

人物追跡手法とマッチングアルゴリズムを用いた路線バスにおける乗客 OD 推定

西浦 翼

北海道大学大学院情報科学院

tsubasanishiura@ist.hokudai.ac.jp

横山 想一郎 山下 倫央 川村 秀憲

北海道大学大学院情報科学研究所

佐藤 好美 長谷川 怜 平澤 幸

株式会社シーズラボ

要旨

日本国内では深刻な少子化による人口減少からコンパクトシティ化が進んでおり、特にバス路線に関して公共交通ネットワークを再編成する必要がある。交通計画を策定するためには住民の移動を集計した OD (Origin-Destination) データが必要である。OD データの収集はその需要と裏腹に、高いコストなどから収集できていない現状にある。そこで本研究では路線バスを対象に、カメラで撮影した動画データからバス乗客 OD を推定するシステムを開発する。様々な要素技術を検討し、物体検出、物体追跡、メトリックラーニング、割当アルゴリズムを組み合わせたバス乗客 OD 推定システムを提案する。データセットとして、実際に運行している路線バスにカメラを取り付けてバスに乗降する人物の動画を撮影し、人物矩形のアノテーションを行った。モデルの選定やモデルの学習を行って推定精度を検証し、精度低下の原因を分析するとともに、精度向上の施策を行った。本研究の結果は、コストや労力をかけずに OD データを収集する手段として貢献することが期待される。

1. はじめに

日本は少子化による人口減少という深刻な問題に直面しており、コンパクトシティ開発に向けた議論が進んでいる。その結果、特にバス路線に関して、公共交通ネットワークを再編成する必要性が高まっている。都市中心部では、主要な幹線道路を走るバス路線が多く重複し、競合している。これを解決するために、幹線と支線を区

別し、各々の役割を明確化することが求められている。また、路線の長さを短くすることで定時性を確保することも重要である。新しいバス交通計画を策定するためには、バス利用者の移動を集計したバス乗客 OD データが必要である。

バス乗客 OD データには、「バス停別乗降者数」と「乗車区間別乗降者数」という 2 つの種類がある。バス停別乗降者数を分析すれば、バス停や時間帯ごとの利用者の多寡を把握でき、バス停周辺の路線の追加・削除や、その時間帯の増便・減便が可能になる。また、乗降区間別乗降者数から人流を分析できれば、利用区間ごとの運賃の見直しなどにも活用できる [1]。

実際に、交通計画策定に向けたバス OD データ収集の需要が高まっており、平成 28 年度に国土交通政策研究所が自治体に対して行ったアンケートで、バス乗客 OD データの必要性が裏付けられている [2]。民間バス事業者の持つデータについて、地方公共団体が必要だと考えるものの回答 684 件のうち、56.4 % にあたる 386 件が「各バス停間の OD」であった。しかし、回答 581 件のうちわずか 11.2 % にあたる 65 件しか、「対象路線の各バス停間の OD」を民間バス事業者から収集できていなかった。このようにバス乗客 OD データは、バス事業者が必要だと考えているにもかかわらず、収集できていない状況にある。

現在の OD データ収集手法としては、パーソントリップ調査などのアンケート調査や調査員を派遣した人手による調査、IC カードの使用履歴データなどがある。アンケート調査や調査員による手法は、費用や工数の面で非常にコストが高い。また、IC カードデータは、IC カードを発行している事業者が管理しており、外部から利用

するのにハードルが高いという問題がある。

このような背景に基づいて、本研究では、バス車内にカメラを設置し、乗降する人物を動画で撮影することで、深層学習を用いた人物追跡技術を適用し、OD データを推定する乗客 OD データ推定システムを開発した。このシステムを利用することで、バス乗客 OD データの収集が効率的に行われることが期待される。これにより、バス路線の最適化や運賃制度の見直しなど、さまざまな交通計画の策定が可能になると考えられる。

本稿の構成を以下に示す。2 章では、バス乗客 OD 推定に必要な技術に関する研究について概説する。3 章では、提案システムの概要を説明する。4 章では、本研究で利用する乗降客の動画データセットを説明する。5 章では、その有効性を確認するための実験について説明し、実験結果を考察する。最後に 6 章で本論文をまとめる。

2. 関連研究

本章では、バス乗客 OD 推定に用いる 4 つの技術、物体検出、物体追跡、メトリックラーニング、割当アルゴリズムに関して概要を説明する。

2.1. 物体検出

Computer Vision のタスクの 1 つに物体検出 (Object Detection) がある。これは入力となる画像の中から定められた物体の位置を推定するタスク (Localization) とクラスを推定するタスク (Classification) が複合したタスクである。特に、人物クラスを検出することを人物検出 (Human Detection) と呼ぶ。

