

# 弱教師あり学習型物体検出器と半教師あり学習型分類器を用いたセルフサービス型店舗における非包装商品の認識

林 慶二 指導教員：小林 邦和

## 1 はじめに

近年,COVID-19(coronavirus disease 2019)の世界的な蔓延で、人同士が触れ合わない非接触の仕組みや製品は増えつつあり、今後の人々の生活様式を変えていくと思われる。その中で、スーパーマーケットやコンビニエンスストアでは、人同士の接触を避ける為の製品としてセルフレジの導入が進んでいる。しかし、セルフレジの導入が小売業、外食産業でも進む中、セルフレジへの対応が難しい商品も存在する。包装されていないパンや野菜、果物等の非包装商品である。それらにはバーコードやICタグなどの識別器は付けることが困難で、セルフレジでの識別は難しい。現状の対策として多くのスーパーマーケットでは、非包装商品に対し買い物客自身が種類を識別し専用のレジ付属のパネルに該当する商品を選択している。しかし、客自身での商品の識別は間違いなどの恐れもあり、優れたシステムとは言い難い。そもそも商品の分類に関して、人の介入は、効率の悪さやヒューマンエラーの恐れもある。そのため、バーコードなどが付けられない非包装商品の分類を、一切の人の介入なく、機械が自動で行うレジシステムの認識機構の開発が必要になっている。

画像の物体認識では、多様な特徴を自動抽出可能な深層学習を用いた物体検出モデルがあり、数多く提案されている。しかし、YOLO(You Only Look Once)[1]等の教師あり学習型物体検出モデルはリアルタイム性や高い検出精度を誇るものの、学習には画像と物体の位置とラベルを表すバウンディングボックス等の対象物への正確なアノテーションデータが必要であり、新しくラベルを作成、認識するには多大なコストがかかる。それらは高性能な物体検出モデルが数多くある中で、実用化の妨げになっていると思われる。従って本研究での目的は、複数物体認識システムを構築する事を大きな目的として、実用化の妨げになっているバウンディングボックス等のアノテーションデータを不要にし、教師あり学習型物体検出モデルと同程度の高速性や高い認識性能を目指す。さらに、実用化に向け、特殊な照明環境を必要としない事やオクルージョンがある物体への対応も目指す。本研究では、バウンディングボックスを学習に使用しない弱教師あり学習型物体検出器と高い分類精度を誇る半教師あり学習型分類器を組み合わせて、目的の実現を目指す。

## 2 提案手法

提案手法 1 では、商品分類のための 2 つのモデルを組み合わせた物体認識手法を提案し、提案手法 2 では、精度向上を目指し提案手法 1 で用いた両モデルと組み合わせたシステム全体に改良を加える。次節から提案手法 1 と 2 の詳細を説明する。

### 2.1 提案手法 1

複数の商品が存在する画像に対して、弱教師あり学習型物体検出器 WSCDN (Weakly Supervised Collaborative Detection Network)[2]を用いて個々の商品画像に分け、IIC(Invariant Information Clustering)[3]を用いた半教師あり学習型分類器で、商品を認識する手法を提案する。両モデルとも学習時にバウンディングボックスが不要であり、画像とラベルのみの学習で複数物体認識が可能である。特徴として WSCDN は、オブジェクト全体を確実に覆うバウンディングボックスを多く生成することが可能であり、物体検出後の分類器での分類を行いやすくする。また教師あり学習型モデルの分類精度を超えると報告された IIC を適用した半教師あり学習型分類モデルで WSCDN から

出力された各物体の画像を分類する。教師あり学習型モデルと比較すると、半教師あり学習、弱教師あり学習型モデルの複数物体認識の性能は大幅に劣る。しかし、物体検出が得意な WSCDN と画像認識が得意な IIC を用いた分類器を組み合わせることで、教師あり学習型モデルと同程度の認識性能を目指す。

