

## (15) 深層学習を用いた群衆画像の 人数推計技術に関する一考察

野村 圭哉<sup>1</sup>・今井 龍一<sup>2</sup>・姜 文淵<sup>3</sup>・山本 雄平<sup>4</sup>・  
神谷 大介<sup>5</sup>・中原 匡哉<sup>6</sup>

<sup>1</sup>学生会員 法政大学大学院 デザイン工学研究科 (〒162-0843 東京都新宿区市谷田町2-33)

E-mail: keiya.nomura.9a@stu.hosei.ac.jp

<sup>2</sup>正会員 法政大学教授 デザイン工学部 (〒162-0843 東京都新宿区市谷田町2-33)

E-mail: ryuichi.imai.73@hosei.ac.jp

<sup>3</sup>正会員 大阪産業大学准教授 工学部 (〒574-0013 大阪府大東市中垣内3-1-1)

E-mail: kyo@ce.osaka-sandai.ac.jp

<sup>4</sup>正会員 関西大学助教 環境都市工学部 (〒564-8680 大阪府吹田市山手町3-3-36)

E-mail: y\_yamamo@kansai-u.ac.jp

<sup>5</sup>正会員 琉球大学准教授 工学部 (〒903-0213 沖縄県中頭郡西原町千原1)

E-mail: d-kamiya@tec.u-ryukyu.ac.jp

<sup>6</sup>正会員 大阪電気通信大学講師 総合情報学部 (〒575-0063 大阪府四條畷市清滝1130-70)

E-mail: nakahara@oecu.jp

イベントや災害時に群衆の人数を計数することは、事故を未然に防ぐために重要である。近年、深層学習を用いて人物を計数する技術が発展し、様々な手法が開発されている。群衆はカメラで撮影されることが多いため、群衆を撮影した動画に深層学習を用いると手軽かつ正確に人数を計数できる可能性がある。しかし、群衆が映る動画は多様なカメラの角度や設置高さで撮影されており、場面に応じた最適な計数手法が明らかとなっていない。そこで本研究では、多くの人が存在する5つの場面を想定して、必要な情報や満たせるニーズを整理した。さらに、群衆人数をより正確に計数するために、整理した5つの場面ごとに4つの画像解析技術を検証した。検証の結果を人物の大きさやオクルージョンの発生に着目して考察し、既存技術の特徴や課題を把握できた。

**Key Words:** deep learning, crowd counting, crowd analysis, obobject detection, density map

### 1. はじめに

群衆の人数を計数することは、イベントや災害時において、群衆の状態を把握し事故を未然に防ぐために重要である。日本で発生した群衆事故の事例としては、2001年に兵庫県明石市で実施された花火大会における事故<sup>2)</sup>があげられる。この事故では、花火の見物客が歩道橋で異常なほどの高密度な状態で滞留したことが原因で、子どもや高齢者を含む11名が亡くなった。この事故から、人々を安全に誘導する重要性和、そのための指標となる群衆の人数の把握が必要不可欠であることがわかる。しかし、群衆の人数を手で計数するには大変な労力が伴うため、手軽に人数を把握することは困難である。こうした背景の下、近年のIoT技術とAI技術の発展により、画像内の人数を自動で計数可能な技術が多数開発されている。これらの技術を適用することで、群衆事

故の防止が期待できる。画像中の人数を計数する既存研究<sup>3)</sup>に着目すると、群衆が映る動画の撮影条件はカメラの角度や設置の高さ、視野角など多種多様であるため、深層学習を用いた様々な人数計数の手法が必要である。しかし、人数を計数する手法は数多く開発されているものの、実社会の様々な場面における群衆の人数を汎用的に計数する最適な手法は明らかとなっていない。また、群衆が発生する場面ごとに必要となる情報や満たせるニーズも整理されていない。

これらの動向を踏まえて、本研究の目的は、群衆が発生する様々な条件・状態ごとに、群衆人数の計数により満たせるニーズをまとめ、それぞれの場面に対する既存技術の計数精度を検証することとした。これにより、既存技術の特徴や課題を整理できるため、場面に応じて最適な手法を切り替えるような、汎用的な群衆人数の計数手法の開発に貢献できる。

## 2. 群衆データの活用方法の整理

既存研究では、深層学習を用いて人物を検出することで、特定のエリアの人数や交通量、混雑度など群衆に関する情報を取得できる。本章では、多くの人物が存在する5つの場面を想定して、必要な情報や満たせるニーズを整理する(表-1)。

