

認識 AI を迅速に賢くするフレームワークの構築 — どんなものも一瞬で認識できる世界を目指して —

1. 背景

昨今の消費者ニーズの変化は極めて速い。ニーズの変化に合わせた多様な商品が日々、新商品として製造される状況である。さらに、人手不足対策や経済効果を狙ったコンビニや物流倉庫の無人化の試みがある。これらの多様な商品に溢れる現場での無人化を実現するためには、「カメラによる商品の高精度な認識」と「商品の迅速な登録」の二点が必要である。現在、一つ目の課題に関しては、深層学習を用いることで、精度の飛躍的向上がなされている。しかし、商品の学習データの準備にかかる手間と時間のために、二つ目の課題が残ったままである。我々が過去に実施した例をあげると、6 つの商品の学習データセットを 500 枚分収集する場合、学習データセットの準備全体で 15 時間以上もかかる。

学習データセットの生成には、様々な見え方で写った商品画像と、それに紐づいた商品情報のデータが必要である。カメラから見た商品が色んな姿勢で写るように商品を配置換えして多視点から画像を撮影する。その画像に対して、商品情報のデータ付けていく、アノテーションと呼ばれる作業を行う。アノテーションでは、図 1 のようにツールを使用して、商品の輪郭を囲むボックス(バウンディングボックス)を描画して、そのボックスの商品が何かのラベル付けを 1 枚 1 枚の画像に対して行う。この二つの作業を人手で行うと、膨大な時間がかかるため、新商品に対応して認識システムを迅速に組み替えることができない。これらの人手作業が商品の迅速な登録を実現する上でのボトルネックとなっている。



図 1. lebellmg (<https://github.com/tzutalin/lebellmg>) を用いたアノテーション作業

2. 目的

本プロジェクトでは、画像認識 AI に必要となる学習データセット作成の手間と時間を削減し、商品の迅速な登録が可能な枠組みの構築を目的とする。実際に、多視点画像撮影とアノテーション作業を自動化することで、人手で 15 時間以上かかる作業時間を大幅に短縮することを目指す。

3. 開発の内容

本プロジェクトでは、ロボットアームと自動回転ステージを用いた多視点画像撮影システムと、高精度視覚マーカを用いた自動アノテーションシステムを開発した。図 2 に本プロジェクトで構築したフレームワークの概要を示す。結果として、これまで 15 時間以上かかっていた学習データの準備の工程を 4 分以内で実現することに成功した。

