Vol.2020-AAC-14 No.4
2020/12/4

ユーザ参加型物体検出データセット構築のためのアプリケーション

馬場 哲晃[1,a)]

概要：我々はこれまで視覚障害者の屋外歩行における移動支援を目的としたデータセット開発を継続してきた。横断歩道や歩行者用信号、歩行者用押し釦等の 39 クラスにおいて 2020 年 11 月 10 日時点で 17,355 の画像に対して 210,701 インスタンスのアノテーションを登録した。独自にデータセットを開発する一方で、これらの作業工程を継続可能な形でデザインするために、ユーザ参画型のアノテーション構築を目指している。様々な地域のユーザが気軽に参画できることで、地域ごとの画像の偏りも解決可能であると考えている。そこで我々はユーザが容易に参画可能なアノテーション手法として、スマートフォン上で画像送信や単純なアノテーション作業が可能なアプリケーションを開発した。

1. はじめに

深層学習による技術向上は様々な分野において state of art な成果を達成している。特に Computer Vision(CV) 分野においてはその成果は著しい。生産工場から自動運転、さらにはデザインや芸術表現まで多くの分野で活用され、アクセシビリティ分野においても同様の流れが見られる。Kayukawa らによる誘導型スーツケース [1] では歩行者検出に深層学習による物体検出を利用し、前方の歩行者を検出している。CV を利用した支援技術は今後のアクセシビリティ分野においても大きく進展が期待される.

近年ではスマートフォン上においてニューラルネットの推論計算が大幅に高速化されており、例えば iOS においては A13 チップ上の Neural Engine を利用することで yolov3 モデル（106 層からなる CNN）[2] の推論が 30fps 程度で動作可能である。従来は大掛かりな装置を利用して実装、実験を行っていた本研究領域においても今後小型化された支援システム開発が容易になる。

一方で、データセット開発に関して国内においてはそれほど多くの視覚障害支援用途のものが見られない。我々はこの問題に対して独自にデータセットを開発してきた. 我々はこれまで視覚障害者の屋外歩行における移動支援を目的としたデータセット開発を継続し，横断歩道や歩行者用信号、歩行者用押し釦等の 39 クラスにおいて 2020 年 11 月 10 日現在で 17,355 の画像に対して 210,701 インスタンスのアノテーションを登録した。図 1 に一覧を示す．一方でこれらの開発工程を継続可能な形でデザインするために、ユーザ参画型のアノテーション構築を目指している。様々な地域のユーザが気軽に参画できることで、地域ごとの画像の偏りも解決可能であると考えている。本稿ではユーザが容易に参画可能なアノテーション手法として開発したスマートフォン上で画像送信や単純なアノテーション作業が可能なアプリケーションについて報告する。

2. 継続可能なデータセット構築

当事者参加型におけるデータセット開発においては、VizWiz がよく知られている [3][4]．VizWiz では主に Visual Question Answering(VQA) と呼ばれる撮影された画像の概要説明を目的としたデータセットが数多く収集されている．それらデータセットは VizWiz ウェブサイト[*1]にて閲覧ができる．我々が対象とするデータセットは物体検出器用途であるため，画像内の学習対象物を矩形領域にて登録する必要がある．画像さえ収集することができれば矩形領域登録自体はクラウドソーシング等でも対応可能であるが，継続的な予算獲得や運営には不確定要素が多く，必ずしも長期に渡ってプロジェクトを継続できるとは限らない．また本データセットは日本国内に特化しているが，歩行者目線からの画像データを日本全国から広く収集するだけでも膨大な労力が必要とされる.

3. アプリケーション

当事者やその介助者らが手軽に画像やアノテーション

1 馬場哲晃
Tokyo Metroplitan University, 6-6 Asahigaoka, Hino, Tokyo 191-0065, Japan
a) baba@tmu.ac.jp