### (1) 自由な歩行空間

駅構内や駅前および商業施設の広場は自由に歩行できる空間であり、方向が定まらない移動が発生する。商業施設における広場や休憩スペースは、イベントや展示の会場となるため、その空間自体が集客の効果を持ち、都市の回遊性を高めている。俣賀<sup>9)</sup>は、広場の併設による効果は商業施設だけでなく、街の活性化に貢献する施策であると述べている。このことから自由な歩行空間において街の活性化を定量的に把握するためには、広場内の人数を計数し滞留箇所および滞留した時間を取得する必要がある。これにより効果的な広場や歩行空間の活用が期待できる。

### (2) 単路

鉄道駅の改札や階段は多くの場面で単路となっており、移動方向が限定的である。このような単路では、断面交通量を計数することで、その場所の利用人数や通過人数を把握できる。そのため、街なかにおける活性化の度合いを測るための指標の1つとして活用<sup>9)</sup>されている。安藤ら<sup>9)</sup>の研究では、歩行者交通量調査への画像解析技術の適用可能性が検証されている。単路は、平常時であれば街の活性化の指標として活用できる場面である。しかし、駅周辺でスポーツや祭りなどのイベントが開催され、短時間に多くの人物が集まるときには、通行のボトルネックとなる可能性が高い。そして、多くの人物が滞留し混雑度が上昇すると、抜け道が無い単路は非常に危険な状態となる。そのため、単路の人数の計数により、一定以上の人数や密度となった場合にアラートを出し、的確な誘導により群衆の動きを制御することが重要となる。

### (3) 交差点・横断歩道

交差点や横断歩道は赤信号による通行の待機が生じるため、単路と同様に通行のボトルネックの1つであるといえる。横断する人数を取得することで、交通の円滑化に向けた信号のサイクルの最適化、街のにぎわいや交差点単位での狭域における交通流動の把握に活用できる。例えば、渋谷のスクランブル交差点では、NTTドコモ社の「モバイル空間統計」を用いて、交差点の人数や混雑状況を分析し、コロナ禍における人出の指標として活用

られている。以上のことから、交差点・横断歩道において、信号待ちで待機している人数や横断している人数の計数は、円滑な交通の実現や街のにぎわいの評価への活用が期待できる。

### (4) イベント会場

イベント会場の多くは、有料の入場チケットがテーマパークやパブリックスペースに設営され、その中では誰でも自由に入場・閲覧可能な展示会など、幅広いジャンルのエンターテインメントが開催されている。イベントを開催するときには、次のイベントに向けて集客の効果を検証するだけでなく、安全なイベント運営に向け、入場制限を設ける必要がある。しかし、有料のチケット制でも指定の席がない場合、イベント会場のどこに何人いるかという詳細な情報までは取得が困難である。さらに、無料のイベントにおいては入場者数の把握に膨大な労力や人件費を要する。このようなイベント会場において、会場を訪れた人数を自動で計数できれば、イベント効果の定量的な把握と安全性の向上が期待できる。

### (5) 待機行列

待機行列は、日常生活におけるバス停、トイレおよび飲食店などの比較的小規模な行列から、イベント会場やテーマパークへの入場待ちなどの大規模な行列まで様々な規模の行列が存在する。既存のサービス<sup>9)</sup>では、深層学習を用いて小規模な行列の人数を計数することで、空いている店舗を把握しており、混雑度が店舗選びの新たな基準となりつつある。このように店舗の混雑度を把握できると、ショッピングモールなどの複合施設においては施設全体の滞在時間の増加に繋がる。さらに、ピークの時間帯が分散されるため、店舗の負担が軽減される効果も期待できる。また、大規模な行列においては、実際に入入りした人数はチケットの枚数で把握できるが、待機時の行列の人数は把握できない。これらの大規模な行列の人数を把握できれば空いている列への誘導や待機可能なエリアを越えた行列の検出が可能となる。

表-1 群衆データの活用方法のまとめ

活用場面	必要な情報	満たせるニーズ
自由な歩行空間	人数 滞在時間 滞在场所	● 広場や公園の利用状況を把握 ● 利便性向上に向けた施策の参考
単路	人数 移動方向 混雑度	● 通行人数から街の活性化を評価 ● 混雑時の的確な誘導
交差点 横断歩道	人数 待機時間	● 信号のサイクルを最適化 ● 横断する人数から街の賑わいを評価
イベント会場	人数 混雑度	● イベントによる集客効果の評価 ● イベントエリアへの入場制限の基準
待機行列	人数 待機エリア	● 空いている列への誘導待機 ● エリア外の人を検出

