

画像認識を用いた車種別交通量計測システムにおける推論精度と速度の関係

八千代エンジニアリング(株) 正会員 ○藤井純一郎, 小篠耕平, 菅原宏明

八千代エンジニアリング(株) 非会員 大久保順一, 岡野将大

1. 背景・目的

近年、道路交通量を自動で計測する方法としてカメラ等の画像から交通量を読み取るシステムが開発されている。筆者らは、路側に設置し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車種別交通量を自動計測するシステムTRAVICを開発してきた。TRAVICでは物体検出モデルとしてSingle Shot MultiBox Detector (以下SSD)²⁾を採用しており、SSDのベースとなる画像特徴量抽出にはVGG16を用いている。SSDによって車両等を検出し、フレームごと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る。そのため、動画の解析時のフレームレートが高いほど、車両の見逃し等が発生せず、交通量のカウント精度が高くなる。

一方で、現地でのリアルタイム計測を求められる場合、高性能なPCを用いることが難しいため、より処理能力の低いエッジデバイスやスマートフォンで動作することが求められる。しかし、VGG16ベースのSSDでは推論時間を要し、高いフレームレートを維持できない。推論速度の速いモデルとしては、MobileNetV1³⁾やMobileNetV2⁴⁾をベースとしたSSDが開発されている。ベンチマークデータを用いた実験では、MobileNetV1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%にまで削減したと報告されている。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。

そこで本研究では、路側からの撮影動画に対する車種別物体検出を対象として、従来のVGG16ベースのSSDに対して、MobileNetV1、MobileNetV2をベースネットとするSSDを同条件下で学習し、それぞれの推論精度と速度を比較する。

2. 実験内容

(1) データセットの作成

本研究では、図-1に示す方法で路側から撮影した動画から切り出した画像に対して、「全国道路・街路交通情勢調査」において採用されている自動車類4車種区分を含む7クラス(乗用車・バス・普通貨物・小型貨物・自転車・その他二輪・歩行者)のアノテーションを行った対象物データセットを作成した。

データセットにおける各クラスのデータ数の比率を図-2に示す。車種別の比率の偏りを可能な限り減らすため、最もアノテーションデータ数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、学習用3168個、検証用1454個、合計4622個の対象物データセットを作成した。



<撮影方法> <動画イメージ>

図-1 入力対象動画

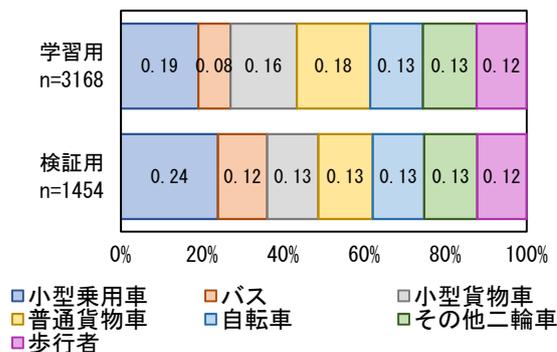


図-2 データセットの比率

(2) 実験方法

図-2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モデルの学習を行った。学習時は、学習用データのみ水平反転、色調変化、拡大、切り抜きのデータ拡張を実施した。ハイパーパラメータ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、検証用データでのmAP (mean Average Precision) とし、IoU (Intersection over Union) が0.5以上のものを正解とした。評価時のモデルは、学習時に検証用データでの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®プロセッサを搭載したデスクトップ環境で推論した際の、画像500枚分の推論時間の平均値によって評価した。

キーワード 交通量, 画像認識, 人工知能, 深層学習, 物体検出

連絡先 〒111-8648 東京都台東区浅草橋5-20-8 CSタワー 八千代エンジニアリング(株) TEL:03-5822-6626

(3) 学習結果

図3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、クラスに依らず高精度と評価できるモデルは無かった。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4にそれぞれのモデルの推論時間を示す。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。

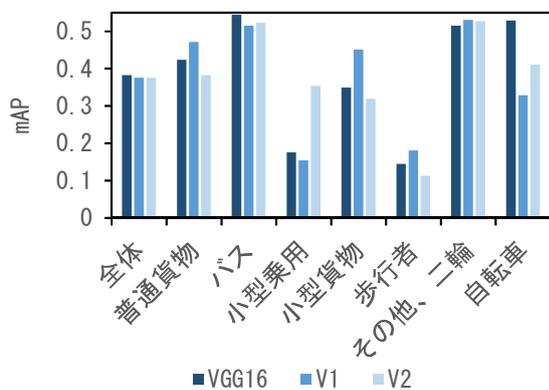


図-3 mAPの計測結果

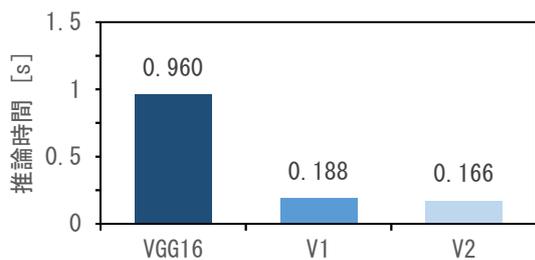


図-4 モデルごとの推論速度

(4) 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。

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。MobileNetV2 では、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い。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。

3. まとめと今後の課題

本研究では、VGG16, MobileNetV1, MobileNetV2 のそれぞれをベースネットにした SSD で、7車種を検出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、実用上はモデルを変更しても問題ない。

(2) MobileNetV1 と比べても精度・推論速度が大きく向上すると予測した 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、原因を分析する必要がある。各クラスで最も高い mAP を示したモデルがどのような検出を行っているか、その他 2つのモデルがどのような誤検出・未検出を生じているか、各画像データに対して分析することが課題である。

また、今回は限られた地点で撮影した動画から教師データセットを作成したが、他地点での撮影データを用いた検証も必要である。複数地点での撮影データを教師データセットに追加することで、本研究と同等の推論精度と速度が得られるか検証することが課題である。

参考文献

- 1) 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.
- 2) Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg: SSD: Single Shot MultiBox Detector, arXiv: 1512.02325, 2016.
- 3) 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.
- 4) Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, Liang-Chieh Chen: MobileNetV2: Inverted Residuals and Linear Bottlenecks, arXiv: 1801.04381, 2018.