

物体検出とロボットアームを用いた 複雑な定型業務に対応する RPA の実現

井崎俊太郎[†] 堀川三好[†] 岡本東[†]

岩手県立大学大学院ソフトウェア情報学研究科[†]

1. はじめに

近年、画像を入力として特定の物体を見つけ出す物体検出技術は小売業界や医療業界において導入が進んでいる。また、ソフトウェアロボットを用いて PC 操作を自動化することで業務効率化を目指す取り組みである Robotic Process Automation (RPA) は、オフィス業務において注目を集めている。これらの技術及び取り組みを活用した事例は、今後さらに拡大していくことが予想される。本研究では物体検出とロボットアームを組み合わせることで、状況によって異なる対応が必要となるような複雑な定型業務や非定型業務に対応する RPA を実現することを目的としている。本稿では共同研究先企業と開発したスマホアプリの稼働を監視するシステムの概要、画像データ収集方法の実験およびシステム信頼性向上のための拡張機能について報告する。

2. 関連事例・研究

RPA の事例には、オフィス業務での PC 操作およびサーバの内部処理の自動化を対象としたものが多い。また、AI 技術を用いたシステムの異常検知に関する研究にはオープンデータセットを使用して正常値を学習し、外れ値を異常として推論する研究が多く見られる。

3. アプリ動作監視システム

3.1. 概要

近年のアプリは、複数のサーバを経由して利用者にコンテンツを提供するものが多い。また頻繁な仕様変更が行われるため、アプリの稼働状態を常に監視する必要がある。これを満たすには、利用者の立場から確認する必要がある。この一連の監視作業を物体検出とロボットアームを用いて自動化するシステムを開発する。今回共同研究先企業から以下に示す三つの要望を満たす形で開発している。

Realizing RPA for complex routine tasks using object detection and a robotic arm

[†] Iwate Prefectural University Graduate School of Software and Information Science

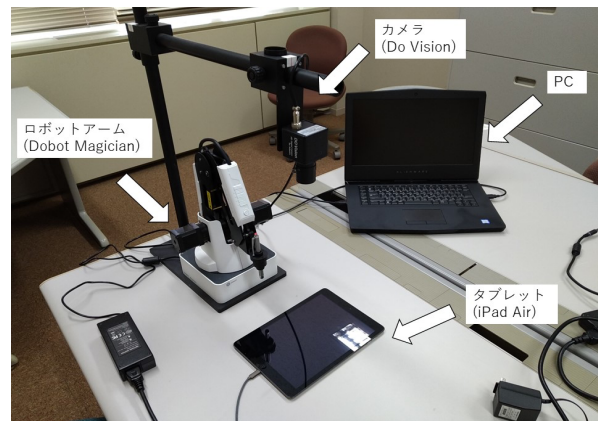


図 1. ハードウェア構成図

- ① 低価格で容易な実務導入が可能
- ② 変動要因が多いオフィス環境で利用可能
- ③ 複数のスマホアプリを同時に監視可能

3.2. システム構成

ハードウェア構成を図 1 に示す。ロボットアームには 10 万円程度と低価格ながら高性能な制御が可能な Dobot Magician を使用する。ロボットアームの先端部分には一般的タブレット用のタッチペンを装備することでタブレット操作を行う。Do Vision を用いてカメラを固定し画像を取得する。タブレットにはバッテリーの安定性に優れた iPad Air を使用する。物体検出用として Intel Core i7-8750H を搭載したノート PC を用いる。

3.3. 処理フロー

提案システムは、対象アプリの表示、動作確認の実行、動作判定および異常判定の 4 つの処理で構成される。複数アプリの監視を想定した場合の例として、LINE チャットボットを対象とした処理フローを以下に示す。

① 対象アプリの表示処理

アプリを切り替えるためのボタンを検出して、タップを行う。事例では LINE のプロフィールアイコンを検出することで切り替えを行う。

② 動作確認の実行処理

各アプリに定めた動作確認ボタンの物体検出を行い、ロボットアームでタップを行う。失敗し

た場合には未検出カウンタを加算して④異常判定処理に移行する。事例では、特定の LINE チャットボットのボタンを検出し、タップする。

③ 動作判定処理

動作結果の画面から各アプリに定めた動作判定アイコンの物体検出を行い、検出結果と動作確認ボタンとの相対的な位置関係からアプリの動作が正常かの判定を行う。異常であるときは未反応カウンタを加算する。事例では、LINE のボタンタップ後の反応画面を対象とする。

