

(2) 動画像を用いた深層学習による 歩行者交通量調査に関する基礎的研究

安藤 祐輝¹・今井 龍一²・姜 文淵³・山本 雄平⁴
・神谷 大介⁵・中原 匡哉⁶

¹学生会員 法政大学大学院 デザイン工学研究科 都市環境デザイン工学専攻
(〒162-0843 東京都新宿区市谷田町 2-33)

E-mail: yuki.ando.5w@stu.hosei.ac.jp

²正会員 法政大学教授 デザイン工学部 都市環境デザイン工学科
(〒162-0843 東京都新宿区市谷田町 2-33)

E-mail: ryuichi.imai.73@hosei.ac.jp

³正会員 関西大学特別任命准教授 先端科学技術推進機構 (〒564-8680 大阪府吹田市山手町 3-3-35)

E-mail: b.kyo@kansai-u.ac.jp

⁴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 3-3-35)

E-mail: y_yamamoto@kansai-u.ac.jp

⁵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 番地)

E-mail: d-kamiya@tec.u-ryukyu.ac.jp

⁶正会員 大阪電気通信大学講師 総合情報学部 (〒575-0063 大阪府四条畷市清瀧 1130-70)

E-mail: nakahara@oecu.jp

現在、我が国の歩行者交通量調査の多くは、現地調査や動画像の目視判断で歩行者を計数している。そのため、ヒューマンエラーによる誤計数や屋外の長時間計測（労働）や熱中症の危険性などの課題が山積している。こうした背景の下、国土交通省では情報通信技術を活用した新道路交通調査体系の検討を進められている。しかし、不特定多数の人が往来する環境では、オクルージョンによる歩行者計数精度の低下の課題が顕在化しており、解決策の確立には至っていない。そこで、本研究では、オクルージョンに対応可能な深層学習による人物認識の技術を検証し、課題点を明らかにすることで、実用可能な歩行者交通量の調査手法を導出した。その結果、人物領域の大半が隠れている場面でも考案手法を適用することで高精度に歩行者を計数できることが示された。

Key Words: person recognition, measurement for people flow, image processing, deep learning

1. はじめに

我が国の歩行者交通量調査の多くは、人海戦術による現地調査および動画像の目視判断で歩行者を計数している。そのため、調査費の制約や調査員の労力を考慮し、調査日、時間帯および調査地点数を限定して実施されるのが一般的となっている。さらに、ヒューマンエラーによる誤計数や屋外での長時間計測による熱中症や日射病の危険性もある。

こうした背景の下、国土交通省では、情報通信技術を活用した新道路交通調査体系の検討¹⁾が進められている。広域の交通流動の計測には、携帯電話基地局の運用デー

タ、GPS データや Wi-Fi パケットセンサデータなどが活用されている。一方、局所的な範囲の交通流動の計測には、レーザカウンタやカメラ動画像が活用されている。

ここで、歩行者交通量調査とは、任意断面を通行した歩行者を計数することを指す。昨今、カメラから取得した動画像に人物の識別処理などを適用し、歩行者数を計測する手法が研究されている。対馬ら²⁾は、鉄道駅構内に設置された防犯カメラを用いて、OpenPose による断面交通量の自動計測の適用可能性を示している。また、不特定多数の人が往来する環境では、オクルージョンによる人物認識精度の低下が顕著である結果をまとめている。

本研究の目的は、オクルージョンに対応可能な深層学

習による人物認識の技術を検証して課題を分析し、実用可能な歩行者交通量の調査手法の導出とした。第2章では、既存の人物領域検出技術を調査・選定し、歩行者交通量調査への適用可能性を検証する。第3章では、人物認識精度を向上させるべく頭部に着目した学習モデルを考案し、第4章にて、考案手法の有用性を検証する。第5章では、本研究を総括し、今後の展望を述べる。

2. 既存の人物領域認識技術の適用可能性の検証

(1) 物体認識技術の選定

本研究では、歩行者交通量調査に適した人物領域認識技術を確認するため、既存の物体認識技術を調査した。今井ら³⁾の研究では、市販のビデオカメラの動画像を用いた車両領域の認識可否を検証し、Object Detection や Instance Segmentation などの物体検出技術に一定の精度があることを明らかにしている。加えて、歩行者交通量調査は、一般的に12時間実施されることから、動画像のデータ量が膨大になると考えられる。現行の調査と同等の調査期間を要件にするならば、高速に処理可能な人物認識技術が適用対象になる。このため、本研究では比較的高速処理ができる Object Detection³⁾に着目して調査を進めた。Alexeyら⁴⁾は、検出精度が同等のSSD (Single Shot Detector) と YOLOv3 (You Only Look Once) とを比較し、YOLOv3の方が約3倍高速であることを実証している。以上より、本研究では、人物領域認識技術として YOLO を選定した。