有名な手法として、2005 年に提案された HOG (Histogram of Oriented Gradients) 特徴 [3] がある。HOG 特徴は計算コストが低く、実際の問題に対して非常に効果的である。しかし、その数年後、深層学習を用いた手法が登場した。2012 年の ILSVRC で AlexNet[4] が優れた結果を残して以降、ResNet[5] をはじめとする畳み込みニューラルネットワーク (Convolutional Neural Network; CNN) が画像認識に広く用いられるようになった。近年では、高い精度を維持しながら効率的な計算を実現する EfficientNet[6] が開発されている。

物体検出に CNN アルゴリズムを応用した先駆けとなったモデルが、2014 年に提案された R-CNN (Region-based CNN) [7] である。CNN は計算量が大きく推論速

度が遅いという課題があったが、Selective Search と呼ばれる物体候補アルゴリズムを導入することで計算を高速化した。その後、Fast R-CNN[8]、Faster R-CNN[9]、SSD (Single Shot Detector) [10] などが提案されてきた。

特に注目されるのが、YOLO (You Only Look Once) [11] シリーズである。YOLO は画像全体をあらかじめグリッド分割し、各グリッドごとに物体の矩形と確信度を出力する方法を採用している。このため、推論速度が高いことが特徴である。YOLO シリーズは現在も様々な研究者や研究チームによって派生モデルが開発されている。

各手法の性能については、YOLO シリーズが高速な推論速度を誇るが、精度面では Faster R-CNN や SSD といった他の手法と比較して劣ることがある。一方で、Fast R-CNN や Faster R-CNN は精度が高いものの、推論速度が遅いことが課題である。

2.2. 物体追跡

物体追跡 (Object Tracking) は動画内に映っている物体や人物を追跡するタスクで、画像内から物体の位置とクラスを推定する物体検出 (Object Detection) と、複数の画像間で検出された同一の物体をマッチングさせる再同定 (Re-Identification) から構成される。物体追跡技術は、カルマンフィルター、パーティクルフィルター、光流法、Mean Shift などの従来手法から、近年の深層学習を用いた手法 (Deep SORT[12] や MDNet[13] など) まで、さまざまなアルゴリズムが提案されている。

一般的な Re-Identification と比べて、画像の時系列情報や画像のどこに矩形が位置していたのかという座標情報が使える点が異なる。Object Detection と Re-Identification のそれぞれに多くの課題が存在し、評価指標も物体追跡を行う目的に応じて変化する。物体追跡の性能評価には、MOT Challenge や VOT Challenge などのデータセットや、MOTA、MOTP、ID スイッチなどの評価指標が利用されている。

また、人物追跡では、追跡の対象となるデータによって難易度が大きく変わる。まず、人物追跡のタスクの難易度は追跡する対象の多寡によって異なる。リアルタイムで対象を追跡するオンラインのタスクでは、現在までのフレームしか使えない。それに対して、全区間のフレームが利用可能なオフラインのタスクでは、ある時点からの未来にあたるフレームも利用することができる。対象が別の物体と重なるオクルージョンが発生すると対

象に同一の ID を振り続けられないため、対策が必要となる。さらに、対象の細かい状態や動作も追跡することが求められるのであれば、対象を矩形で検出するのではなく、セグメンテーションすることも必要となる。

物体追跡技術は、自動運転、スポーツ分析、監視カメラなど、さまざまな分野で応用されている。本研究では、バス停別乗降者数の推定と乗降区間別乗降者数の推定に向けて、乗車口と降車口のそれぞれで同一人物群の生成を目的としている。また、リアルタイムにバス停別乗降人数や乗降区間別乗降者数が必要とされていないため、オフラインで使用でき、追跡対象が数人規模のモデルを用いる。このような状況に適した物体追跡アルゴリズムを選択・適用することが、本研究の重要なポイントとなる。

2.3. メトリックラーニング

乗車人物と降車人物の同一判定においては、日々同様の乗客が乗り降りするとは限らない。よって機械学習アルゴリズムによって人物を認識するためには訓練データには存在しない未知の乗客の画像を同一人物か否かを判定する必要がある。しかし CNN を使用したクラス分類は事前に学習したクラス以外の分類に対しての適用はできず、2つの入力データに対して類似度を計算する構造を持っていない。

このような未知のデータが類似しているか否かを学習する手法としてメトリックラーニングがある。メトリックラーニングで似た特徴を持つデータ間の類似度を近く、異なる特徴を持つデータ間の類似度を遠くなるよう学習が行われ、これまでの研究において車両の認証 [14] や顔の認証 [15] において優れた結果を残してきた。

