# 動画解析による公共空間の行動検出 Action detection in public spaces by video analysis

岡野 将大\*1 Masahiro Okano

緒方 陸\*1 Riku Ogata

大久保 順一\*1 Junichi Okubo

藤井 純一郎\*1 Junichiro Fujii

安野 貴人\*1 Takato Yasuno

\*1 八千代エンジニヤリング株式会社 Yachiyo Engineering Co., Ltd. #1

Assessing the value of public spaces is an issue in urban planning. People's staying time in the target space is used as one of the indexes to quantify the value of public space. Until now, surveys have been conducted by people visually checking videos, but labor is required and work efficiency is required. Therefore, in this study, we examined a method to automatically evaluate the staying time by utilizing RTFM (Robust Temporal Feature Magnitude learning) shown by Tian et al. RTFM is a model for detecting abnormal behavior of surveillance cameras, and this method was applied to detect specific human behavior in public spaces. In this paper, we conducted RTFM learning on images taken with a handy camera in a public space, and compared the accuracy when the loss function and activation function were changed.

# 1. はじめに

#### 1.1 背景

近年,車両中心の空間から人中心の空間への転換が求めら れており、公共空間価値の定量評価においては、歩行者交通 量といった流れの指標と、滞在のしやすさなど行動の指標の二 種類ある[国土交通省 2019]. 評価方法としては現地で人手に よるカウント調査か, 監視カメラで撮影した動画から目視で人流 や行動をカウントするビデオ調査が主流である.しかしこれらの 調査は,膨大な手間と時間がかかることが課題であり,作業自 動化の達成は急務である.

#### 1.2 既往研究

公共空間価値の評価を自動化した研究事例は,評価指標の 一つとなる人流について、[高森 2021]の Re-Identification の手 法を用い, 通行人数を定量的に評価した例がある. しかし, 歩 行者の行動を定量的に評価(ベンチ利用者数の把握など)した 研究例はこれまでに無い.公共空間以外においては,深層学 習の登場により、AIを用いて動画から行動認識を行う研究が急 激に増えている. 近年では, NAS (Neural Architecture Search)を 取り入れた行動認識手法である AssembleNet が Charades など のベンチマークデータセットで高精度を発揮した[Ryoo 2020]. また, [杉村 2020]は不審行動などの検知を目的とした手法を 構築している.

したがって,動画から行動認識を行う技術を公共空間評価に も適用し, 効率的かつ効果的に人の行動を定量的に評価する 手法の確立が課題となっている.

#### 1.3 本研究の目的

将来的に人の行動を指標に基づき映像から自動検知・評価 することを見据え、深層学習により「滞在しやすさ」の指標の一 つである「ベンチ利用行動」を定量的に評価する技術を開発し た.本研究では、精度の向上を目指し、損失関数や活性化関 数を変更した場合の精度比較を実施した.

精度が向上することで本手法の活用により得られるメリットを 最大限にすることが出来ると期待される.

## 2. 実験概要

# 2.1 学習データ

学習用および検証用データとして, 滋賀県内のサービスエリ ア A にて撮影された動画データを用いた.この動画データは, ベンチの利用状況を調査するために撮影したものである.現状 の調査では、人手により、図1の赤枠で示すようなベンチ利用 者を動画から手作業で確認している. 本研究ではこれらの動画 データを用い,「ベンチ利用行動」有無判定のため,学習・推論 を行った.用いた動画は 212 本であり,各動画の長さは数十秒 ~1分半程度,高さ×幅は720×1280,フレームレートは15 fps である. 歩行者がベンチを利用しているフレームを positive, 利 用していないフレームを negative とし、教師データを作成、推論 を実施した.

#### 2.2 手法

本研究では[Tian 2021]の RTFM (Robust Temporal Feature Magnitude learning)をベースとしたモデルを構築した. なお, RTFM は行動認識で用いられる Two-stream ConvNets, TSN, I3D, Non-local block などの手法を活用しており, 人間の異常行 動を検知する目的として提案された手法である.本研究ではこ れを改良し、特定の行動を検知する目的として利用した.

#### 2.3 精度評価

図1の左側の二つはベンチの利用状況を撮影した動画から 切り出したフレーム画像を示す.対して右側の図は、縦軸はモ デルにより算出したスコア,横軸は左図の動画のフレーム数を 示している. 右側各図は動画内で「ベンチ利用行動」が確認さ れると高いスコアを示す.

図より、「ベンチ利用行動」が確認できる場合にはスコアが反 応している(図1上側)一方で、「ベンチが利用行動」が確認で きない状況下ではほとんどスコアが反応していない(図1下側). 以上の結果から「ベンチ利用行動」を定量的に評価できることを 確認した.

