Vol.2018-AAC-7 No.8
2018/8/25

深層学習による物体検出を用いた視覚障害者の屋外活動支援システムにおけるデザイン指針の検討とプロトタイピング

馬場 哲晃[1,a)]　渡邉 英徳[2]　釜江 常好[3]

概要：本稿では，深層学習を利用したリアルタイム物体検出を，視覚障害者の屋外活動支援システムに応用する．物体検出にはいくつかのアルゴリズムがある中，検出精度と高速な実行時間のバランスを取る必要がある．まずはいくつかの検出アルゴリズムを試した後，SSD および YOLO を利用した物体検出プロトタイプをスマートフォン上で実装した．ユーザはスマートフォンと白杖を利用して，遠方の物体情報をスマートフォンを利用することで実時間取得が可能である．さらにデータセット自体をユーザ参加型で作成可能にする他，GPS 情報と連動した重みファイルの共有機能を開発することで，ユーザの地元（Local）における最適化（Optimization）を重みデータに対して実現可能であると考えている．本支援システム開発の初期段階として，検出アルゴリズムやデータセット構築，インタラクションデザインに関して検討を行い，体験価値を提供可能な初期プロトタイプまでのデザインプロセスに関して注意深く述べる．

キーワード：視覚障害，支援技術，アクセシビリティ，深層学習，物体検出，プロトタイピング

Tetsuaki Baba[1,a)]　Hidenori Watanave[2]　Tsuneyoshi Kamae[3]

1. 背景

本研究で対象とする視覚障害者の屋外活動支援においては，盲導犬やガイドヘルパーによって当事者の支援が可能であるが，育成問題 [1] や介助者への気遣い等の，社会福祉法人日本盲人会連合のアンケートによれば，単独で外出できると回答した視覚障害者の 53 %は弱視であり，特に全盲の障害者に対してこの単独歩行支援は重要な問題である [2]．視覚障害者の歩行支援デバイスの観点からは Electronic Travel Aids（ETAs）に関する研究が 1960 年代より報告されている．超音波センサやレーザーによるセンシング機能により障害物情報を音声や振動情報としてユーザに提示するものが一般的であるが，多くはユーザビリティの低いものが多く，実際に利用されているものは少ない一方で，90 年代から Computer Vision（以下 CV）を活用した ETAs の研究報告がなされている [3]．これまでは複数のセンサを利用して周辺情報を取得しているのに対し，CV ベースな ETAs の場合，プロセッサとカメラがあれば基本的なシステム設計ができる点に利点がある．本研究ではこの点に着眼し CV ベースな ETAs デバイスをスマートフォンで代用することで，当事者が手軽に使えるシステムを目指す．

近年の深層学習による発展を振り返ると，2012 年に Deep Learning による画像識別手法が他の機械学習手法よりも高スコアを獲得したことで [4]，とりわけ CV 領域において物体検出，画像キャプション生成，スタイル変換，画像生成等多くの手法が実用性を伴って発表されている．CV を利用した Assistive Technology（以下 AT）に関する論文もいくつか報告され始めており，近年では深層学習により，AT 分野が大きく進展する可能性が示唆されている [5]．

深層学習においてデータセットの作成が重要であることはよく知られているが，すでに ImageNet[*1] や COCO[6] 等に代表されるデータセットを学習させることで，汎用的な物体検出器開発は比較的容易になった．一方で著者らの想定するユーザシナリオでは，横断歩道，歩行者用押しボタ

[1] 首都大学東京
Tokyo Metropolitan University, Asahigaoka, Hino, Tokyo 191–0065, Japan
[2] 東京大学
The University of Tokyo
[3] 東京大学/スタンフォード大学
The University of Tokyo/Stanford University
[a)] baba@tmu.ac.jp

[*1] http://imagenet.stanford.edu

ン，改札機等を検出する必要があるだけでなく，屋外での活動であることから誤認識に対して注意深く扱う必要がある．そこで我々はこれらデータセットの構築から始め，ユーザが積極的にデータセット構築に参加可能な持続可能デザインをあわせて提案する．

