

インスタンスセグメンテーションを用いた トラックバース状況の分析

横山 皓大¹ 片山 晋¹ 浦野 健太¹ 米澤 拓郎¹ 河口 信夫^{1,2}

概要：インターネットサービスの普及により通販の需要は増加し、物流倉庫では業務量の増加や労働量不足が問題となり、作業効率化や作業支援が必要とされている。物流倉庫には、トラックを停車して荷積み、荷下ろしを行う場所であるトラックバースと呼ばれるスペースがある。トラックバースでは、空きバースが無いことによる荷待ち時間を緩和するための予約システムなどが提供されている。一方、荷物の積み下ろしにおいては、荷下ろしスペースも含めた管理は十分に行われているとは言えない。そこで、本研究では、バースの利用効率や作業効率向上を目的として、トラックバースにおける荷物の積み下ろし状況の分析を行う枠組みを提案する。具体的な手法としては、固定カメラを用い、トラック、荷下ろしスペース、作業員の状況を把握するために、インスタンスセグメンテーションによる物体検出を実施した。大量の実環境データを用いて学習データのアノテーションを行い、荷物の積み下ろし状況の把握が可能になった。

Analysis of truck berth status using instance segmentation

KODAI YOKOYAMA¹ SHIN KATAYAMA¹ KENTA URANO¹ TAKURO YONEZAWA¹
NOBUO KAWAGUCHI^{1,2}

1. はじめに

インターネットサービスの普及に伴い、通販の需要は増加している。物流倉庫では業務量の増加や少子高齢化も相まって、労働量不足が問題となっている。この問題を解決するために、倉庫内の作業効率化や作業支援が必要とされている。物流倉庫には、トラックを停めて荷積み、荷下ろしを行う場所であるトラックバースと呼ばれるスペースがある。この場所においても作業効率化、作業支援の取り組みがいくつか行われている。バースでは、多くの荷待ち時間が発生しており、国土交通省の行った調査 [1] によると、荷待ち発生件数の内、半数以上で 1 時間以上の荷待ち時間が発生しており、3 時間以上も約 1 割を占める。現在この荷待ち時間の緩和のため、荷待ち時間の記録やトラックバース事前予約システムによるバースの混雑の解消などのサービスが提供されている [2][3]。これらサービスにより、

トラックの入退管理に関しては効率化、作業支援が実現している。

一方、荷物の積み下ろしにおいては、荷下ろしスペースも含めた管理は十分に行われているとは言えない。そこで、本研究では、バースの利用効率や作業効率向上を目的として、トラックバースにおける荷物の積み下ろし状況の分析を行う枠組みを提案する。具体的な手法としては、固定カメラを用い、トラック、荷下ろしスペース、作業員の状況を把握するために、インスタンスセグメンテーションによる物体検出を実施した。大量の実環境データを用いて学習データのアノテーションを行い、荷物の積み下ろし状況の把握が可能になった。

2. 関連研究

画像内のオブジェクトを検出する手法として物体検出 [7]、セマンティックセグメンテーション [8]、インスタンスセグメンテーションなどが存在する。物体検出とは、画像の中から特定の物体の位置、種類、個数などの情報をバウンディングボックスで囲って認識する技術であるが、正確な形状までは把握できない。セマンティックセグメンテ

