

物体検知・識別器による自動車交通量調査システムの試作と評価

Development and Evaluation of Traffic Volume Survey System

Using Object Detector and Classifier

筑波大学大学院 峯岸朋弥, ネットワーク情報学部 吉田享子

University of Tsukuba Tomoya MINEGISHI, School of Network and Information Kyoko YOSHIDA

Keywords: Traffic volume survey, Object detection, Object classification, VGG16, YOLO, TensorFlow

Abstract

In the current automobile traffic survey, there are some ways to count, for example, a mechanical counting with a loop coil embedded in the lane and a visual measurement by the investigator, but these ways have problems that introduction and operation costs are high.

In this paper, to solve the problems of the current traffic survey, we developed the prototype system that can detect objects and identify cars, using deep learning image application technology from the video, and evaluated it.

1. はじめに

一般交通量調査は、全国道路・街路交通情勢調査（道路交通センサス）の一環として、全国の道路の交通量及び道路現況等を調査し、道路の計画、建設、維持修繕その他の管理などについての基礎資料を得ることを目的に実施しているものである[1]。現在の自動車の交通量調査は、調査員の目視による測定の方法が主流であるが、その他、車線に埋め込んだループコイルや超音波により計測するなどして機械的にカウントする方法、歩道橋上にレーザーを用いた機械を設置しカウントする方法などがある。しかしこれらのいずれも、作業コストや導入・運用コストが高いという点が問題であると考えられる。

本研究では、既存の交通量調査の課題を解決することを目的として、画像識別器を用いた簡易な交通量調査自動化システム Automatic Traffic Volume Survey System (ATVS v.1.0)を試作した。評価としては、片側一車線の複線路をカメラで撮影し、撮影した動画から抽出した画像を入力画像として、自動車の分類ごとの集計結果を評価した[2]。ATVS v.1.0では動画から自動車が見える画像の抽出において物体識別モデルである Visual Geometry Group 16-layer [3]（以下 VGG16）を用いたが、抽出精度に問題が多く残り、また自動車分類処理においても、画像識別器の精度に改良が必要であった。

今回は、このシステムの自動車画像抽出処理に改良を

加えることを目的として、リアルタイムオブジェクト検出手法である You Look Only Once（以下 YOLO）[4]を用いて動画から自動車が写る画像の抽出を行い、システムの評価を行った（ATVS v.2.0）。

本稿では ATVS v.1.0 の内容の説明と評価結果の後、ATVS v.2.0 の改良内容と評価結果を説明する。

2. 関連研究

交通量調査の自動化に関する関連研究としては、ビデオカメラの映像から路面との間で幾何学変換を行い自動車の位置を決定する方法[5]や、ネットワークカメラを設置し映像を解析する画像処理手法を構築し、交通流計測を行う方法[6]がある。またカメラを使用することの欠点を克服したマイクロフォンによる車両カウントシステム[7]などもある。これらは通過した自動車の検出や通過数のカウントについては有効であるが、現在の交通量調査で行われている自動車の分類（例えば乗用車や貨物自動車など）ごとの自動的な集計は実現できていない。またクラウドソーシングを用いた交通量調査手法を提案する研究[8]もある。これは動画から静止画を手動で抽出し、その静止画を不特定多数の人に業務を委託するかたちで自動車数をカウントしてもらうものであり、自動車類の分類が実現できる点では有効であると考えられる。しかし分類を担当した人により精度が異なることや、動画からの画像抽出間隔により正解率が異なる