### 3. 既存の画像解析技術の適用可能性の検証

#### (1) 実験概要

本章では、表-1で整理した条件の動画像に既存の画像解析技術<sup>9)~12)</sup>を適用し、検証の結果から既存技術の特徴と課題を整理する。検証においては、人物がなるべく多く存在する瞬間の1フレームを切り出し、複数人で数えた人数の平均値を正解人数、既存技術により推定された人数の正解人数に対する割合を検出率とする。

#### (2) 検証に用いる動画像の選定

既存の画像解析技術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。撮影地点は、2章で整理した5つの場面を参考に選定する。群衆人数の計数を目標としているため、なるべく多くの人物が存在する場所・時間に撮影した。動画像は、4K解像度に統一し、撮影地点の高さや撮影角度に縛りを設けず、交差点やイベント会場などの人物が多く存在するエリア全体が映るように俯瞰して撮影した。撮影した動画像の中でも、特に多くの人物が往来する鉄道駅の改札（地点2）や横断歩道（地点3）、イベント会場（地点4）などの合計5地点（表-2参照）を検証に用いる。

#### (3) 検証する画像解析技術の選定

本研究では、2章で分類した5つの場面のそれぞれに適した計数手法を選定するために、人物の検出方法の違いに着目して、代表的な4つの画像解析技術（図-1参照）を検証する。具体的には、人物を検出して領域を矩形で囲む物体検出の手法であるYOLOv4<sup>9)</sup>、人物領域をピクセル単位で検出して色塗りをする手法であるMask R-CNN<sup>10)</sup>、人物の関節の動きを推定する人物姿勢推定の手法であるOpenPose<sup>11)</sup>、人物が多い場所をヒートマップとして表現する手法であるAMRNet<sup>12)</sup>の4つを用いる。人物を検出するモデルは、インターネットで公開されている学習済みモデルを使用し、人物以外も検出された場合、その物体の検出結果を非表示にする。

#### (4) 既存の画像解析技術を適用した結果と考察

画像解析技術を適用した結果を表-3に示す。地点1は比較的人物がまばらに映っており、オクルージョンが少ない画像といえる。そのため、YOLOv4、Mask R-CNN、

OpenPose<sup>11)</sup>において、画面手前側の人物は正確に検出できており、誤検出や検出漏れは発生していない。一方で、画面奥側の小さく映っている人物やオクルージョンが発生している人物に対しては、検出漏れや2人を1人として誤検出する傾向があり、他の4地点でも同様の傾向がみられた。これは、公開されている学習済みのモデルは、群衆の状態における人物の検出には特化しておらず、小さな人物やオクルージョンが発生している人物の特徴を十分に学習できていないことが原因と考えられる。特にOpenPoseは、人物の関節点を検出した後、それぞれの人物を構成する関節点をまとめることで人物を特定するボトムアップの手法である。そのため、小さな人物の関節の検出が困難となり、他の手法よりも検出漏れが増加したと予想される。

次に、AMRNetの検出結果を確認すると、地点3、4、5の人物が存在しない背景の部分において、ヒートマップの色が濃くなっている。背景の中でも特に植栽を人物や群衆として検出しており、結果として人数を過剰に計数する傾向がみられる。AMRNetの学習済みモデルはShanghaiTech<sup>13)</sup>という混雑度が非常に高いデータセットを学習して構築されている。そのため、植栽の小さな葉を人物として誤検出し、過剰に計数したと考えられる。ただし、局所的に誤検出が発生しているため、人手による簡易な修正を加えることで、検出率が向上する可能性が高い。地点5のようにオクルージョンが多く発生している場面では、AMRNetのみが奥側の人物を計数できており、YOLOv4やMask R-CNNには検出が難しいタスクだと考えられる。今後はより混雑度が高い画像を対象に、AMRNetの限界を調査する必要があると考えられる。

















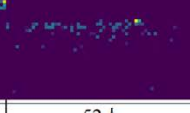
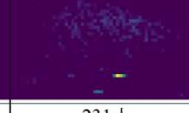
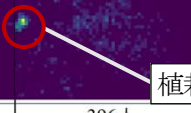



図-1 選定した画像解析技術

表-2 検証に用いる動画像の概要

撮影地点番号	1	2	3	4	5
撮影映像					
活用場面	自由な歩行空間	単路	交差点・横断歩道	イベント会場	待機行列
撮影地点	大阪府大阪市 大阪駅周辺	神奈川県鎌倉市 改札	大阪府大阪市 大阪駅周辺	東京都千代田区 丸の内	神奈川県鎌倉市 神社
撮影機材	iPhone12	GoPro Hero9	iPhone8	Sony a7Ⅲ	GoPro Hero9

