(43) 複数の人数計数手法の組み合わせによる群衆人数の推計に関する一考察

野村　圭哉[1]・今井　龍一[2]・山本　雄平[3]・姜　文渊[4]・
中原　匡哉[5]・神谷　大介[6]

[1]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

[2]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: imai@hosei.ac.jp

[3]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

[4]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

[5]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

[6]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

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．これらの手法を群衆を撮影した動画像に適用することで，手軽かつ正確に人数を自動で計数できる可能性がある．しかし，群衆の動画像は様々な条件で撮影されているため，自動で最適な人数計数手法を選択して適用することは困難である．

本研究では，人数計数手法の検出結果の傾向の違いに着目し，回帰分析を用いて人数を推計する手法を考案した．検証の結果，回帰分析によって人数計数手法ごとの弱点を補完し合うことで，稀に発生する極端な誤検出や検出漏れが緩和され，より正確に人数を推計できることが判明した．

Key Words: *crowd counting, crowd analysis, deep learning, regression analysis*

1. はじめに

イベントや災害時において群衆の状態を把握し，事故を未然に防ぐという観点から，群衆の人数を計数することは非常に重要[1]である．さらに，群衆を形成するまでの小規模な群衆人数の変化から，イベントなどの特定のエリアに人が集まる過程を把握する必要がある．しかし，長時間かつ多様な条件において群衆の人数を人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 と AI 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，的確に群衆の人数を把握でき，交通整理や入場規制など混雑状況に合わせた最適な手立てを打つことで群衆事故の防止が期待[2]できる．画像中の人数を計数する既存研究[3]では，群衆が写る動画像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とされている．そこで今井ら[4]は，複数の人数計数手法を比較し，群衆の人数規模ごとに手法の精度が異なる点と，人数規模に応じた最適な手法を明らかにした．しかし，単一の入力画像から撮影条件や群衆の状態を推測できず，最適な人数計数手法を選択するのは困難である．

これらの動向を踏まえて，本研究の目的は，1 枚の画像に対して複数の人数計数手法を適用し，検出結果から回帰的に人数を推計する手法の開発とした．回帰分析を用いて複数の検出結果から総合的に人数を推計することで，撮影条件や群衆の状態など，群衆を撮影する上で発生する不確定要素による影響を抑えることができる．さらに，人数計数手法ごとの弱点を補完し合うことで，様々な群衆の状態に対応した汎用的な群衆人数の推計手法の開発が可能になると考えた．

2. 既存の人数計数手法の適用可能性の検証

本研究では，動画像に既存の人数計数手法を適用し，その結果から手法ごとの特徴と課題を分析する．本研究で比較する人数計数手法は2手法あり，個々の人物の位置を検出する手法と密度から推計する手法を用いる．検証結果を定量的に評価するために，正解人数と検出・推計人数の差を表す人数（以下，「誤差人数」とする．）を算出した．そして，正解人数に対する誤差人数の絶対数の割合から平均を求めた値（以下，「誤差率の平均」とする．）を用いて各手法を比較した．

(1) 検証に用いる動画像の選定

既存の人数計数手法の特徴や課題を明らかにするため，人数規模が異なる群衆の画像を選定する．本研究では，既存研究[4]で選定された10地点を代表的な群衆の動画像として設定した（**表-1**参照）．検証用の画像は，**表-1**と同等規模の人数かつ俯瞰する角度で撮影された画像を各地点20枚ずつ，計210枚を既存のデータセットから選定する．利用するデータセットは，比較的人数規模が小さい画像で構成されるFDST[5]，群衆用のデータセットで人数規模が大きい画像で構成されるUCF-QNRF[6]，JHU-CROWD++[7]およびNWPU-Crowd[8]の4種類とした．

(2) 検証する手法と学習モデルの選定

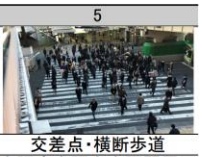
本研究では，検出した人物の領域を矩形で囲む物体検出の手法であるYOLOv4[9]，YOLOv6 v3.0[10]，YOLOv7[11]，VisDrone[12]，PPhuman[13]およびヒートマップとして密集を表現する手法であるAMRNet[14]の6手法を用いる．人物を検出するモデルは，公開されている学習モデル6つと既存研究[4]において CrowdHuman データセット[15]をYOLOv4，v6およびv7に学習させて作成した頭部を検出するモデル（以下，「CH」とする．）3つの計9つを検証に用いる．

