スマートフォンカメラによる道路上のマルチオブジェクト検出

前田紘弥・樫山武浩・須藤明人・関本義秀

Abstract:近年，自動運転に代表されるように，カメラ映像から道路上の状況を把握しようという試みが多くなされているが，それらは比較的高性能なカメラ等を用いることが多い．一方で，スマートフォンカメラなど簡易なデバイスのみによる道路上のオブジェクト検出の試みもあるが，精度等の面で物足りない．そこで本研究では，深層学習を用い，高精度でスマートフォンカメラだけで道路上の状況把握をする手法を開発した．

Keywords: スマートフォン（smartphone），深層学習（deep learning），マルチオブジェクト検出（multi object recognition）

1. はじめに

近年，パターン認識の分野で深層学習と呼ばれる多層ニューラルネットワークを用いた機械学習手法が，従来手法よりも高い精度を示すなど，注目を集めている．このような状況の中で，自動運転に代表されるように，高精細なカメラ映像から道路上のマルチオブジェクトを自動検出するという研究が数多くなされている（史ほか, 2010など）．また，行政のインフラ維持管理業務においても，膨大な時間的・財政的コストを軽減するため，カメラ映像からインフラの損傷を把握しようとする試みも見受けられる（CHUN ほか, 2015など）．

一方で，スマートフォンなどモバイル端末の性能の進歩は著しく，昨今では高度な演算を高速で実行することが可能になっている．また，スマートフォン端末の普及率は極めて高く，総務省白書によると日本国内では2015年度には6割を超え，世界でも50%を超えるとされている．

そこで，世界中で広く普及している高性能なスマートフォン端末上で深層学習による高精度な画像処理技術を用いることが応用としてまず考えられる．例えば，世界中で広く普及しているスマートフォン端末のみで，深層学習等を利用した高精度なインフラ維持管理を実現できれば，行政にとって大きな予算削減に繋がるだろう．

しかし，基本的に深層学習の分野では大規模GPU サーバーによる処理を前提としており，スマートフォンなどハードウェアリソースが限られるデバイスでの実行は難しい．例えば，Chen ら（2015）では，標識等をスマートフォンカメラでリアルタイム認識するために，取得した画像を一時的にサーバーに転送し，処理を行っている．

本研究では，深層学習による画像認識システムをスマートフォン側でうまく実行・処理することを目指す．特に，車載スマートフォンカメラの映像から道路上の標識というオブジェクトの自動検出に焦点を当てて研究を進める．

2. 手法の概略

標識をはじめとした道路上のオブジェクト検出としては，標識の色に注目したものや（Xiong ほか， 2012），機械学習手法を適用したものなど多くの研究がなされている（Brkic ほか, 2010）．深層学習を適用したものとしては，Sermanet ら（2011）や CireşAn ら（2012）などがある．これら論文によると，99%を超える極めて高い精度で判定可能であると報告されているが，スマートフォン上での実行速度に言及した例はない．

前田紘弥 〒153-8505 東京都目黒区駒場 4-6-1
東京大学生産技術研究所関本研究室
Phone: 03-5452-6406
E-mail: maedahi@iis.u-tokyo.ac.jp

2.1 深層学習について

近年，畳込ニューラルネットワーク(CNN)を用いた画像認識と画像分類における研究が数多くなされている．CNN の良い点は特徴量の重みだけでなく，特徴量そのものも学習することができる点である．CNN はここ最近で最も高い判定精度を実現している．深層学習を用いた画像処理などについては，Srinivas ら（2016）に詳しい．

ただし，CNN は，学習させるパラメータ数が多く，学習済みのモデルが数百 MB を超えることも珍しくない．また，パラメータ数の分だけ，学習にも判定にも時間がかかってしまう．よって，学習済みモデルをスマートフォン上で実行させることを考えると，大きな容量・パラメータ数がボトルネックとなり，処理が遅くなってしまう．このような背景で CNN の構造をできるだけ軽量化しようという試みが多くなされている（Iandola ほか， 2016 など）．本研究では画像処理に特化した深層学習の一般的なフレームワークである Caffe を用いてモデルを学習させる．

2.2 モデル圧縮と高速化

モバイル端末上で，深層学習による学習済みモデルを使用するためには，モデルそのものの圧縮と演算処理の高速化が必須である．モデルそのものの軽量化については Gong ら（2014）や Han ら（2015）などがある．Gong らは vector 量子化によるパラメータ圧縮を行い，ほぼ精度を落とすことなく 20 倍程度モデル容量を圧縮した．畳込層の行列演算の速度向上については，Courbariaux ら（2015）などがある．Courbariaux らは順伝播・逆伝播における重みを 2 値化することで，必要な演算を 1/3 にした．このように多くの研究がなされているが，本研究では画像サイズが 32×32 ピクセルと小さかったため，モデル圧縮をせずとも，畳込層のフィルタ数と全結合層のニューロン数のみを調整するだけで，スマートフォン上で十分な速度を出すことができた．ただし今後はモデル圧縮・高速化なども検討していきたい．

