

庫内作業員を対象としたトラッキングによる検出漏れ抽出に基づく MLOps の運用評価 Worker-Oriented MLOps Based on Missed Detection Mining through Tracking: Empirical Evaluation in a Warehouse

加納 一馬[†] 森 裕輝[†] 片山 晋[†] 浦野 健太[†] 米澤 拓郎[†] 河川 信夫[†]
Kazuma Kano Yuki Mori Shin Katayama Kenta Urano Takuro Yonezawa Nobuo Kawaguchi

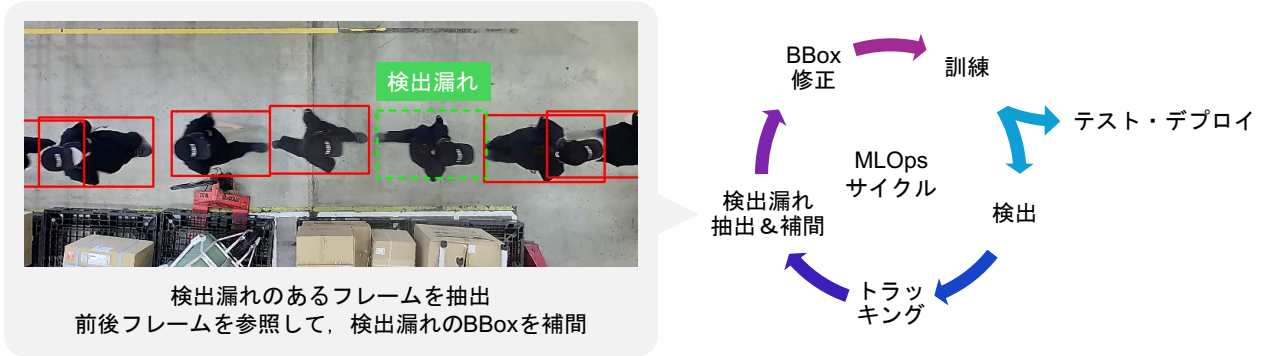


図 1 トラッキングされた物体の検出漏れ抽出に基づく MLOps フレームワーク

1. はじめに

カメラによる物体検出は警備や介護などあらゆる場面で利用されており、産業現場における生産性や Quality of Work (QOW) の向上を目的とした作業分析への応用も期待されている[1]。現在の物体検出技術の多くは深層学習に基づいて高い精度を実現しているが、一般に公開されている汎用モデルは、産業現場のような環境、広角カメラによる歪み、真上からの視点などに対して十分に対応できず、ファインチューニングやドメイン適応が必要になる。特に、物体検出システムの恒常運用においては、偶発的な事象や経時的な変化が問題になりやすい。例えば物流倉庫では、荷物の種類や配置が流動的であり、対象の大部分が遮蔽されていて検出しにくい状況などが偶然に起こりうる。また、時期による服装の違いやカメラレンズの汚れなども、ドメインシフトを引き起こし検出性能の低下を招く可能性がある。このように、あらゆる状況をあらかじめ想定してモデルを作成するのは困難であるため、検出性能を維持・改善するには継続・反復的なモデルの更新が重要となる。そこで、本研究ではモデルの再訓練や検証を含む一連のプロセスを円滑化する Machine Learning Operations (MLOps) に取り組む。

物体検出モデルの訓練における実用上の課題として、正解データのアノテーションにかかる人的労力が挙げられる。特に、モデルは典型的なデータよりも例外的なデータに対する推論を苦手とする傾向があるため、訓練データを無作為に増やすだけでは精度の底上げにつながりにくい。コストを抑えつつモデルの性能を向上させるには、学習効果の高いサンプルに対して重点的にアノテーションする能動学習が有効である。そこで本研究では、図 1 に示すように、物体検出モデルの MLOps に能動学習を取り入れる。本フレームワークでは、トラッキング結果に基づいて、検出漏れのあるフレームを自動的に抽出し、検出漏れのバウンデ

ィングボックスを前後フレームの情報から線形に補間する。その後、検出または補間したバウンディングボックスについて人が確認しズレ等を修正したものを正解データに加える。これにより、モデルが検出に失敗したデータが優先的に再訓練に使用され、効果的な性能向上が期待できる。また、完全に手でアノテーションするのとは比べ、少ない労力で正解データを作成できる。我々は非技術者による関与を想定し、ウェブ UI 等の機能を備えた MLOps システムを構築した。物流倉庫における作業員の検出モデルを対象として、訓練と正解データ作成のサイクルを実践したところ、サイクルを通じた検出精度の向上を確認できた。