連絡先:岡野将大,八千代エンジニヤリング株式会社,東京都 台東区浅草橋 5-20-8 CS タワー 技術創発研究所, TEL:03-5822-6028, Mail:ms-okano@yachiyo-eng.co.jp



図 1 ベンチ利用状況とスコア (左上:ベンチ利用時フレーム,右 上:ベンチ利用動画のスコア,左下:ベンチ未利用時フレーム,右 下:ベンチ未利用動画のスコア)

Frame

#### 2.4 精度向上に向けた検討

#### (1) 活性化関数

精度向上へ向けた検討として,活性化関数を GELU に変更し,実験を行った.

図 2 に示す通り、GELUは ReLUよりも入力値 0 付近で滑ら かな形状をとり、0 付近の数値を学習できると考えた.また、 [Hendrycks 2020]は GELUのノイズに対するロバスト性が確認し ている.現実のデータを扱う場合は、特にノイズを考慮して進め るため、ノイズに強いと思われる GELUを使用することで精度が 向上するのではないかと考えた.

### (2) 損失関数

精度向上へ向けた検討として,損失関数をHuber loss に変更 した. Huber loss の特徴として,損失が大きい場合には MAE に 似た機能を持ち,損失が小さい場合には MSE と同様の機能と なる(図 2). この特徴から,外れ値に寛容でありながら MAE の 欠点を克服ができ精度向上に貢献すると考えた.



図2 (左)活性化関数の比較 (右)損失関数の比較

# 3. 結果と考察

## 3.1 精度比較

図 3 の左に活性化関数の精度比較を示す. ここでは AUC (Area Under the Curve)で評価を行った. 活性化関数を GELU に変えたことにより, ROC の AUC が 0.825 から 0.878 に精度が 向上した. これは, 今回使用したデータにノイズが含まれており, 活性化関数を GELU に変えたことによりノイズに惑わされること なく学習できたと推察した.

図4の右に損失関数の精度比較を示す. MAEから Huber に 変えたことにより, ROC の AUC が 0.825 から 0.860 に精度が向 上した. これは, Huber に変えたことにより外れ値などの誤差を 学習したこと、また 0 付近で微分を行えることが影響していると 推察した.



## 4. おわりに

#### 4.1 本研究の成果

深層学習により「滞在しやすさ」の指標の一つである「ベンチ 利用行動」を定量的に評価する技術を開発することが出来た.ま た、本研究では行動認識を AUC で 0.878と良好な精度で実現 した.

以上から、本手法は実用に充分な精度を有し、作業効率化 にも貢献できる可能性を示した.

## 4.2 今後の課題

今後の課題を次に整理する.まず,本研究で対象としたフィ ールドは1地点のみであり,他のフィールドで本手法を適用した 場合に同様の精度が得られるか,頑健性を確認する必要がある. また,現状のモデルでは,「特定の行動が認識される/認識さ れない」の2クラスで分類を行っている.実用にあたっては複数 の行動を一度に検知できることが望ましく,多クラス分類を行う ためにはモデルの改良が必要である.

#### 謝辞

本論文を執筆するにあたり,株式会社 高速道路総合技術研 究所 緑化技術センターより実験動画をご提供いただきました. 厚く御礼を申し上げ,感謝の意を表します.

#### 参考文献

- [国土交通省 2019] 国土交通省:「都市の多様性とイノベーションの創出に関する懇談会」中間とりまとめ報告書, 2019.
- [Ryoo 2020] Michael S. Ryoo, AJ Piergiovanni, Mingxing Tan, Anelia Angelova: AssembleNet: Searching for Multi-Stream Neural Connectivity in Video Architectures, arXiv:1905.13209v4, 2020.
- [杉村 2020] 杉村由花,内田大輔,鈴木源太,遠藤利生:映像 から人の様々な行動を認識する「行動分析技術 Actlyzer」, 第 34回全国大会, 2020.
- [高森 2021] 高森真紀子,大久保順一,藤井純一郎:都市空間での人流解析における深層学習の応用,AI・データサイエンス論文集,2021.
- [Hendrycks 2020] Dan Hendrycks, Kevin Gimpel: GAUSSIAN ERROR LINEAR UNITS (GELUS), arXiv:1606.08415v4, 2020.
- [Tian 2021] Yu Tian, Guansong Pang, Yuanhong Chen, Rajvinder Singh, Johan W. Verjans, Gustavo Carneiro: Weakly-supervised Video Anomaly Detection with Robust Temporal Feature Magnitude Learning, ICCV, 2021.