著者らは過去の情報処理学会アクセシビリティ研究会での発表 [7] の後，スマートフォンを利用した物体検出及びフィードバックシステムに関して議論を重ねることで，システムのみならずユーザ参加型のシステムデザイン仕様をプロトタイピングから明確化し，ユーザがデータセット作成に能動的に関わる持続可能なデザインを本研究のゴールとした．本稿では特に初期デザインのベースとなる議論及び第一ステップのプロトタイピングプロセスに関して述べる．

2. 関連研究

視覚障害者の支援技術として深層学習を活用している事例が近年報告され始めているが，それ以前から OCR（Optical Character Recognition）を始め，Vison-based な支援技術に関しては多くの研究がなされてきた．RFID を利用した屋内外のナビゲーション [8] や深度センサ，画像処理を利用した周辺環境認識支援に関する報告があり，条件を整える必要があるが，当事者に対して品質の高いナビゲーションを提供可能である．この他画像認識を利用した支援技術として，スマートフォンカメラを利用した紙幣認識 [9] や，特徴点抽出による障害物検出 [10] 等が報告されている．支援技術には様々な手法が存在するなか，近年はスマートフォン利用や，画像認識が多く活用されている．これは Plos らが提案する，Assistive Technology に必要なデザイン指針の観点から，今後も重要な点であると言える [11].

視覚障害者を対象とした研究ではないが，AlexNet をベースにした Convolutional Networks を Visual Place Recognition に応用した研究が 2015 年に報告されている．自立ロボットにおける位置認識（ローカライゼーション）を目的としている．Baljit ら [12] は視覚障害者支援を目的として，通常の USB カメラに測距センサを追加し，RGB+Depth の 1 チャンネルを追加し Faster-RCNN ベースのネットワークを設計した．また，周辺情報（人や車）等を音声ナビゲーションを通じてフィードバックを行った．Mulfari らはシングルボード PC（Rasberry Pi 3）に tensorflow 環境を構築し，メガネに搭載したカメラから物体検出を行うシステム実装を行い，視覚障害者支援に関する可能性を議論している [13]．Chaudhry ら [14] は顔認識を利用した視覚障害者のための人物特定支援システムを開発している．このように機械学習及びスマートフォンの処理性能向上によって，支援技術の領域においてこれから実用的なサービスが登場する可能性が高い．

一方で，商品やサービスとしてすでにユーザに提供されている視覚障害者向け支援技術も数多くある．Microsoft 社は Seeing AI というプロジェクト名で，機械学習を活用した視覚障害者支援アプリをすでにリリースしている．上記で述べた OCR や Place Recognition, 紙幣認識などの機能が備わっている．認識処理をサーバサイドで行うため実行速度にはタイムラグが生じるが，その分精度が高い[*2]．東京都障害者 IT 地域支援センターウェブサイトでは，スマートフォンアプリケーションを対象に，障害のある人に便利なアプリの情報を提供している[*3]．

3. 基本設計

ここまでの調査を元に，本プロジェクトにて開発するシステムの基本的な設計をまとめる．現時点では実働するプロトタイプシステムも存在しないため，まずは 1st プロトタイプに関わる仕様をまとめることとする．