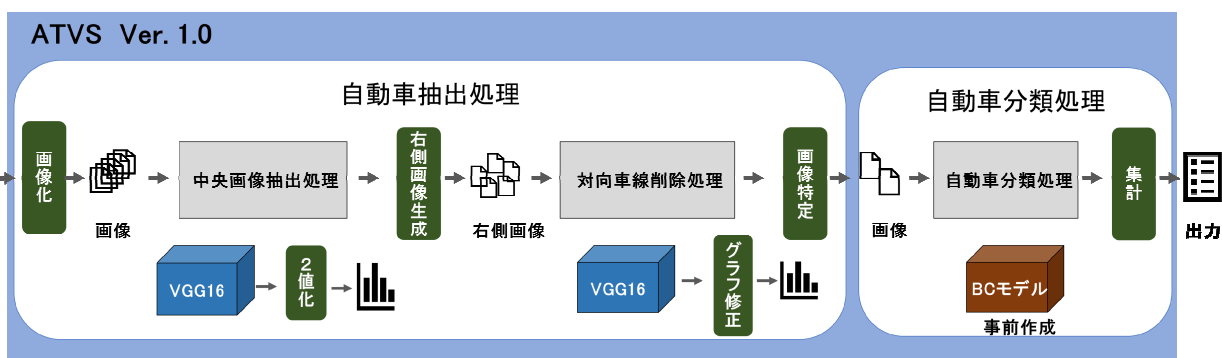


図 1 ATVS v.1.0 概要図

という問題がある。

本研究では、このような課題を解決することを目的として、Deep Learning の画像応用技術を用いた簡易な交通量調査自動化システムの開発を試みた。

3. ATVS v.1.0

3.1. システム概要

ATVS v.1.0 は、道路脇から道路を走行する車両を撮影した動画をもとに、自動的に車両ごとに分類された交通量を集計するものである。ATVS v.1.0 システムは、撮影された動画から自動車が中央に写る画像を抽出する処理（自動車抽出処理）と、画像に写る自動車を種類ごとに分類して集計する処理（自動車分類処理）の2つに分かれている。

まず、撮影された動画から自動車が中央に写る画像を切り出す自動車抽出処理では、全画像フレームを物体識別モデルである VGG16 に入力することで、中央に写る画像を抽出した。その後自動車分類処理で、抽出した画像を今回構築した学習済みモデル（分類器）である BC (Bodytype Classifier) モデルに入力し、通過した自動車を分類して集計した (図 1)。

なお、ATVS v.1.0 は、片側一車線の道路の交通量調査を想定しており、事前に撮影する動画は、カメラを自動車が走行する車線と垂直に設置し、自動車側面が映るように撮影することとした (図 2)。

また、動画に写る自動車を抽出した後、その自動車などのボディタイプに属するか分類・集計したが、分類は、以下の2つのパターンで行った。

- (1) 3種：現在行われている交通量調査と同様に、3種（バス、乗用車、貨物自動車）で分類する。
- (2) 5種：(1)を詳しくし、5種（3box, バス, ハッチバック, ステーションワゴン, 貨物自動車）で分類する。

3.2. 自動車抽出処理

動画から抽出した全画像は、VGG16 に入力して車両であるか判定を行った。VGG16 はオックスフォード大学 Visual Geometry Group により構築された 16 層のニ

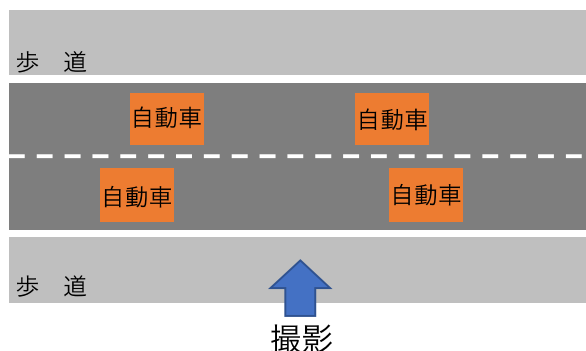


図 2 動画の撮影方法

ューラルネットワークのことであり、ImageNet (大規模な画像データセットで約 1000 クラスを持つ) を使った訓練済みモデルである。この分類器に画像を入力すると、画像に写る物体を識別してその判定結果 (確率) として上位 5 つが出力される。その中から自動車 (beach wagon, cab, jeep 等) が写っていると識別された画像をリストアップし、出力された識別の確率を用い、自動車が写っている可能性がある画像であるか否かの 2 値 (0 と 1) に分けた。また、撮影した動画から 1 台の自動車が通過するフレーム数 (今回は 30 と設定) をカウントし、30 フレーム毎に VGG16 による判定結果の確率を調べた。30 フレームのうち一定数が自動車であると判定された場合、自動車が通過しているものとして処理を行った。この 30 フレームのうち 15 フレーム目の画像が、中央に自動車が写っている可能性が高いと考え、自動車が中央で写る画像を特定した。