2. 関連研究

2.1 MLOps

モデルの訓練やバージョンニング、データの収集やアノテーションなど、機械学習における開発から保守までのライフサイクルを一貫して支援する MLOps は、システムを恒常的に運用する上で重要である[2]。Kukkaro らは工場における蒸気の検出において、フォグデバイス上での継続学習の有効性を実証した[3]。ただし、現時点の MLOps に関する事例は企業によるものが主であり、実際にシステムを構築した研究報告は比較的少ない。また、MLflow [4]やKubeflow [5]など、オープンソースの MLOps 向けツールも開発されているが、プロジェクトの規模や他システムとの兼ね合いによっては自前で用意の方が適切な場合もある。

2.2 自己教師あり表現学習と転移学習

限られた正解データで高い汎化性能を実現するために、アノテーションのない大量のデータでバックボーンを事前に訓練するアプローチがしばしば用いられる[6]。He らは Masked Auto Encoder (MAE) によって優れた視覚表現を獲得できることを示した[7]。Oquab らは 1 億 4 千万枚以上の画像で訓練された DINOv2 の重みを公開している[8]。なお、これらの手法を物体検出などの下流タスクに適用するには、

[†]名古屋大学 大学院工学研究科
Graduate School of Engineering, Nagoya University

依然としてアノテーションされた正解データを要する。また、産業現場のような特殊な状況では、一般に公開されている基盤モデルの有効性は限定的である。

2.3 アノテーションの部分自動化

多くの研究が様々な切り口でアノテーションにかかる人的労力の削減に取り組んできた。あるモデルによる予測結果を別のモデルのための正解データとして利用するアプローチは疑似ラベリングと呼ばれる[9]。Chen らは物体上の一点のアノテーションからバウンディングボックスの疑似ラベルを予測する Point DETR を提案した[10]。Li らは画像ごとに 1 クラス当たり 1 つの物体にのみアノテーションし、その他の物体には類似性に基づいて疑似ラベルを付与する SIOD を提案した[11]。これらの手法は弱教師あり学習に該当し、簡素な正解情報でモデルを訓練できる。しかし、視覚特徴に依存した疑似ラベリングでは、オクルージョンにより手がかりが少なかったり、歩いている人と屈んでいる人のように同じクラスでも見た目が異なったりする場合に、データが不正確になりやすい。

いくつかの研究では、画像のアノテーションにおいて動画の時系列情報を積極的に活用している。Mahmud らは、複数の連続したフレームに対して教師モデルでバウンディングボックスを予測し、オプティカルフローの結果と整合が取れているもののみ、生徒モデルの疑似ラベルとして採用する SSVOD を提案した[12]。しかし、教師モデルの重みを生徒モデルの指数移動平均として更新するため、教師モデルが誤って予測すると生徒モデルが間違っただけで学習し教師モデルの性能が劣化する悪循環（いわゆる確認バイアス）に陥る可能性がある。Sivapuram らはあるフレームにおいてアノテーションされた物体に対してシングルオブジェクトトラッキングを行い、信頼度の高いフレームにおけるバウンディングボックスのみを疑似ラベルとして採用する SA-LfV を提案した[13]。なお、複数の物体が同時に映り込む状況は想定されていない。また、疑似ラベルの精度は追跡器の性能に依存する。

一方、我々はこれまでに、オプティカルフローで動体を抽出し、その埋め込み表現をクラスタリングによってグループ化する、半自動アノテーションフレームワークを開発した[14, 15]。バウンディングボックスの自動決定と複数物体への一括ラベリングにより、少ない労力で信頼性の高い正解データの合成が可能になった。しかしその反面、合成されるデータには静止した状態の対象物体が含まれないという課題があった。実際、この手法で庫内作業員のデータセットを作成し検出モデルを訓練したところ、モデルは検品のために静止している作業員を見落とす傾向があった。

2.4 能動学習

能動学習は、モデルの訓練に要する正解データ量を削減するために、アノテーションすべきサンプルを自動的に選定するアプローチである。Park らは物体検出モデルにとっての確信度を追加のヘッドで予測するとともに、データ自体の難しさをバウンディングボックスの構造に基づいて算出した[16]。Yang らは物体検出モデルの構造によらずにデータの不確かさと多様性を定量化できる PPAL を提案した[17]。その他、能動学習をアノテーションの部分自動化などと組み合わせた先行研究も存在する[18, 19]。

