Vol.2018-AAC-8 No.1
2018/11/30

ユーザ参加型アノテーションにおけるUI及びデータオーグメンテーションのデザイン

石曽根 奏子[1,a)]　馬場 哲晃[1,b)]　渡邉 英徳[2]　釜江 常好[3]

概要：本研究では深層学習を利用した物体検出をベースに，視覚障害者向け屋外移動支援システムを開発している [1]．その中で持続可能デザインとして，ユーザ参加型アノテーションサービスを開発しており，本稿では一般ユーザが参画可能で精度向上に効果的なアノテーションデザインに関する手法を議論する．

キーワード：視覚障害，深層学習，物体検出，アノテーション，データオーグメンテーション

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

1. 背景

深層学習での物体検出には膨大なデータセットが必要となる．今現在では，多くの種類のデータセットが公開されている．物体検出においてVOC[*1]やCOCO[2]データッセットは評価用データセットとして頻繁に用いられる．数多くカテゴライズされている中で，必要なデータセットが提供されていない場合がある．本研究である，視覚障害者向け屋外移動支援システム開発に必要な横断歩道や歩行者用信号機といったデータセットも提供されていない．データセットを開発するために，一般的にはimgLab[*2]やBBox-LabelTool[*3]，VoTT[*4]等の公開されているアノテーションツールを利用する．これらの既存のアノテーションツールでは動画または画像に対して手作業で登録しなければならず，この作業には膨大な時間を要する．少ないデータセットから画像データを自動で拡充し，認識精度を向上させるデータオーグメンテーション（Data Augmentation）という方法もあるが，必ず対象物のデータセットを準備する必要がある．

本研究の中で，持続可能なデザインとして，一般ユーザからデータセットを提供できるように，ユーザ参加型のアノテーションサービスを議論している．一般ユーザを対象にすることから手作業で時間を要する作業はできるだけ減らすことが望ましい．そこで本稿ではアノテーション作業をリアルタイムで行う方法を提案する．

2. アノテーションアプリケーション

2.1 UIの検討

データセットを開発するにあたり，以前ヒューマンエラーを減らすために機能を絞ったアノテーションツールを自作した[*5]．このアノテーションツールでは画像と動画から一枚ずつアノテーションを行う仕様になっている．画像は動画から数秒毎で静止画を書き出してから，動画では任意の箇所でアノテーションを行う．このとき登録枚数に大きな偏りがあった [3]．一枚ずつ手作業のアノテーション作業は膨大な時間を要するため，対象物をリアルタイムで登録をすることでアノテーション作業を高速化できないかと考えた．スマートフォンで撮影しながら対象物をバウンディングボックスで囲み，画像とバウンディングボックスの位置情報をリアルタイムで書き出していくというものである．図 1.

手作業の登録で数時間かかっていたものが数十秒でデー

[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)] ishisone-kanako@ed.tmu.ac.jp
[b)] baba@tmu.ac.jp

[*1] http://host.robots.ox.ac.uk/pascal/VOC/
[*2] https://github.com/davisking/dlib/tree/master/tools/imglab
[*3] https://github.com/puzzledqs/BBox-Label-Tool
[*4] https://github.com/Microsoft/VoTT
[*5] ofxYolov2: https://github.com/TetsuakiBaba/ofxYolov2

タセットを生成することが可能になる．また画像や動画を撮影してアノテーションツールに取り込む手間も省くことができる．しかしながらデメリットとして手作業で一枚ずつ登録する場合ほど正確に対象物を登録できないことが挙げられる．実際にこのアノテーションツールで作成したデータセットで，どの程度の認識精度が得られるのか確認するために実験を行った．

