

基盤モデルを用いた都市公園利用者数の自動計測手法の提案

○岡野 将大[†], 武田 陸[‡], 藤井 純一郎[†]

[†]: 八千代エンジニアリング株式会社 技術創発研究所

[‡]: 八千代エンジニアリング株式会社 九州支店 都市デザイン部

ms-okano@yachiyo-eng.co.jp

概要:本研究では,都市公園における利用者数の自動計測手法を提案する.少子高齢化が進む中,都市公園の管理運営最適化は重要な課題となっているが,従来の人手計測は時間的・人的コストが高い.CNN ベースモデルは高精度を達成しているが,大量の教師データが必要で複数地点展開時の汎用性に課題がある.基盤モデルによるゼロショット手法は追加学習なしで使える汎用性が期待されるが,初期実験では樹木や建造物による背景誤検知と密度変化時の精度変動が確認された.そこで,SAM2 による背景除去,CLIP-EBC によるゼロショット群衆カウント,密度閾値処理を統合した 3 段階処理手法を開発した.都市公園 6 地点で実験を行い,密度閾値調整により閾値なしの MAE 34.67 から 11.67 へと約 66%の誤差削減を達成し,追加学習なしで基盤モデルの汎用性を活かした実用的な手法の有効性を確認した.

<キーワード> 群衆カウント, 公共空間, 基盤モデル

1. はじめに

少子高齢化が進む中で,都市公園の利活用や管理運営の最適化は重要な社会的課題となっている[1].利用者数の把握は,管理最適化の基礎データとなるが,現在の人手による計測手法では,時間的・人的コストが高く,継続的な計測に課題がある[2].

これらの課題に対して,画像処理技術を活用した自動計測手法の導入が検討されている.画像処理技術の中で,人数を計測するタスクとして群衆カウントタスクがあり,CSRNet[3]などの CNN ベースのモデルが高い計測精度を実現している.しかし,これらのモデルを屋外空間に適用する際には,樹木や建造物などの背景要素が誤カウントを生む課題がある.また,新しい環境で高い精度にするためには大量の教師データとアノテーション作業が求められ,複数地点へ展開するには汎用性に課題がある[4].

そこで,基盤モデルによるゼロショット手法に着目する.基盤モデルは大規模データで事前学習されており,追加学習なしに様々なシーンで安定して高い精度が得られることが期待される.しかし,初期実験では都市公園環境において以下の問題が確認された:(1) 樹木や建造物による背景誤検知,(2) 閑散時と混雑時の利用者密度変化での精度変化.これらの課題により安定した人数把握が難しいことが確認された.

そこで本研究では,都市公園環境における背景

誤検知抑制と密度変化に対応した,基盤モデルによるゼロショット人数計測手法を提案し,その計測精度を評価することを目的とする.



図 1 : 撮影画像例

※対象が画像全体と比べ小さいため切り取り画像



図 2 : CLIP-EBC の密度マップ

2. 人数計測基盤モデル (CLIP-EBC)

2.1. CLIP-EBC の動作原理

CLIP-EBC (CLIP Can Count Accurately through Enhanced Blockwise Classification)[6]は、大規模視覚言語モデル CLIP を基盤として、ブロック単位での密度分類により群衆カウントを実現する手法である。従来の群衆カウント手法が密度マップ回帰を用いるのに対し、CLIP-EBC は画像を複数のブロックに分割し、各ブロックの人数を分類問題として解く。この手法の利点は、異なる密度スケールに対して柔軟に対応できる点と、大規模データで事前学習された CLIP の汎用的な特徴表現を活用しながら、追加学習なしで群衆カウントを可能にしている点にある。

2.2. 都市公園環境での初期実験

提案手法の基礎となる CLIP-EBC の性能を評価するため、都市公園 6 地点で撮影した画像データセットを用いて初期実験を実施した。実験では、CLIP-EBC をそのまま適用し、追加の前処理や後処理を行わない状態での計測精度を評価した。各地点から 30 分間隔で撮影した画像計 120 枚 (各地点 20 枚) を使用し、人手による計測値を Ground Truth として精度を検証した。

2.3. 初期実験結果の分析と課題

表 1: CLIP-EBC 単体での地点別計測精度

地点	MAE (人)	RMSE (人)	MAPE (%)	σ (人)	Max (人)
1	30.40	34.51	371.62	16.34	54
2	12.40	14.15	40.21	6.81	23
3	2.40	3.41	11.53	2.89	8
4	15.00	16.38	66.03	6.59	24
5	30.80	38.15	99.95	22.51	80
6	10.50	12.04	10.66	7.64	18

表 1 に地点別の計測精度、表 2 にデータ全体での計測精度を示す。全体での MAE は 16.92 人、RMSE は 23.35 人となり、ある程度の計測は可能であることが確認された。しかし、MAPE が 99.99% と極めて高く、特に利用者数が少ない場面での相対誤差が大きいことが判明した。さらに、地点間での精度にも大きなばらつきがある。