3.3. 反対車線の自動車の画像削除

ATVS v.1.0 では、動画から VGG16 を用いて片側一車線の道路の撮影場所から近い車線を通る自動車を抽出することを試みた。しかし、撮影した動画には反対車線を走行する自動車も映り込んでしまい、VGG16 はこれらも自動車として識別してしまうため、反対車線を走行する自動車のフレームを削除することが必要となった。このため、すべての画像を右側の半分のみ画像に切断し、それらをもう一度 VGG16 に入力し判定させる

こととした。日本では左側通行のため、右側から左側に向かって自動車が通過することになり、動画では先に右側にある自動車が左に移動するものを抽出すればよいことになる。右側の画像に自動車が写っていれば抽出し、写っていなければ抽出しない処理を追加した(図3)。この処理により反対車線を走行する自動車が写る画像も抽出されてしまう問題を解決した。



図3 抽出すべき画像と削除すべき画像

3.4. 自動車分類処理

中央に写る自動車の画像を抽出した後は、自動車を分類・集計する自動車分類処理を行った。ATVS v.1.0では、自動車を分類するために事前にBCモデルを構築した。BCモデルとは、画像に写る自動車のボディタイプを分類する学習済みモデル(分類器)である。BCモデルの構築にはTensorFlowによるImage Retrain技術を使用した。Image Retrainは、ImageNetで訓練された強力な画像分類器の特徴抽出機能を再利用し、最上位に新しい分類層を訓練するもので、下位レイヤーに変更無しで、新たな分類を行えるよう再学習を行うものである[9]。今回は、Image Retrainによる画像識別器をさらに自動車識別用に特化させるため、自動車側面の画像を集め約3万枚入力することによって、最上位層を訓練することで作成した。



図4 不適切な画像抽出例

4. ATVS v.1.0 の評価

作成したシステムを、埼玉県戸田市にある市役所前通りで撮影した動画を使用して評価した。動画は11分3秒で、総フレーム数は19,898である。

4.1. 自動車抽出処理の結果

表1, 2はATVS v.1.0に、撮影した動画を入力してシステムを一通り実行し、(1)3種, (2)5種で分類した結果である。動画を筆者が目視して通過した自動車の台数と種類を確認し、分類して集計したものを正解としている。

目視(正解)では、(1)3種の分類では、バス1台, 乗用車74台, 貨物自動車8台で通過した自動車数は合計83台であった。また(2)5種の分類では、3BOX11台, バス1台, ハッチバック55台, ステーションワゴン8台, トラック8台であった。

ATVS v.1.0の結果としては、(1)の分類ではバス2台, 乗用車65台, 貨物自動車7台で合計74台の自動車を検出した。また(2)の分類では3BOX12台, バス3台, ハッチバック38台, ステーションワゴン18台, トラック3台であった。

ATVS v.1.0は自動車抽出と自動車分類に処理工程がわかれているため、さらに工程ごとの結果を検討した。

4.2. 自動車抽出処理の検討

動画から抽出すべき画像数は83枚であったが、ATVS v.1.0の自動車抽出処理の結果としては74枚の画像しか抽出できなかった(表3)。さらに詳しく画像を検討した結果、抽出した74枚の画像のうち自動車が中央に写る適切な画像は53枚であった。撮影した道路を通過した30台の自動車を適切に抽出できていないことが分かった。

適切に抽出できなかった画像の例を図4に示す。図4の左上は何も写っていない画像を抽出した例である。右上は自転車を抽出した例である。左下は歩道を通行する歩行者を抽出した例である。右下は車道を通るバイクを抽出した例である。

4.3. 自動車分類処理の検討

自動車分類処理に用いたBCモデル単独の精度について評価するため、筆者が抽出した自動車が中央に写る83枚の画像をBCモデルに入力した。BCモデルを用いて(1)3種, (2)5種の2パターンで分類した結果が、表4, 5である。(1)3種の分類では、バス2台, 乗用車69台, 貨物自動車12台であった。また(2)5種の分類では、3BOX10台, バス2台, ハッチバック43台, ステーションワゴン16台, トラック12台であった。