3. 設計と実装

3.1 フレームワークの設計

物体検出モデルの性能を効率的に維持・改善するため、トラッキングによる検出漏れ抽出に基づく MLOps フレームワークを提案する。アノテーションの部分自動化と能動学習により、人的労力を削減しつつ学習効果の向上を図る。本フレームワークは、対象物体がトラッキング可能である限り、任意の環境における任意の検出モデルに適用でき、事前学習済みモデルの転移学習とも併用できる。モデルの訓練から正解データの作成までのサイクルの流れを以下に示す。各ステップについて 3.1.1~3.1.5 項で詳しく述べる。

- (1) 正解データセットを学習用と検証用に分割。
- (2) 検出モデルをファインチューニング。
- (3) 動画の各フレームに対して検出を実施。
- (4) 検出した物体に対してトラッキングを実施。
- (5) 検出漏れが含まれるフレームを抽出。
- (6) 検出漏れのバウンディングボックスを補間。
- (7) 不確かまたは類似したフレームを除外。
- (8) バウンディングボックスを人が確認・修正。
- (9) フレーム画像を正解データセットに追加。

3.1.1 物体検出モデルの訓練

ステップ(1)では、過去のサイクル分を含む全てのアノテーション付き画像のデータセットを用意し、学習用と検証用のサブセットにランダムに分割する。ステップ(2)では、それらのデータを用いて検出モデルをファインチューニングする。このとき、学習用サブセットに対する損失が小さくなるように重みを更新し、検証用サブセットに対する精度が最大となったエポックにおける重みを採用する。なお、サイクルの度に重みを初期化してしまうと、損失の収束に時間がかかるだけでなく、モデルの検出精度の分布が変化して能動学習の効果が損なわれる可能性がある。そのため、前回のサイクルの重みを引き継いで追加学習を行う。

3.1.2 物体の検出

ステップ(3)では、ファインチューニング後のモデルを用いて動画の各フレームに映っている対象物体を検出する。データのパターンを増やすために、動画は過去のサイクルと異なる日時のものが望ましい。

3.1.3 物体のトラッキング

ステップ(4)では、動画に映っている対象物体を全フレームに渡ってトラッキングする。このとき、あらかじめ検出されたバウンディングボックスをフレーム間で繋げる Tracking-by-Detection (TBD) 型のマルチオブジェクトトラッキング手法を用いる[20]。これにより、物体の軌跡はその物体が検出されたフレームにおけるバウンディングボックスの系列として表される。本稿では TBD 型の代表的な手法である ByteTrack [21]を複数広角カメラ向けに独自に拡張した。トラッキングはグローバル座標系で行い、その軌跡を各カメラのローカル座標系に射影した。

3.1.4 検出漏れの抽出・補間

ステップ(5)では、トラッキング結果に基づいて検出漏れが含まれるフレームを抽出する。ある軌跡 τ に紐づけられたバウンディングボックスが存在するフレーム番号の集合を I_{det}^{τ} 、全ての軌跡の集合を \mathcal{T} とおく。検出漏れが含まれるフレーム番号の集合 I_{miss} は次のように与えられる。

$$I_{miss} = \bigcup_{\tau \in \mathcal{T}} \{i \in \mathbb{N} \mid \min I_{det}^{\tau} < i < \max I_{det}^{\tau} \wedge i \notin I_{det}^{\tau}\} \quad (1)$$

ステップ(6)では、検出結果とトラッキング結果に基づいて検出漏れのバウンディングボックスを線形に補間する。すなわち、それ以前に検出された最後のフレームとそれ以降に検出された最初のフレームにおけるバウンディングボックスを内分したものを、検出されなかったフレームにおけるバウンディングボックスとみなす。ある軌跡 τ に紐づけられた i 番目のフレームにおけるバウンディングボックスの左上の座標を $(x^{\tau,i}, y^{\tau,i})$ とおく。 m 番目のフレームにおける補間されたバウンディングボックスの左上の座標 $(\tilde{x}^{\tau,m}, \tilde{y}^{\tau,m})$ は以下のように求まる。右上、左下、右下の座標についても同様である。

$$l = \max \{i \in \mathbb{N} \mid i < m \wedge i \in I_{det}^{\tau}\} \quad (2)$$

$$n = \min \{i \in \mathbb{N} \mid i > m \wedge i \in I_{det}^{\tau}\} \quad (3)$$

