

飲食店の天井カメラによる空き皿検出

細野 真衣[†] 中島 克人[‡]

東京電機大学未来科学部情報メディア学科^{†‡}

1 はじめに

国内の飲食店業界では、人件費の高騰やアルバイトの高い離職率に加え、COVID-19 感染対策のための消毒作業など、追加業務による人手不足が深刻である。そのため、限られた人員による効率的な店舗運営が求められる。そこで我々は店内の天井カメラにより利用客のテーブル上で空になった食器類を検出し、店員に通知するシステムを提案する。通知を受けた店員が空き皿回収のために客のテーブルに行く事は追加注文を受ける機会にもなり、売上向上に寄与し得ると考える。

本研究では、システムの一部とする空き皿の検出について、物体検出器に基づいた「皿の検出」と、その検出した皿が空き状態か否かの判別を行う「空き皿の判別」の2段階方式の空き皿検出器の手法を提案し、評価を行う。

2 関連研究

七五三掛らは、1段階方式による空きジョッキ検出器を提案している[1]。COCO dataset で事前学習済の YOLOv5x [2]と EfficientDet-D4, D5 それぞれに空きジョッキと中身在りジョッキの映ったシーン画像で追加学習し、空きジョッキの検出器を実装している。EfficientDet-D4, D5 と比較し、YOLOv5x は検出率が高く、適合率 100%、再現率 74%と、実用に耐えうる精度を実現した。

対象とするジョッキは国内で一般的にビール等に用いられる透明のもので、バリエーションは大きく無いため、空きジョッキ検出は1クラスで行われている。

3 提案手法

飲食店の店内の天井のカメラを用いた映像を入力し、物体検出で空き皿の検出を行う。図1に飲食店内カメラ映像の例を示す。

飲食店で使われる皿は多種多様であるため、空き皿検出は1段階・1クラスの検出器では困難と考えられる。そこで今回は、物体検出器 YOLOv5x に基づいた皿の検出と、その皿が空きか否かの判別を行う Siamese Network による空き皿判別の2段階方式とした。図2に処理の流れを示す。



図1 飲食店内カメラ映像の例

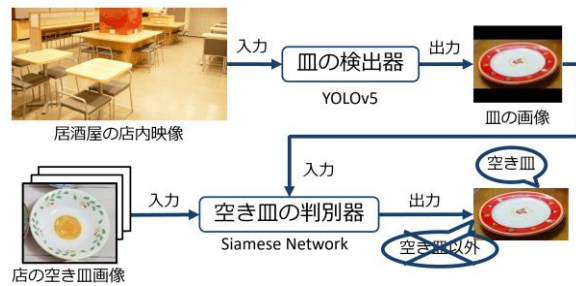


図2 空き皿検出処理の流れ

後者の空き皿判別器は、前者の検出器が検出した皿とその飲食店が使用する各皿とを順次比較し、空き皿かどうかを判別する。

4 実験と評価

4.1 皿検出器用のデータセット

皿が1枚以上映るシーン画像50枚を学習データの元画像とし、皿531枚にアノテーションを行った。元画像のうち、14枚は実際の居酒屋で撮影した広角のシーンで画像サイズ1920×1080px、36枚はインターネットから収集し、画像サイズは150×150~1920×1080pxである。訓練データは±25%の範囲の色相変換で増分した132枚と検証データ5枚に元画像を分割した。テストデータは136枚とし、黒と白の丸皿を茶色の机で撮影した動画から69枚と、4.2に示す空き皿判別器用のデータセットにおける皿検出器への入力画像の67枚である。

4.2 空き皿判別器のデータセット

合計17種類の皿が映る画像を皿検出器に入力し、検出した矩形に切り取って出力し、それらを「空き皿」と「中身有り皿」の2クラスに分け、空き皿判別器の学習の元画像とした。図3にラベル付けした画像例を示す。空き皿判別を実施する際の撮影時の皿の傾きを許容するため、1種類の皿に

Empty plate detection in pubs and restaurants using a ceiling camera

[†] Mai Hosono · Tokyo Denki University

[‡] Katsuto Nakajima · Tokyo Denki University



図3 皿データセットの例

対して、中身の有無が確認できる範囲で水平から垂直方向の異なる角度から撮影している。これらを 224×224 px にリサイズした上で、±30 度の範囲での回転と±80 の範囲でのチャンネルシフトを行って約 7 倍に増分した。「空き皿」が 56 枚、「中身有り皿」が 59 枚含まれる。