*1 https://vizwiz.org/

ⓒ 2020 Information Processing Society of Japan

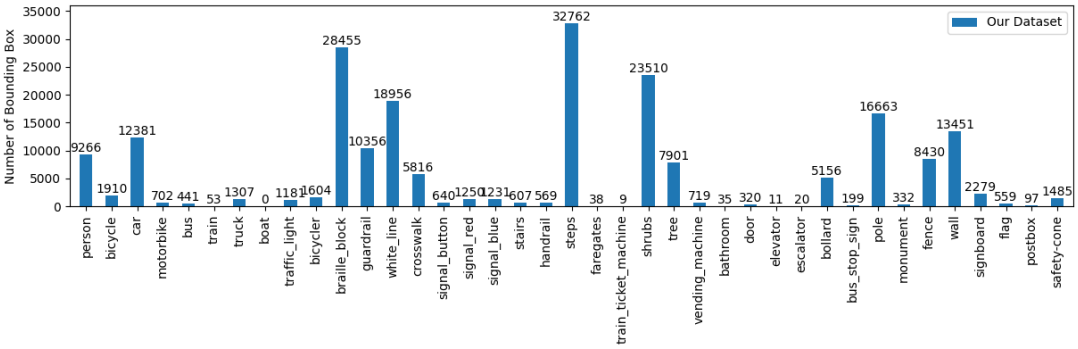

図 1　2020 年 11 月 10 日現在で登録されている各クラスのインスタンス数

データを逐次手軽にクラウド上にアップロード可能とすることで継続可能なデータセット構築を目指し，次のようなユーザシナリオを想定した．

(1) アプリケーション上では学習済みモデルを利用し物体検出がおこなえるため，実際の通学，通勤経路に対して介助者が事前に認識状況を確認する．
(2) 認識結果が良くない（認識されない物体が多い）箇所があった場合，即座に画像を撮影し，必要に応じてアノテーションを行いサーバに送信することができる．
(3) 送信されたデータは学習用データに自動で共有され，品質等を確認した後にデータセットに組み込み，学習済みモデルを生成する．
(4) アプリケーション側の更新によりデータセットが最新版に更新される．

アプリケーションは iOS 上にて実装した．データサーバには Firebase[*2]を利用した．アプリケーション動作時のスクリーンショットを図 2 に示す．

本アプリケーションを開発チーム内で実際に利用したところ，「アノテーション作業はスマートフォン上では通常のアノテーションソフトのように精確に矩形領域を指示できない」，「アノテーションまでやると結構時間を取られる」等の意見が挙がった．実際の利用状況はアノテーションをするために撮影をする，というより，「この場所必要だな」と思いスマートフォンを取り出し撮影を行うことが多いのでは，という議論になった．以上の事前議論から，アノテーションがなされていない場合でもデータアップロードは可能とした．

4.　今後の展望

今回開発したアプリケーションは TestFlight[*3]を利用した開発チーム内のテスト配布のみであったが，今年度内でのリリースを現在計画している．現時点では利用していないが，撮影時のスマートフォン姿勢及び GPS 座標も取得しサーバ上には JSON 形式で送信している．今後データ数が増加した場合に，地域に特化した学習モデルの提供等が可能になると考えている．

*2　Google 社が提供する mBaaS の一つ
*3　Apple 社により提供されている，ユーザ指定型アプリ配布システム

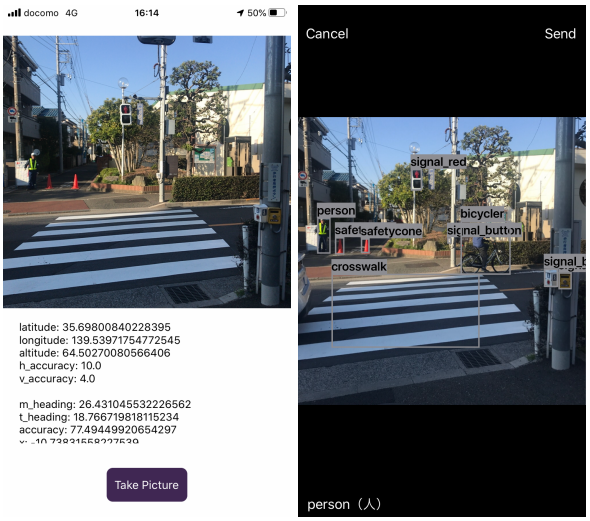

図 2　開発した参加型アノテーションソフトウェア。左：カメラ撮影モード、右：アノテーションモード時のスクリーンショット。

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

参考文献

[1] Kayukawa, S., Ishihara, T., Takagi, H., Morishima, S. and Asakawa, C.: Guiding Blind Pedestrians in Public Spaces by Understanding Walking Behavior of Nearby Pedestrians, *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Vol. 4, No. 3 (online), DOI: 10.1145/3411825 (2020).
[2] Redmon, J. and Farhadi, A.: YOLOv3: An Incremental Improvement, *arXiv* (2018).
[3] Gurari, D., Li, Q., Stangl, A. J., Guo, A., Lin, C., Grauman, K., Luo, J. and Bigham, J. P.: VizWiz Grand Challenge: Answering Visual Questions from Blind People, *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (online), DOI: 10.1109/cvpr.2018.00380 (2018).
[4] Gurari, D., Zhao, Y., Zhang, M. and Bhattacharya, N.: Captioning Images Taken by People Who Are Blind (2020).

ⓒ 2020 Information Processing Society of Japan