$$\tilde{x}^{\tau,m} = \frac{(n-m)x^{\tau,l} + (m-l)x^{\tau,n}}{n-l} \quad (4)$$

$$\tilde{y}^{\tau,m} = \frac{(n-m)y^{\tau,l} + (m-l)y^{\tau,n}}{n-l} \quad (5)$$

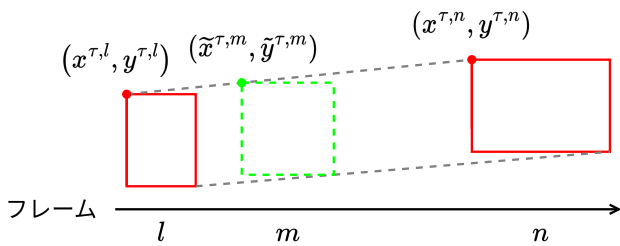


図2 検出漏れのバウンディングボックスの線形補間

ステップ(7)では、不確かさと類似性の2つの観点から、検出漏れが含まれるフレームのうち正解データに加えるものと加えないものを選択する。第一に、検出されない期間が長く続くと、補間されたバウンディングボックスが実際の物体と大きく乖離する可能性がある。データの質やアノテーションの効率の低下を防ぐために、ある期間 T_{miss} 以上検出されずに補間されたバウンディングボックスを含むフレームは採用しない。第二に、時間的に近接したフレームが持つ情報は互いに重複している傾向がある。モデルの学習効率を高めるために、採用したフレームからある期間 T_{near} 以内のフレームは採用しない。

3.1.5 バウンディングボックスの修正

ステップ(8)では、採用されたフレームの画像とバウンディングボックスを人が見て確認する。このとき、補間されたバウンディングボックスが実際の物体から大きくずれていたり、対象物体でないものにバウンディングボックスが付与されていたりした場合は修正する。これはサイクルの中で唯一の人手を伴う作業であるが、完全に手動でアノテーションするのとは比べて楽であり、人為的な見落としによるアノテーション漏れも起こりにくいと考えられる。ステップ(8)では、確認・修正されたアノテーション付きフレーム画像を正解データに加える。これらのデータは今回のサイクルのモデルにとって苦手なパターンを含んでいると期待される。以上のサイクルを繰り返して正解データを増強しモデルの汎化性能の向上を図る。

3.2 システムの実装

先述したフレームワークに従い円滑にサイクルを遂行するための MLOps システムを構築した。全体の構成を図3に示す。黒色のノードはユーザがやり取りする外部のインターフェース、青色のノードはシステムが内部で処理するバックエンドを表している。灰色のブロックはそれぞれ独立したコンテナとして動作し、ネットワークを介して通信する。

3.2.1 インターフェース

プロセスの実行や監視に関する主要な操作はウェブアプリとして実装されたダッシュボードから行う。これはプログラミングなどに馴染みのない非技術者との協働を意図している。参考としてサイクル監視用の画面を図4に示す。プロセスが終了するとユーザは Slack でプッシュ通知を受ける。フレーム画像とバウンディングボックスの確認および修正には、図5に示すような FastLabel というクラウド型プラットフォームのアノテーションツールを利用する。