メトリックラーニングは Person Re-Identification という人物を認証するタスクに対しても研究されており [16]、本研究では、その中で優れた結果を残している BoT [17] という手法を使用する。メトリックラーニングを用いた Person Re-Identification においては基準となるアンカーに対して異なるクラスであるネガティブなデータよりも同クラスであるポジティブなデータとの距離をより近くなるような学習を行う triplet loss を用いた手法 [18] があるが、BoT は triplet loss を用いたベースラインに、提案されている BNNeck に加え、学習率の Warmup, Random Erasing Augmentation [19], LabelSmoothing [20], そして Last Stride [21] の複数の学習トリックを加えること

で、精度の向上を図っている。学習トリックに含まれる Center Loss [22] と、triplet loss と人物 ID の分類損失である ID loss の 3 つを損失を組み合わせている。

2.4. 割当アルゴリズム

代表的な組み合わせ最適化問題の一つに割当問題がある。割当問題とは集合 A の要素を集合 B の要素のどれに割り当てるかを決定する問題である。割当問題はタスクによって、ボトルネック割当問題、一般化割当問題、2次割当問題など多くの種類に分類され、定式化の仕方や解法が異なる。

本研究は乗車人数と降車人数をシステムで推定するため、同一人物であってもオクルージョン等により途中で ID が切り替わる (ID スイッチ) ことで重複して数えることや、人物を追跡できず欠損するなど、乗車人数と降車人数が必ずしも一致するとは限らない。よって、M 人の乗車人物と N 人の降車人物の二部グラフと考えられる。M > N として、重複を許して M 人を N 人に割り当てて M ペアを作成する場合、ID スイッチが発生しても正しいペアを作成できる可能性があるが、特定の降車人物 1 人に多くの乗車人物が割り当てられてしまう可能性もある。重複を許さず N 人を M 人に割り当てて N ペアを作成する場合、乗降人物が欠損しても残りの人数分はペアを作成することができる。本研究では人数が少ない方のペアを作成することで、マッチング問題として定式化することができる。

重み付き二部グラフのマッチング問題の解法として、ハンガリアン法が存在する。計算量は $O(n^3)$ であるため、中規模の割当問題に対して効率的であるが、サイズが大きい問題には不適切である。本研究の対象はバス車内という限定的な空間であるため、十分に利用可能である。

3. バス乗客 OD 推定システム

本システムでは、バス停別乗降者数と乗車区間別乗降者数の推定を行う。システムの概要を図 1 示す。処理は大きく 2 段階に分かれる。1 段階目では乗降口における個人識別を行うことで各バス停で何人乗降したかを推定する。乗降口で撮影された動画を入力として人物検出と人物追跡を行うことで乗降する個人を軌跡として識別するとともに、バス停別乗降者数を出力する。2 段階目ではその乗車人物の軌跡と降車人物の軌跡の間で同一人物

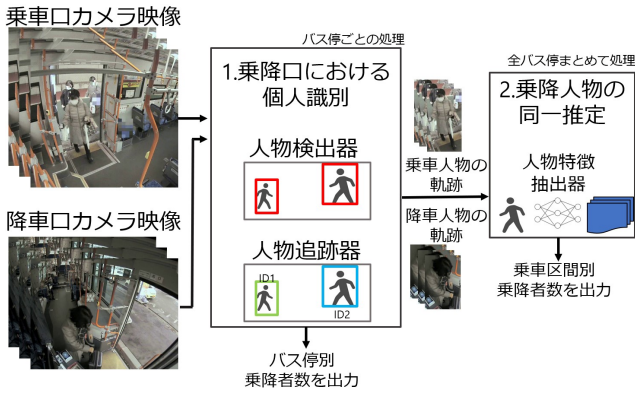


図 1. バス乗客 OD 推定システムの処理の流れ

の推定を行うことでどこで乗った人物がどこで降りたのかを推定する。メトリックラーニングを用いて各人物の軌跡を特徴空間に埋め込んで類似度を求め、割当アルゴリズムを用いて乗車・降車ペアの作成を行って乗車区間別乗降者数を出力する。これら 2 つの処理について以下で説明する。

3.1. 乗降口における個人識別

乗客が乗降する動画を入力としてバス停別乗降者数を出力することを目指す。候補となる要素技術には、顔認識、骨格推定、頭部検出、人物検出、人物追跡などが挙げられる。4 章でデータセットの詳細を説明するが、個人情報保護の観点から動画内の乗客の顔には全てモザイク処理が施されているため、顔認識により個人を識別することはできない。また、バス車内、特に乗降口付近は人物同士のオクルージョンが多く、カメラで体全体をとらえることができないため、骨格推定による個人識別も難しい。車内全体をカメラで撮影できれば頭部検出による車内の乗客数の変化から乗降者数を推定することも可能であるが、バス車内はカメラの設置位置に制限があり全体が撮影できなかつたり手すりとのオクルージョンがあったりと、こちらも現実的でない。そこで本研究では人物検出と人物追跡を用いて乗降口における個人識別を行う。