(3) 既存の人数計数手法を適用した結果と考察

1地点21枚の画像ごとに算出した誤差率の平均と210枚における誤差人数の結果を**図-1**および**図-2**に示す．まず，**図-1**のグラフより4つのグループに分けられる．正解人数の増加に伴い，誤差率の平均が徐々に低下するVisDrone（グループ1），大幅に低下するAMRNet（グループ2），徐々に上昇するYOLOv6-CHとYOLOv7-CH（グループ3），大幅に上昇するYOLOv4，YOLOv4-CH，YOLOv6-L，YOLOv7-Xおよび PPhuman（グループ4）の4つのグループに分けることができる．

次に，**図-2**の正解人数が200人程度のエリアでは，グループ1と2において誤差人数がプラスの側に多くの点が

表-1 代表的な群衆の画像

撮影地点番号	1	2	3	4	5
入力画像					
活用場所	待機行列	単路	自由な歩行空間	単路	交差点・横断歩道
撮影地点	神奈川県鎌倉市	静岡県静岡市改札	大阪府大阪市大阪駅周辺	神奈川県鎌倉市	大阪府大阪市大阪駅周辺
撮影機材	GoPro Hero9	GoPro Hero9	iPhone12	GoPro Hero9	iPhone8

撮影地点番号	6	7	8	9	10
入力画像					
活用場所	イベント会場	自由な歩行空間	待機行列	自由な歩行空間	自由な歩行空間
撮影地点	東京都千代田区丸の内	UCF-QNRF	神奈川県鎌倉市	UCF-QNRF	UCF-QNRF
撮影機材	Sony α7Ⅲ	-	GoPro Hero9	-	-

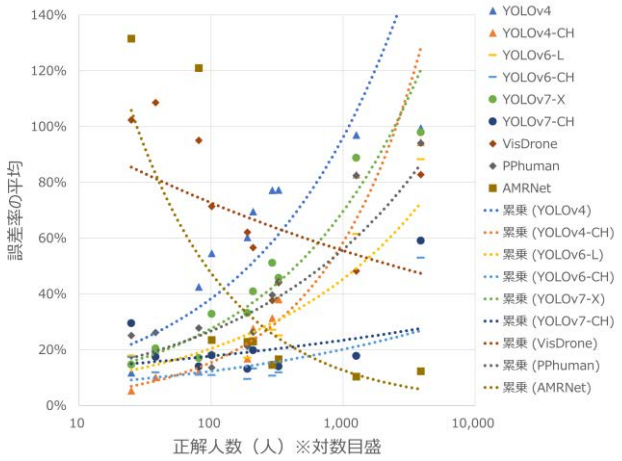


図-1 正解人数と各手法の誤差率の平均

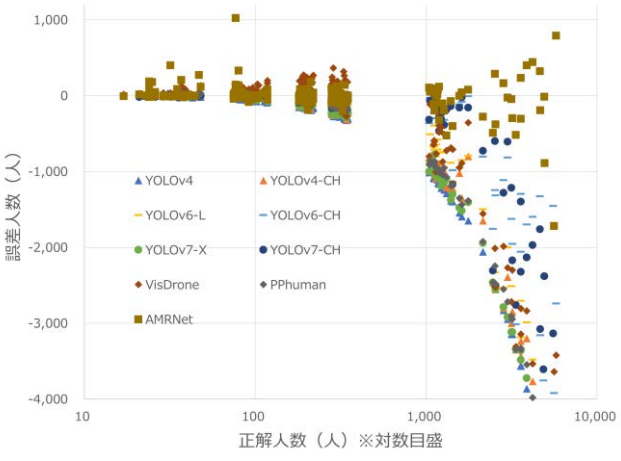


図-2 正解人数と各手法の誤差人数

存在しており，過剰検出が発生していることがわかる．また，グループ3と4では誤差人数がマイナスの値であり，検出漏れが発生している．ただし，VisDroneの実態は正解人数が数百人の画像での過剰な検出であり，常に50%以上の誤差率である点に留意する必要がある．そして，グループ3では，正解人数が1,000人を越えると検出漏れが発生し，4,000人規模の画像において検出漏れが50%程度発生している．

以上より，人数計数手法ごとに検出結果の傾向がそれぞれ異なるため，群衆の人数の規模に応じて適切な手法を使い分けることが有用であるといえる．しかし，単一の入力画像から撮影条件や群衆の状態を推測できず，最適な人数計数手法を手軽に選択することは困難である．

3. 回帰を用いた群衆人数の推計手法の検証

2章の結果より，人数計数手法ごとに検出結果の傾向が異なるという特性を明らかにすることができた．人数規模に応じた手法の切り替えは困難であるが，手法ごとに特徴的な検出結果になるため，回帰モデルを活用することで，正確かつ手軽に人数を推計できると考えた．そこで，6つの回帰モデルを比較し，より正確に人数を推計できる手法を明らかにする．

