2U-03

歩行者流を考慮した 人流密度推定ニューラルネットワークの検討

羽原 丈博 †　　小島 諒介 ‡
京都大学工学部電気電子工学科 †　京都大学医学研究科 ‡

1 背景

ある領域内の人の数をカウントする群衆数え上げや，領域内の群衆の密度分布をヒートマップなどで可視化する密度推定は群衆解析と呼ばれている．この群衆解析は監視カメラやイベント会場での誘導，交通ルートの設計といった分野で使われていおり，特に最近では新型コロナウイルスの流行における影響から，イベント会場などでの人の分布を可視化し，評価する目的でも注目されている．群衆解析を用いることで部屋の中にいる人の数やイベント会場等などでの人の分布を可視化することができ，感染症対策に直接貢献する．また群衆解析は粒子の運動の解析や野生動物の行動解析といった人間以外を対象とした応用も期待されている．

群衆数えの手法として，群衆の数を直接求める手法があるが, 近年では，より多くの情報が得られることから直接群衆の数ではなく，密度推定を用いた研究 [5][2] も広く行われている．これらの研究では，密度を推定し，その総和として間接的に人数を計算している．

人流の密度推定では，単に，密度を推定するほかに，人の移動している方向を流入および流出量として計算する方法 [3] が提案されている．この手法 [3] では Context-Aware-Network(CAN) モデル [2] を用いて人の時間変化から，流入力と流出量を考慮した損失関数を導入している．しかしこれらの損失関数はグリッド間の流入・流出量しか考慮していないため，群衆の人間らしさは取り入れられておらず，推定した結果が拡散しやすいという問題がある．

2 提案手法

既存の手法 [3] はグリッド間の流入・流出量しか考慮していない 2 種類の損失関数 L_{cycle} と L_{flow} を導入していた．入力は時刻 t, $t+1$ の群衆画像であり，出力は上下左右斜めの 8 方向と移動しない量と外部からの流入量の 10 チャンネルの 8 ピクセル四方を一つのグリッドマップとしたものを考える．まず L_{cycle} は式 (1) で表され，時刻 t から $t+1$ と変化したときのあるグリッドへの流入量は，時刻 $t+1$ から t へ変化したときのそのグリッドからの流出量と一致することを意味する．$I_{i,j}(t,t+1)$ は時刻 t から $t+1$ にかけて i,j グリッドへ流入する量，$O_{i,j}(t,t+1)$ はこのグリッドから流出する量を示す．

$$L_{cycle} = \sum_{i,j} \left(I_{i,j}(t,t+1) - O_{i,j}(t+1,t)\right)^2 \qquad (1)$$

また L_{flow} は式 (2) と表され，時刻 $t-1$ から t へのあるグリッドへの流入量の和と時刻 $t-1$ から t へのそのグリッドからの流出量の和は一致することを意味する．

$$L_{flow} = \sum_{i,j} \left(I_{i,j}(t-1,t) - O_{i,j}(t,t+1)\right)^2 \qquad (2)$$

我々は推定結果が群衆の人らしさを表現するために歩行者流 [6] の性質を取り入れた新たな損失関数を提案する．歩行者流とは同じ方向に進むもの同士が形成する集団のことであり，後続の集団は前の人に追従する性質が知られている．我々はこの性質を式 (3) として正則化項 $L_{directional}$ を提案する．

$$L_{directional} = \sum_{c} \sum_{i,j} \left(y_{c,i,j} - y_{c,i+\alpha,j+\beta}\right)^2 \qquad (3)$$

i,j は出力グリッドマップの位置を示す．c は上下左右斜めの 8 方向を表す変数であり，α_c，β_c はその方向のグリッドへの相対座標である．例えば c が左斜め上を示すとき，$(\alpha_c, \beta_c) = (-1,-1)$ を示す．これは隣接したグリッドは同じ方向に同じ流量進みやすいという現象を表している．これにより歩行者の追従性を考慮する正則化が可能になる．

3 ベンチマークによる評価

CAN[3] モデル，L_{cycle}，L_{flow} 損失関数に加えて $L_{directional}$ 正則化項を用いて学習を行った．使用したデータセットは CrowdFlow[4]，Venice[2]，FDST[1] である．CrowdFlow はシミュレーションによって作られた群衆が歩行する 5 つの人工動画データである．Venice と FDST は群衆が公道を歩行するデータセットである．

CrowdFlow, Venice, FDST のデータセットによる評価の結果をそれぞれ表 1, 表 2, 表 3 に示す． MAE, RMSE はそれぞれ画像内の推定された人数と正解人数の平均絶対値誤差と二乗平均平方根誤差である．pix-MAE，pix-RMSE はグリッドごとに計算した正解人数との MAE と

Human Flow Density Estimation Neural Network Considering Pedestrian Flow
† Takehiro Habara, Kyoto University Undergraduate School of Electrical and Electronic Engineering
‡ Ryosuke Kojima, Kyoto University Graduate School of Medicine