¹ 名古屋大学大学院 工学研究科
Graduate School of Engineering, Nagoya University

² 名古屋大学 未来社会創造機構
Institutes of Innovation for Future Society, Nagoya University

ションは、画像全体や画像の一部の検出ではなくピクセル1つ1つに対して、そのピクセルが示す意味をラベル付ける技術であるが、同クラス間で重なりがある場合、同クラスの領域として認識するため、物体ごとの認識・カウントができない。インスタンスセグメンテーションは物体検出の個体ごとの検出と、セマンティックセグメンテーションのピクセル単位での検出の両方の特徴を合わせ持つ技術である。本研究では、バース内の混在するオブジェクトを個別に検出し、その形状情報を利用するにあたって、インスタンスセグメンテーションが適していると考え、この技術を使用した。しかし、物体検出は、パターン認識とコンピュータビジョンの分野において古くから研究されてきた技術であり、その応用に関しても多岐に渡る。その一つとして、物体追跡に応用した手法がある。動画に写っている複数の物体を追跡する Multiple Object Tracking (MOT) において、Tracking by Detection[9] と呼ばれる手法が物体検出を応用した手法である。これは、物体検出モデルによって動画の各フレームから追跡対象の物体を検出した後、フレーム間ごとに同じ物体に対して同一の ID を与える。動画の各フレームで検出された同じ物体のバウンディングボックスに対して、ID を割り振ることにより、物体を追跡する。この MOT において、YOLO をサポートするモデルが多く存在する [10][11][12]。そこで本研究では MOT への応用や高速な処理が可能である YOLO を物体検出モデルとして使用した。

物流倉庫の物体検出やセグメンテーションを使用した作業効率化に関して、以下のような研究が行われている。商品のピッキング作業では、物体検出やセグメンテーションを使用することで、倉庫内の商品やパッケージを正確に識別し、ロボットアームや自動化システムを制御して適切にピッキングする技術が研究されている [13][14]。また荷物の搬送において、無人搬送車、無人搬送ロボットを用いた自動化による作業効率化が進んでいる。このロボットのナビゲーションを行うために、柵脚やタグを物体検出により認識し、最適経路を導き出す研究も行われている [15]。

3. トラックバース状況の認識

YOLO などで提供されている既存のオブジェクト認識器は、人や車といった一般的な形状に対してはある程度有効であるが、本研究が対象とするトラックバースにおいては、対象物が特殊であるため、十分な精度を出すことが困難である。そこで、図 1 に示すように、トラックバースに設置したカメラ画像に対し、アノテーションを行い、認識モデルを学習した。