(1) 検証方法

まず，2章で選定した210枚のうち，表-1以外の200枚における9手法の検出結果を入力，200枚それぞれの正解人数を出力とする．モデルは統計学の手法の１つである回帰分析を用いる．次に，作成した回帰モデルに表-1の10枚を入力して推計される人数の結果を比較する．

(2) 検証に用いる回帰モデルの概要

線形回帰と非線形回帰の両方を用いて計6つの手法を検証する．線形回帰ではLinear Regression，Ridge，Lasso，ElasticNetの4つのモデルで直線的な関係をモデル化する．Ridge回帰はL2正則化で過学習を防ぎ，Lasso回帰はL1正則化で変数選択を行う．また，ElasticNet回帰はL1正則化とL2正則化を組み合わせ，柔軟なモデリングが可能である．非線形回帰ではDecision TreeとRandom Forestの2つの手法で非線形な関係をモデル化する．Random Forestは条件分岐によりデータの特徴を捉え，アンサンブル学習による精度向上が図られている手法である．

(3) 回帰モデルで群衆人数を推計した結果と考察

6つの回帰モデルを200枚で機械学習した際のモデルの重みを表-2に示す．ここで，表-2における重みとは，線形回帰では偏回帰係数(coefficient)，非線形回帰では特徴量の重要度(feature importance)である．また，残りの10枚における9つの人数計数手法による検出結果を入力として，回帰モデルによって推計された人数を表-3に示す．

表-2および表-3より，線形回帰の4手法の重みは一部マイナスの値が含まれており，更に推計人数もマイナスになっているため，多重共線性が発生していることが考

表-2　人数計数手法と回帰モデルごとの重み

人数計数手法	Linear Regression	Ridge	Lasso	ElasticNet	Decision Tree	Random Forest
YOLOv4	-0.01354	-0.01354	-0.01209	-0.01355	0.00008	0.00096
YOLOv4-CH	-0.75844	-0.75844	-0.75837	-0.75821	0.00058	0.00074
YOLOv6-L	1.30934	1.30934	1.30906	1.30896	0.00010	0.00066
YOLOv6-CH	-0.03663	-0.03663	-0.03668	-0.03670	0.08373	0.07625
YOLOv7-X	-1.08366	-1.08366	-1.08352	-1.08299	0.00375	0.00045
YOLOv7-CH	-0.43193	-0.43193	-0.43177	-0.43175	0.00018	0.00647
Visdrone	-0.08040	-0.08040	-0.08031	-0.08031	0.00730	0.00320
PPhuman	1.88354	1.88353	1.88296	1.88278	0.00076	0.00259
AMRNet	0.99921	0.99921	0.99921	0.99921	0.90353	0.90869

表-3　9手法による検出結果を入力した際の回帰モデルごとの推計人数（人）

撮影地点番号	入力：[YOLOv4, YOLOv4-CH, YOLOv6-L, YOLOv6-CH, YOLOv7-X, YOLOv7-CH, Visdrone, PPhuman, AMRNet]	正解人数	Linear Regression	Ridge	Lasso	Elastic Net	Decision Tree	Random Forest
1	[16, 22, 25, 43, 21, 10, 34, 31, 53]	24	-14.28	-14.28	-14.30	-14.27	26.00	25.90
2	[21, 35, 27, 33, 21, 19, 54, 37, 28]	33	-40.39	-40.39	-40.41	-40.39	38.00	33.10
3	[39, 100, 77, 86, 70, 62, 99, 103, 81]	87	75.57	75.57	75.55	75.56	80.00	85.10
4	[33, 90, 54, 83, 28, 46, 160, 116, 51]	96	95.26	95.26	95.23	95.23	77.00	66.40
5	[59, 173, 103, 177, 89, 103, 217, 156, 230]	185	251.57	251.57	251.56	251.55	192.00	203.00
6	[41, 270, 176, 257, 139, 117, 293, 179, 395]	212	412.74	412.74	412.70	412.74	282.00	288.30
7	[138, 265, 279, 285, 278, 278, 489, 209, 235]	283	209.76	209.76	209.87	209.84	308.00	316.50
8	[23, 138, 51, 154, 55, 109, 82, 134, 475]	335	459.83	459.83	459.78	459.81	280.00	257.10
9	[59, 272, 586, 1084, 195, 995, 1151, 210, 1179]	1165	1250.39	1250.39	1250.42	1250.42	1277.00	1179.10
10	[66, 428, 714, 2596, 156, 2226, 662, 216, 4638]	4195	4261.63	4261.63	4261.72	4261.69	3884.00	4674.30