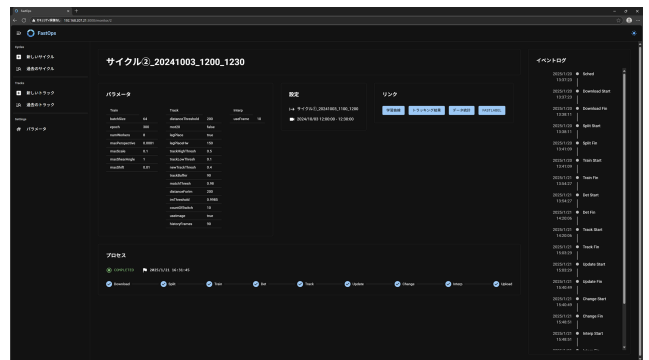


図4 ダッシュボードアプリ

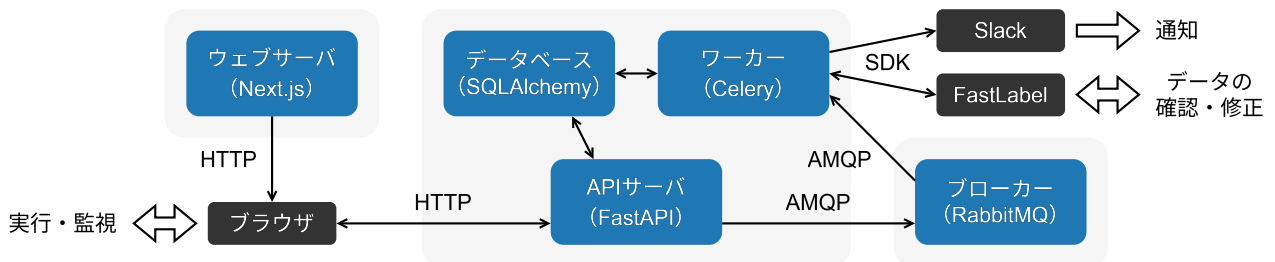


図3 MLOps システムの構成

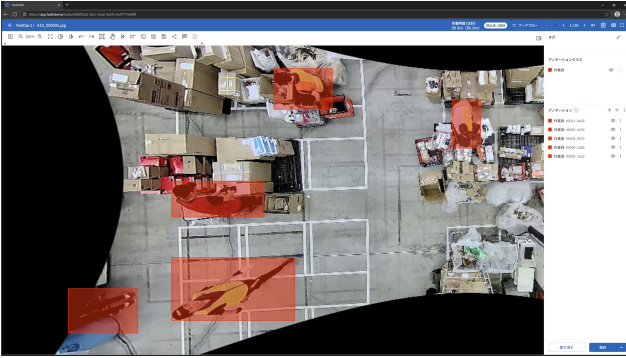


図5 FastLabelのアノテーションツール



図7 歪み補正後のフレーム画像の例

3.2.2 バックエンド

API サーバはブラウザから新しいサイクルの開始要求を受け取ると、ブローカーが管理するキューにタスクを登録する。ワーカーはキューに登録されているタスクを順に実行し、最後のタスクが完了またはエラーが発生したら Slack のボットでメッセージを送信する。なお、ここでいうタスクとは、修正後のデータを FastLabel からダウンロードしてモデルを訓練してから、検出漏れを抽出・補間して修正前のデータを FastLabel へアップロードするまでの一連の自動的な処理を指す。

4. 運用評価

4.1 使用するモデルとデータ

物流倉庫における作業員の検出モデルを対象として、MLOps システムの有効性を検証する。図6に示すように、入出荷場の天井に広角カメラ（H.View 製 HV-800G2A5）が鉛直下向きで設置されている。検出モデルには YOLOv8 の extra-large サイズを用いる[22]。モデルは手動でアノテーションされた 8515 枚の画像と半自動で合成[14, 15]された 8802 枚の画像からなる正解データで事前にファインチューニングされている。今回の実験ではこれらのデータに加えてある平日における 19 台分の映像を用いる。これらの映像はフル HD、8000 [kbps]、5 [fps]で記録され、Double Sphere カメラモデル[23]で歪みを補正されている。歪み補正後のフレーム画像の例を図7に示す。

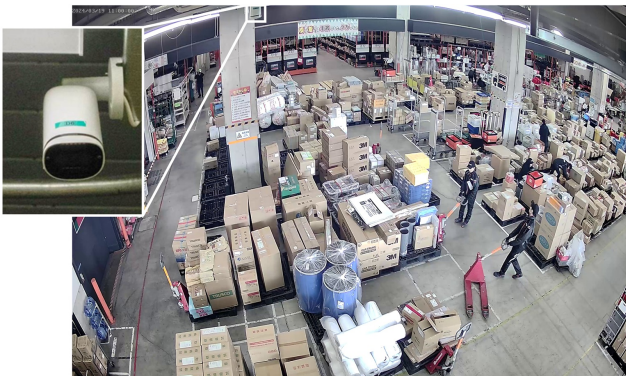


図6 入出荷場の様子

4.2 MLOps サイクルの実践

検出モデルの訓練から正解データの作成までのサイクルを 8 回実施した。正解データセットを学習用と検証用に分割する際の比は 9 対 1 とした。また、学習用のサブセットのみ hue, saturation, brightness, translation, scale, shear, perspective, flip についてランダムに拡張した。訓練時のバッチサイズは 64、最大エポック数は 300 とした。データの信頼性と多様性を決定するパラメータである T_{miss} と T_{near} はどちらも 2 秒間とした。各サイクルで使用した動画時間とそこから作成された正解データ量を表1にまとめる。動画時間当たりの正解データ量がサイクルによって大きく異なるのは、検出モデルの精度に加えてそもそも動画に映っていた作業員の数が異なるためだと推定される。

