上で確認できる交通量観測結果の可視化の実装例を示している。サーバ上のデータベース（DB）に保存された解析結果の一覧が表示され、画面をタップすることで交通量の調査結果の詳細を確認することができる。円グラフ等のUIによって車種別の台数表示やスマートフォンのGPS機能によって位置情報が取得できている

表-1 車種の区分

ナンバープレートの頭数字	クラス名
3/5/7	乗用車
4/6	小型貨物
1/8/9/0	大型貨物
2	バス
-	自動二輪車
-	自転車

表-2 車両検知モデルの評価結果

#	学習	評価	mAP	mAR
1	交差点A 8時台	交差点A 8時台	0.547	0.355
2		交差点A 11時台	0.696	0.478
3		交差点B 8時台	0.391	0.298
4		交差点C 8時台	0.342	0.214
5	交差点B 8時台	交差点A 8時台	0.452	0.300
6		交差点B 8時台	0.745	0.593
7		交差点B 12時台	0.756	0.552
8		交差点C 8時台	0.271	0.207
9	交差点C 8時台	交差点A 8時台	0.594	0.381
10		交差点B 8時台	0.749	0.485
11		交差点C 8時台	0.510	0.381
12		交差点C 11時台	0.676	0.523

場合は、位置情報も含めて表示される。

3. 実験と精度検証

本節では開発した交通量調査技術のうち車両検知と車種推定の検証および車両追跡の解析技術に対して検証を行う。

(1) 実験条件

車両検知および車種推定手法の精度検証を行うため、片道1~3車線をもつ東京都内の交差点（図-4）で撮影した動画をを用いた交通量推定結果と人手観測結果を比較する。図-4に示す画像は交差点に設置した市販の一眼レフカメラで撮影したものであり、これらの固定された画角で撮影した動画をを用いて検証する。ここで、解像度は4K、フレームレートは30FPSである。撮影日時は2020年9月の平日のうち、通勤時間帯で交通量が多い8時台（8:00-8:59）と日中の交通量が少ない11時台（11:00-11:59）の2時間帯とした。

車両検知および車種推定を担うDETRは、車種6種類と歩行者1種類をクラスとした計7種類で学習した。此処で、車種6種類の区分け（表-1）は道路交通センサスで

表-3 車両追跡の評価結果

#	学習	評価	特異度	感度
1	交差点A 8時台	交差点A 8時台	0.934	0.960
2		交差点B 8時台	0.924	0.938
3		交差点C 8時台	0.907	0.908



図-5 類似度検証の車両画像

表-4 類似度検証の評価結果

	A	B	I	J
A	1.000	0.958	0.030	0.060
B	0.958	1.000	0.033	0.058
I	0.030	0.033	1.000	0.576
J	0.060	0.058	0.576	1.000

表-5 交差点A (8時台)

	乗用車		小型貨物		大型貨物		バス		自動二輪	
	真値	推定値	真値	推定値	真値	推定値	真値	推定値	真値	推定値
右折	2	0	0	0	0	0	0	0	1	0
直進	15	19	7	0	0	0	0	0	0	0
左折	0	3	1	0	0	1	0	1	0	0
総数	17	22	8	0	0	1	0	1	1	0

採用される区分に準拠している。ただし、歩行者と自転車に関しては通行方向が複雑となるため、表-1に示す車種6種類のうち自転車を除外した5種類に関して評価することとした。

(2) 解析結果

a) 車両検知モデルの評価結果

DETRによる車両検知および車種推定の性能を表-2に示す。ここで、mAPは式(1)に示すmean Average Precision、mARは式(2)に示すmean Average Recallを意味する。mAPは推定結果の正確さである適合率を、mARは推定結果

表-6 交差点A (11時台)

	乗用車		小型貨物		大型貨物		バス		自動二輪	
	真値	推定値	真値	推定値	真値	推定値	真値	推定値	真値	推定値
右折	3	2	0	0	0	0	0	0	0	0
直進	19	14	4	1	0	0	0	0	0	0
左折	5	6	3	0	0	1	0	0	0	0
総数	27	22	7	1	0	1	0	0	0	0

表-7 交差点B (8時台)