乗降口における個人識別のアルゴリズムを図 2 に示す。入力となる動画に対してまず人物検出器を適用することで各フレームに写る乗客を矩形として推定する。次に推定された矩形情報を入力として、人物追跡器を適用することでフレームをまたいだ同一人物の矩形に対して同一の ID を付与することで人物の軌跡を出力する。最後に

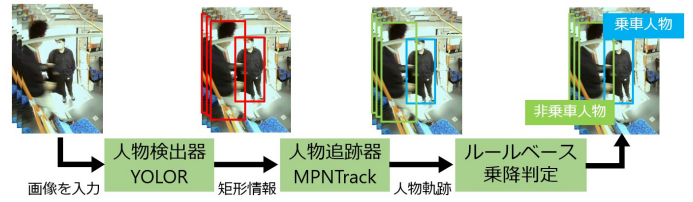


図 2. 乗降口における個人識別のアルゴリズム

出力された人物の軌跡が、乗車(降車)してきた人物なのか元々バス車内にいた人物なのかをルールベースで判定する。

人物検出器は YOLOR[23] と YOLOv8[24] の精度を比較する。バス停別乗降者数の推定はオフラインで行われることを想定しているため、既存研究におけるオフラインの人物追跡器として MPNTrack[25] を用いる。MPNTrack は追跡途中で対象人物の軌跡が入れ替わってしまう ID スイッチの少ないモデルで、2020 年に行われたマルチオブジェクトトラッキングのベンチマークとなっている MOTChallenge¹ の MOT20 において最も ID スイッチの発生回数が少なかったモデルである。現在の研究において、用いられているモデルとしては Siamese CNN[26] をベースとしたモデルが挙げられるが、MPNTrack はマルチオブジェクトトラッキングの定式化に Message Passing Network [27] を応用したモデルである。

乗車口、降車口の画像例を図 3 に示す。乗降の判定に関しては、それぞれ以下の判定方法を適用する。乗車口の画像では、ドアが開く前から映っている人物は既にバスに乗っている乗客である。また、降車しようとするバス後方にいる乗客はバス後方(画像の左)から現れて、バス前方(画像の右)に移動する。そのため、乗車の判定に関しては、ある乗客の軌跡が、i) ドアが開いた時間以降に現れている、ii) 最初のフレームの矩形の中心座標が乗車口の範囲に入っている、という 2 つの条件を満たしている場合に乗車と判定する。乗降口の範囲は図 4 のように画像の赤い矩形部分とする。

降車の判定に関しては、ある乗客の軌跡が、i) ドアが閉まる時間以前に軌跡が消失している、ii) 最後のフレームの矩形の中心座標が降車口の範囲に入っている、という 2 つの条件を満たしている場合に降車と判定する。

また、乗降者数推定の精度を下げる要因の 1 つとして

¹<https://motchallenge.net/>



図 3. 乗車口と降車口の画像例

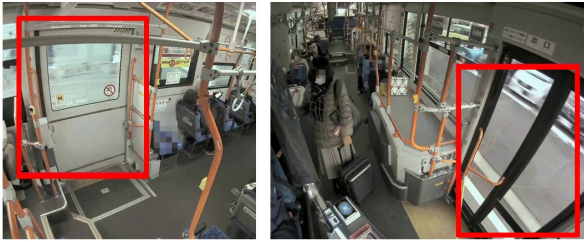


図 4. 乗降判定の範囲

ID スイッチがある。これを抑制するために人物検出によって出力される矩形に対していくつかの条件を設け、人物追跡のノイズとなる矩形を取り除く。

3.2. 乗降人物の同一推定

乗車区間別乗降者数の推定は、乗降口における個人識別によって作成された乗車人物の軌跡群と降車人物の軌跡群で同一人物のペアを推定することで各人物の乗車地点と降車地点を特定することで行う。

乗降人物の同一推定のアルゴリズムを図5に示す。乗車口における個人識別によって推定される乗車人数と降車人数には差異があるため、推定した乗車人物を M 人、推定した降車人物を N 人とする。ある乗車人物 A の軌跡が i フレーム、ある降車人物 B の軌跡が j フレームあるとき、 A と B に対してメトリックラーニングモデル BoT を適用することで $i \times j$ の特徴量距離行列を出力することができる。各軌跡は乗客の移動時間や人物検出・人物追跡の精度によって撮影されているフレーム数が数枚から数十枚と幅広いこと、撮影された人物の角度や明るさによって同一人物間でも特徴量距離に大きな差があることから、 $i \times j$ の特徴量距離行列から上位5つを抽出して平均したものを乗車人物 A と降車人物 B の特徴量距離とする。