表1 使用した動画時間と作成された正解データ量

サイクル	動画時間 [分]	画像数	アノテーション数
1	60	3807	19732
2	60	1678	7940
3	30	1201	5029
4	30	660	2069
5	30	1271	6166
6	30	1241	6152
7	30	852	3530
8	30	939	4147

4.3 テストデータの作成

サイクルで使用していない合計 120 分間の映像から、評価用のテストデータを手動で作成した。ただし、モデルの検出精度の分布によらず公平に評価するために、検出漏れが含まれるフレームではなく、1 分おきに一樣にサンプリングされたフレームに対してアノテーションした。テストデータに含まれる画像数は 2280 (= 19 × 120) 枚、アノテーション数は 6251 個となった。

4.4 評価結果と考察

テストデータに対する各サイクルのモデルによる平均適合率 (AP: Average Precision) を表2にまとめる。また、これらの推移をプロットしたものを図8に示す。なお、 j 回目のサイクルのモデルとは、 $j-1$ 回目以前のサイクルで作成したデータを用いて j 回目のサイクルでファインチュ

ーニングした後のモデルを指す。1 回目のサイクルのモデルは実験前の状態と同じである。これらの結果より、サイクルを経るにつれて平均適合率が増加する傾向が見られた。我々の MLOps システムはアノテーションコストを抑えながらも確かに検出精度の向上に寄与するといえる。

表 2 テストデータに対する平均適合率

サイクル	AP50	AP50:95
1	0.850	0.634
2	0.913	0.677
3	0.918	0.678
4	0.927	0.684
5	0.926	0.691
6	0.934	0.687
7	0.936	0.682
8	0.938	0.686
9	0.939	0.685

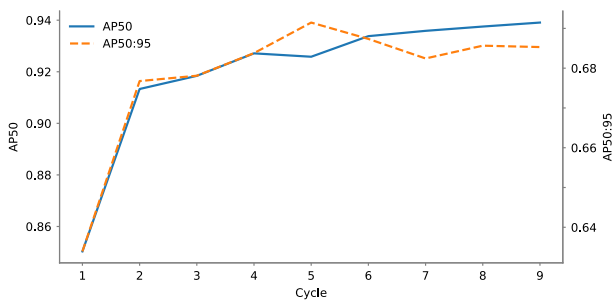


図 8 テストデータに対する平均適合率

ただし、安定的に増加していた AP50 と比べて AP50:95 は不安定であった。この要因として、作成された正解データに含まれるバウンディングボックスのズレが考えられる。本研究では検出漏れのバウンディングボックスを線形に補間したが、それらは必ずしも実際の物体の位置および大きさと一致しない。ズレがどれだけ明らかであったときに修正するのにはアノテータ次第であり、より厳密にズレを修正すれば AP50:95 のさらなる向上が期待できるが、その分だけアノテーションにかかる労力は増加する。バウンディングボックスのズレを低減するには、スプライン補間のような非線形な補間法や、物体の境界を予測するセグメンテーションなどが有効な可能性がある。とはいえ、バウンディングボックスの正確さがそれほど問題にならない用途も存在する。例えば、物流倉庫において作業員がどのパレットを検品しているのかを分析するには、パレット 1 個分に相当する 1.1 [m]程の粒度で測位できれば十分である。

図 9 は各サイクルのモデルによる PR 曲線である。なお、視認性のため一部の領域のみを表示している。この結果より、サイクルを経るにつれて再現率 (Recall) と適合率 (Precision) の両方が向上する傾向が見られた。本研究では検出漏れ (偽陰性) を抑制するように能動学習を行った。これが Recall の向上につながったと推察される。一方で、本研究は誤検出 (偽陽性) の抑制には焦点を当てていない。正解データに含まれるパターンの充実により、自然と正検出 (真陽性) が増加し Precision も向上したと考えられる。

