社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

［ショートペーパー］歩行者検出におけるグランドチャレンジ

藤吉　弘亘†　　山下　隆義†　　山内　悠嗣†

† 中部大学 工学部　〒 487–8501 愛知県春日井市松本町 1200
E-mail: †{hf,yamashita}@cs.chubu.ac.jp, ††yuu@isc.chubu.ac.jp

あらまし　大規模画像認識のコンペティション ILSVRC における 1,000 クラス画像分類問題では，2012 年より深層学習によるアプローチが上位を占め，2015 年に発表された 152 層からなる ResNet は人間の能力を超えた認識性能を獲得した．歩行者検出においても同様に深層学習のアプローチが数多く提案されている．歩行者検出問題は，歩行者クラスと背景に分離する 2 クラス分類問題であるが，その認識性能は人間の性能に達していない．本稿では，これまでの歩行者検出技術を振り返りつつ，歩行者検出タスクにおける問題を共有し，歩行者検出におけるグランドチャレンジについて議論したい．

キーワード　歩行者検出，深層学習，グランドチャレンジ

1. はじめに

大規模画像認識のコンペティション ILSVRC における 1,000 クラス画像分類問題では，2012 年より深層学習によるアプローチ [1]〜[3] が上位を占め，2015 年に発表された ResNet [3] は 152 層の構造を持ち，そのエラー率は 3.56%である．同様のタスクを人間が行った場合，そのエラー率は 5.1%であり，深層学習のアプローチは人間の能力と同等以上の認識性能を獲得した．また，顔照合問題においても深層学習ベースの DeepFace [4] は，440 万枚を用いた学習により人間と同等の照合精度を達成したことが報告されている．このように，多くの画像認識タスクにおいて，深層学習のアプローチが人間と同等以上の認識性能を獲得するようになった．さらには，一般物体認識の最終タスクともいえる画素単位にカテゴリ識別を行うセマンティックセグメンテーションも深層学習により解かれようとしている [5]．

本稿で扱う歩行者検出は物体検出の一タスクであり，対象カテゴリの物体の位置と大きさを検出する問題である．歩行者検出においても同様に深層学習のアプローチが数多く提案されている．本稿では，これまでの歩行者検出技術を振り返りつつ，歩行者検出タスクにおける問題を共有し，歩行者検出におけるグランドチャレンジについて述べる．

2. 深層学習による歩行者検出

2005 年に Dalal らは歩行者の勾配に着目する特徴量 Histogram of Oriented Gradient (HOG) [6] を提案し，Support Vector Machine (SVM) と組み合わせることで歩行者検出を実現した．2009 年には，HOG 特徴量だけでなく，色情報やエッジ情報を特徴量に加えるチャンネル特徴量を用いた ICF が提案された [7]．2013 年以降に，深層学習を用いた手法が数多く提案され始めた．深層学習による歩行者検出の先駆けとなる Joint Deep Learning [8] は，入力画像をネットワークに入力することで特徴マップを獲得し，パーツ領域ごとに後段の Restricted Boltzmann Machine (RBM) へ入力することで歩行者を検出している．その後，Luo らは歩行者検出に有効なパーツ領域を学習により選択する Switchable Deep Network (SDN) [9] を提案した．2015 年には，深層学習と複数の歩行者検出法を組み合わせた Complexity Aware Cascade Training (CompACT) [10] が提案された．2016 年には，Region Proposal を用いた歩行者検出法 [10]〜[12] や，Single Shot MultiBox Detector (SSD) [13] をベースにした歩行者検出法 [14] が提案された．図 1 に歩行者検出性能の遷移を示す．深層学習による歩行者検出の詳細については文献 [15] を参照されたい．

3. Caltech Pedestrian Benchmark における評価と課題

Caltech Pedestrian Benchmark [16] は，車載カメラで撮影した走行データから構成されているデータセットである．Caltech Pedestrian Dataset は，一般道や高速，トンネルなどの場所で撮影された約 6 時間の走行映像から構築されている．データセットには，歩行者が約 35 万枚存在している．データ数やデータに対するラベル情報，評価ツールが充実しているため，主要な歩行者検出のデータセットとして用いられている．

図 2 に主な手法の DET (Detection Error Tradeoff) カーブを示す．各手法の miss rate は，FPPI (False Positives Per Image) = 0.1 で比較する．グラフ中の Human baseline は，文献 [17] により示された人間による未検出率であり，1%以下であることがわかる．一方，深層学習ベースの歩行者検出法の未検出率は約 10%であり，人間と比べて約 9%低い．物体検出タスクは歩行者とそれ以外の背景である 2 クラス分類問題として解かれているが，背景クラスには 1,000 クラス以上のカテゴリが含まれる可能性があり，難しい問題であることを示している．対象カテゴリのみを検出する問題は，検出対象と背景を分離す