次に、公開されている YOLOv3 と YOLOv4 の学習済みモデル⁵⁾を5分間の歩行者交通の動画像を用いて比較検証した。結果として、人物領域認識の処理速度は、YOLOv3 と YOLOv4 とともに入力した動画像の2倍程度の約11分であった。また、図-1のような認識精度の違いがあり、YOLOv4の方が比較的高精度に人物領域を認識できることが明らかとなった。

(2) 動画像の撮影条件の設定

本節では、YOLOv4の学習済み人物領域認識モデルを用いて、撮影条件ごとに検出精度を確認し、最適な条件を導き出す。まず、本技術に適用する歩行者交通量調査の動画像を撮影(図-2参照)し、オクルージョン発生時の要因となる撮影角度に着目して人物領域認識を検証した。撮影角度は15°・30°・45°・60°とした。撮影カメラはJVC社のGZ-RY980-Aを用いて4K解像度で撮影した。

人物領域の認識結果の一例を図-3に示す。撮影角度15°は、広範囲を撮影できるが、オクルージョンの発生頻度が他の撮影角度と比べて高いことがわかった。一方、撮影角度60°は、人物領域の半分以上が隠れるオクル

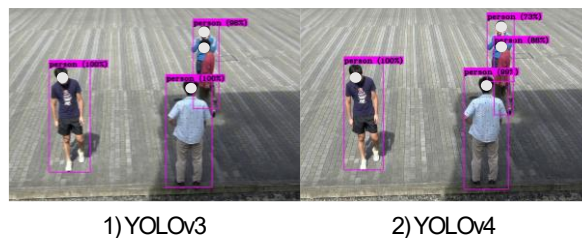


図-1 オクルージョン発生時の人物領域の認識結果



図-2 撮影地点の様子

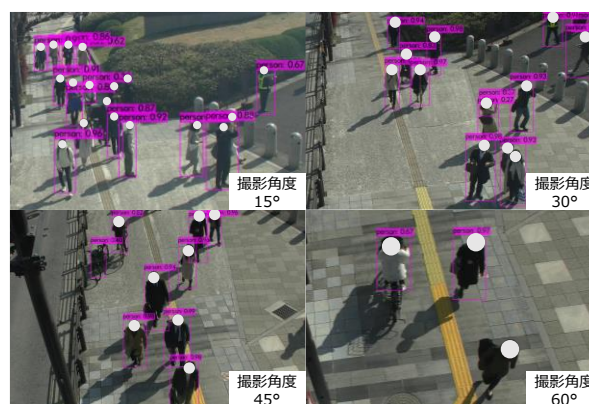


図-3 各撮影角度の人物領域の認識結果の一例

ージョンの発生は少ないが、頭部が大きく映ってしまって人物を認識できない場面が散見された。また、撮影角度30°および45°は、撮影角度15°と比較するとオクルージョンの発生頻度が低く、比較的安定して人物領域が認識できる。以上より、本研究で考案する歩行者交通量の調査手法への適用条件として撮影角度は30°および45°程度が適しているといえる。

(3) 既存の人物領域認識技術の課題

前節では、歩行者交通量調査を想定した環境下で既存の人物領域認識技術の適用可能性を検証した。そして、撮影角度によって人物の見方が大きく変化し、認識結果の大きな要因になることを示した。また、歩行者の交通量が多く密度が高い場面では、歩行者同士のオクルージョンが発生し、人物領域の半分以上が隠れてしまう場面が散見された。一方、オクルージョンが発生している場面を分析すると、頭部が隠れる場面は少ない傾向にあることが確認できた。そこで、本研究では、より高精度に人物領域認識を実装するため、既存の人物領域全体を認識する手法の代替となる頭部のみを認識する手法を考案することとした。

3. 頭部に着目した歩行者認識手法の考案

本研究では、歩行者の交通量が多くてもオクルージョンの影響を受けにくい手法を考案した。具体的には、第2章の検証により得られたオクルージョンの影響を受けにくい頭部に着目した新たな頭部認識モデルを考案した。頭部認識モデルの作成手順を図-4に示す。まず、図-2のような撮影環境下にカメラを設置し動画を撮影する。撮影カメラはJVC社のGZ-RY980-A、Sony社のFDR-AX40およびApple社のiPhone11 Proのカメラを用いて4K解像度で撮影した。次に、得られた動画をフレームごとに切り出す。そして、切り出し画像群から頭部が映っている画像を選定し、図-5に示すように、頭部領域を手作業でラベリングする。さらに、ラベリングして取得した頭部の座標値および頭部を含む画像を学習データとする。最後に学習データをYOLOに学習させて頭部認識モデルを構築する。