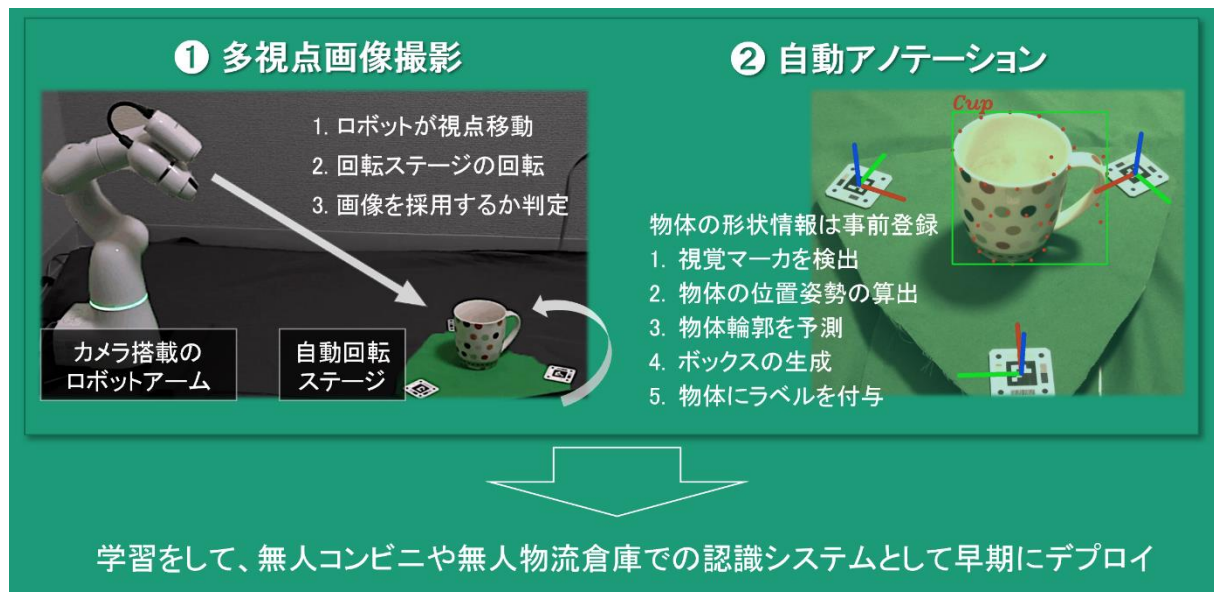


図 2. 学習データの準備が迅速に完了するフレームワーク

3.1. 多視点画像撮影

本プロジェクトでは、まず様々な見え方の商品を自動で撮影するためのシステムを開発した。システム構成としては、株式会社デンソーウェーブのカメラ搭載のロボットアーム COBOTTA とシグマ光機株式会社の自動回転ステージ OSMS-60YAW を用いている。自動回転ステージの回転プレートは、SOLIDWORKS で設計した。6 自由度のロボットアームを用いて、様々な方向から商品を撮影する。ロボットアームの可動範囲では撮影できない死角領域も撮影するため、自動回転ステージを用いて商品自体を回転させる。各撮影画像に対して、画像を採用するかの判定を行い、学習する際に有用なデータのみを蓄積していく。

3.2. 自動アノテーション

自動アノテーションを実現するために、本プロジェクトでは AR マーカを使用している。AR マーカは、画像中で検出できれば、その ID と位置姿勢を知ることができる。それを応用して、事前に商品と紐づけておいた AR マーカを画像中で検出することで、商品の位置姿勢も推定できる。また、商品の形状情報を事前に与えておくことで、AR マーカの検出情報からその輪郭位置も推定することができ、輪郭を囲むボックスの生成も可能となっている。最後に AR マーカの ID と紐づけた商品のラベルを画像中のボックスに対して付与することで、学習データセットを生成していく。AR マーカを用いた自動アノテーションに関しては、自動画像撮影部分も含めて特許取得済みである（特許第 6474179 号）。

自動アノテーションで用いる AR マーカについては、今回は国立研究開発法人産業技術総合研究所が開発した LentiMark を用いている。LentiMark は通常のマーカよりも位置姿勢

の検出精度が高く、検出ロストが少なく、マーカ表面に対して平行な方向から見た場合でも検出可能であるという特徴を持つ。これにより学習データセットとして収集可能な撮影視点のバリエーションが増えている。

しかし、商品を囲むボックスの中に AR マーカが写り込んでしまった場合、そのマーカの画像中の特徴が、認識 AI の学習を混乱させてしまう問題があった。AR マーカの持つ画像特徴を商品の特徴だとして学習してしまうのである。そのような学習をしてしまうと、認識 AI は AR マーカがなくては、商品を認識できない。従って、本プロジェクトでは、そのようなケースが起こる場合は、画像中の AR マーカの領域にノイズ画像を上書きする手法をとる。マーカがボックス内部に含まれていても、ノイズ画像で消せば、認識 AI はマーカを学習することなく、商品のパッケージの特徴に注目して学習するようになり、認識精度も向上する(図 3、図 4)。