3.1 カメラによる画像の取得

トラックバースは図 2 のように 5 ヶ所のスペースがそれぞれ区切られた状態になっている。その 5 ヶ所のスペース

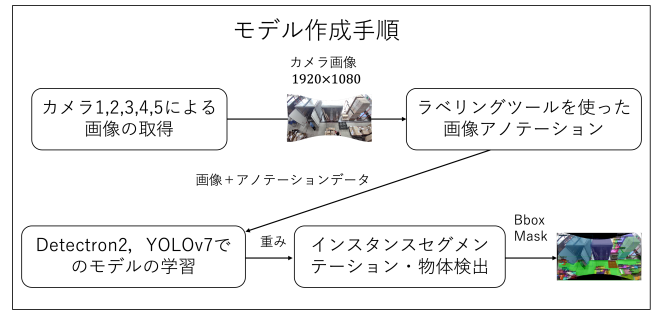


図 1 モデル作成手順

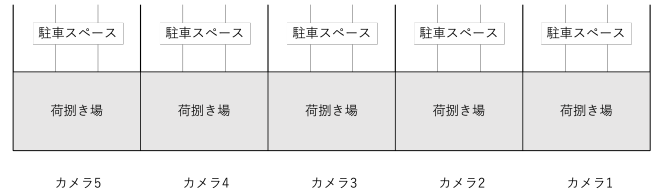


図 2 バースイメージ図

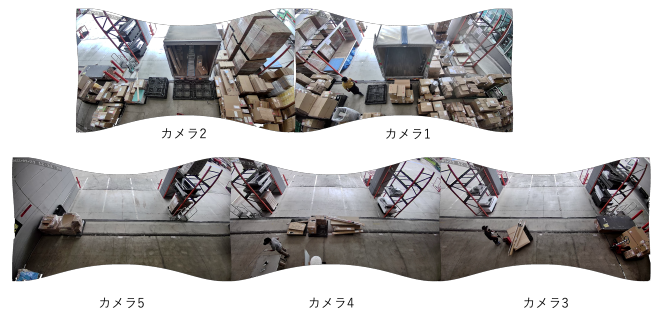


図 3 カメラ画像

の天井にカメラ 1-5 を設置することで、5 視点から画像を取得した。それぞれカメラ 1-5 によって取得した画像を図 3 に示す。

3.2 データセット

トラックバース内の 5 つのカメラから取得した画像を使用した。画像 1791 件に対し、179324 件のアノテーションを行い、データセットを作成した。図 4 に示す 14 クラスの分類を行った。今回は、アノテーションツールとして FastLabel を採用した [16]。また、トレーニング用とテスト用の画像件数を 8:2 となるように分割しており、1432 件の画像をトレーニング用に、359 件の画像をテスト用に使用した。アノテーションの詳細な件数に関して表 1 にまとめており、それぞれトレーニング、テスト、合計の 3 つのアノテーション件数についてまとめている。

3.3 トレーニング

前節で記載したトレーニングデータセットを使って検出用モデルの学習を行う。初めに、Detectron2 での学習に取り組んだ。Detectron2 は Facebook AI Research によって開発され、Pytorch で実装されたフレームワークである。



図 4 分類クラス

表 1 アノテーション件数

クラス名	train	test	total(train+test)
梱包カート	610	191	801
ロードコーン	374	76	450
ハンドパレット	870	211	1081
カート	1951	499	2450
平台車	323	75	398
トラック	1994	508	2502
作業員	2152	536	2688
パレット	28596	7061	35657
コンテナ	13643	3201	16844
梱包コンテナ	1010	297	1307
床	1428	358	1786
緩衝材	1561	378	1939
ケージトロリー	1002	207	1209
ダンボール	87779	22433	110212

これには, Fast R-CNN, Faster R-CNN, Mask R-CNN などのいくつかの物体検出モデルの実装が含まれており, ここでは, バックボーンアーキテクチャとして ResNet-50 を備えた, インスタンスセグメンテーションが可能である Mask R-CNN を使用した. 次に, YOLOv7 での学習に取り組んだ. YOLOv7 は, 物体検出タスクにおけるディープラーニングモデルの一つであり, You Only Look Once (YOLO) アーキテクチャの改良版である. YOLOv7 は, 高速かつ精度の高い物体検出を可能にするため開発された物体検出モデルである. これら検出用モデルの学習をそれぞれバッチサイズ 8, エポック数 400 に設定し, 損失が収束するまで行った.

4. 検出用モデルの評価, 考察

本章では, インスタンスセグメンテーション, 物体検出の評価指標を用いた定量的評価と, 実際に検出された画像での考察を踏まえた定性的評価を行う. インスタンスセグメンテーション (Detectron2) と物体検出 (Detectron2, YOLOv7) の評価をそれぞれ行った. また, インスタンスセグメンテーションの過程において, 物体検出によるバウンディングボックスの取得が行われる. そのため, このバウンディングボックスの評価については Detectron2 と YOLOv7 でそれぞれ算出した値を比較した.

4.1 インスタンスセグメンテーションの評価

インスタンスセグメンテーションに使用される標準的な

表 2 Detectron2. セグメンテーションのクラス毎の AP

クラス名	AP
梱包カート	0.101
ロードコーン	0.488
ハンドパレット	0.413
カート	0.307
平台車	0.133
トラック	0.736
作業員	0.505
パレット	0.223
コンテナ	0.467
梱包コンテナ	0.508
床	0.649
緩衝材	0.369
ケージトロリー	0.579
ダンボール	0.441

