

## 動画像を用いた群衆人数の計数手法に関する基礎的研究

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

群衆人数の計数は、多くの人が集まる場所の安全性の確保、まちのにぎわいの評価や安全管理のために必要不可欠である。特に、群衆に起因する事故は世界中で発生しており、群衆の安全な誘導の指標として群衆人数を数える必要がある。一方、群衆人数の計数には大変な労力が伴うため、手軽に計数することができない。通常、群衆はカメラで撮影されることが多いため、群衆の写った動画像に深層学習を適用すると、手軽かつ正確に人数を計数できる可能性がある。既存研究に着目すると、群衆の写った動画像はカメラの撮影角度や視野などの条件によって多種多様であるため、様々な深層学習を用いた人数計数の手法が開発されていた<sup>2)</sup>。また、撮影条件毎の群衆の動画像に適した人数計数の手法が体系立てられていない状況であった。

以上より、本研究の目的は、群衆を撮影した動画像に人数計数ができる画像解析技術を適用し、計数精度を検証することとした。本検証により各画像解析技術の特徴や課題が整理されるため、撮影条件毎の群衆の動画像に適用する人数計数の手法の開発に貢献できる。

### 2. 検証に用いる動画像の概要

各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。まちづくりにおいてにぎわいを定量的に取得し、地域活性化の効果の把握が推進<sup>3)</sup>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、交差点・横断歩道の中でも、多くの人が往来する阪急うめだ周辺の横断歩道（地点1）および銀座周辺の交差点（地点2）において群衆の動画像を撮影した。群衆の動画像を撮影する際には、撮影地点の高さや撮影角度に縛りを設けず、交差点や横断歩道全体が写る

ように俯瞰して撮影した。また、撮影機材は地点ごとに異なるが、4K解像度に統一して撮影した。

### 3. 既存の画像解析技術の選定

既存の画像解析技術を調査し、各技術の特徴を把握することで、群衆人数の計数に適した技術を明らかにする。本研究では、人物の検出方法の違いに着目し、各技術における代表的な4種類の画像解析技術（図-1参照）を使用する。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>4)</sup>、人物領域をピクセル単位で検出して色塗りする画像領域分割の手法であるMask R-CNN<sup>5)</sup>、人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>6)</sup>、人が多い場所をヒートマップとして表現する密度マップの手法であるAMRNet<sup>7)</sup>を使用した。人物の検出には、インターネット上に公開されている学習済みモデルを使用し、人物以外も検出するYOLOv4やMask R-CNNでは、人物以外が検出された場合は検出結果を非表示にした。



図-1 選定した画像解析技術

キーワード 群衆, 群衆人数, 深層学習, 画像解析技術, 人物認識, 安全管理

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

#### 4. 選定した画像解析技術を適用した結果

本章では、交差点で撮影した群衆の画像に 3 章で選定した画像解析技術を適用し、その結果から判明した各画像解析技術の特徴と課題を整理する。検証においては、人が多く存在している瞬間の 1 フレームを切り出し、正解人数は対象の画像を複数人で数えた値の平均、検出率は推定人数を正解人数で除算した値とした。各画像解析技術を適用した結果を図-2 に示す。地点 1 と地点 2 を比較すると、地点 2 は地点 1 よりも写っている人数が多く、人物が小さく写っているといえる。

まず、YOLOv4 と Mask R-CNN は、画面手前側の人物を比較的正確に検出できているが、画面奥側で小さく写っている人物やオクルージョンが発生している人物の検出ができていないことがわかる。その理由としては、現状のモデルでは群衆の状態における人物の学習データが少なく、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考えられる。また、OpenPose は、画像の中の人物を検出するのではなく、人物の関節点を一気に検出した後に、それぞれの人物を構成する関節点をまとめあげるボトムアップ方式であり、小さな人物の関節の検出が困難であることから、検証した画像解析技術の中で最も検出率が低くなったと考えられる。

次に、AMRNet は、地点 2 の手前側と奥側で人の密集度は大きく変化していないにも関わらず、視覚的に奥側は人が集まっているように見えるため密度マップの色が濃く表現された。そして、実際より密度マップが濃く表現されたことで人数を過大に計数し、検出率が 100%を超える場面が多くなったと考えられる。

#### 5. おわりに

本研究では、群衆を撮影した動画に 4 つの画像解析技術を適用し、特徴や課題を整理することで、群衆人数の計数への適用可能性を検証した。検証結果より、YOLOv4 と Mask R-CNN は人物が大きく写っている場面での使用に適しており、AMRNet は密集度が高い場面での使用に適している。一方、小さい人物の検出が困難である OpenPose は群衆人数の計数には適していないといえる。今後は、物体検出や画像領域分割の手法を用いて、小さく写っている人物やオクルージョンが発生している人物でも検出が可能な学習モデルの作成を目指し、検出するクラスの工夫や学習データの追加等、より正確に群衆人数を計数できる手法を模索する。

撮影地点		1	2
入力画像			
撮影地点		大阪府大阪市 阪急うめだ周辺	東京都中央区 銀座周辺
撮影機材		iPhone8	iPhone12
正解人数		186人	558人
物体検出	YOLOv4	出力画像	
		推定人数 検出率	59人 31.7%
画像領域分割	Mask R-CNN	出力画像	
		推定人数 検出率	92人 49.5%
人物姿勢推定	Open Pose	出力画像	
		推定人数 検出率	41人 22.0%
密度マップ	AMRNet	出力画像	
		密度マップ	
		推定人数 検出率	231人 124.2%
		出力画像	
		密度マップ	
		推定人数 検出率	727人 130.4%

図-2 画像解析技術を適用した結果

**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所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) Jiang, L., Chenqiang, G., Deyu, M. and Alexander, G. H.: DecideNet: Counting varying density crowds through attention guided detection and density estimation, *Cornell University arXiv:1712.06679v2*, pp.1-10, 2018.
- 2) Guangshuai, G., Junyu, G., Qingjie, L., Qi, W. and Yunhong, W.: CNN-based Density Estimation and Crowd Counting: A Survey, *Cornell University arXiv:2003.12783v1*, pp.1-25, 2020.
- 3) 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, まちの活性化と歩行者量, <<https://www.mlit.go.jp/common/001282666.pdf>>, (入手 2022.3.30).
- 4) Alexey, B., Chien-Yao, W. and Hong-Yuan, M. L.: YOLO v4: Optimal Speed and Accuracy of Object Detection, *Cornell University arXiv:2004.10934v1*, pp.1-17, 2020.
- 5) 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.
- 6) Zhe, C., Gines, H., Tomas, S., Shih-En, W. and Yaser, S.: OpenPose: Realtime Multi Person 2D Pose Estimation using Part Affinity Fields, *Cornell University arXiv:1812.08008v2*, pp.1-14, 2019.
- 7) Xiyang, L., Jie, Y. and Wenrui, D.: Adaptive Mixture Regression Network with Local Counting Map for Crowd Counting, *Cornell University arXiv:2005.05776v2*, pp.1-16, 2020.