

物体検出を用いた交通量の自動調査に向けたモデルの比較

Comparison of object detection models for automatic traffic surveys

岡野 将大*¹
Masahiro Okano

大久保 順一*¹
junichi Okubo

小篠 耕平*¹
Kouhei Ozasa

菅原 宏明*¹
Hiroaki Sugawara

藤井 純一郎*¹
Junichiro Fujii

*¹ 八千代エンジニアリング株式会社
Yachiyo Engineering Co.,Ltd

Traffic surveys are carried by manual. For automation, traffic counting application of vehicle types using object detection technology has been developed. In previous studies, SSD300 with VGG16 as the base network was adopted, however assuming actual operation, the required speed for data processing and accuracy differ depending on the survey condition, and thus it is desirable to use the model properly according to the situation. This study compared the performance (the inference time and mAP) of learning the same data set with SSD models with three networks (VGG, MobileNetV1 and MobileNetV2). MobileNetV1 and V2 won the VGG with similar mAP, however V1 and V2 had similar inference time.

1. はじめに

人による交通量調査は、調査員の人件費や人員不足といった観点から、調査の実行が困難となっている[国総研 2002]。一方、機械による調査には、簡易型トラフィックカウンターを設置する方式があるが、機器が数十万円と高価で、車種の判別が不可能なため、費用対効果が小さい。そこで、路側に置いた市販のビデオカメラの映像から、ディープラーニングの物体検出技術によって交通量を自動計測する手法が開発された[菅原 2020] [大久保 2020]。従来の自動計測手法の物体検出モデルには、VGG16 をベースネットとした SSD300 を使用していたが、処理時間が長く、スマートフォン等の現場で簡便に計測できるデバイス上に、システムを実装することが困難であった[Wei Liu 2016][大久保 2020]。

従来手法では、SSD によって車両等を検出し、フレームごとに移動するバウンディングボックスの座標を追跡することで、交通量をカウントしていた[菅原 2020] [大久保 2020]。そのため、動画を解析時のフレームレートを下げると、車両の見逃し等が発生し、交通量カウント精度が低下する。よって処理性能の低いデバイスを用いて、現場でリアルタイムに解析するシステムを開発することを想定すると、可能な限り高いフレームレートでの処理を実行可能にしなければならない。したがって、より推論速度が速い検出モデルへの転換を検討する。

推論速度の速いモデルとしては、MobileNetV1 や MobileNetV2 をベースとした SSD が開発されている[Howard 2017] [Sandler 2018]。MobileNetV1 では、VGG16 ベースの SSD より 1% 程度の正解率の低下で、畳み込み演算の総計算量を約 10% にまで削減している。さらに、MobileNetV2 では、MobileNetV1 と比較して最大 30% の推論時間を削減可能であることが示されている。MobileNetV1 と MobileNetV2 では、推論時間削減のための機構が異なるため、本研究では既往研究で用いた VGG16 ベースの SSD300 に対して、MobileNetV1、MobileNetV2 をベースネットとする SSD300 を同条件下で学習し、それぞれの推論精度と速度をデスクトップ環境で比較する。

2. 実験内容

2.1 データセット

本研究で用いるデータセットは、道路行政関連調査[国総研 2002]で使用されている車種区分を考慮し、7 クラスとする。データセットにおける各クラスのデータ数の比率を図-1 に示す。データセットの作成は、[大久保 2020]を基に、比率の偏りを可能な限り減らした。すなわち、最もアノテーションデータの少なかったバスの数に合わせて、アンダーサンプリングし、Train 対象物データ 3168 個、Validation 対象物データ 1454 個、合計 4622 個の対象物データセットを作成した。

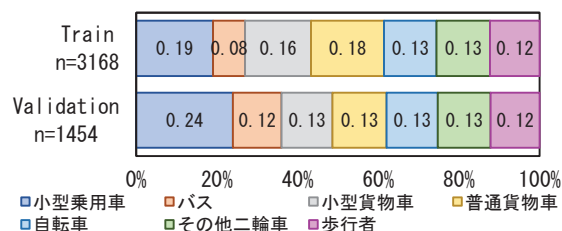


図-1 データセット区分割合

2.2 実験方法

図-1 のデータセットによって、VGG16, MobileNetV1, MobileNetV2 をベースネットとした 3 つの SSD モデルの学習を行う。学習時は Train データのみ水平反転、色調変化、拡大、切り抜きの Augmentation を実施した。その他の学習条件は実験条件等を統一する目的で、同じとした。具体的にバッチサイズは 32, epoch 数は 500, Loss 関数は localization Loss と confidence Loss の重み付き和、最適化関数は学習率 0.001 の SGD とした。

推論精度の評価指標は validation データでの mAP (mean Average Precision) とし、IoU (Intersection over Union) が 0.5 以上のものを正解としている。評価時のモデルは、学習時に Validation の loss が最も低下したものを採用した。推論速度は、インテル® Xeon® スケールブル・プロセッサによって推論した際の、画像 500 枚分の推論時間の平均値によって評価する。