図 3. 学習用画像のサンプル

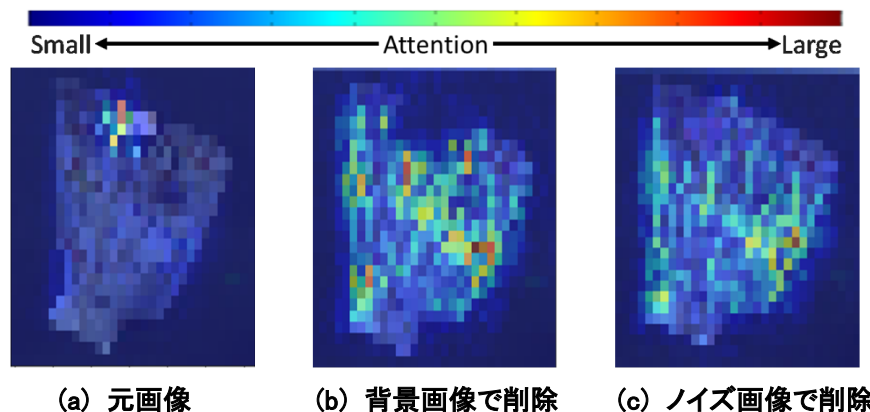


図 4. 物体特徴の学習時における、認識 AI の各画像領域の注目度合いを示すヒートマップ

4. 従来の技術(または機能)との相違

学習データセットの迅速な生成に関する従来技術を大きく二つ紹介して、それらと比較することで、本プロジェクトにおける技術の特徴を明らかにする。

4.1. 学習データセットの生成

一つは、既存の画像データを加工することにより学習データセットを増量する手法がある。例えば、画像の切り取り、回転、拡大縮小の処理を施すことや、新しい背景画像の合成を行うことで、新しい画像データを作成する。もう一つの手法は、商品の 3D CAD モデルのレ

ンダリング画像を使用する手法である。しかしながら、これらの手法では、実環境の商品のような変形や照明環境の再現は困難であり、我々のアプローチのような実環境に写った商品画像を迅速に収集する方が有用である。

4.2. 効率的なアノテーション

実環境の画像データを効率的に収集するための手法がいくつか提案されている。一つは、学習済みの検出器を用いる方法である。検出器によるラフなアノテーションに対して、人が良く付けられているかそうでないかの最終チェックをしている。そのため、依然として、人の作業が必要である。他のアプローチは、アノテーションを容易にする何らかのツールを使用する方法である。物体のバウンディングボックスを囲む作業を改善し、物体の特徴的な4点のみをクリックしていくツールを使う手法や、ヒューマノイドロボットに搭載の物体の追従システムを使用して、人がロボットの目の前の物体を動かしてあげることで色んな環境の物体を撮影してアノテーションする手法がある。しかしながら、これらの手法は、アノテーション作業に関して完全に自動化されておらず、人手の作業が必須である。

4.3. 本プロジェクトで開発した技術の特徴

従来のアプローチと比較して、本プロジェクトのアプローチの特徴は、現実世界の画像に対して完全に自動化されたアノテーション手法であり、かつ多視点の画像の自動収集が可能であることである。そこが本プロジェクトの技術の強みであり、学習データセット生成とその迅速な収集に関して、新しく有効な特徴である。

5. 期待される効果

本プロジェクトで開発したプロダクトにより学習データの準備が迅速に完了することができるため、無人コンビニや無人物流倉庫において、高頻度の品替えにも対応できるようになる。実際に、ロボット導入を推進している物流メーカの、アスクル株式会社にヒアリングを行い、実用化のニーズがあることを確認している。まずは、産業用ロボットシステムの分野における物流工場や食品工場などで普及することが期待でき、その先として無人コンビニや完全に自動化された物流倉庫の普及に貢献できると考える。

6. 普及(または活用)の見通し

まずは、物流現場での業務フローに導入できる形で本プロダクトをリリースする。また、直接プロダクトとして売り込むだけではなく、機械学習に必要な画像データセットの共通プラットフォーム化をサービスとして開始することも検討を進めている。

7. クリエータ名(所属)

清川 拓哉(奈良先端科学技術大学院大学)

友近 圭汰(株式会社リクルート住まいカンパニー)