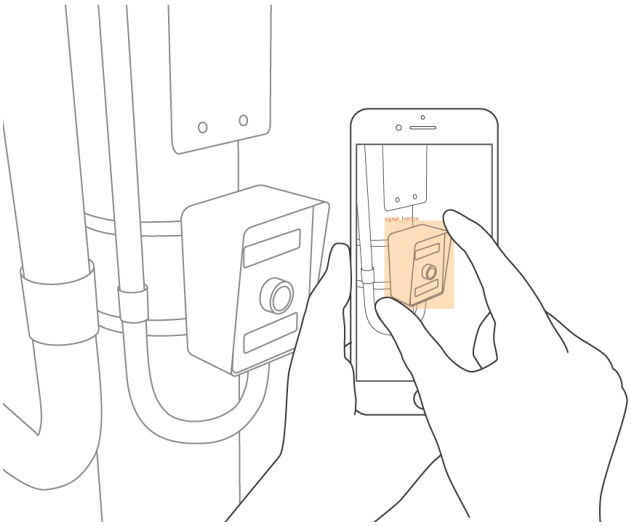

図 1　アノテーションアプリケーションの操作イメージ

2.2　テンプレートマッチング

リアルタイムでアノテーションする際に，書き出すタイミングを検討した．データセットとしては似通った画像ではなく，アングルや大きさ色味等が少しずつ異なった画像セットが好ましい．一つの案として，既存のアノテーションツールである VoTT の機能としても備わっている，一定時間毎で書き出す方法がある．しかし，あまり変化のない画面が続いた場合には，ほぼ変わらない画像が何枚も生成されてしまう．加えて，変化の大きい箇所では適切なタイミングを設定しなければ，本来データセットとして書き出したい箇所での生成ができない，または足りないという事態が起こりうる．したがって対象物を囲んでいるバウンディングボックス内でテンプレートマッチングを試すことにした．方法としては，バウンディングボックス内をグレースケールに変換し，解像度を 64x64 に変更する．一つ前のフレームと 64x64 ピクセル値の合計の差分を求める．この差分が一定の値を超えたときに書き出しが行われる仕組みである．

1st プロトタイプでは，データセットを自作することが目的であるので，公開されているデータセットでは提供されていないものを選定する必要がある．そこで，sunlemon[*6]のキャラクタである，カモノハシを模したぬいぐるみのアノテーションを行った．画像サイズは 720x720，画像枚数 682 枚，バウンディングボックス数は図 1 からもわかる通り，画像に対して一枚なので画像枚数と同数の 682 個のデータセットを作成した．

[*6] http://www.sunlemon.co.jp

3.　学習

作成したデータセットを元に YOLO[4] ネットワークにて学習を行った．現在 YOLO は Version3 まで公開されているが，今回は Version2 の yolov2-tiny のネットワークを用いた．batch size は 64, subdivisions は 2 とし，17,000 回のイテレーションで学習を行った．

3.1　結果

データセットを作成した場所で，スマートフォンを用いて新たに撮影した画像に対して認識処理を行った．図 2 のように画像サイズに対して比較的大きく写っている場合に関しては検出が確認できた．

図 2　1st プロトタイプでの認識処理結果

一方で対象物との距離が離れ，画像サイズに対して比較的小さい場合には検出されなかった．スマートフォンの画面上を 2 本指で触れてバウンディングボックスを引く際に，小さい対象物では画面を触れている指で隠れて見えにくくなってしまい，登録が難しいという点が挙げられる．必然的にタッチしやすい，画面上で大きく写るデータセットが集まったと考えられる．そのうえカメラの入力サイズである，解像度 720x720 をスマートフォンの画面上に全体が表示されるようにスマートフォンの画面幅に縮小表示している．これにより，画像に対して比較的大きく写るデータセットになってしまったと考えられる．

4.　2nd プロトタイプ

結果から，画像内で比較的小さく写る対象物を認識させるために，データオーグメンテーション（以下，DA）を用いて，今回作成したデータセットを元にデータセットの拡

充を行い，精度向上を図った．アノテーションアプリケーション上で機能を追加することも考えられるが，ユーザの操作性を損なわないために，コンピュータ上での処理で解決することとした．

4.1 データオーグメンテーション

深層学習を用いた画像処理には，多くのデータだ必要になる．そのため，既存のデータセットを加工してデータセットを拡充する方法として一般的に DA が用いられる．例えば画像の反転，回転，切り取り，RBG 値の変換等の方法が報告されている [5]．本稿では，1st プロトタイプで作成したデータセットを元に画像の縮小を行った．画像数 680 枚のデータセットを，元の解像度である 720x720 から，360x360，180x180，90x90，45x45 の 4 種類の変換を行い，合計で画像数 3,410 枚に拡充したデータセットを作成した．