2.アンドロイド上への実装

本研究では深層学習による学習済みモデルをアンドロイドアプリに組み込むために，Java Native Interface (JNI)を通して，C++のコードを呼び出す OpenCV のライブラリを用いる． OpenCV には学習済み Caffe モデルを読み込むためのライブラリが用意されている．

3. 適用例

3.1 教師データ作成

本手法では，標識の教師画像として Ruhr-Universitt Bochum の Real-time Computer Vision グループが公開している The German Traffic Sign Recognition Benchmark(GTSRB)の画像を用いる． GTSRB データベースはドイツの道路標識（43 クラス）を含み，合計 50,000 枚を超える画像が格納されている．標識の外接枠の座標データも公開されているので，この座標データを用いて標識そのもの以外の背景を除去した画像を教師データとして使用する．特に本研究で使用したのは，GTSRB に含まれる図- 1 に示す 6 種類の道路標識である．各クラスが 1000 枚ずつ（学習用 750 枚，評価用 250 枚）になるようにランダムに画像を抽出した．また，標識でない画像データとして，図- 2 のように，別途車載スマートフォンカメラで撮影した画像から，ランダムに 1000 枚の画像をクロップし，教師データとした（学習用 750 枚，評価用 250 枚）．最後にすべての教師データを 32×32 ピクセルになるようにリサイズした．

図- 3　教師データ例（6 クラス，各 1000 枚ずつ）

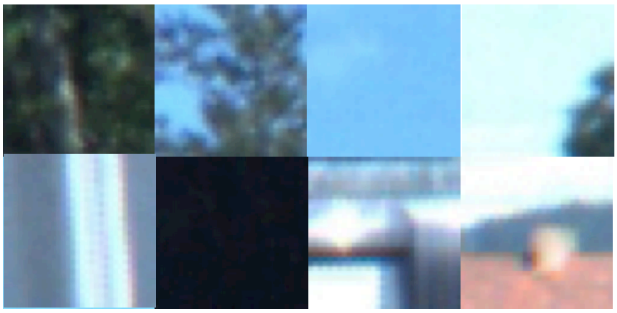

図- 4　教師データ例
（標識ではない画像データ 1000 枚）

3.2CNN の構造とモデルの学習

CNN には一般物体認識のベンチマークとしてよく使用されている CIFAR-10 という画像データセットの分類でよく使われるモデルを使用する．より詳細に説明すると，32×32 ピクセルの入力画像に対して，畳み込み・プーリングの処理を 3 回ずつ行う 7 層の CNN である（表- 1）．標識検出には色情報も重要であると考えられるため，RGB のカラー画像を入力としている．

表- 1　　CNN の構造

Layer	Type	# maps	kernel
0	input	3 maps	
1	convolution	32 maps	5*5
2	max pooling	32 maps	3*3
3	convolution	32 maps	5*5
4	average pooling	32 maps	3*3
5	convolution	64 maps	5*5
6	average pooling	64 maps	3*3
7	fully connected	7 neurons	

3.3 結果

上述の学習データ 5250 枚（7 クラス，各 750 枚）でモデルを学習させ，評価データ 1750 枚（7 クラス，各 250 枚）で評価を行った時の様子が図- 3 に示されている．最終的なテスト精度は 98%であった．およそ 200iteration 前後で精度が収束している様子がわかる．学習済みモデルは 350KB であった．またこの時 1000iteration 学習を行ったモデルをアンドロイド上で実行させた場合，32×32 ピクセルの画像 1 枚を判定するのに，45 ナノ秒を要した．本研究で使用したアンドロイド端末は Motorola Moto G である．

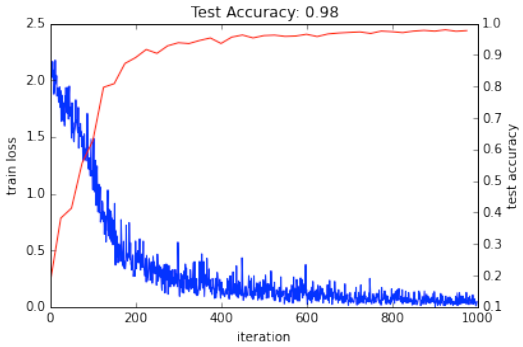

図- 5　学習誤差（青）とテスト精度（赤）

3.3 結果の評価

本研究では，比較的軽量なモデルを用いることで，45 ナノ秒という非常に速い処理速度で標識を判別することができた．一般に機械学習手法を用いたオブジェクト抽出では，学習させたオブジェクトを入力画像内で窓を動かして探索していく．リアルタイムで処理を行うことを考えると，30fps 程度以上の処理速度が求められる．45 ナノ秒で判定可能ということは，10^5 回以上判定をしてもリアルタイム処理できる条件を満たしていると言える．本研究では取り組むことができなかったが，処理速度を向上させるため，図-4 のオレンジ色の部分だけ探索するというように探索領域を限定する工夫をすることが考えられる．スマートフォンカメラを用いた白線検出に関しては Chanawangsa ら（2012）などがある．