この処理を乗車人物 M 人と降車人物 N 人に対して総当たりで実行して、 $M \times N$ の特徴量距離行列を求める。

最後に図6に示すように割当アルゴリズムを適用して乗車人物と降車人物のペアを推定する。乗車人物と降車人物のペアを出力する問題は二部グラフのマッチング問題と捉えられるため、乗車人物の軌跡と降車人物の軌跡の特徴量距離を割り当てコストとみなし、ハンガリアン法を適用する。ハンガリアン法の適用において、軌跡間の特徴量距離をユークリッド距離行列の行ごとに順位で置き換えた後にハンガリアン法を適用する手法が考えられる。本稿では、こうした手法をランキング法と呼び乗車区間別乗降者数の推定精度を比較する。

ここで、乗車したバス停より前に降車した人物とのペアが作成される可能性があり、実際にはあり得ないペアが出力されるという問題が発生しないように制約として、ハンガリアン法を適用する前に距離行列に対して簡単なバス停ルールを適用する。バス停ルールでは、乗車人物の軌跡群と降車人物の軌跡群のバス停の関係性を考慮し、乗車したバス停が降車したバス停と同じか降車したバス停より後の場合、該当する距離行列の要素を無限大に置換する。これにより組み合わせが最小となるように計算するハンガリアン法において実際に起こりえないペアの作成を防ぐ。出力されるペアを乗車・降車ペアとし、乗車・降車ペアの数は乗車人物の軌跡群と降車人物の軌跡群を比較したときに要素数が少ない方と等しくなる。

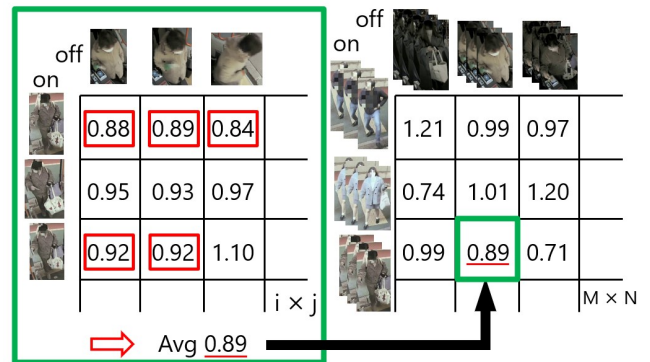


図 5. 特徴量距離行列の作成

4. データセット

本システムの検証のため、実際に札幌市内を走行する路線バスにカメラを取り付け、2021年12月5日と2021

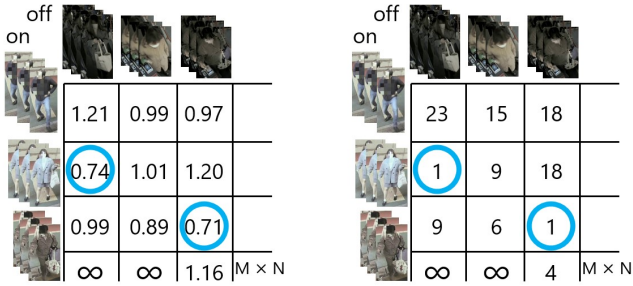


図 6. ハンガリアン法による割当 (左) とランキン
グ法による割当 (右)

年 12 月 6 日に乗客が乗降するデータを収集した。撮影したバスの路線は全部で 5 本あり、動画データをバス 1～5 とする。各データには乗車カメラの動画と降車カメラの動画が含まれ、停車したバス停数分の動画がある。データセットの詳細を表 1 に示す。バス 1 とバス 2 の動画は全て画像に切り出し、乗降人物に対して人物矩形のアノテーションを行った。バス 3～5 の動画は目視で乗降人数のみ計測した。

図 7 に示されるようにバス車内の乗車口・降車口にカメラを設置して、4fps で撮影した。乗客が簡単に触ることができない天井近くにカメラを設置した。撮影はドアが開く 10 秒前から撮影が開始され、ドアが閉まってから 10 秒後に撮影を終了している。個人情報の取り扱いの観点から、バスに動画を撮影していることを示す張り紙をし、動画に関しては撮影と同時に顔にモザイク処理を施してから保存している。