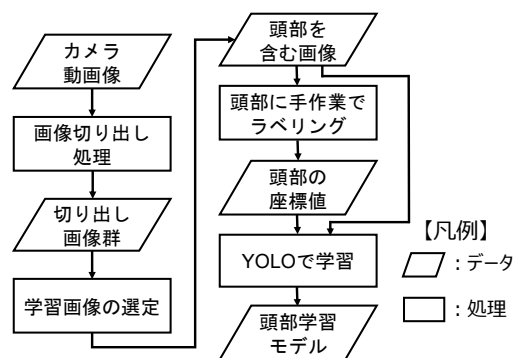


図-4 頭部認識モデルの作成手順



図-5 頭部ラベリングの一例

4. 考案手法の有用性の検証

(1) 検証の概要

本研究では、考案した頭部認識モデルの有用性を検証する。既存の人物領域認識モデルでは頭部が映らない場面でも認識可能である。しかし、頭部認識モデルでは、頭部が映っていることが条件となる。そのため、歩行者として計数する対象を、画面内に全身が完全に映っている歩行者かつ歩道内に位置している人物とした。検証には、図-3に示す撮影角度が異なる4種類の動画画像を使用する。撮影角度ごとに、30fpsで約4分間（計7,200枚）を撮影しており、その中の100フレーム分を検証の対象とした。対象画像は2種類用意した。1つ目は、各動画画像内でオクルージョンが発生している場面のみ100フレームを抽出した。なお、撮影角度60°ではオクルージョンの発生個所が少なかったため検証の対象外とする。2つ目は、恣意的な要素を取り除くために、等間隔となる72フレームごとに抽出し対象画像を選定した。

(2) 検証に用いる頭部学習データの構築方法

一般的に、深層学習は下記の点に留意する必要がある。

- 学習データのバリエーション

バリエーションが多くなると、髪型の違いおよび頭部の大きさや見え方の違いによる精度低下を防止できる。

- 学習データの枚数

学習データの数が多くなると、よりロバスト性のある学習モデルの構築が可能である。

- 学習回数

高精度な頭部認識モデルを生成するには、学習時の損

表-1 頭部学習データの構築

撮影地点数	画像枚数	解像度	学習回数	損失係数
9地点	994枚	4K解像度	50,000回	0.1前後

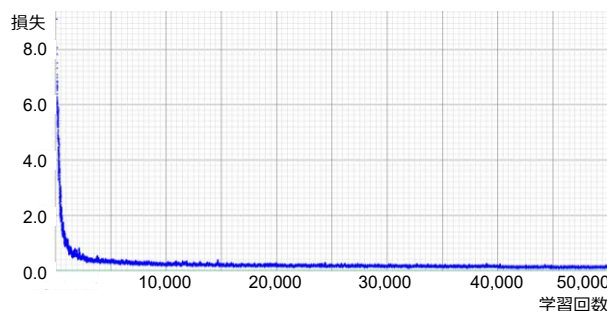


図-6 学習回数に対する損失の推移

失（図-6の縦軸）を収束させる必要がある。そのため、学習回数は損失を見ながら決定する必要がある。

本研究は、上記の3点を考慮し、表-1に示す学習データの構築方法を使用する。また、バリエーションを確保するため、9地点の映像を利用する。学習回数は、過学習を考慮して損失が0.1前後となった際の学習モデルを使用した。

(3) 検証方法

本検証では、画像内に映っている歩行者の認識精度を適合率、再現率およびF値で確認する。適合率は、各モデルが認識した人数のうち、正しく人物領域（頭部）を計数できていた割合を示した値である。再現率は、各モデルが認識した人数のうち実際に人物領域（頭部）であった人数と、正解人数の割合を示した数字であり、取り

こぼしがどの程度あるかを確認できる。F 値は、適合率と再現率の調和平均の算出結果となっており、1.000 に近いほど高精度な検出結果であるといえる。

(4) 検証結果及び考察

オクルージョンの発生個所を抽出した場合の精度検証結果を表-2、動画像を等間隔に抽出した場合の精度検証結果を表-3に示す。表-2より、人物領域にオクルージョンが発生している場面において、頭部認識モデルの方が適合率、再現率およびF値の値が高い結果となった。また、図-7に示すような、オクルージョンに対するロバスト性を確認できた。さらに、表-2および表-3の各モデルの適合率を比較すると、頭部認識モデルの方が高い傾向にあるといえる。人物領域認識モデルでは、街灯、車両および標識などの物体を人物領域と誤認識する例が見られたが、頭部認識モデルでは、頭部以外を頭部と誤認識する例が少なかったため適合率の値が高い傾向となった。このことから、考案手法が有用であることがわかった。

