

物体検出手法を用いた群衆計数に関する一考察

法政大学大学院 学生会員 ○野村 圭哉
法政大学 正 会 員 今井 龍一
琉球大学 正 会 員 神谷 大介

関西大学 正 会 員 山本 雄平
大阪産業大学 正 会 員 姜 文淵
大阪電気通信大学 正 会 員 中原 匡哉

1. はじめに

大規模なイベントや災害時において、群衆を計数して状態を把握することは事故を未然に防ぐために重要である。一方、群衆を手で一人ひとり計数するには大変な労力が必要である。近年、カメラで撮影した動画像に深層学習を適用し、手軽に群衆の人数を把握する技術の開発が期待されている。既存研究²⁾では、物体検出や密度マップを用いた手法を群衆が発生する場面ごとに使い分ける必要性を明らかにしている。特に物体検出手法の開発は日進月歩であり、群衆計数に有効な手段の1つである。このため、各手法の適用限界や特性を明らかにしておくことが群衆計数に活用できる。

以上より、本研究では、群衆を撮影した動画像に複数の物体検出手法を適用し、計数精度を比較することで、各手法の特徴を明らかにする。本研究により物体検出手法の違いによる特徴や課題が整理でき、より高精度な群衆の計数手法の開発に貢献できると考えられる。

2. 物体検出手法の検証の概要

(1) 物体検出手法の選定

本研究では、物体検出手法におけるデファクトスタンダードとなっている YOLO の中から手法を選定し、特徴を把握することで、群衆計数に適した手法を明らかにする。YOLO とは、動画像から人物を検出して領域を矩形で囲む物体検出手法の1つである。原型となる YOLO³⁾は Joseph Redmon 氏により開発されたが、YOLOv4 以降は様々な研究機関が独自の手法を開発している。そのため、本研究では YOLOv4⁴⁾と YOLOv4 より後に開発された YOLOv6v3.0⁵⁾、YOLOv7⁶⁾および YOLOv8⁷⁾の4種類の YOLO を使用する。各手法において、既存研究²⁾でも使用された CrowdHuman データセット⁸⁾で頭部を検出するモデルを作成し検証に用いる。

(2) 検証に用いる動画像

物体検出手法ごとの特徴や課題を明らかにするため、検証に用いる群衆の動画像を撮影する。まちづくり

においてにぎわいを定量的に計測し、地域活性化の効果の把握が推進⁹⁾されていることから、人が集まりやすい中心市街地の交差点・横断歩道を撮影地点とした。本研究では、静岡市内の交差点においてイベント開催日の10時から18時頃にかけて動画像を撮影した。建物の3階部分、高さ約9mより、水平角度約20度で設置し、交差点全体が映るように4K解像度のビデオカメラで撮影した。左右の道幅はそれぞれ約17mと約13.5m、交差点の面積は約230m²である。検証では、約9時間分の動画から1分に1枚の画像を切り出した530枚を使用する。530枚のうち比較的人数の増減が発生している40枚を選定し、手で数えた値を正解人数として用いる。

3. 物体検出手法の検証結果

530枚の画像に物体検出手法を適用したすべての結果を図-1、昼間と夜間において人数が多い画像の結果を図-2 および図-3 に示す。全体の結果として正解人数の増減に合わせて、どの手法においても大まかな傾向は掴めていることがわかる。一方、青色の線で示した YOLOv4 は波形の増減の振れ幅が大きく、正解人数よりも多く検出されている時間帯が散見される。これは、図-2 のように画像の上側の端において小さく映る人物が密集している場合、1人に対して検出結果が複数発生したことが原因である。人数を過剰に計数することで検出数の平均28%は誤検出であった。また、YOLOv7 においては一定地点より奥側は人物が検出されず、誤検出が平均0.6%と4つの手法の中で最も誤検出が少なかった。さらに、YOLOv6 と YOLOv8 は YOLOv7 よりも奥側の人物まで検出ができて、より正解人数に近い値が得られた。しかし、YOLOv6 は図-3 ように人が密集している場合に枠が大きく発生するなど誤検出が平均8.8%、YOLOv8 は夜間において人物が存在しない箇所やマンホールの誤検出が平均7.1%含まれていた。以上より、YOLOv4 以外の3手法が群衆計数に適用できる

キーワード 群衆計数, 画像処理, 物体検出, 深層学習, YOLO

連絡先 〒162-0843 東京都新宿区市谷田町2-33 法政大学 TEL: 03-5228-1347 Email: keiya.nomura.9a@stu.hosei.ac.jp