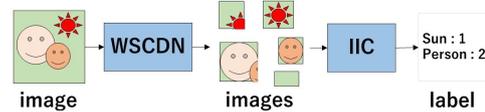


図 1: 提案手法

### 2.1.1 構築方法

提案手法の構築方法について説明する。WSCDN は物体検出だけでなく、物体認識も可能である。しかし、弱教師あり学習のため教師あり学習と比較すると mAP 等の物体認識精度は劣り、1 つの物体に対し複数のラベルが付く場合もある。そのため、WSCDN で分類されたラベル情報は使用せずに WSCDN は物体検出機能のみを使用する。また、ラベル情報をなくすことにより NMS[4] の効果が上がる。NMS は 1 つの物体に対し複数生成された領域候補から最もスコアが高い候補を選ぶアルゴリズムで、同種類のラベルを持つ領域候補しか適用できなかった。しかし、ラベルがない為、間違っただけのラベルにも適用でき、間違っただけの領域候補の削除が可能になった。そして、WSCDN で使用されている NMS を、改良された Soft-NMS[5] に変更する。Soft-NMS は NMS と違い領域候補を削除するのではなく、領域候補のスコアを減らすアルゴリズムで物体検出の再現率が NMS より高いのが特徴である。そのため、正しく検出された多くの領域候補を分類器に渡し提案手法全体の正解率向上を図る。次に IIC を用いた半教師あり学習型分類器の構築方法を説明する。まず IIC とは、画像  $x$  と  $x$  を加工した画像  $x'$  を作成し、2 つのデータの相互情報量を最大化することで、共通部分に着目し主要な特徴を抽出可能な学習方法である。画像認識モデルである VGG16 と画像  $x$  と  $x'$  を用意し、IIC の学習方法を用いて教師なし学習を行う。次に、学習後の VGG16 を少量の画像とラベルでファインチューニングすることで半教師あり学習を行い、精度の高い分類器を構築している。

### 2.2 提案手法 2

#### 2.2.1 D-NMS

WSCDN で使用される NMS の閾値 IoU は 2 つのバウンディングボックスの面積の積を和で割った値で、感覚的にパラメータ調整が難しい。また IoU では、2 つのバウンディングボックスの距離間が考慮されておらず、接触しない場合、2 つの距離が近い場合でも遠い場合でも、等しく IoU の値は 0 になる。そこで、そもそもの閾値を物体間の距離に変更した D-NMS を提案する。これにより、閾値の設定が視覚的に分かるようになり、より細かな設定ができること期待できる。以下に手順を示す。

1. 2 つのバウンディングボックスの中心座標を取得
2. 中心座標間のユークリッド距離を計算
3. 事前に設定した閾値より近い場合、スコアの低い方を削除

#### 2.2.2 入力画像の生成

本研究で使用する分類器の入力画像の形状は正方形である必要がある。しかし、WSCDN によって切り出された画像のアスペクト比は一定ではない。そのため画像を入力サイズにただ合わせるだけでは、アスペクト比の影響で画像が大きく歪む恐れがあ

る。そこで、検出時のバウンディングボックスの時点で正方形にする事で、変形時の画像の歪みをなくし、元々の画像と変わらない形で特徴を抽出する手法を提案する。正方形に画像を切り出すことで対象物体の周辺特徴も学習させることで、より高い精度で学習できると考えている。手順は、WSCDN が検出したバウンディングボックスの短辺と長辺を取得し、短辺の長さを、短辺の中心から均等に長辺の長さまで伸ばし正方形のバウンディングボックスにする。その際、画像からはみ出た部分はゼロパディングする。それによって歪みのない画像にする。

### 2.2.3 両モデルで分類

提案手法 1 では、WSCDN で物体を検出、IIC を用いた半教師あり学習型分類器で検出した画像を認識し、複数物体認識を行っていた。しかし、両モデルが得意な分類クラスや画像は異なると考えられる。さらに、WSCDN の方が IIC を用いた分類器より分類精度が高いクラスがある場合、そのクラスは WSCDN で認識した方がシステム全体の認識結果は向上する。そこで、WSCDN で高い分類スコアの画像のみを認識し、分類スコアが低い苦手な画像は IIC を用いた分類器で認識する両モデルで認識するハイブリッド手法を提案する。方法としては、本研究の実験では、分類スコアに閾値を設定し、WSCDN の分類スコアが閾値以上ならば、WSCDN で認識させ、閾値以下の画像を分類器で認識させる。

### 3 計算機シミュレーション

本章では、背景が白色トレイで設定されたパンの商品画像とパン画像より困難な課題である自然画像 PASCAL VOC 2007 を使用し、画像内の物体検出から物体認識までの工程を問題設定とする。提案手法 1 の検証は、3.1 節の提案手法 1 の性能検証で行う。2 つのモデルを組合せた物体認識のシステム全体の検証と Soft-NMS の有用性について検証する。また、提案手法 2 の検証は 3.2 節の提案手法 1 と 2 の比較検証で行う。D-NMS、入力画像の生成、両モデルで分類それぞれの有用性について検証する。比較手法として、代表的な教師あり学習型物体検出モデルである YOLOv3、弱教師あり学習モデルの代表的なモデルでベンチマークにもなる WSDDN(Weakly Supervised Deep Detection Network)、提案手法で物体検出器として利用した WSCDN を通常通り検出、分類器として使用する。