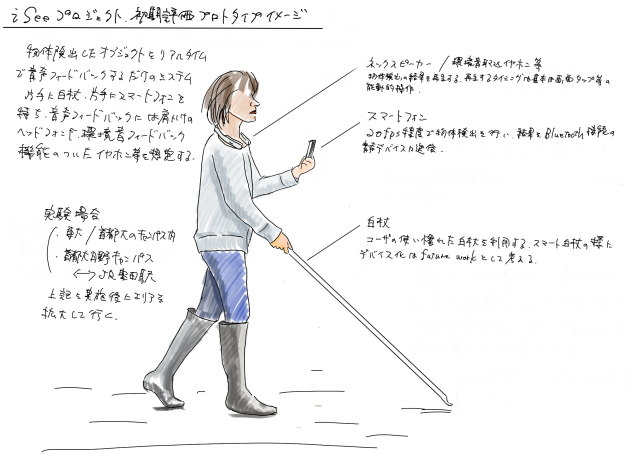

図 1 初期評価プロトタイプシステムのスケッチ．白杖を持ちながらスマートフォンをかざし，周辺情報は音声でフィードバックする

図 2 に 1st プロトタイプ利用時の様子を示す．ユーザは白杖及びスマートフォンをかざし，周辺情報が肩がけのヘッドセットから検出した物体を音声フィードバックする．今回は text-to-speech により発音することとする．なお，この音声フィードバックに関しては，Mascetti ら [15] が示すように，言語の発話より，Sonification によるフィードバックが好まれる場合もあるため，今後の検討事項とする．

3.1 データセット構築

画像認識においては機械学習のデータ元となる，データセットを作成を最初にする必要があるが，Pascal VOC Challenges でよく知られる VOC データセット [16] や，Microsoft 社が提供する COCO データセット [17] がアルゴリ

*2 https://www.microsoft.com/en-us/seeing-ai
*3 http://www.tokyo-itcenter.com/index.html

ズム評価用途のデータセットとして利用される他，Google 社による Open Image Dataset V4[*4]や，研究者が活用するデータセットとして世界最大の ImageNet[*5]が大規模データセットとしてよく知られている．これらを利用することですでに登録されているクラスであれば，容易に学習用データセットを作成可能であるが，本研究で必要となるデータセットを検討した結果，上記データセットデータベースでは必要なクラスが多く不足していることがわかった．特に歩行者用押しボタンや横断歩道の信号，さらには改札機や改札口などの日本固有の学習データが極めて不足している．初期評価プロトタイプ制作にあたり，これらデータセットをまずは開発する必要があることが明確となった．

3.2 局所最適化のための地元データ

今回の初期評価では議論のみであるが，データセットを開発するにあたり，初期評価ではプロジェクト運営側にてある一定のデータセット構築を行うが，本研究は情報支援システムであると同時に，常にデータセットも更新されていくことが好ましい．このような継続的運営にするためには，核となるデータセット開発をユーザ参加型に切り替えて行く必要がある．それにより，介助者や家族が当事者支援のために自らデータセットを提供することで，自宅周辺の単独歩行やちょっとした買い物に出かけるなどの可能性が自然発生的に生じると考えられる．

現時点で開発している簡単なシステム構成を図 2 に示す．

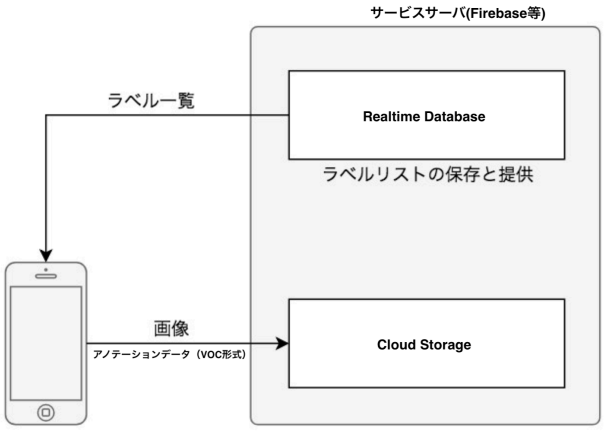

図 2 ユーザがスマートフォンを利用し，アノテーションを行った結果が GPS 付与され，サーバ上のデータセットに追加される．

3.3 物体検出手法

物体検出（Object Detection）はカメラ画像中における任意物体がどこにあるかを求める手法で，Haar-Like 特徴量を利用した顔検出 [17] や，HOG 特徴量を利用した人物検出 [18] はよく知られたアルゴリズムである．その中で近年深層学習を用いた Faster-RCNN[19] や SSD(Single Shot Multibox Detector)[20]，YOLO(You Only Look Once)[21] といったアルゴリズムが物体検出において広く知られるようになった．特に SSD や YOLO においては実行速度と認識精度（mAP）のバランスがよく，プロトタイピングのみならず，スマートフォン，エッジデバイスへの組み込みにおいて頻繁に利用されている他，これらをベースにした高速化手法や [22] 重みファイルの軽量化手法 [23] などが次々と報告されている．