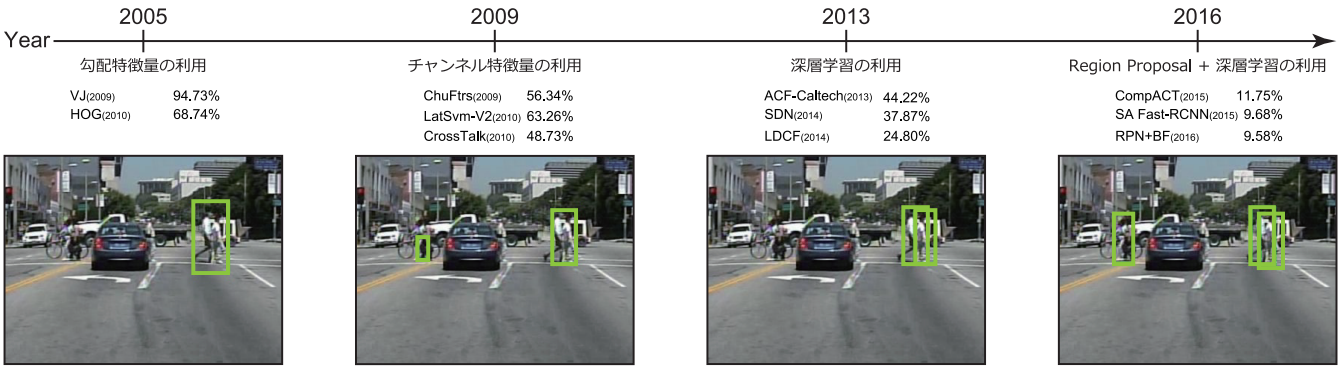

図 1　歩行者検出性能の遷移.

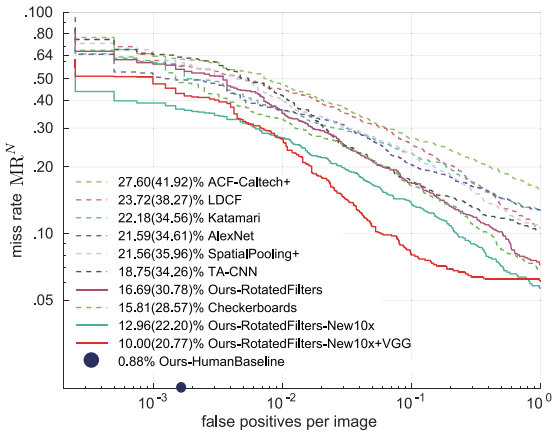

図 2　Caltech テストデータでの歩行者検出性能 (文献 [17] より引用).

る問題である．ILSVRC の多クラス分類において，人間よりも深層学習の認識性能が高いのは，人が多クラス分類に向いていないからとも言える．検出対象が 1 カテゴリであれば，人間による検出は画像中の歩行者をほぼ検出できる．今後，この差がどのような要因で発生しているのかを解明していくことは大きな課題であろう．Caltech Pedestrian dataset の学習サンプル数は，約 8 万枚である．この学習サンプル数がさらに 10 倍となれば解決できるのか，もしくは現在の深層学習のフレームワークで人間の認識性能に迫ることは可能なのであろうか．

4. ま と め

本稿では，これまでの歩行者検出技術を振り返りつつ，歩行者検出タスクにおける問題を明確にした．グランドチャレンジとしては地味な課題であるかもしれないが，歩行者検出における人間の認識性能との差の解明と，この差を縮めていくことに取り組むことが必要であろう．

文　献

[1] A. Krizhevsky, I. Sutskever, and G.E. Hinton, “Imagenet classification with deep convolutional neural networks,” Neural Information Processing Systems, pp.1097–1105, 2012.

[2] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” IEEE Conference on Computer Vision and Pattern Recognition, pp.1–9, 2015.

[3] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” IEEE Conference on Computer Vision and Pattern Recognition, pp.770–778, 2016.

[4] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, “Deepface: Closing the gap to human-level performance in face verification,” IEEE Conference on Computer Vision and Pattern Recognition, pp.1701–1708, 2014.

[5] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation,” arXiv preprint arXiv:1511.00561, 2015.

[6] N. Dalal and B. Triggs, “Histograms of Oriented Gradients for Human Detection,” IEEE Conference on Computer Vision and Pattern Recognition, vol.1, pp.886–893, 2005.

[7] P. Dollár, Z. Tu, P. Perona, and S. Belongie, “Integral Channel Features,” British Machine Vision Conference, pp.91.1–91.11, 2009.

[8] W. Ouyang and X. Wang, “Joint deep learning for pedestrian detection,” IEEE International Conference on Computer Vision, pp.2056–2063, 2013.

[9] P. Luo, Y. Tian, X. Wang, and X. Tang, “Switchable Deep Network for Pedestrian Detection,” IEEE Conference on Computer Vision and Pattern Recognition, pp.899–906, 2014.

[10] Z. Cai, M. Saberian, and N. Vasconcelos, “Learning Complexity-Aware Cascades for Deep Pedestrian Detection,” IEEE International Conference on Computer Vision, pp.3361–3369, 2015.

[11] J. Li, X. Liang, S. Shen, T. Xu, and S. Yan, “Scale-aware Fast R-CNN for Pedestrian Detection,” arXiv preprint arXiv:1510.08160, 2015.

[12] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” Neural Information Processing Systems, pp.91–99, 2015.

[13] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-y. Fu, and A.C. Berg, “SSD : Single Shot MultiBox Detector,” European Conference on Computer Vision, pp.1–15, 2016.

[14] X. Du, M. El-Khamy, J. Lee, and L.S. Davis, “Fused DNN: A deep neural network fusion approach to fast and robust pedestrian detection,” arXiv preprint arXiv:1610.03466, 2016.

[15] 福井宏，小芝駿王，山下隆義，山内悠嗣，藤吉弘亘，“[サーベイ論文] Deep Learning を用いた歩行者検出の研究動向,” パターン認識・メディア理解研究会，2016.

[16] P. Dollar, C. Wojek, B. Schiele, and P. Perona, “Pedestrian detection: An evaluation of the state of the art,” IEEE transactions on pattern analysis and machine intelligence, vol.34, no.4, pp.743–761, 2012.

[17] S. Zhang, R. Benenson, M. Omran, J. Hosang, and B. Schiele, “How Far are We from Solving Pedestrian Detection?,” IEEE Conference on Computer Vision and Pattern Recognition, pp.1259–1267, 2016.