学習ネットワークは 1st プロトタイプと同様の yolov2-tiny モデルを使用し，batch size は 64, subdivisions は 4, イテレーション数は 481,700 回である．

4.2 学習結果の比較

物体検出では，mAP（mean Average Precision）を測ることで，どれほど正確に検出できているかの指標になる．AP は検出処理画像に対してそれぞれの recall での最大精度の平均であり，mAP とは AP の全てのクラスの平均値である．本稿では 1 クラスのみの学習であるので，AP と mAP は等しくなる．

精度を測るために，データセットと同じ場所で撮影した動画から一枚ずつ手作業でアノテーションした 114 枚のテストデータ作成した．1st プロトタイプでの mAP は 42.07 %の精度であった．2nd プロトタイプの mAP はイテレーション数に対する mAP を図 3 で示した．イテレーション回数によって差はみられるが，最適箇所のイテレーション数の重みデータを用いることで，1st プロトタイプより高い精度が得られることがわかった．

4.3 問題点

2nd プロトタイプでは 1st プロトタイプでは検出できない小さい対象物に対して，検出範囲が対象物を外れて広く検出されることが多々確認された．今回の DA の方法では，対象物の画像サイズは縮小されたが，画像に対するバウンディングボックスの比率は大きさに起因していると考えられる．そこは今後の課題とする．

4.4 歩行者用ボタンでのプロトタイプ

本研究は，視覚障害者を対象とするため単独歩行に必要な公共物を対象にしている．2.1 節で述べたように，以前のアノテーションツールで極端に登録数が少なかった歩行者用ボタンのアノテーションを再度行った．登録場所は首

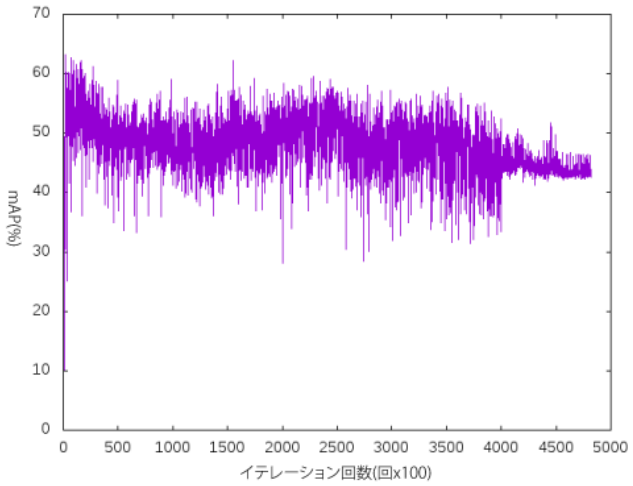

図 3　2nd プロトタイプでのイテレーション回数に対する認識精度

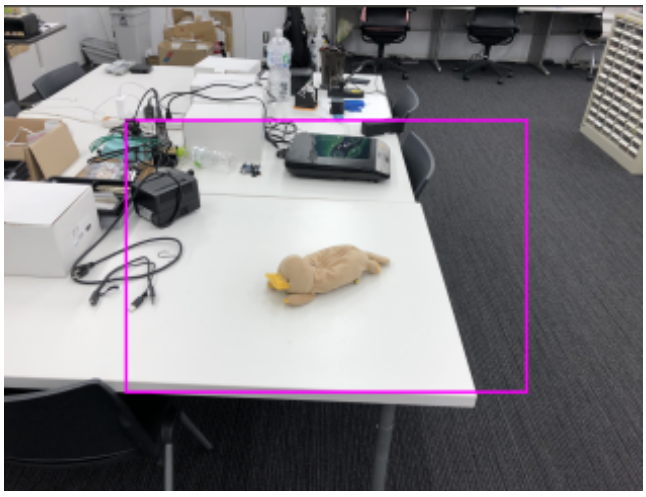

図 4　新たに撮影したぬいぐるみの認識処理結果. ぬいぐるみの大きさに対して検出範囲が広くなっていることがわかる．

都大学東京日野キャンパスの正門側の横断歩道にて，横断歩道の両側の歩行者用ボタンの 2 個を登録対象とした．

4.4.1 1st プロトタイプ

登録枚数は 2,285 枚でぬいぐるみと同じネットワーク構成を用いて，イテレーション数は 22,300 回で学習を行った．同じ時刻に新たに撮影した画像の認識処理結果を図 5 に示す．