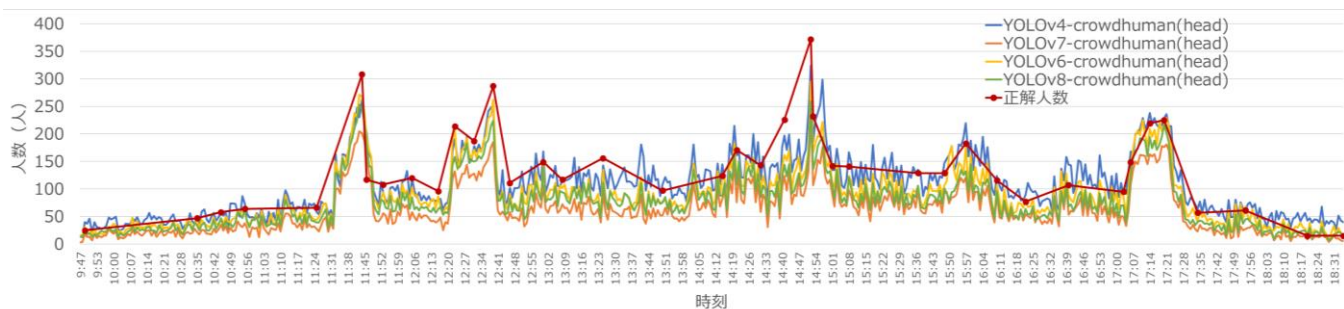


図-1 530 枚の画像に 4 つの物体検出手法を適用した結果



図-2 昼間のピークの 14 時 52 分頃の検出結果



図-3 夜間のピークの 17 時 14 分頃の検出結果

と判明した。特に、YOLOv7 は、どの画像においても誤検出が少なく、画像の中央付近においては検出漏れも非常に少ないことから、本検証においては YOLOv7 が最も良い結果であるといえる。YOLOv7 で検出ができた人物において、手前側では頭部のサイズが 40~60 ピクセル、奥側では 18~25 ピクセルであった。したがって、高さ 9m から撮影した場合、YOLOv7 で人物を検出するには頭部サイズが 20 ピクセル以上必要であり、約 1,145m² が撮影範囲の限界であると考えられる。

4. おわりに

本研究では、群衆を撮影した動画画像に 4 つの物体検出手法を適用し、特徴や課題を整理することで、群衆計数への適用可能性を検証した。検証結果より、YOLOv4 以降の YOLO を用いることで人数増減の傾向を把握できることが判明した。しかし、誤検出が多く発生している手法もあり、対象範囲を限定して検出する手法の考案が必要である。また、検出可否が分かれたエリアにどのような違いが存在するのかを調査することで正確な群衆計数に向けた手法を模索する必要がある。

謝辞：本研究にて検証で用いた動画画像の収集にあたっては、静岡市および昭和設計社にご協力を賜った。ここに記して感謝の意を表す。

参考文献

- 1) Yuqian, Z., Tao, W., Guohui, L., Jun, L. and Luyang, W.: Crowd Counting using DMCNN, *International Conference on Innovation in Artificial Intelligence*, pp.138-144, 2019.
- 2) 今井龍一, 山本雄平, 姜文淵, 中原匡哉, 神谷大介, 野村圭哉: 群衆の状態に応じた最適な人数計数手法に関する一考察, 土木学会論文集 F3 (土木情報学), 7 巻, 22 号, pp.1-13, 2023.
- 3) Redmon, J., Divvala, S., Girshick, R. and Farhadi, A.: You Only Look Once: Unified, Real-Time Object Detection, *Computer Vision and Pattern Recognition*, pp.779-788, 2016.
- 4) Bochkovskiy, A., Wang, C. and Liao, H.: YOLOv4: Optimal Speed and Accuracy of Object Detection, <<https://arxiv.org/pdf/2004.10934.pdf>>, (accessed 2023.3.30).
- 5) Li, C., Li, L., Geng, Y., Jiang, H., Cheng, M., Zhang, B., Ke, Z., Xu, X. and Chu, X.: YOLOv6 v3.0: A Full-Scale Reloading, <<https://arxiv.org/pdf/2301.05586.pdf>>, pp1-7, (accessed 2023.3.31).
- 6) Wang, C., Bochkovskiy, A. and Liao, H.: YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, <<https://arxiv.org/pdf/2207.02696.pdf>>, pp1-15, (accessed 2023.3.31).
- 7) Github: ultralytics, <<https://github.com/ultralytics/ultralytics>>, (accessed 2023.3.31).
- 8) Shao, S., Zhao, Z., Li, B., Xiao, T., Yu, G., Zhang, X. and Sun, J.: CrowdHuman: A Benchmark for Detecting Human in a Crowd, <<https://arxiv.org/abs/1805.00123>>, (accessed 2023.3.31).
- 9) 国土交通省観光庁: 観光入込客統計に関する共通基準 調査要領, <<https://www.mlit.go.jp/common/000995212.pdf>>, (入手 2023.3.31) .