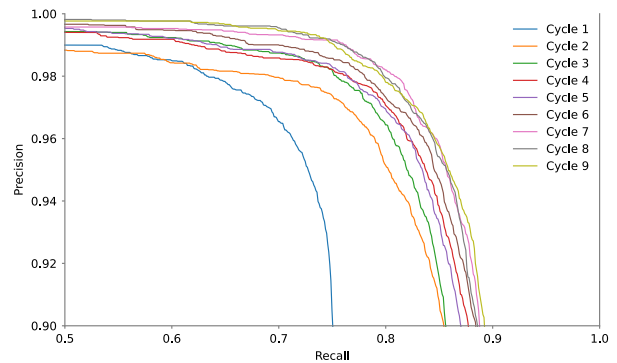


図 9 テストデータに対する PR 曲線

本研究では検出モデルが過去のサイクルで学んだ内容を忘れないよう、サイクルの度に全ての正解データを用いてファインチューニングした。この方法は本実験では十分に機能したものの、サイクルを経るにつれて訓練にかかる時間と計算量が増える上に、新しく追加された正解データの影響が弱まるという課題がある。継続学習[24]の文脈では、破滅的忘却を防ぎつつ新しいデータに適応させるために、過去の重要なデータと新しいデータの混合[25, 26]や、重要な重みパラメータの更新制限[27, 28]といったアプローチが提案されている。MLOps システムを長期的に運用していくために、今後はこれらの技術の適用も検討する。

5. おわりに

本研究では物体検出モデルの性能を効率的に維持・改善するための MLOps に取り組んだ。我々の MLOps フレームワークは、検出または補間されたバウンディングボックスを活用し、アノテーションにかかる人的労力を削減する。また、検出漏れのあったフレーム画像を再訓練に使用する能動学習により、学習効果の向上が期待できる。本研究ではこのフレームワークをシステムとして実装し、その仕様についても紹介した。さらに、庫内作業員の検出モデルを対象として、訓練とデータ作成のサイクルを実践した。その結果、サイクルを通じた検出精度の向上を確認できた一方で、バウンディングボックスのズレに関する課題も示唆された。我々は今後も本システムを改良し継続的に運用していく予定である。

謝辞

本研究の一部は JSPS 科研費 (JP22K18422)、NEDO 委託研究 (JPNP23003)、内閣府 CSTI SIP3 (JPJ012495)、JST BOOST (JPMJBS2422) の支援を受けています。また、実験環境を提供いただいたトラスコ中山株式会社様に感謝いたします。

参考文献

- [1] Nobuo Kawaguchi, Yusuke Asai, Kazuma Kano, Kairi Takaki, Yuki Mori, Yuma Suzuki, Kisho Watanabe, Yuki Gushi, Shin Katayama, Kenta Urano, Takuro Yonezawa, Shintaro Hashiguchi, "Digitization Methods for a Logistics Warehouse Towards Digital Twin-Driven Optimization," 2025 IEEE International Conference on Smart Computing (SMARTCOMP) (2025).
- [2] Abdullah Ikram Ullah Tabassam, "MLOps: A Step Forward to Enterprise Machine Learning," arXiv preprint (2023).
- [3] Ari Kukkaro, Sergio Moreschini, David Hästbacka, "Continuous Training vs. Transfer Learning on Edge and Fog Environments: A Steam Detection Use Case," 2024 50th Euromicro Conference on