テストデータには、学習に用いた皿の内の 9 種類が映るシーン画像 1 アングル 22 枚を用いた。シーン画像サイズは 1,920×1,080 px であり、「空き皿」が 39 枚、「中身有り皿」が 25 枚含まれる。

4.3 皿検出器の実験結果

皿検出器には COCO dataset で事前学習済の YOLOv5 の 4 つのモデル x, l, m, s の中で低速度だが高精度な YOLOv5x を使用し、4.1 節のデータセットで追加学習を行う。学習パラメータは、入力サイズは 640×640 px, オプティマイザは Adam[3]とした。

皿検出の評価指標を表 1 に、確信度の閾値ごとの検出精度 (適合率 = TP / (TP+FP), 再現率 = TP / (TP+FN)) を表 2 に示す。

皿と誤検出されたもの (FP) も後段の空き皿判別で排除するため、ここでは再現率が 97.6% と十分高い閾値 0.25 を採用するのが妥当である。

表 1 皿検出の評価指標

True Positive (TP) 皿を「皿」と検出	False Positive (FP) 皿以外を「皿」と検出
False Negative (FN) 皿を「皿」と検出できず	True Negative (TN)

表 2 皿検出器の精度

閾値	適合率 (%)	再現率 (%)
0.80	98.1	74.7
0.40	86.3	91.8
0.25	80.4	97.6

4.4 空き皿判別器の実験結果

Siamese Network [4] は類似度を算出するため、他の深層学習ネットワークに比べ、比較的少ない学習データで高い判別精度を得ることが可能である。そのため、本研究では皿検出器で検出した皿の画像が空き皿か否かの判別を Siamese Network で行うこととした。画像の特徴抽出部には、ImageNet で事前学習済みの ResNet-50 [5] を用いた。ネットワークの構成図を図 4 に、また、空き皿判別の評価指標を表 3 に示す。

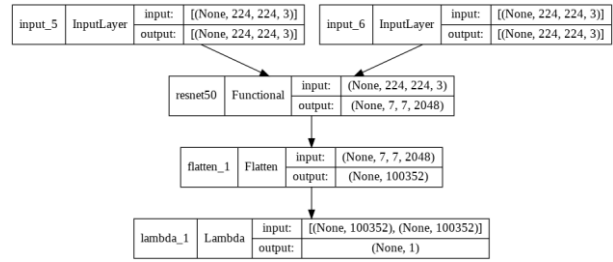


図 4 Siamese Network の構成図

表 3 空き皿検出の評価指標

True Positive (TP) 空き皿を「空き皿」と検出	False Positive (FP) 空き皿以外を「空き皿」と検出
False Negative (FN) 空き皿を「空き皿以外」と検出	True Negative (TN) 空き皿以外を「空き皿以外」と検出

4.2 のデータセットで学習と評価を行った結果、空き皿の識別精度は、テスト対象の学習済の皿 9 種それぞれに対して求め、全体で適合率は 16.8%、再現率は 79.5%、正解率は 72.0% となった。

適合率低下の原因となる誤検出には、良く似た色や形の別の空き皿を「空き皿」と判定した場合が含まれており、これを正解と見なす場合は、適合率は 59.2% であった。再現率低下の主な原因は、学習データには無いような皿の傾き、および、醤油で汚れていたり箸がのっているような空き皿の検出ミスであった。

5 考察とまとめ

本研究では、YOLOv5x を利用した皿検出器と Siamese Network を利用した空き皿判別器を作成し、2 段階方式で空き皿検出を行った。

システム全体の精度としては、評価指標は表 3 同じであり、4.4 節に示した数値に 4.3 節の「皿検出の再現率」を乗じたものとなる。

空き皿判別器の精度改善には前節に述べたような検出ミスを低減する必要があり、学習データの増強や Siamese Network の特徴抽出部の見直しなどを検討する必要がある。前者については、少ない撮影枚数であっても傾きや回転に関するバリエーションは画像処理で増強できよう。醤油などの汚れの付着等の許容は更なる検討が必要である。

参考文献

- [1] 七五三掛 豊和 他, “天井のカメラを用いた空きジョッキ検出の精度向上,” 第 83 回情処全大, 2021.
- [2] “YOLOv5,” <https://github.com/ultralytics/yolov5>, 2021 年 10 月 18 日参照.
- [3] D. P. Kingma, et al., “Adam: A Method for Stochastic Optimization,” arXiv:1412.6980, 2014.
- [4] G. Koch, et al., “Siamese Neural Networks for One-shot Image Recognition,” ICML deep learning workshop, 2015.
- [5] K. He, et al., “Deep Residual Learning for Image Recognition,” CVPR, 2016.