図 8, 9 に示される動画は実際に撮影した乗車口、降車口の映像である。解像度は乗車口で縦 882 ピクセル・横 886 ピクセル、降車口で縦 1,024 ピクセル・横 768 ピクセルとなっている。動画のフレーム数は 100～200 フレーム程度だが、最初のバス停の動画は発車時刻までドアを開けたままにして乗客が徐々に乗ってくるため長くなっている。

5. 実験と考察

5.1. 人物検出器の精度比較

人物検出器 YOLOv8 と YOLOR の精度比較実験を行う。どちらのモデルもオープンデータセット MSCOCO

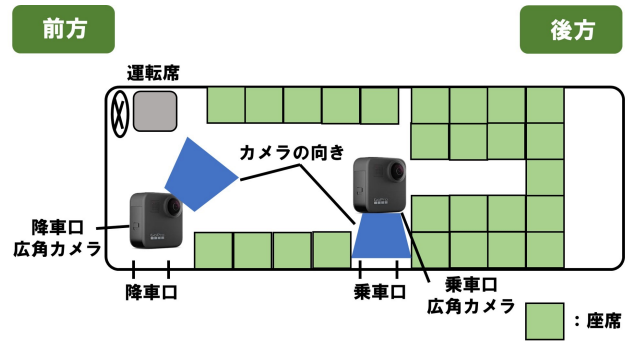


図 7. カメラの設置箇所

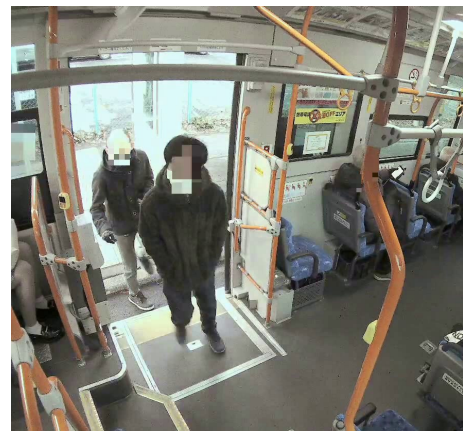


図 8. 乗車口カメラの画像

で学習させ、バス 1、バス 2 の動画に対して適用して人物検出の精度を比較する。評価指標は物体検出において最も一般的な指標である AP(Average Precision) を用いる。推論時の確信度の閾値を 0.2～0.6 の間で 0.1 ずつ変化させて AP を計測する。その結果を表 2 に示す。今回の実験では YOLOR を確信度 0.4 に設定した場合が AP 0.61 と最も高くなった。推定した矩形を確認すると、図 10 に示すように YOLOv8 はオクルージョンによって後ろに小さく写っている人物も検出できているのに対して、YOLOR はオクルージョンに弱い結果となっている。しかしアノテーションの段階で、体の一部が見切れている人物は軌跡を出力する際にノイズとなる恐れがあり、アノテーションを行わなかったため YOLOv8 の AP の方が低くなる結果となった。

表 1. データセットの詳細

	乗客の数	バス停の数	乗車動画の数	降車動画の数	アノテーション
バス 1	25	19	6	12	✓
バス 2	32	31	13	15	✓
バス 3	25	17	9	8	
バス 4	26	24	7	14	
バス 5	22	16	11	4	



図 9. 降車口カメラの画像

表 2. 人物検出器の平均適合率

Confidence	0.2	0.3	0.4	0.5	0.6
YOLOv8	0.60	0.60	0.52	0.58	0.55
YOLOR	0.48	0.55	0.61	0.58	0.54

5.2. バス停別乗降者数推定

乗降口における個人識別によるバス停別乗降者数の推定を行う。人物検出器については 5.1 の実験に基づいて YOLOR を確信度 0.4 で適用する。人物追跡器に関しては MPNTrack をバス 2 の動画で学習させて用いる。評価に関してはバス 1, バス 3, バス 4, バス 5 の動画を用いる。精度評価は以下の計算で求められる誤差率を用いる。誤差率は各バス停ごとに計算し、それを全バス停で平均した値を最終的な誤差率として計算する。

$$\text{誤差率} = \frac{|\text{実際の人数} - \text{推定人数}|}{\text{実際の人数}}$$

推定結果を表 3 に示す。乗車口では全体平均 13.3%, 降車口では全体平均 5.24% となった。乗車口では降りるために車内を移動する人物と乗車してくる人物との間にオクルージョンが発生し ID スイッチが起きやすくなっている。特に乗車人物が少ない時に ID スイッチが発生



図 10. YOLOv8(左) と YOLOR(右) の推定結果

して推定人数が増えると誤差率が極めて大きくなる傾向にある。

表 3. バス停別乗降者数の誤差率

	乗車口	降車口
バス 1	8.33%	8.33%
バス 3	12.5%	0.00%
バス 4	7.38%	10.0%
バス 5	25.0%	2.63%
平均	13.3%	5.24%