一方、表-3の結果より、動画像全体を検証対象にすると、既存の人物領域認識モデルの方がF値の値が高い傾向にあることが示された。これは、帽子を被っている人や背景が頭部と似た色の場合に、頭部認識モデルの認識精度が低下するためである。

以上より、オクルージョンの生じた画像に、本研究で提案する頭部認識モデルを適用すると、既存の人物領域認識モデルよりも高精度に歩行者計数が可能であることが示された。一般的に歩行者交通量調査は、通勤時間帯の駅構内など歩行者同士のオクルージョンが常時発生している環境下で実施されることから、本手法は有用となる可能性がある。さらに、頭部認識モデルに含まれる頭部のバリエーションは既存の人物領域認識モデルよりも少なくモデルとしての成熟度が低いので、学習データの拡張により認識精度の向上が見込まれる。

5. おわりに

本研究では、歩行者交通量調査の自動化を目指し、既存の物体認識手法を用いた人物領域認識技術の適用可能性を検証した。そして、より高精度な歩行者認識の実現を目指すべく、オクルージョンの影響を受けにくい頭部に着目したモデルを構築し、有用性を検証した。その結果、オクルージョンの生じた画像に対してはF値0.950以上の高精度な歩行者計数が可能であることを示した。

今後は、頭部画像のバリエーションを増やし、頭部認識モデルの精度を向上させる。加えて、歩行者交通量調査を想定したオクルージョンが多い場面で頭部認識モデルの有用性を検証する。

表-2 オクルージョン発生個所における精度検証結果

撮影角度	正解人数(人)	認識手法	モデル認識計数(人)	正しい計数(人)	適合率	再現率	F値
15°	531	人物	651	526	0.808	0.991	0.890
		頭部	548	531	0.969	1.000	0.984
30°	646	人物	654	602	0.920	0.932	0.926
		頭部	595	595	1.000	0.921	0.959
45°	606	人物	599	585	0.977	0.965	0.971
		頭部	602	602	1.000	0.993	0.997

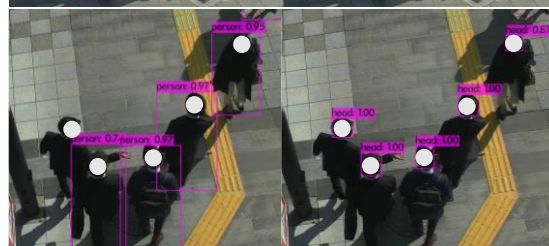
表-3 等間隔に抽出した画像における精度検証結果

撮影角度	正解人数(人)	認識手法	モデル認識計数(人)	正しい計数(人)	適合率	再現率	F値
15°	539	人物	555	510	0.919	0.946	0.932
		頭部	419	415	0.990	0.770	0.866
30°	853	人物	890	851	0.956	0.998	0.976
		頭部	767	754	0.983	0.884	0.931
45°	251	人物	266	247	0.929	0.984	0.956
		頭部	249	240	0.964	0.956	0.960
60°	71	人物	71	69	0.972	0.972	0.972
		頭部	89	69	0.775	0.972	0.863

事例
①



事例
②



1) 人物領域認識モデル 2) 頭部認識モデル

図-7 オクルージョン発生個所での実行結果の一例

謝辞：本研究を遂行するにあたり、関西大学総合情報学部の中田研究室の皆様には分析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る。

参考文献

- 1) 国土交通省：ICT を活用した新道路交通調査体系検討会、<<https://www.mlit.go.jp/road/ir/ir-council/ict/index.html>>、(入手 2021.6.7)。
- 2) 対馬銀河、石突光隆、柴田宗典：鉄道駅構内における断面交通量の自動計測に関する基礎検討、土木計画学研究発表会論文集、土木学会、Vol.62, No.17-01, 2020。
- 3) 今井龍一、神谷大介、山本雄平、田中成典、中原匡哉、中畑光貴：交通量調査の適用に向けた深層学習による移動体認識技術の調査研究、土木情報学シンポジウム講演集、土木学会、Vol.44, pp.173-176, 2019。
- 4) Alexey, B., Chien-Yao, W. and Hong-Yuan, L.: YOLOv4: Optimal Speed and Accuracy of Object Detection, Cornell University, 2020。
- 5) GitHub : AlexeyAB/darknet, <<https://github.com/AlexeyAB/darknet>>、(入手 2021.6.7)。