地点 3 では MAE 2.40 人、MAPE 11.53% と良好な精度を示した一方で、地点 1 では MAE 30.40 人、MAPE 371.62% と大きな誤差が発生している。Ground Truth (GT) の総人数では、地点 1 は 100 人と最も少なく、地点 6 は 1014 人と最も多い。この

利用者数と MAPE の関係から、利用者密度が低い環境ほど CLIP-EBC の相対誤差が大きくなることを確認した。

特に地点 1 の MAPE 371.62% という値は、平均的に実測値の 3.7 倍以上の誤差が生じていることを示している。図 3 に地点 1 での低密度時の画像例を示す。実際の利用者は数名程度であるにも関わらず、図 4 に示すように、CLIP-EBC の密度マップでは樹木の葉や枝の先端や、物体などが高い活性値を示しており、これらを人として誤検知していることが確認できる。地点 5 においても最大誤差 80 人という大幅な過大推定が確認され、これらの結果から、基盤モデルをそのまま適用するだけでは、特に低密度環境での安定した人数計測は困難であることが分かった。



図 3 : 地点 1 の低密度時の画像例 (後で変更)
※対象が小さいため切り取り画像

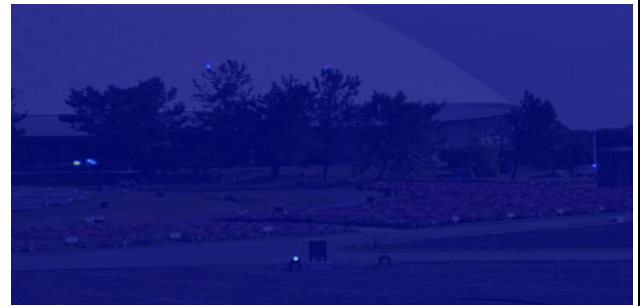


図 2 : 地点 1 の低密度時の誤検知例

3. 提案手法

都市公園環境での適用時における課題に対し、基盤モデルを用いた 3 段階処理手法を提案する。

1 SAM2 による背景誤検知の抑制

固定カメラ環境の利点を活かし、事前に樹木や建物にポイントプロンプトを設定して SAM2 (Segment Anything Model 2)[5]でマスク画像を生成し、人数計測の対象外領域として設定することで背景誤検知を抑制する。

2 基盤モデルによる汎用性の向上

基盤モデルである CLIP-EBC[6]によるゼロショット群衆カウントを行う。

3 密度変化の課題への対応

CLIP-EBC 出力における密度環境変化時の精度変動を抑制するため、密度閾値処理を導入する。各ブロックの予測人数に対し閾値を設定し、閾値以下の予測値は 0 に補正することで、低密度時の誤検知を除去する。最終的な人数は補正後の全ブロックの人数の総和として算出する。

本手法により、追加学習なしで屋外特有の背景誤検知と密度変化の両課題に対応した。

4. 実験概要

使用したデータセットの事例を図 1 に示す。本研究では、都市公園内の 6 地点で固定設置カメラを用いて 30 分間隔で撮影した画像をデータセットとして作成した。各地点は樹木密度や建物などが異なり、多様な環境条件となっている。

実験は人手計測値を Ground Truth (GT)として評価した。実験では以下の 3 つを行った。

- 1) 密度閾値を適用しない (閾値なし)
- 2) 密度閾値を一定値に設定 (閾値一定)
- 3) 密度閾値を画像ごとに調整 (閾値調整)

これにより、密度閾値処理が計測精度に与える影響を定量的に検証し、都市公園環境における提案手法の有効性を評価した。

精度評価には、MAE (Mean Absolute Error)および RMSE (Root Mean Square Error) を採用した。MAE は平均的な誤差の大きさを、RMSE は大きな誤差への敏感性を評価でき、両指標により提案手法の安定性と精度を検証した。

表 2 : 全データでの精度評価

	MAE	RMSE	MAPE	σ	Max
CLIP-EBC	16.92	23.35	99.99	16.40	80
閾値なし	5.73	9.85	19.69	9.85	55
閾値一定	10.18	13.39	26.15	10.95	41
閾値調整					

5. 結果

6. 考察

7. まとめ

謝辞 本研究は山口キララ博記念公園様よりデータのご提供をいただきました。厚く御礼申し上げます。ここに感謝の意を表します。

参考文献

- [1] OpenConf: URL: <https://www.tc-iaip.org/ViEW/2025/openconf/>
- [2] <https://www.tc-iaip.org/view/2025/>
- [3] 課題好一郎 : 広大なインターネット空間を対象とする FA 手法の提案, 日本 FA 学会論文誌, Vol.3, No.4, pp.124-129, 1999.
- [4] 令和太郎 : 外観検査課題の検討, ビジョン学会論文誌, Vol.6, No.4, pp.98-105, 1996.

令和太郎 : 各著者の簡単な略歴・研究歴, 現在の研究興味などを書いてください。ホームページの URL などもある有益な情報です。

ウィリアム・テイラー : Soho Inc. CEO. 多くの外観検査の自動化システムの設計・試作を行う。代表システムは, IESS, YTI, KEE など。最近では自律ロボットビジョンモジュールの開発を進めている。