5.3. 乗降区間別乗降者数推定

乗降口における個人識別において推定された乗車人物の軌跡群と降車人物の軌跡群を用いて乗降人物の同一推定を行うことで、乗車区間別乗降者数を推定する。メトリックラーニングモデル BoT は DukeMTMC-ReID で学習済みのモデルを使用する。ハンガリアン法で乗車・降車ペアを作成する場合、人数が少ない方と同数のペアが出力される。ID スイッチ等で乗車人数も降車人数も実際の人数と異なっている場合、出力されるペア数が実際の人数と異なる可能性がある。しかし、5.2 の実験から乗車人物の誤差率も降車人数の誤差率も十分小さいと

判断した。マッチングアルゴリズムには通常のハンガリアン法と、人物特徴量距離の値を行ごとの順位で置き換えた後にハンガリアン法を適用するランキング法とを比較する。テストデータにはアノテーションが付与されたバス1の動画を用いる。精度評価には以下の式で表される再現率、適合率、F値を用いる。

$$\text{再現率} = \frac{\text{正解した乗車・降車ペア数}}{\text{実際の全乗車・降車ペア数}} \quad (1)$$

$$\text{適合率} = \frac{\text{正解した乗車・降車ペア数}}{\text{推定した全乗車・降車ペア数}} \quad (2)$$

$$\text{F 値} = \frac{2 * \text{再現率} * \text{適合率}}{\text{再現率} + \text{適合率}} \quad (3)$$

実験結果を表4に示す。ハンガリアン法では全乗降者25人中18人分の、ランキング法では14人分のODペア作成に成功した。推定した乗車人物が26人、降車人物が24人であったため、少ない方の24人分のODペアが出力され、適合率の分母にあたる推定した全乗車・降車ペア数は24となっている。今回の実験では通常のハンガリアン法がどの指標でも0.7を超える値を示すことができた。

また、精度を下げている原因として、乗車人物の候補がIDスイッチにより増加していること、降車人物の候補が人物追跡の段階で1人追跡できなかったことが第一に挙げられる。これは、バス停別乗降者数の推定精度を向上させる必要がある。次に、乗車人物と降車人物の映り方の差異が挙げられる。図11に示すように乗車口では自然光が多く入り比較的明るかったり白飛びした画像になったりする傾向にあるが、降車口ではバス中央の通路から料金箱付近では暗くなる傾向にある。こういった明るさの違いが乗車人物と降車人物との特徴量距離に影響を与えていることが考えられる。さらに図12に示すように、乗車口では乗車人物を正面で撮影した画像が多いのに対して、降車口では通路部分では正面、料金箱付近では頭上、ドア付近では後方から撮影され、体の角度も斜めになることがある。今回学習に用いたDukeMTMC-ReIDは人物を正面から撮影した画像が多いため、タスクに適したデータセットを作成・学習させる必要がある。

表4. 乗車区間別乗降者数の推定結果

	再現率	適合率	F 値
ハンガリアン法	0.72	0.75	0.74
ランキング法	0.56	0.58	0.57



図11. 乗車口(左)と降車口(右)での明るさの違い

6. おわりに

人物検出器 YOLOR, 人物追跡器 MPNTrack, メトリックラーニングモデル BoT, マッチングアルゴリズムハンガリアン法を組み合わせた路線バスにおける乗客OD推定手法を開発し、バス停別乗降者数と乗車区間別乗降者数の推定を行った。実データを用いた検証においては、バス停別乗降者数の推定では乗車口でも降車口でも誤差は1人ずつと十分な精度で推定できることを確認して、誤差の原因の分析とその対策を考察した。また、乗車区間別乗降者数の推定では再現率、適合率、F値が全て0.7を超えることを確認して、精度を低下させている要因が乗車口と降車口で明るさや撮影角度の違いであることを確認した。今後は更なる精度向上に向けてバス車内という環境でのデータセット作成及びモデルの学習を行っていく。