文献 [20] に示されている Faster-RCNN，SSD, YOLO の認識精度，速度比較一覧を一部抜粋，追記したものを表 1 に示す．YOLO に関しては現在 Version が 3 まである他，SSD も MobileNet といった最新の比較にはなっていないが，それぞれの性能の指標として示す．検証には VOC2007 データセットを，グラフィックカードに Titan X with cuDNN v4，CPU に Intel Xeon E5-2667v3@3.20GHz を利用している．

表 1 それぞれの物体検出アルゴリズムの精度（mAP:mean Average Prevision），および実行速度（FPS:Frame Per Second），入力画像サイズを示す．

Method	mAP	FPS	Input Resolution
Faster R-CNN(VGG16)	73.2	7	1000 x 600
Tiny YOLO(v.1)	52.7	155	448 x 448
YOLO(v.1)	66.4	21	448 x 448
SSD300	74.3	46	300 x 300
SSD512	76.8	19	512 x 512

本研究において，プロトタイプの段階からスマートフォンでの動作を前提としているため，SSD 及び YOLO を利用した実装を行うこととした．これら検出手法の中で，SSD では MobileNet モデルを，YOLO では yolov2-tiny モデルを利用した初期評価アプリケーションを実装することとした．それぞれのモデルを同環境で評価した実験結果がないため，今後は実装したデバイス上での比較検討も行う．

4. まとめ

本稿では視覚障害者の屋外歩行支援を目的とした物体検出システムをベースに，ユーザがデータセット開発に参加可能な仕組みづくりを含めた基礎設計を議論した．すでにデータセットのプロトタイプを行っており，それを元にした認識システムを開発できている．この初期評価プロトタイプを元に当事者からのヒアリングなどをおこなうことで，具体的なユーザインタフェースやインタラクションの開発につなげていく．初期評価プロトタイプの詳細に関しては同研究会内の発表「視覚障害者の屋外移動支援に向けた物体検出データセットの基礎検討とプロトタイピング」を参照されたい．

謝辞 本研究は JSPS 科研費 JP18H03486 の助成を受けたものです。

*4 https://storage.googleapis.com/openimages/web/index.html
*5 http://imagenet.stanford.edu

参考文献

[1] 福井良太：世界から見た日本の盲導犬育成事業，日本補助犬科学研究， Vol. 2, No. 1, pp. 22–25（オンライン），DOI: 10.3373/jssdr.2.22 (2008).

[2] 社会福祉法人日本盲人会連合：視覚障害者の移動支援の在り方に関する実態調査　報告書 (2015).

[3] Terven, J. R., Salas, J. and Raducanu, B.: New Opportunities for Computer Vision-Based Assistive Technology Systems for the Visually Impaired, *Computer*, Vol. 47, No. 4, pp. 52–58 (online), DOI: 10.1109/MC.2013.265 (2014).

[4] Krizhevsky, A., Sutskever, I. and Hinton, G. E.: ImageNet Classification with Deep Convolutional Neural Networks, *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, USA, Curran Associates Inc., pp. 1097–1105 (online), available from ⟨http://dl.acm.org/citation.cfm?id=2999134.2999257⟩ (2012).

[5] Hu, F., Tang, H., Tsema, A. and Zhu, Z.: Chapter 1 - Computer Vision for Sight: Computer Vision Techniques to Assist Visually Impaired People to Navigate in an Indoor Environment, *Computer Vision for Assistive Healthcare* (Leo, M. and Farinella, G. M., eds.), Computer Vision and Pattern Recognition, Academic Press, pp. 1 – 49 (online), DOI: https://doi.org/10.1016/B978-0-12-813445-0.00001-0 (2018).

[6] Lin, T., Maire, M., Belongie, S. J., Bourdev, L. D., Girshick, R. B., Hays, J., Perona, P., Ramanan, D., Dollár, P. and Zitnick, C. L.: Microsoft COCO: Common Objects in Context, *CoRR*, Vol. abs/1405.0312 (online), available from ⟨http://arxiv.org/abs/1405.0312⟩ (2014).