図 5　歩行者用ボタンの 1st プロトタイプでの認識処理結果

こちらもぬいぐるみと同じく画像サイズ内で比較的大きく写っている場合，高確率で検出を確認できた．また mAP は 40.83 %であった．mAP に用いたテストデータは学習データと同時刻に撮影した動画から手作業でアノテーションを行った 98 枚を使用している．

4.4.2 2nd プロトタイプ

1st プロトタイプで作成した 2,285 枚に対して，4.1 節と同じ手法で DA を行い，データセットを画像数 11,409 枚に拡充した．ネットワーク構成は変更せず，イテレーション数は 263,300 回で学習を行った．イテレーション数に対する mAP を図 6 に示した．2nd プロトタイプでは 1st プロトタイプと同等または，それ以上の精度が出ていることが確認できた．

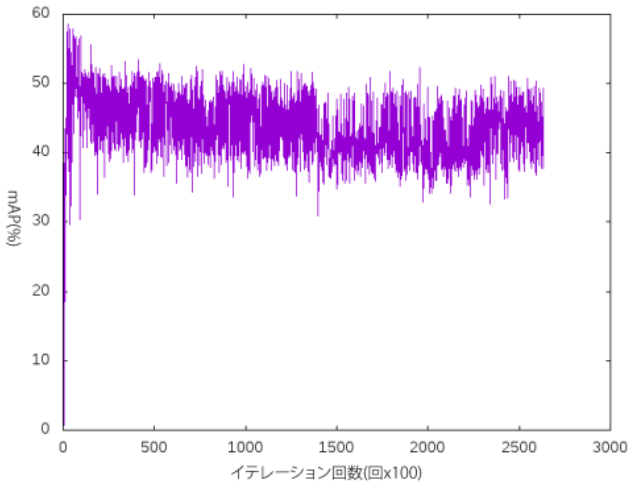

図 6 歩行者用ボタンの 1st プロトタイプでのイテレーション回数に対する認識精度

また，ぬいぐるみの学習結果と同じく，画像サイズに対して対象物が小さく写る場合に，検出範囲が大きくなってしまうことも確認された．視覚障害者向けの屋外移動支援システムとしては，検出範囲が対象物に対して余分に大きいと，位置情報としては信頼度が下がってしまうことが問題として挙げられる．

5. まとめ

リアルタイムでアノテーション作業を行うアプリケーションを作成した．また作成したデータセットを元に DA を行い，認識精度の向上を試みた．結果としてリアルタイムによるアノテーションアプリケーションは物体検出において有効な手段であり，作業時間も大幅に短縮することができた．画像の縮小による DA を用いることで精度の向上に効果的であることを示した．今後の課題として，画像の縮小方法を見直し，検出範囲の問題改善を行っていく．また一般ユーザに提供するサービスとして，アノテーションアプリケーションのユーザビリティも今後検討していく．

参考文献

[1] 馬場哲晃，渡邉英徳，釜江常好：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，研究報告アクセシビリティ（AAC），Vol. 2018-AAC-7, No. 8, pp. 1–4 (2018).
[2] 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).
[3] 石曽根奏子，馬場哲晃，渡邉英徳，釜江常好：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，研究報告アクセシビリティ（AAC），Vol. 2018-AAC-7, No. 9, pp. 1–4 (2018).
[4] Redmon, J. and Farhadi, A.: YOLOv3: An Incremental Improvement, *arXiv* (2018).
[5] Taylor, L. and Nitschke, G.: Improving Deep Learning using Generic Data Augmentation, *ArXiv e-prints* (2017).