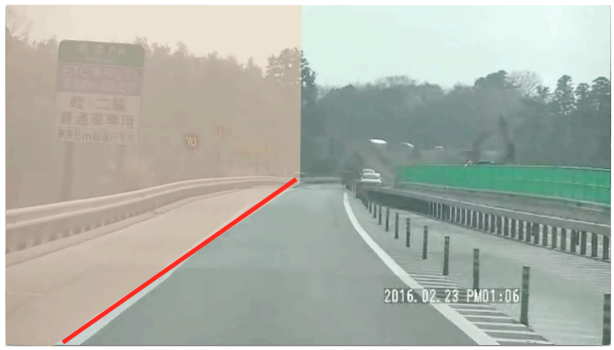

図- 6　手法イメージ

4. おわりに

本研究では，3.1 節で述べたような 7 クラス分類モデルを作成した．ただし，本来標識の種類はより多岐に渡る．認識可能な標識の種類を増やすことと，本研究で得られた知見を活かし実用に耐えうるスマートフォンアプリを開発することが今後の課題である．世界中の多くの人が保有するスマートフォンのみで道路上のオブジェクトを検出できることは，自動運転のみならず，昨今のインフラ維持管理を取り巻く厳しい情況を打破する突破口となりうるものである．

参考文献

史中超，史云，& 柴崎亮介．2010，車載ステレオ画像とレーザデータの融合による道路標識の自動抽出手法の開発．*写真測量とリモートセンシング*，*49*(2)，75-82.

CHUN，P. J.，& 橋本和明．2015，アスファルト舗装撮影画像からのひび割れ半自動検出システムの開発．*土木学会論文集 E1（舗装工学）*，*71*(3)，I_31-I_38.

Chen，T. Y. H.，Ravindranath，L.，Deng，S.，Bahl，P.，& Balakrishnan，H.，2015，Glimpse: Continuous，real-time object recognition on mobile devices. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems* (pp. 155-168). ACM.

Xiong，B.，& Izmirli，O.，2012，A road sign detection and recognition system for mobile devices. In *2012 International Workshop on Image Processing and Optical Engineering* (pp. 83350B-83350B). International Society for Optics and Photonics.

Brkic，K.，2010，An overview of traffic sign detection methods. *Department of Electronics，Microelectronics，Computer and Intelligent Systems Faculty of Electrical Engineering and Computing Unska*，*3*，10000.

CireşAn，D.，Meier，U.，Masci，J.，& Schmidhuber，J. (2012). Multi-column deep neural network for traffic sign classification. *Neural Networks*，*32*，333-338.

Srinivas，S.，Sarvadevabhatla，R. K.，Mopuri，K. R.，Prabhu，N.，Kruthiventi，S. S.，& Babu，R. V.，2016，A Taxonomy of Deep Convolutional Neural Nets for Computer Vision. *arXiv preprint arXiv:1601.06615*.

Iandola，F. N.，Moskewicz，M. W.，Ashraf，K.，Han，S.，Dally，W. J.，& Keutzer，K.，2016，SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 1MB model size. *arXiv preprint arXiv:1602.07360*.

Krizhevsky，A.，Sutskever，I.，& Hinton，G. E.，2012，Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

Gong，Y.，Liu，L.，Yang，M.，& Bourdev，L.，2014，Compressing deep convolutional networks using vector quantization. *arXiv preprint arXiv:1412.6115*.

Han，S.，Mao，H.，& Dally，W. J. (2015). Deep compression: Compressing deep neural network with pruning，trained quantization and huffman coding. *CoRR，abs/1510.00149，2*.

Courbariaux，M.，Bengio，Y.，& David，J. P.，2015，Binaryconnect: Training deep neural networks with binary weights during propagations. In *Advances in Neural Information Processing Systems* (pp. 3123-3131).

Chanawangsa，P.，& Chen，C. W. (2012，February). A new smartphone lane detection system: realizing true potential of multi-core mobile devices. In *Proceedings of the 4th Workshop on Mobile Video* (pp. 19-24). ACM.