[7] 常好釜江，富夫小出，哲夫野口：視覚障碍者支援は最新の ICT や AI 技術を必要としている-自動運転で開発された AI 技術から学ぼう-，技術報告 11，東京大学／スタンフォード大学, クリエートシステム開発株式会社, クリエートシステム開発株式会社 (2017).

[8] Sato, D., Oh, U., Naito, K., Takagi, H., Kitani, K. and Asakawa, C.: NavCog3: An Evaluation of a Smartphone-Based Blind Indoor Navigation Assistant with Semantic Features in a Large-Scale Environment, *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '17, New York, NY, USA, ACM, pp. 270–279 (online), DOI: 10.1145/3132525.3132535 (2017).

[9] Liu, X.: A Camera Phone Based Currency Reader for the Visually Impaired, *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '08, New York, NY, USA, ACM, pp. 305–306 (online), DOI: 10.1145/1414471.1414551 (2008).

[10] Tapu, R., Mocanu, B., Bursuc, A. and Zaharia, T.: A Smartphone-Based Obstacle Detection and Classification System for Assisting Visually Impaired People, *2013 IEEE International Conference on Computer Vision Workshops*, pp. 444–451 (online), DOI: 10.1109/ICCVW.2013.65 (2013).

[11] Plos, O., Buisine, S., Aoussat, A., Mantelet, F. and Dumas, C.: A Universalist strategy for the design of Assistive Technology, *International Journal of Industrial Ergonomics*, Vol. 42, No. 6, pp. 533 – 541 (online), DOI: https://doi.org/10.1016/j.ergon.2012.09.003 (2012).

[12] Kaur, B. and Bhattacharya, J.: A scene perception system for visually impaired based on object detection and classification using multi-modal DCNN, *CoRR*, Vol. abs/1805.08798 (online), available from ⟨http://arxiv.org/abs/1805.08798⟩ (2018).

[13] Davide Mulfari, A. P. and Fanucci, L.: USING TENSORFLOW TO DESIGN ASSISTIVE TECHNOLOGIES FOR PEOPLE WITH VISUAL IMPAIRMENTS, IADIS International Conference Big Data Analytics, Data Mining and Computational Intelligence 2017 (part of MCCSIS 2017), iadis, pp. 110–116 (2017).

[14] Chaudhry, S. and Chandra, R.: Design of a Mobile Face Recognition System for Visually Impaired Persons, *ArXiv e-prints* (2015).

[15] Mascetti, S., Picinali, L., Gerino, A., Ahmetovic, D. and Bernareggi, C.: Sonification of guidance data during road crossing for people with visual impairments or blindness, *ArXiv e-prints* (2015).

[16] Everingham, M., Eslami, S. M. A., Van Gool, L., Williams, C. K. I., Winn, J. and Zisserman, A.: The Pascal Visual Object Classes Challenge: A Retrospective, *International Journal of Computer Vision*, Vol. 111, No. 1, pp. 98–136 (2015).

[17] Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L. and Dollár, P.: Microsoft COCO: Common Objects in Context, *ArXiv e-prints* (2014).

[18] Dalal, N. and Triggs, B.: Histograms of oriented gradients for human detection, *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Vol. 1, pp. 886–893 vol. 1 (online), DOI: 10.1109/CVPR.2005.177 (2005).

[19] Ren, S., He, K., Girshick, R. and Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, *ArXiv e-prints* (2015).

[20] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y. and Berg, A. C.: SSD: Single Shot MultiBox Detector, *ArXiv e-prints* (2015).

[21] Redmon, J. and Farhadi, A.: YOLOv3: An Incremental Improvement, *arXiv* (2018).

[22] Li, Y., Li, J., Lin, W. and Li, J.: Tiny-DSOD: Lightweight Object Detection for Resource-Restricted Usages, *ArXiv e-prints* (2018).

[23] Xu, J., Wang, P., Yang, H. and López, A. M.: Training a Binary Weight Object Detector by Knowledge Transfer for Autonomous Driving, *ArXiv e-prints* (2018).