#### 3.1 提案手法 1 の性能検証

##### 3.1.1 問題設定

提案手法 1 の正解率を確認する。WSCDN は、パン画像と PASCAL VOC 2007 の訓練、評価画像集合で学習を行ったモデル、IIC を用いた分類器は、訓練、評価画像集合から切り出された物体画像で学習を行ったモデルを使用する。そして、複数の物体が写ったテスト画像を提案手法 1 に入力し、画像中の物体をどれだけ正しく分類できたか正解率を出す。各データセットの訓練、評価画像は、図 2,3 の左半分に、テスト画像は右半分にある。



図 2: パン画像の画像例



図 3: PASCAL VOC 2007 の画像例

##### 3.1.2 結果と考察

表 1 に各手法の正解率を示す。表 1 に示すように WSCDN と IIC を組み合わせることにより WSCDN や WSDDN より高い正解率は示せた。また、YOLOv3 には及ばなかったものの特にパン画像では、WSCDN 単体より大きく改善できた。このことから、WSCDN と IIC のモデルを組み合わせる有用性はあると考えられる。またパン画像では、2 つのモデルを組み合わせること、Soft-NMS の効果を確認できた。テスト画像にオクルージョンや物体同士の距離が近い状態が多くあった為、Soft-

NMS の効果が発揮されたと考えている。しかし、PASCAL VOC 2007 では、組み合わせた効果は得られなかった。元々 PASCAL VOC 2007 での分類性能が低い事が起因していると思われる。また、NMS の方が性能が良いという結果になり、自然画像では物体同士の距離が離れている事や同じ物体が重なった状態が少ない事が原因であると思われる。データセットの特徴をとらえ、NMS、Soft-NMS を使い分ければ、精度向上は見込められると思われる。

表 1: 物体認識モデルとの比較 (正解率 [%])

手法	パン画像	PASCAL VOC 2007
提案 1(NMS)	57.46	36.60
提案 1(Soft-NMS)	64.55	32.81
YOLOv3	72.20	77.64
WSCDN	54.07	54.07
WSDDN	13.11	4.807

#### 3.2 提案手法 1 と 2 の比較検証

##### 3.2.1 問題設定

次に各モデルに改善を加えることによる精度向上を検証した。2.2 章で説明した D-NMS、入力画像の生成、両モデルでの分類の有用性の検証し、精度向上が見られた両モデルで分類のみを使用した提案手法 2 と、提案手法 1 の結果を比較する。

##### 3.2.2 結果と考察

結果を表 2 に示す。パン画像では提案手法 1,2 で Soft-NMS が最も有用だと分かり、両モデルで得意なクラスが違った為、両モデルでの分類は有用であった。PASCAL VOC 2007 では、提案手法 1,2 で NMS が最も有用で、これは PASCAL VOC 2007 が自然画像のため離れた物体が多く、Soft-NMS の効果があまりないことに起因している。また、WSCDN の方が正解率が高い為、得意なクラスを重点的に分類することで全体の正解率は上昇すると思われる。

表 2: 提案手法 1,2 と物体認識モデルとの比較 (正解率 [%])

手法	パン画像	PASCAL VOC 2007
提案手法 1	64.55	36.60
提案手法 2	65.22	52.75

## 4 まとめと今後の展望

画像とラベルだけで商品进行分类するレジシステムの認識技術を開発を行った。提案手法では色情報を用いた画像処理を使用しない為、物体検出時に特殊な照明環境を必要としないメリットがある。また、物体同士の距離が近いパン画像には Soft-NMS は有用であることや、提案手法 2 は各モデルが得意な画像进行分类可能で両データで有用であることが分かった。精度としては 1 枚の商品画像を 1 人の客と想定した時、客 1 人に対し平均 65.22% の正解率でトレイ上のパンを正しく分類可能である。展望は、現状追加学習を行うことが出来ない為、新商品が出た際初めから全商品を学習する必要があり時間がかかる。そのため、オンライン機械学習を使用し追加学習をすることによって始めから学習し直す手間をなくしたいと思っている。

## 参考文献

- [1] Joseph Redmon and Ali Farhadi: "Yolov3: An incremental improvement", CoRR, *abs/1804.02767*,(2018)
- [2] J.Wang, J.Yao, Y.Zhang, and R.Zhang: "Collaborative learning for weakly supervised object detection", In IJCAI,(2018)
- [3] Ji.X.,Henriques,J.F.,Vedaldi,A.: "Invariant Information Clustering for Unsupervised Image Classification and Segmentation", In ICCV,(2019)
- [4] R. Girshick, J. Donahue, T. Darrell, and J. Malik. "Rich feature hierarchies for accurate object detection and semantic segmentation." In CVPR, (2014)
- [5] Bodla, N., Singh, B., Chellappa, R., Davis, L.S.: "Soft-nms improving object detection with one line of code." In ICCV. pp. 55625570. (2017)