④ 異常判定処理

未検出カウンタと未反応カウンタの結果から、予め設定した閾値以上である場合に、異常を通知する。事例では、閾値を5回に設定し、サーバ異常が約5分以上回復されないときに通知する。

3.4. 監視システムの異常検知

提案システムを長期稼働すると、アプリ異常と関係なくシステム異常が生じる。原因の大半は対応が容易だが、種類が多い。そのため、直前の動画を用いて原因を推論する学習モデル生成し、信頼性向上を目指す。本稿では、ロボットアーム異常を検知する学習モデルを生成する。

4. 評価実験

4.1. データセット比較実験

2つのLINEチャットボットアプリを対象に実験を行う。この事例では、アプリにより位置やデザインが変わるため、ロボットアーム制御が必要となる。そのためデータセットをいくつか作成し、mAP(mean Average Precision:各物体ごとの平均適合率の平均)を評価指標として用いて精度検証を行う。先行研究[1]では、YOLOv3を用いて物体検出を行ったが、実務適用を想定して、クラウド上で容易に学習モデルを生成可能なGoogle Cloud AutoML Visionを用いる。テスト画像には300枚のランダムにタブレットを水平回転させた画像を用いる。学習用データセットには回転角度、30度の垂直傾斜の有無を組み合わせ4つのデータセットを作成した。

- ① D1 (回転角度 90・傾斜無・12枚)
- ② D2 (回転角度 45・傾斜無・16枚)
- ③ D3 (回転角度 90・傾斜有・20枚)
- ④ D4 (回転角度 45・傾斜有・40枚)

実験結果を表1に示す。結果からデータセットには、D2が最適であると思われる。さらにD2の収集方法の方針で、画像枚数の増加や回転角度の増減による精度への影響を検証した。結果として、画像枚数を16枚から増やす、または回転角度を45度以上細かくすることによる精度向上は、ほとんど見られなかった。今回の事例にお

表 1. 精度検証実験結果

	D1	D2	D3	D4
画面切替 アイコン(AP)	0.985	0.983	0.778	0.980
動作確認 ボタン(AP)	0.530	0.993	0.525	0.990
動作判断 アイコン(AP)	0.253	0.895	0.179	0.820
mAP	0.589	0.957	0.494	0.930

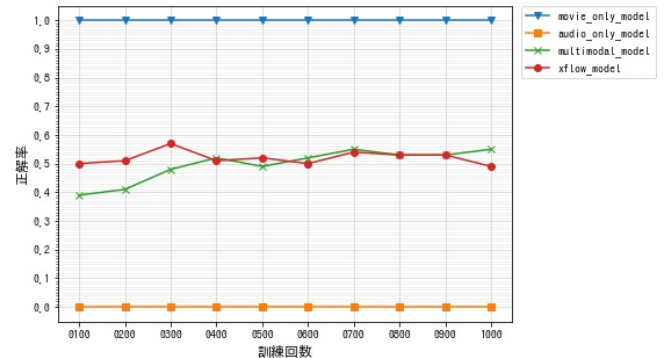


図 2. 訓練 100 回数ごとの精度グラフ

いては、データセット D2 のように回転角度 45 度程度ずつ回転させながら 16 枚の画像を用いるのが適切である。これらの指針をもとにデータセットを作ることで、1 時間程度で画像収集から学習モデル作成まで行うことが可能となった。

4.2. 監視システムの異常検知検証実験

動画と音声マルチモーダル学習する方法として、XFlow[2]を用いる。精度検証のため、動画または音声のどちらか入力としたモデル、XFlowにおいて、特徴量交換有り・なしのモデルの4つを生成する。ロボットアーム稼働時に収集される動画 100 本に対して正常・異常の推論を行い精度比較する。結果を図2に示す。ロボットアームの異常検知については、動画のみのモデルの精度が高い結果となった。他の異常原因に対して、どの学習モデルが有効であるかは、今後の検証が必要である。

5. おわりに

本研究ではアプリ稼働監視システムを開発し、データセットに関する検証とシステム異常を検知する学習モデルについて報告した。本システムは共同研究先で事業化を行っており、今後は他アプリや未適用分野への適用を検討する。

参考文献

[1] ロボットアームと物体検出による新たな業務自動化の実現, 井崎俊太郎, 堀川三好, 岡本東, 第23回画像の認識・理解シンポジウム: IS1-1-3(2020)

[2] XFlow: Cross-modal Deep Neural Networks for Audiovisual Classification, Cătălina Cangea, Petar Veličković, Pietro Liò, arXiv:1709.00572(2019)