	乗用車		小型貨物		大型貨物		バス		自動二輪	
	真値	推定値	真値	推定値	真値	推定値	真値	推定値	真値	推定値
右折	4	11	3	0	2	28	0	1	0	0
直進	26	24	20	0	1	3	2	2	3	1
左折	3	1	2	0	0	1	1	0	0	0
総数	33	36	25	0	3	34	3	3	3	1

表-8 交差点C (8時台)

	乗用車		小型貨物		大型貨物		バス		自動二輪	
	真値	推定値	真値	推定値	真値	推定値	真値	推定値	真値	推定値
右折	8	3	7	4	0	1	0	0	1	1
直進	56	62	45	1	11	8	0	0	11	1
左折	10	9	4	0	3	5	0	0	0	0
総数	74	74	56	5	14	14	0	0	12	2

の抜け漏れを表す再現率である。Cは車種の区分数であり、推定車種cに対するPrecisionを $Precision_c(r)$ 、Recallを $Recall_c(p)$ とした。PrecisionとRecallはそれぞれ0.0~1.0の値をとり、1.0に近いほど性能が高いことを意味する。このときPrecisionとRecallをそれぞれPとRの分割数で分散化している。PrecisionとRecallはトレードオフの関係があり、Precision-Recall Curveを描くことによって車種別にモデルの推定性能を確認することができるが、推定する車種が複数ある場合はこれらを分割数と車種数で平均化したmAPとmARが用いられる。これらの値もPrecisionとRecallと同様に0.0~1.0の値をとり、1.0に近いほど性能が高いことを示す。

$$mAP = \frac{1}{C} \frac{1}{R} \sum_c \sum_{r=0.0}^{1.0} Precision_c(r) \quad (1)$$

$$mAR = \frac{1}{C} \frac{1}{P} \sum_c \sum_{p=0.0}^{1.0} Recall_c(p) \quad (2)$$

表-2の評価では、動画撮影開始時刻から6割に相当するデータを学習用、続く2割のデータを検証用、最後の2割のデータを評価用として利用した。なお、検証用

データとは学習時に利用されるものであるためその結果は表記していない。交差点A (#1~4) と交差点B (#5~8) の交通量調査結果のうち、最も精度が高いのは、学習と評価時が同一な交差点である#2と#6となった。交差点Cの撮影動画を学習用データとした場合 (#9~12) は、交差点Bに対する評価結果#10が最も精度が高い結果となった。また、交差点Aまたは交差点Bでの撮影動画を学習用データとして交差点Cに対して評価した場合、#4と#8はそれぞれ最も低い精度となった。これは車線数が多く遮蔽等の影響が発生しやすい交差点Cの撮影動画を学習用データとすることで、モデルの堅牢性が向上した可能性であると考えられる。また、車線数が少ない交差点の撮影動画を学習用データにすると、交差点Cに対しては十分な状況を学習できないことに起因し、精度が低い結果となったと考えられる。

b) 車両追跡結果および人手観測との比較結果

ここでは交差点Aの撮影動画データを利用して車両追跡モデルの評価を行う。学習用および評価用のデータは2枚の画像を1セットとし、異なる画像から抽出した同一対象物体の画像セットを正例、異なる画像から抽出した異なる対象物体の画像セットを負例とする。正例と負例の画像セットを500ペアずつ用意して評価を行った(表-3)。特異度は、異なる対象物体を入力したとき、異なるものであると判断できた割合を意味し、感度は同一対象物体を同一なものであると判断できた割合を意味する。両指標がともに0.0~1.0の値をとり、この値が高いほど性能が良いことを示している。表-3の結果から9割以上の精度があることがわかる。さらにAdacosによる類似度の結果を確認するため、評価用データから2つの車両画像のセット(図-5)をサンプリングし、0.0~1.0の値をとる類似度を表-4に示す。この値は1.0に近いほど車両間が類似していることを意味する。例えば、車両Aと車両Bは異なる画像から抽出した同一車両であり、車両Iと車両Jの関係性も同様である。このとき、表-5の対角線に並ぶ数値は同一画像の同一車両間の類似度の場合には1.0の値をとる。異なる画像の同一車両である車両Aと車両B間の類似度および同様な条件である車両Iと車両J間の類似度はともに高い結果であることが確認できる。一方で、同一車両ではない画像の類似度は0.0に近い値となっておりAdacosによる手法を活用することで同一車両の同定ができていたことを確認した。

c) 車種推定結果および人手観測との比較結果

車両検知と車種推定を行った後に、車両追跡によって方向別の交通量観測を実施した結果を表-5から表-8に示す。数値は車両台数である。これらの評価においても、該当するデータのうち、動画撮影開始時刻から6割を