表 3 Detectron2, YOLOv7. 全クラスの bbox の平均 AP

	AP	AP50	AP75
Detectron2	0.599	0.748	0.667
YOLOv7	0.707	0.891	—

評価指標には, 平均適合率 (Average Precision, AP) が用いられる. これは, COCO セグメンテーションチャレンジ [17] における最先端手法のランキングに使用されている指標である. 本研究では, 公式の COCO 評価ツールと同じように AP を計算し, その値を表示している. AP は, 特定の IoU (Intersection over Union) 閾値で計算された適合率-再現率曲線の下面積である. また AP は通常, 複数の IoU 閾値 (通常は 0.05 ステップの間, 隔 [0.5, 0.95]) でそれぞれ計算し, 平均をとることで算出する. mAP は, データセットの全クラスの AP の平均値である. 以下では, IoU とクラスで平均化した値を AP と略記する. 特定の IoU 閾値 (例えば 0.5) に対するクラス平均の AP に言及する場合は, AP50 と表記する. Detectron2 で学習したモデルにおいて, セグメンテーションでの AP の算出を行った. セグメンテーションの全クラスの平均 AP は, AP : 0.432, AP50 : 0.652, AP75 : 0.478 が得られた.

4.2 物体検出の評価

前節同様に AP に関して評価を行った. Detectron2, YOLOv7 のそれぞれでバウンディングボックス (bbox) の AP の算出を行い, 表にまとめた.

4.3 検出結果の可視化

ここではインスタンスセグメンテーションで得られた結果をクラス単位で可視化し, 正解データとの比較を定行的に行う. 正解データに検出できている例を成功例, 逆に正解データと異なる検出結果となった例を失敗例とした. 図はそれぞれ左図が正解データ, 右図が検出結果を表している.

表 4 Detectron2,YOLOv7.bbox のクラス毎の AP

クラス名	Detectron2 AP	YOLOv7 AP
梱包カート	0.314	0.528
ロードコーン	0.547	0.724
ハンドパレット	0.571	0.653
カート	0.656	0.743
平台車	0.434	0.595
トラック	0.958	0.989
作業員	0.576	0.465
パレット	0.513	0.682
コンテナ	0.576	0.699
梱包コンテナ	0.570	0.752
床	0.905	0.892
緩衝材	0.503	0.719
ケージトロリー	0.744	0.77
ダンボール	0.523	0.692



図 5 成功例 a(左：正解データ，右：検出結果)

図 5,6,7 ではそれぞれ成功例を表している。成功例 a は、対象の物体が一番上に積まれている場合であり、重なりにより隠れていないような場合である。成功例 b は、同じ状態のアノテーションが多く含まれている場合であり、長時間移動が行われていない物体が多い。成功例 c は、閉まっている状態のトラックであり、開いているトラックよりも精度良く検出できている。

図 8 14 はそれぞれ失敗例を示している。失敗例 a は、梱包してあるコンテナの認識について、梱包されていないコンテナと認識される場合である。失敗例 b は、作業員の服装が他の物体と誤検出される場合である。失敗例 c は、ビニールで包まれて、まとめられているダンボールに対して、それぞれ別々のダンボールと検出された場合である。失敗例 d は、梱包カートの検出において、正解データの無い箇所に対して検出された場合である。失敗例 e は、カートの検出に対して、検出がされない場合である。失敗例 f は、作業員の下半身が背景のオブジェクトに一体化して、検出できなかった場合である。失敗例 g は、重なりあう物体に対して、検出が行えていない場合である。

4.4 考察

ここでは前節での評価を踏まえ、考察を行う。まず、セグメンテーションの AP に関して、トラックが最も高く、0.736、続いて床が 0.649 であった。また、梱包カートは最