2.3 学習結果

図-2 にそれぞれのモデルの mAP (mean Average Precision) の計測結果を示す。それぞれのクラスにおいて、mAP が最高となるモデルは異なり、一貫して高精度と評価できるモデルはなかった。全クラスの mAP の平均では、VGG16, MobileNetV1, MobileNetV2 という順で高精度となったが、その差は 1% 以下であり、目立った精度低下は見られなかった。また、いずれのモデルにおいても歩行者の精度が低く、0.2 を下回った。VGG16 と比べ MobileNetV1, MobileNetV2 は、各クラスの mAP も小型乗用と自転車を除けば概ね同じ精度を得ていた。

図-3 にそれぞれのモデルの推論速度を示す。MobileNetV1 および MobileNetV2 をベースとする SSD は、VGG16 を使用した SSD に比べ、推論速度が約 5 倍向上した。しかし、MobileNetV1 と V2 の間に推論速度の大きな差はなかった。

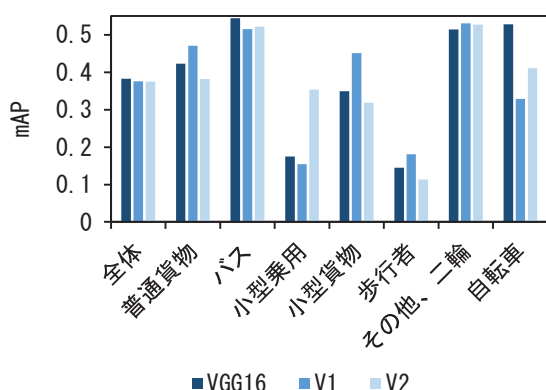


図-2 mAP の計測結果

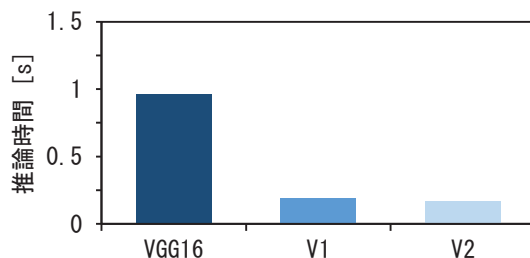


図-3 モデルごとの推論速度

3. 考察

今回の実験では、ベースネットワークを MobileNet に変更することで、精度を維持し、推論速度を上げることができた。しかし、各クラスの mAP が最も高いモデルはそれぞれ異なる結果となった。この結果から、それぞれのモデルでクラス間の特徴量が明確に分離できておらず、学習が安定していないと予想される。例えば、歩行者の mAP が低下したのは、自転車に乗る人との特徴が混同したためと考えられる。また、SSD の特徴である小さい Object の検出のしづらさもこの結果に結びついていると考えられる。

MobileNetV2 の推論速度の向上が芳しくなかった理由として、推論環境の違いがある。[Sandler 2018]は、推論環境を Android の端末に最適化した形で検証を行っており、デスクトップ環境での検証は主眼とされていなかった。そのため、本環境での速度向上には繋がらなかったと考えられる。

4. 最後に

本研究では、VGG16, MobileNetV1, MobileNetV2 のそれぞれをベースネットにした SSD300 で、道路行政関連調査に用いられる 7 車種を検知するモデルを作成し、それぞれの精度と推論速度を検証した。以下に得られた知見を示す。

- (1) MobileNetV1 と MobileNetV2 をベースにした SSD は、VGG16 をベースにした SSD よりも推論速度が約 5 倍向上した。mAP は 1% 以下の差であり、実用上でモデルを変更することが推奨される。
- (2) MobileNetV1 と比べても精度・推論速度が大きく向上すると考えられた MobileNetV2 は、あまり変わらない結果となった。

今後の課題として、各モデルで最も mAP の高いクラスが大きく異なる点について、原因を分析することが考えられる。各クラスで最も高い mAP を示したモデルがどのような検知を行っているか、その他 2 つのモデルがどのような誤検知・検知漏れを生じているか、各画像データに対して、分析することが考えられる。

データセットを拡張した際の検証も必要がある。今回、各クラスのアノテーション数は 800~2400 程度であったことから、車種の偏りも考えられる。多地点での撮影データを追加することで、今回の再現性を確認すべきである。

実際の運用を考慮すると、ポータブルデバイスでの実装試験も行う必要がある。このことから、実際にポータブルデバイス上で本実験を再試験し、[Sandler 2018]での条件で改善効果が再現可能か検証を要すると考えられる。

参考文献

- [大久保 2020] 大久保 順一, 菅原宏明, 藤井純一郎, 小篠耕平: 車種別交通量調査における物体追跡法の改良, 土木学会 AI シンポジウム, 2020.
- [国総研 2002] 国土技術政策総合研究所: 道路センサス (全国道路・街路交通情勢調査) の変遷と改善に関する調査, 第 3 章, pp.12-15, 2002.
- [菅原 2020] 菅原宏明, 小篠耕平, 藤井純一郎, 大久保順一, 小早川悟: 路側に設置した簡易撮影機器の動画による交通量自動計測に関する基礎的研究, 第 40 回交通工学研究発表会論文集, 2020.
- [Howard 2017] Andrew G. Howard, Menglong, Zhu Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam: Mobilenets: Efficient convolutional neural networks for mobile vision applications, arXiv:1704.04861, 2017.
- [Liu 2020] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg: SSD: Single Shot MultiBox Detector, arXiv:1515.02325, 2016.
- [Sandler 2018] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, Liang-Chieh Chen: MobileNetV2: Inverted Residuals and Linear Bottlenecks, arXiv:1801.04381, 2018.