学習用、続く2割を検証用、最後の2割を評価用として使用した。交差点によって画角や遮蔽の発生頻度が異なるが、いずれの交差点映像においても乗用車の検出精度が高い結果となった。一方で、小型貨物と自動二輪の検出漏れが目立つ結果となった。また、交差点Bでは大型貨物の検出結果では誤検出が多く発生し、特に左折台数の推計誤差が大きいことに起因している。推計誤差が大きい理由として、交差点の背景に存在する店舗では金属製シャッターを使用しており、金属製シャッターを貨物車両と誤検知した可能性が挙げられる。しかし、交差点Aの撮影動画を学習用データとして使用した場合でも、交差点Cでは通過する乗用車を精度良く捉えたことから、学習用データに多く含まれる車両に関しては遮蔽が起こりやすい交差点においても検出できる可能性が高い。小型貨物や大型貨物、自動二輪に関してはデータ中に出現する頻度が少ないことから、データの不均衡性に対する対処を行うことでさらなる精度向上の余地が考えられる。

4. まとめ

本稿では自治体に限らず民間企業や個人においても利用可能となる交通量調査システムの検討および提案を行った。多くの自治体で活用されているCCTVカメラを用いた動画解析手法は従来研究において一定の精度が確認できており、実用化が近いと考えられる。一方で、民間企業や個人が時間や空間的に疎である環境への適用という観点では、設置場所が限られており、必ずしも高所にカメラが設置できるとは限らない。筆者らが提案する交通量調査システムは三脚が接地する高さにも適用できることが特徴であり、低所における動画撮影は車両どうしの重なりによって遮蔽が発生しやすいという課題を解決する車両検出および車種推定手法を検討した。判定する車両区分のうち、乗用車は学習するデータ量が多いため、高い精度で観測できることを確認したが、貨物車両が検知されないデータが多く存在した。車両追跡には従来研究と同様にDeepSortを使用し、前段の処理にNMSせず、コサイン類似度ではなくAdacosが提唱する距離尺度を学習することにより、一定の精度で交通量を推計できることを明らかにした。矩形を用いて交通量を集計する手法ではなく、検知線を設定することで交通量を集計したが、十字交差や丁字以外の複雑な交差点形状、膨大な交差点数が存在する場合は、効果的な検知線の提案や検知線の自動設定が実用上必要になると考える。また、スマートフォンによって簡易的に確認できる情報の内容選定や、三脚固定ではなく、ユーザが手持ちで撮影した際の手振れによる影響を低減する手法の検討も今後の課題である。

参考文献

- 1) Bui, K.H.N., Yi, H. and Cho, J. : A Multi-Class Multi-Movement Vehicle Counting Framework for Traffic Analysis in Complex Areas Using CCTV Systems, *Energies*. Vol. 13, No. 8:2036, 2020.
- 2) Redmon, J. and Farhadi, A. : YOLOv3: An Incremental Improvement, *ArXiv abs/1804.02767*, 2018.
- 3) Wojke, N., Bewley, A. and Paulus, D. : Simple online and realtime tracking with a deep association metric, In *Proceedings of the 24th International Conference on Image Processing (ICIP)*, pp.3645–3649, 2017.
- 4) Huang, X., Ge, Z. , Jie, Z. and Yoshie, O. : NMS by Representative Region: Towards Crowded Pedestrian Detection by Proposal Pairing, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
- 5) Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A. and Zagoruyko, S. : End-to-End Object Detection with Transformers, *ArXiv abs/2005.12872*, 2020.
- 6) Xiao, Z. and Rui, Z., Yu, Q., Xiaogang, W. and Hongsheng, L.: AdaCos: Adaptively Scaling Cosine Logits for Effectively Learning Deep Face Representations, *ArXiv 1905.00292*, 2019.

(2021. 10. 01 受付)