図 6 成功例 b

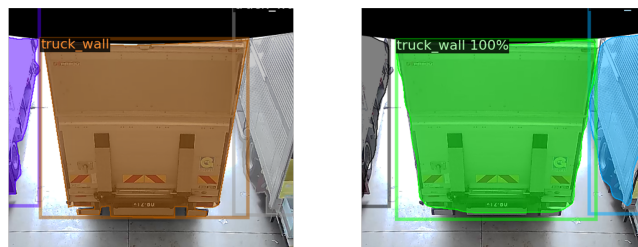


図 7 成功例 c



図 8 失敗例 a



図 9 失敗例 b

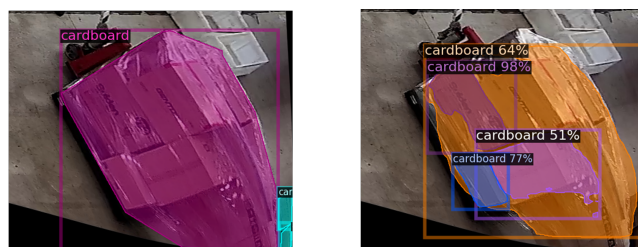


図 10 失敗例 c

も低く、0.101、続いて平台車が 0.133 であった。AP が最も高いトラックに関しては、データセットの中で、場合によっては最も大きいオブジェクトであり、物体が上に乗る、重なりが発生することもなく、駐車位置や角度もある程度決まっているため、精度良く検出できたと考えられる。これに対し、床は荷物や人のいない状態であれば、最も大き

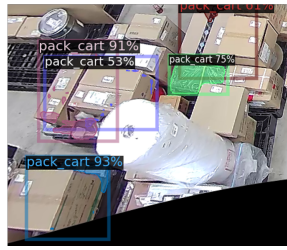


図 11 失敗例 d

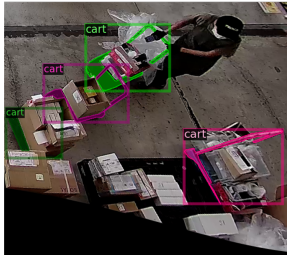


図 12 失敗例 e

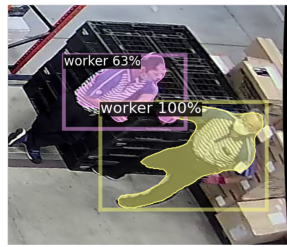


図 13 失敗例 f

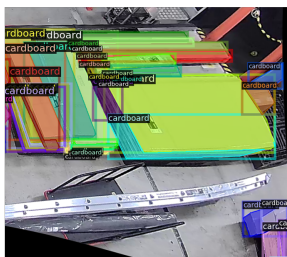


図 14 失敗例 g

く、動かずに固定されているオブジェクトであるが、荷物が置かれたり、人が通ったりすることで頻繁に形が変わる。そのため、ある程度検出しやすいオブジェクトであるが、トラックよりは精度は低くなったと考えられる。また AP が最も低い梱包カートに関しては、図 11 の失敗例 d のように一部しか写っていない場合が多く、さらにオブジェクト自体が比較的小さかった。図 14 の失敗例 g のように細かく重なりあった状態の場合も数多く存在した。検出できている場合としては、図 5 の成功例 a のようなオブジェクトが一番上に積まれている場合であった。この傾向は、平台車やパレットにも同様に見られ、AP が低い傾向となった。なお平台車に関しては、アノテーション件数が最も小さいことも原因として考えられる。