表 1　CrowdFlow データセットの評価

	MAE	RMSE	pix-MAE	pix-RMSE
文献値 [3]	97.8	**112.1**	—	—
再現実験	142.4	175.0	0.025	0.067
提案手法	**94.66**	123.5	0.023	0.067

表 2　venice データセットの評価

	MAE	RMSE	pix-MAE	pix-RMSE
文献値 [3]	15.0	19.6	—	—
再現実験	8.87	11.9	0.027	0.08
提案手法	**8.81**	**10.7**	0.026	0.08

表 3　FDST データセットの評価

	MAE	RMSE	pix-MAE	pix-RMSE
文献値 [3]	2.17	2.62	—	—
再現実験	**1.96**	**2.61**	0.003	0.030
提案手法	2.05	2.68	0.003	0.029

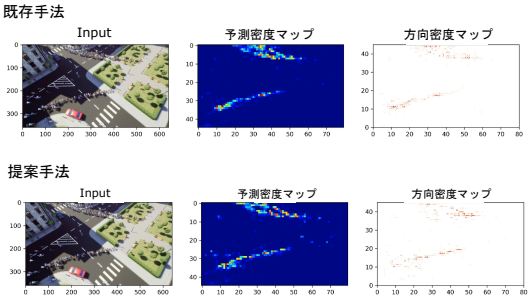

図 1　CrowdFrow のデータセットによる推測結果: 左が入力画像，中心が人の密度マップ，右が方向密度マップ (色が濃いほど多い)

RMSE を表す．この結果から提案手法はグリッドごとの精度は変わらないが，全体の人数の観点からみると精度がよくなっていると言える．

また図 1 における，画像上部の二列に歩行者が歩いている領域において，既存手法では拡散によって 2 列の間も人がいると予測しているが，提案手法では拡散せずに 2 列を強調して予測できていることが分かる．

4　外部データによる評価

学習したモデルの汎化性能を調べるために，我々は簡単な黒点動画のデータセットを構築し，人以外の外部データへの汎化の可能性を評価した．黒点は 3 × 3，4 × 4，5 × 5 の群集をなし，原点に対し点対称の位置に 2 群配置し，原点に関し時計回りに動く設定とした (図 2)．

CrowdFlow のデータセットを用いて学習したモデルを用いて予測した結果を表 4 に示す．また前節と同様に入力画像に対する密度マップと方向密度マップを図 2 に示す．これらの結果より両者とも数え上げや密度推定に関してはある程度の汎化性能を示すことが分かった．一方、群衆の流れる方向に関してはうまく予測できていないことがわかる。これは，モデルは方向を推定する際に人の顔の向きといった情報を利用していることが原因として考えられる．

表 4　黒点データの評価

	既存		提案	
	MAE	RMSE	MAE	RMSE
18 個	1.06	1.18	3.69	3.76
32 個	13.3	13.3	11.9	11.9
50 個	28.2	28.2	23.8	23.8

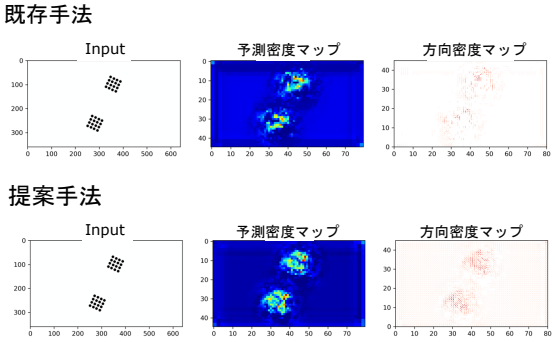

図 2　32 個の点群のデータによる推測結果

5　結論

我々は群衆数え上げニューラルネットワークモデルの学習において歩行者流の数式モデルを正則化関数として加えることで，より効果的な群衆数え上げをなすことができることを示した．

また人工点群のデータによる実験から密度推定に関してはある程度の汎化性能を示すことが分かった。今後、歩行者流に関するさらに詳細な性質をモデルに取り入れることで、さらなる精度向上が期待できる。

謝辞

本研究は JSPS 科研費 No.20H00475, 19KK0260 の助成を受けた．

参考文献

[1] Fang, Y., Zhan, B., Cai, W. and et al.: Locality-constrained spatial transformer network for video crowd counting, *2019 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 814–819 (2019).
[2] Liu, W., Salzmann, M. and Fua, P.: Context-aware crowd counting, *CVPR 2019*, pp. 5099–5108 (2019).
[3] Liu, W., Salzmann, M. and Fua, P.: Estimating people flows to better count them in crowded scenes, *ECCV*, pp. 723–740 (2020).
[4] Schröder, G., Senst, T., Bochinski, E. and Sikora, T.: Optical flow dataset and benchmark for visual crowd analysis, *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–6 (2018).
[5] Thanasutives, P., Fukui, K.-i., Numao, M. and et al.: Encoder-Decoder Based Convolutional Neural Networks with Multi-Scale-Aware Modules for Crowd Counting, *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 2382–2389 (2021).
[6] 兼田敏之：歩行者流のエージェントシミュレーション，計測と制御，Vol. 43, No. 12, pp. 944–949 (2004).