- Software Engineering and Advanced Applications (SEAA), pp.138-141 (2024).
- [4] Andrew Chen et al., “Developments in MLflow: A System to Accelerate the Machine Learning Lifecycle,” Proceedings of the Fourth International Workshop on Data Management for End-to-End Machine Learning, no.5 (2020).
- [5] Aditya Pandey, Maitreya Sonawane, Sumit Mamtani, “Deployment of ML Models using Kubeflow on Different Cloud Providers,” arXiv preprint (2022).
- [6] Tobias Uelwer, Jan Robine, Stefan Sylvius Wagner, Marc Höftmann, Eric Upschulte, Sebastian Konietzny, Maike Behrendt, Stefan Harmeling, “A survey on self-supervised methods for visual representation learning,” Machine Learning, vol.114, no.4 (2025).
- [7] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, Ross Girshick, “Masked Autoencoders Are Scalable Vision Learners,” 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.15979-15988 (2022).
- [8] Maxime Oquab et al., “DINOv2: Learning Robust Visual Features without Supervision,” Transactions on Machine Learning Research Journal (2024).
- [9] Patrick Kage, Jay C. Rothenberger, Pavlos Andreadis, Dimitrios I. Diochnos, “A Review of Pseudo-Labeling for Computer Vision,” arXiv preprint (2025).
- [10] Liangyu Chen, Tong Yang, Xiangyu Zhang, Wei Zhang, Jian Sun, “Points as Queries: Weakly Semi-supervised Object Detection by Points,” 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.8819-8828 (2021).
- [11] Hanjun Li, Xingjia Pan, Ke Yan, Fan Tang, Wei-Shi Zheng, “SIOD: Single Instance Annotated Per Category Per Image for Object Detection,” 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.14177-14186 (2022).
- [12] Tanvir Mahmud, Chun-Hao Liu, Burhaneddin Yaman, Diana Marculescu, “SSVOD: Semi-Supervised Video Object Detection with Sparse Annotations,” 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp.6759-6768 (2024).
- [13] Arun Kumar Sivapuram, Prashanth Komuravelli, Rama Krishna Sai Gorthi, “SA-LfV: self-annotated labeling from videos for object detection,” Machine Learning, vol.114, no.1 (2025).
- [14] Keisuke Higashiura, Kodai Yokoyama, Yusuke Asai, Hironori Shimosato, Kazuma Kano, Shin Katayama, Kenta Urano, Takuro Yonezawa, Nobuo Kawaguchi, “Semi-Automated Framework for Digitalizing Multi-Product Warehouses with Large Scale Camera Arrays,” 2024 IEEE International Conference on Pervasive Computing and Communications (PerCom), pp.98-105 (2024).
- [15] Kazuma Kano, Yuki Mori, Keisuke Higashiura, Tahera Hossain, Shin Katayama, Kenta Urano, Takuro Yonezawa, Nobuo Kawaguchi, “Composite Image Generation Using Labeled Segments for Pattern-Rich Dataset without Unannotated Target,” Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp.507-512 (2024).
- [16] Younghyun Park, Wonjeong Choi, Soyeong Kim, Dong-Jun Han, Jaekyun Moon, “Active Learning for Object Detection with Evidential Deep Learning and Hierarchical Uncertainty Aggregation,” The Eleventh International Conference on Learning Representations (2023).
- [17] Chenhongyi Yang, Lichao Huang, Elliot J. Crowley, “Plug and Play Active Learning for Object Detection,” 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.17784-17793 (2024).
- [18] Yuting Wang, Velibor Ilic, Jiatong Li, Branislav Kisačanin, Vladimir Pavlovic, “ALWOD: Active Learning for Weakly-Supervised Object Detection,” 2023 IEEE/CVF International Conference on Computer Vision (ICCV), pp.6436-6446 (2023).
- [19] Mingfu Liang, Jong-Chyi Su, Samuel Schuller, Sparsh Garg, Shiyu Zhao, Ying Wu, “AIDE: An Automatic Data Engine for Object Detection in Autonomous Driving,” 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.14695-14706 (2024).
- [20] Zhiyu Guan, Zhaofa Wang, Gan Zhang, Luwei Li, Miaomiao Zhang, Zhiping Shi, Na Jiang, “Multi-object tracking review: retrospective and emerging trend,” Artificial Intelligence Review, vol.58, no.8, pp.235 (2025).
- [21] Yifu Zhang, Peize Sun, Yi Jiang, Dongdong Yu, Fucheng Weng, Zehuan Yuan, Ping Luo, Wenyu Liu, Xinggang Wang, “ByteTrack: Multi-object Tracking by Associating Every Detection Box,” Computer Vision – ECCV 2022: 17th European Conference, pp.1-21 (2022).
- [22] Glenn Jocher, Ayush Chaurasia, Jing Qiu, “Ultralytics YOLOv8” (2023).
- [23] Vladyslav Usenko, Nikolaus Demmel, Daniel Cremers, “The Double Sphere Camera Model,” 2018 International Conference on 3D Vision (3DV), pp.552-560 (2018).
- [24] Haoxuan Qu, Hossein Rahmani, Li Xu, Bryan Williams, Jun Liu, “Recent Advances of Continual Learning in Computer Vision: An Overview,” IET Computer Vision, vol.19, no.1, pp.e70013 (2025).
- [25] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, Christoph H. Lampert, “iCaRL: Incremental Classifier and Representation Learning,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.5533-5542 (2017).
- [26] David Lopez-Paz, Marc'Aurelio Ranzato, “Gradient episodic memory for continual learning,” Proceedings of the 31st International Conference on Neural Information Processing Systems, pp.6470-6479 (2017).
- [27] Friedemann Zenke, Ben Poole, Surya Ganguli, “Continual learning through synaptic intelligence,” Proceedings of the 34th International Conference on Machine Learning - Volume 70, pp.3987-3995 (2017).
- [28] Rahaf Aljundi, Francesca Babiloni, Mohamed Elhoseiny, Marcus Rohrbach, Tinne Tuytelaars, “Memory Aware Synapses: Learning What (not) to Forget,” Computer Vision – ECCV 2018: 15th European Conference, pp.144-161 (2018).