画像分類について詳細に検証するため、さらに入力画像1枚ごとに分類結果を確認した(表6, 表7)。

(1)3種の分類では、バスは1台通過したが、それに対し「バス」と正しく判定することができなかった。よって正解数0, 誤分類数1である。また通過した乗用車

表 1 ATVS v.1.0 による分類 (3種) の結果

	正解	ATVS v1.0
バス	1	2
乗用車	74	65
貨物自動車	8	7
合計	83	74

表 2 ATVS v.1.0 による分類 (5種) の結果

	正解	ATVS v1.0
3BOX	11	12
バス	1	3
ハッチバック	55	38
ステーションワゴン	8	18
トラック	8	3
合計	83	74

表 3 VGG16 による自動車抽出処理結果

	正解	VGG16
抽出した画像数	83	74
自動車が中央で写る画像数	83	53

表 4 BC モデルによる自動車分類処理(3種)結果

	正解	BC モデル
バス	1	2
乗用車	74	69
貨物自動車	8	12

表 5 BC モデルによる自動車分類処理(5種)結果

	正解	BC モデル
3BOX	11	10
バス	1	2
ハッチバック	55	43
ステーションワゴン	8	16
トラック	8	12

表 6 BC モデルによる分類(3種)詳細結果

	正解数	誤分類数
バス	0	1
乗用車	62	12
貨物自動車	1	7

表 7 BC モデルによる分類(5種)詳細結果

	正解数	誤分類数
3BOX	1	10
バス	0	1
ハッチバック	26	29
ステーションワゴン	0	8
トラック	2	6

74 台のうち、62 台に対して正しく「乗用車」と判定したが、12 台に対しては乗用車と判定することができなかった。通過した 8 台の貨物自動車のうち、1 台に対して正しく「貨物自動車」と判定したが、7 台に対しては

貨物自動車と判定することができなかった。

(2) 5 種の分類では、通過した 3BOX 11 台のうち、1 台に対して正しく「3BOX」と判定したが、10 台に対しては 3BOX と判定することができなかった。また通過したバス 1 台に対して「バス」と判定することができなかった。通過した 55 台のハッチバックのうち、26 台に対して正しく「ハッチバック」と判定したが、29 台に対してはハッチバックと判定することができなかった。通過した 8 台のステーションワゴンのうち、正しく「ステーションワゴン」と判定できたものはなかった。通過した 8 台のトラックのうち、2 台に対して正しく「トラック」と判定したが、6 台に対してはトラックと判定することができなかった。

4.4. 評価のまとめ

ATVS v.1.0 では、自動車抽出処理のための VGG16 を用いた検出は自動車が中央で写る画像が 83 枚中 53 枚であった。また、自動車のボディタイプの分類を行う BC モデルは、まだ誤分類が多いことがわかった。自動車抽出処理、自動車分類処理の両方にさらに改良が必要であることが分かった。

特に自動車抽出処理については、自動車が手前の車線を走行しているか、反対車線を走行しているか、判断が難しかったため、画像右半分を再度識別器に入力することで、走行車線の判別を行うことを狙った。この処理により反対車線を走行する自動車の削除を行うことができたが、反対車線を走行する自動車が画像右半分に写っていた場合には対処できず、この点も改良が必要であることがわかった。

そこで、VGG16 の代わりとして Image Retrain を再度利用し、自動車の側面が写った画像を特定するモデルを新たに作成して評価を実施したが、正解率は約 58.5% となり良い結果を得ることができなかった。

5. ATVS v.2.0

VGG16 を用いた ATVS v.1.0 は、動画から自動車が中央に写る画像を抽出する自動車抽出処理と、自動車分類処理の両方に問題があったが、ATVS v.2.0 ではまず自動車抽出処理の改良を試みた。

5.1. 自動車抽出処理

自動車抽出処理において、今回は片側一車線の道路の交通量調査を想定しているため、手前を通過する自動車を抽出しなければならない。自動車が手前を通過したものの反対車線を通過したものと判定するため、自動車を囲うのに必要な矩形の位置座標 (バウンディングボックス) を取得することとした。

取得された座標のうち y 座標が一定値内であればそ



図 5 y 座標による車線判別



図 6 抽出すべき車両 (右)

の自動車は画面手前を通過していると考えられる。また取得された x 座標を使用することで、自動車が中央で写るフレームの抽出が可能となる (図 5, 図 6)。

以上を実現するために、物体検出と座標検出を同時に行うことが可能な YOLO を使用した。YOLO は物体検出手法の一つであり、他手法と比べ高速であるという利点を持つ[4]。同時に訓練済みのクラスに対し複数の物体検出・座標検出が可能である。