参考文献

- [1] 福本雅之：公共交通利用実態データの調べ方と活用方法 あなたのまちの公共交通、把握・分析、しっかりできていますか？～地域公共交通セミナー（秋期）



図 12. 乗車口 (左) と降車口 (右) での角度の違い

- ～, <https://wwwtb.mlit.go.jp/chubu/aichi/kikaku/seminafukumotosennsei.pdf> (2019).
- [2] 国土交通省国土交通政策研究所：地方公共交通事業における収益性と効率性に関する調査研究, <https://www.mlit.go.jp/pri/houkoku/gaiyou/kkk138.html> (2016).
- [3] Dalal, N. and Triggs, B.: Histograms of Oriented Gradients for Human Detection, *2005 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2005).
- [4] Krizhevsky, A., Sutskever, I. and Hinton, G. E.: ImageNet Classification with Deep Convolutional Neural Networks, *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, NIPS'12*, p. 1097–1105 (2012).
- [5] He, K., Zhang, X., Ren, S. and Sun, J.: Deep Residual Learning for Image Recognition, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778 (2016).
- [6] Tan, M. and Le, Q.: EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, *Proceedings of the 36th International Conference on Machine Learning*, Vol. 97, pp. 6105–6114 (2019).
- [7] Girshick, R., Donahue, J., Darrell, T. and Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation (2014).
- [8] Girshick, R.: Fast R-CNN, *CoRR*, Vol. abs/1504.08083 (2015).
- [9] Ren, S., He, K., Girshick, R. B. and Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks., *NIPS* (Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M. and Garnett, R., eds.), pp. 91–99 (2015).
- [10] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S. E., Fu, C.-Y. and Berg, A. C.: SSD: Single Shot MultiBox Detector., *ECCV (1)* (Leibe, B., Matas, J., Sebe, N. and Welling, M., eds.), Lecture Notes in Computer Science, Vol. 9905, Springer, pp. 21–37 (2016).
- [11] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A.: You Only Look Once: Unified, Real-Time Object Detection (2015). cite arxiv:1506.02640.
- [12] Wojke, N., Bewley, A. and Paulus, D.: Simple Online and Realtime Tracking with a Deep Association Metric (2017).
- [13] Santellani, E., Sormann, C., Rossi, M., Kuhn, A. and Fraundorfer, F.: MD-Net: Multi-Detector for Local Feature Extraction (2022).
- [14] He, S., Luo, H., Chen, W., Zhang, M., Zhang, Y., Wang, F., Li, H. and Jiang, W.: Multi-Domain Learning and Identity Mining for Vehicle Re-Identification, *Proc. CVPR Workshops* (2020).
- [15] Wang, H., Wang, Y., Zhou, Z., Ji, X., Gong, D., Zhou, J., Li, Z. and Liu, W.: CosFace: Large Margin Cosine Loss for Deep Face Recognition, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2018).
- [16] Wojke, N. and Bewley, A.: Deep Cosine Metric Learning for Person Re-identification, *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, pp. 748–756 (2018).

- [17] Luo, H., Gu, Y., Liao, X., Lai, S. and Jiang, W.: Bag of Tricks and a Strong Baseline for Deep Person Re-Identification, *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1487–1495 (2019).
- [18] Hermans, A., Beyer, L. and Leibe, B.: In Defense of the Triplet Loss for Person Re-Identification, *CoRR*, Vol. abs/1703.07737 (2017).
- [19] Zhong, Z., Zheng, L., Kang, G., Li, S. and Yang, Y.: Random Erasing Data Augmentation, *Proceedings of the AAAI Conference on Artificial Intelligence* (2020).
- [20] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z.: Rethinking the Inception Architecture for Computer Vision, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2818–2826 (2016).
- [21] Sun, Y., Zheng, L., Yang, Y., Tian, Q. and Wang, S.: Beyond Part Models: Person Retrieval with Refined Part Pooling (and A Strong Convolutional Baseline), *Proceedings of the European Conference on Computer Vision (ECCV)* (2018).
- [22] Wen, Y., Zhang, K., Li, Z. and Qiao, Y.: A Discriminative Feature Learning Approach for Deep Face Recognition, *Computer Vision – ECCV 2016* (Leibe, B., Matas, J., Sebe, N. and Welling, M., eds.), pp. 499–515 (2016).
- [23] Wang, C.-Y., Yeh, I.-H. and Liao, H.-Y. M.: You Only Learn One Representation: Unified Network for Multiple Tasks, *arXiv preprint arXiv:2105.04206* (2021).
- [24] Jocher, G., Chaurasia, A. and Qiu, J.: YOLO by Ultralytics (2023).
- [25] Brasó, G. and Leal-Taixé, L.: Learning a Neural Solver for Multiple Object Tracking, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6247–6257 (2020).
- [26] Koch, G., Zemel, R. and Salakhutdinov, R.: Siamese Neural Networks for One-shot Image Recognition, Vol. 37 (2015).
- [27] Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O. and Dahl, G. E.: Neural Message Passing for Quantum Chemistry, *Proceedings of the 34th International Conference on Machine Learning* (Precup, D. and Teh, Y. W., eds.), Proceedings of Machine Learning Research, Vol. 70, PMLR, pp. 1263–1272 (2017).