表-3 既存の画像解析技術を適用した結果

撮影地点番号		1	2	3	4	5
正解人数		87人	96人	185人	212人	335人
YOLOv4	出力画像					
	推定人数	39人	32人	59人	41人	23人
	検出率	44.8%	33.3%	31.9%	19.3%	6.9%
Mask R-CNN	出力画像					
	推定人数	64人	52人	92人	107人	69人
	検出率	73.6%	54.2%	49.7%	50.5%	20.6%
OpenPose	出力画像					
	推定人数	39人	38人	41人	2人	32人
	検出率	44.8%	39.6%	22.2%	0.9%	9.6%
AMRNet	密度マップ					
	推定人数	81人	52人	231人	396人	476人
	検出率	93.1%	54.2%	124.9%	186.8%	142.1%

#### 4. おわりに

本研究では、群衆を撮影した5つの地点に4つの画像解析技術を適用し、それぞれの技術の特徴や課題を整理した。そして、群衆人数の計数手法への適用可能性を検証した。その結果、すべての地点で精度良く検出するには、複数の画像解析技術を使い分ける必要があることがわかった。また、既存の学習済みモデルでは群衆の中にある人物の検出が難しいと判明したため、群衆内の人物計数により特化したモデルを新たに作成する必要がある。加えて、今回検証した4つの他にも、より高精度に物体を検出することを目的として数多くの画像解析技術が開発されているため、これらのアルゴリズムやモデルを追加で検証する予定である。

**謝辞：**本研究を遂行するにあたり、関西大学大学院総合情報学研究科の松尾龍平氏には検証に係わる貴重なご意見とご協力を賜った。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) 明石市：第32回明石市民夏まつりにおける花火大会事故調査報告書（概要版），2002.1, <<https://www.city.akashi.lg.jp/anzen/anshin/bosai/kikikanri/jikochosa/documents/gaiyou.pdf>>,（入手2022.6.13）。
- 3) Guangshuai, G., Junyu, G., Qingjie, L., Qi, W. and Yunhong, W.: CNN-based Density Estimation and Crowd

Counting: A Survey, <<https://arxiv.org/abs/2003.12783>>,（入手2022.6.13）。

- 4) 俣賀真由美：都心商業地におけるパブリックスペースの場所性と滞留者との関係に関する研究，九州大学大学院人間環境学府，2010.
- 5) 国土交通省：まちの活性化を測る歩行者交通量調査のガイドライン, <<https://www.mlit.go.jp/common/001282666.pdf>>,（入手2022.6.13）。
- 6) 今井龍一, 山本雄平, 姜文淵, 神谷大介, 中原匡哉, 安藤祐輝：頭部に着目した歩行者交通量調査に関する研究, 土木学会論文集 F3（土木情報学）, Vol. 78, No.2, pp.1\_82-1\_92, 2022.
- 7) NTT ドコモ モバイル空間統計：緊急事態宣言前後の人口変動分析, <[https://mobaku.jp/covid-19/report/#area\\_30](https://mobaku.jp/covid-19/report/#area_30)>,（入手2022.6.13）。
- 8) VACAN Maps：近くの空いているお店を探せるサイト, <<https://vacan.com/>>,（入手2022.6.13）。
- 9) Bochkovskiy, A., Wang, C. and Liao, H.: YOLOv4: Optimal Speed and Accuracy of Object Detection, <<https://arxiv.org/pdf/2004.10934.pdf>>,（入手2022.6.13）。
- 10) Kaiming, H., Georgia, G., Piotr, D. and Ross, G.: Mask R-CNN, *IEEE International Conference on Computer Vision*, Vol.42, No.2, pp.386-397, 2017.
- 11) Zhe, C., Gines, H., Tomas, S., Shih-En, W. and Yaser, S.: OpenPose: Realtime Multi Person 2D Pose Estimation Using Part Affinity Fields, *Computer Vision and Pattern Recognition*, pp.7291-7299, 2017.
- 12) Xiyang, L., Jie, Y. and Wenrui, D.: Adaptive Mixture Regression Network with Local Counting Map for Crowd Counting, *European Conference on Computer Vision*, pp.241-257, 2020.
- 13) Yingying, Z., Desen, Z., Siqin, C., Shenghua, G. and Yi M.: Single-Image Crowd Counting via Multi-Column Convolutional Neural Network, *Computer Vision and Pattern Recognition*, pp.589-597, 2016.