YOLO を使用し画面手前を通過する自動車を検出することを試みたが、YOLO は自動車を含め歩行者、自転車、バイクなども検出するため、自動車以外を排除する処理を加えた。また手前を走行する自動車と反対車線を走行する自動車を比較した場合、手前を走行する自動車の方が大きく写ることが考えられる。そのため、座標だけではなく、自動車の大きさによる判定も行い、画面手前を通過する自動車を検出する画像特定処理を加えた。

また YOLO の利点としては、動画の入力が可能であることがあげられる。VGG16 の入力には画像である必要があるため動画を 1 フレーム毎画像化する処理が必要であった (図 1) が、その処理が不要となった。

YOLO によって 1 画面内で複数台の自動車の検出が可能となったことから、各自動車の座標の特定、自動車の大きさ判定により、自動車が写るフレームを特定する処理が必要となった。図 7 は処理を図化したものである。改良システムは、図 1 の ATVS v.1.0 で行っていた動画

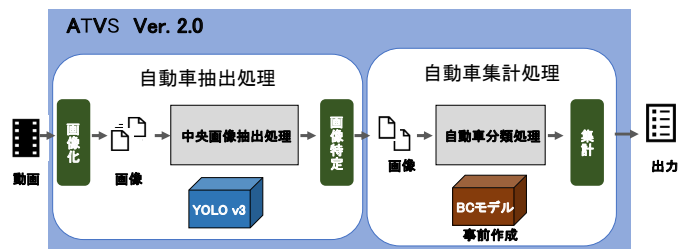


図 7 YOLO を使用した ATVS v.2.0

の画像化, 対向車線削除処理などが不必要になったため, さらにシステムの実行時間の高速化が実現された。

6. ATVS v.2.0 の評価

YOLO を用いた自動車抽出処理を評価するため、ATVS v.1.0 と同じ動画を使用して ATVS v.2.0 との結果を比較した。

6.1. ATVS v.2.0の結果

ATVS v.2.0 に、撮影した動画を入力し、自動車抽出処理と自動車分類処理 (3 種) を通して実行した。表 8 は正解と ATVS v.1.0 と ATVS v.2.0 の結果である。

正解は、バス 1 台、乗用車 74 台、貨物自動車 8 台で合計 83 台の自動車である。ATVS v.1.0 の結果ではバス 2 台、乗用車 65 台、貨物自動車 7 台で合計 74 台の自動車を検出した。ATVS v.2.0 の結果ではバス 1 台、乗用車 71 台、貨物自動車 13 台で合計 85 台の自動車を検出した。ATVS v.2.0 の方が、正解に近い結果が得られた。

6.2. ATVS v.1.0とv.2.0の比較検討

自動車抽出処理について詳細に評価を行った結果、中央で写る画像の正解は 83 枚であったが、VGG16 を使用する ATVS v.1.0 では 53 枚、YOLO を使用する ATVS v.2.0 では 83 枚の画像が抽出された (表 9)。正解数としては、ATVS v.2.0 が正解と同じ数となった。ただ ATVS v.2.0 では、撮影した道路を通過した全ての自動車は適切に抽出できていたが、不適切な画像が 2 枚含まれており合計 85 枚の画像を抽出していた。

自動車分類処理については、ATVS v.2.0 は ATVS v.1.0 からの変更はなく同じ BC モデルの分類器を用いているが、入力に使用した画面によって BC モデルの結果が異なるかを比較した (表 10)。比較として筆者が目視で画像抽出から分類まで全て行った結果 (これを正解とする) と目視で画像を抽出し、分類は BC モデルを使用した結果 (正解画像+BC) を載せる。

ATVS v.1.0 の場合、適切に抽出できた画像のうち、正しく分類まで行うことができた自動車は 47 台であった。

表 8 ATVS v.2.0 による分類 (3種) の結果

	正解	ATVS v1.0	ATVS v.2.0
バス	1	2	1
乗用車	74	65	71
貨物自動車	8	7	13
合計	83	74	85

表 9 自動車抽出処理結果の比較

	正解	VGG16	YOLO
抽出した画像数	83	74	85
自動車が中央で写る画像数	83	53	83

表 10 ATVS v.1.0 と ATVS v.2.0 の比較

	ATVS v.1.0	ATVS v.2.0	正解	正解画像+BC
自動車が中央で写る画像数	53	83	83	83
中央で写る画像・正しく分類	47	63	83	63
中央で写る画像・誤分類	6	20	0	20
不適切な画像抽出	21	2	0	0