表-4　AMRNet の検出人数が極端な画像における回帰モデルごとの推計人数（人）

入力：[YOLOv4, YOLOv4-CH, YOLOv6-L, YOLOv6-CH, YOLOv7-X, YOLOv7-CH, Visdrone, PPhuman, AMRNet]	正解人数	Linear Regression	Ridge	Lasso	Elastic Net	Decision Tree	Random Forest
[20, 21, 36, 20, 24, 18, 74, 34, 211]	24	156.00	156.00	155.99	156.01	29.00	25.40
[28, 24, 28, 23, 28, 23, 77, 30, 7]	24	-75.29	-75.29	-75.29	-75.28	29.00	26.00

えられる．入力に用いた検出結果は，手法ごとに傾向が異なり，1つの線形では表せないことを示している．一方，**表-2**より，非線形回帰の2手法においては，AMRNet の重みが他の人数計数手法と比較して極端に大きくなっているものの，**表-3**のとおり，正解人数に近い値が推計されていることがわかる．したがって，複数の人数計数手法の検出結果から人数を推計するには非線形回帰が適しているといえる．

ここで，非線形回帰における AMRNet の影響を確かめるため，AMRNet の検出人数が極端に正解から外れている画像で人数を推計した結果を**表-4** に示す．結果を確認すると，正解人数が 24 人の画像に対して AMRNet による検出人数が 211 人や 7 人である場合でも，非線形回帰においては推計人数への影響が少なく，20~30 人前後に収まることが判明した．特徴量の重要度の値をみると，推計値が AMRNet による影響を大きく受けると考えられるが，検証の結果としては，それ以外の 8 つの手法によって値が補正されていることを確認した．

以上のように，仮に人数計数の結果が極端に外れた場合においても検出結果に非線形回帰を用いると，推計人数を補正できることを明らかにした．

4. おわりに

本研究では，人数計数手法の検出結果の違いに着目し，回帰的に群衆の人数を推計する手法を考案した．本手法を用いることで各手法の弱点の補完や不得意な条件による影響を補正できると判明した．今後発表される新たな人数計数手法による結果を回帰分析の特徴量として用いることで，正解人数により近い値が得られるかを検証する必要がある．また，本研究では9つの人数計数手法の結果をすべて入力していたが，計算コストをより削減するため，非線形回帰に入力するべき手法を絞り込む必要がある．そのため，回帰モデルに用いる人数計数手法の最適な組み合わせを検証する予定である．

参考文献

1) Zhang, Y., Wang, T., Li, G., Lei, J. and Wang, L.: Crowd Counting using DMCNN, *International Conference on Innovation in Artificial Intelligence*, pp.138-144, 2019.
2) 兵庫県警察：雑踏警備の手引き，<https://www.police.pref.hyogo.lg.jp/zattou/tebi_data/all.pdf >，（入手2023.6.10）．
3) Gao, G., Gao, J., Liu, Q., Wang, Q. and Wang, Y.:CNN-based Density Estimation and Crowd Counting: A Survey, <https://arxiv.org/abs/2003.12783>, (accessed 2023.6.10).
4) 今井龍一，山本雄平，姜文渊，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.
5) Fang, Y., Zhan, B., Cai, W., Gao, S. and Hu, B.: Locality-Constrained Spatial Transformer Network for Video Crowd Counting, *International Conference on Multimedia and Expo*, pp.1-6, 2019.
6) Idrees, H., Tayyab, M., Athrey, K., Zhang, D., Al-Maadeed, S., Rajpoot, N. and Shah, M.: Composition Loss for Counting, Density Map Estimation and Localization in Dense Crowds, *European Conference on Computer Vision*, pp. 544-559, 2018.
7) Sindagi, V., Yasarla, R. and Patel, V.: JHU-CROWD++: Large-Scale Crowd Counting Dataset and A Benchmark Method, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.14, No.8, pp.1-17, 2020.
8) Wang, Q., Gao, J., Lin, W. and Li, X.: NWPU-Crowd: A Large-Scale Benchmark for Crowd Counting and Localization, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.43, No.6, pp.2142-2149, 2020.
9) Bochkovskiy, A., Wang, C. and Liao, H.: YOLOv4: Optimal Speed and Accuracy of Object Detection, <https://arxiv.org/pdf/2004.10934.pdf>, (accessed 2023.6.10).
10) 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 2023.6.10).
11) 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 2023.6.10).
12) GitHub: VisDrone, <https://github.com/PaddlePaddle/PaddleDetection/tree/release/2.4/configs/VisDrone>, (accessed 2023.6.10).
13) GitHub: pphuman, <https://github.com/PaddlePaddle/PaddleDetection/tree/release/2.4/configs/pphuman>, (accessed 2023.6.10).
14) Liu, X., Yang, J. and Ding, W.: Adaptive Mixture Regression Network with Local Counting Map for Crowd Counting, *European Conference on Computer Vision*, pp.241-257, 2020.
15) 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 2023.6.10).