bbox の AP に関しては、全てのクラスでセグメンテ

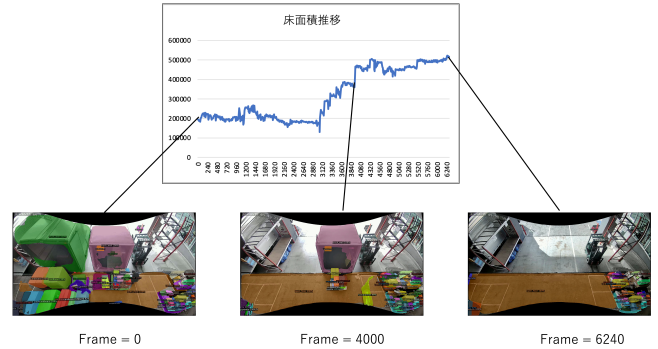


図 15 トラックバース状況の分析（床面積）

ションの AP を上回る結果となった、

5. トラックバース状況の分析

インスタンスセグメンテーション、物体検出の実現によりトラックバース状況の分析が可能となった。その一環として、本研究では床面積の検出による荷物状況推移の分析に取り組んだ。具体的には、あるトラックバースを写した 5fps の動画の各フレームに対して、インスタンスセグメンテーションを行い、床として認識されたピクセル数を算出した。その結果を図 15 に示す。グラフは、時間軸（フレーム）に対する床面積のピクセル数を表している。それぞれ 0 フレーム、4000 フレーム、6240 フレーム地点でのインスタンスセグメンテーションの出力を描画している。5fps の動画であるため、実時間では 0 秒、800 秒、1248 秒に当たる。ここでは床面積が増加することで、荷物の積み下ろしが終わり、トラックが出発した時には床上にあった荷物が全て無くなっていることを示している。

6. まとめ

本研究では、バースの利用効率や作業効率向上を目的として、トラックバースにおける荷物の積み下ろし状況の分析を行う枠組みを提案した。そのために、トラックバースで頻出する 14 クラスのオブジェクトのデータセットの作成を行い、インスタンスセグメンテーションモデルの学習を行なった。このモデルを利用し、バース状況分析の一環として、床面積の検出を行い、荷物状況の推移の分析に取り組んだ。

今後の展望として、床面積だけでなくトラック、作業員、荷物など様々なオブジェクトに関して、それらの相関関係の分析を行い、より高度なトラックバース状況の分析に繋がっていききたい。

謝辞

本研究の一部は、JSPS 科研費 JP22H03696, JP22K18422, NICT 委託研究 22609, JST CREST JPMJCR21F2, トラスコ中山株式会社に支援頂いています。

参考文献

- [1] 国土交通省. 荷待ち時間の記録義務付けについて. https://www.tb.mlit.go.jp/shikoku/content/vol16_kagawa_file07.pdf.
- [2] 株式会社 cec トラック積降効率化 ict ソリューション「logipull (ロジプル)». <https://logistics.cec-ltd.co.jp/>.
- [3] パース管理システム トラック簿. <https://www.info-sense.co.jp/service/truck-book.html>.
- [4] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. Detectron2. <https://github.com/facebookresearch/detectron2>, 2019.
- [5] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*, 2022.
- [6] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.
- [7] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 6, pp. 1137–1149, 2017.
- [8] Evan Shelhamer, Jonathan Long, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 4, pp. 640–651, 2017.
- [9] Zdenek Kalal, Krystian Mikolajczyk, and Jiri Matas. Tracking-learning-detection. *IEEE transactions on pattern analysis and machine intelligence*, Vol. 34, No. 7, pp. 1409–1422, 2011.
- [10] Nir Aharon, Roy Orfaig, and Ben-Zion Bobrovsky. Bot-sort: Robust associations multi-pedestrian tracking. *arXiv preprint arXiv:2206.14651*, 2022.
- [11] Mikel Broström. Real-time multi-camera multi-object tracker using yolov7 and strongsort with os-net. https://github.com/mikel-brostrom/Yolov7_StrongSORT_OSNet, 2022.
- [12] yolov7-deepsort-tracking. <https://github.com/deshwalmahesh/yolov7-deepsort-tracking>.
- [13] Rico Jonschkowski, Clemens Eppner, Sebastian Höfer, Roberto Martín-Martín, and Oliver Brock. Probabilistic multi-class segmentation for the amazon picking challenge. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1–7, 2016.
- [14] Andy Zeng, Kuan-Ting Yu, Shuran Song, Daniel Suo, Ed Walker, Alberto Rodriguez, and Jianxiong Xiao. Multi-view self-supervised deep learning for 6d pose estimation in the amazon picking challenge. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1386–1383, 2017.
- [15] Yiyu Sun, Tonghua Su, and Zhiying Tu. Faster r-cnn based autonomous navigation for vehicles in warehouse. In *2017 IEEE International Conference on Advanced Intelligent Mechatronics (AIM)*, pp. 1639–1644, 2017.
- [16] Fastlabel. <https://fastlabel.ai/>.
- [17] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft coco: Common objects in context. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, pp. 740–755, Cham, 2014. Springer International Publishing.