ATVS v.2.0 の場合、中央に自動車が写る画像のうち、正しく分類まで行うことができた自動車は 63 台であった。

中央に自動車が写る画像のうち、誤分類してしまったものは、ATVS v.1.0 の場合 6 枚、ATVS v.2.0 の場合 20 枚であった。ATVS v.2.0 は v.1.0 と比較し、自動車の抽出処理において大きく精度を向上できたことがわかる。

また ATVS v.2.0 と正解画像 + BC モデルの「中央で写る画像・正しく分類」、「中央で写る画像・誤分類」の数が一致しているため、今後 BC モデルを改良することが求められることがわかった。

6.3. 評価のまとめ

改良を行った ATVS v.2.0 は、自動車が中央で写るフレームを抽出する処理に YOLO を用いることで、抽出処理に関する精度を向上させることができた。改良前のシステムは自動車の座標を特定することができないため、画像右側を再度 VGG16 に入力する処理が必要であったが、YOLO によって画像の複数の物体の座標取得が可能となったため、処理が簡素化され、適切な自動車が中央に写る画像抽出枚数の増加、また不適切画像数が減少できた (表 9)。

しかし、ATVS v.2.0 にはまだ自動車の分類を行う BC モデルによる画像分類に問題がある。BC モデルに入力された画像それぞれの分類結果を確認したが、ATVS v.2.0 による結果と目視で作成した正解画像 + BC モデルによる結果に大きな差がないことから、BC モデルに問題があることがわかり、今後 BC モデルの改良が必要であることがわかった。

7. おわりに

本論文では、既存の交通量調査の課題を解決することを目的として、画像識別器を用いた簡易な交通量調査自動化システム ATVS v.1.0 と、さらに改良を加えた ATVS v.2.0 を試作し、その評価結果について述べた。

ATVS v.2.0 は、自動車抽出処理に関しては実用で使用可能なレベルの結果を出すことができたが、自動車分類処理に関してまだ精度が低く今後の改良が必要となっている。

また、本システムは、手前の車線、反対車線を走行する自動車に対する交通量調査を同時に行う処理を追加することで今後さらに柔軟な交通量調査を可能にすることができると考えられる。また、今回試作したものは、非常に簡易なシステムであるが、今後自動車分類の精度を上げることができれば、防犯カメラなど別の目的で設置されたカメラ映像や、歩道橋など上部から撮影した映像からも交通量調査に応用できると考えられる。

参考文献

- [1] “一般交通量調査について”. <http://www.mlit.go.jp/road/census/h22-1/data/kasyorep.pdf>, (参照 2018-06-04).
- [2] 峯岸朋弥, 吉田享子. 画像識別器による簡易な自動車交通量調査システムの試作と評価. 情報処理, 2019, 2019-IS-147, no. 3, p. 1-4
- [3] “Very Deep CNNs for Large-Scale Visual Recognition”. http://www.robots.ox.ac.uk/~vgg/research/very_deep/, (参照 2018-12-17).
- [4] Redmon Joseph, Divvala Santosh, Girshick Ross and Farhadi Ali. You only look once: Unified, real-time object detection. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016, 2016-December, 779-788.
- [5] 谷育馬, 横井昭, 佐治斉. 幾何学的制約を用いた車両抽出. 画像電子, 2014, Vol.43, No.4, p. 579-587.
- [6] 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, 2015, 平成 27 年度電気・情報関係学会九州支部連合大会 (第 68 回連合大会) 講演論文集, p. 19-20.
- [7] 石田繁巳, 劉嵩, 三村晃平, 田頭茂明, 福田晃. Dynamic Time Warping を用いた路側設置マイクロフォンによる車両カウントシステムの設計と初期的評価. 情報処理, 2016, 2016-ITS-64, no. 6, p. 1-6.
- [8] 鈴木翔太, 小坂隆浩. クラウドソーシングを用いた交通量調査手法の提案. 情報処理, 2018, 2018-IS-143, no. 1, p. 1-2
- [9] “Tensor Flow”. <https://www.tensorflow.org/>, (参照 2018-12-17).