

飲食店の天井カメラによる空き皿検出手法の評価

細野 真衣[†] 中島 克人[‡]

東京電機大学大学院 未来科学研究科 情報メディア学専攻^{†‡}

1 はじめに

国内の飲食業界では、COVID-19 の影響による大幅な売り上げ低下の以前から、人員不足や人件費の高騰が問題として挙げられる。そこで我々は、居酒屋等での限られた人員による効率的な店舗運営を目的として、天井カメラ映像から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している。その一環として、以前に試作および評価を行った空き皿検出手法は、皿を検出後、その皿が空き皿かどうかを判定する 2 段階方式としていた。

本稿では、個々の空き皿を独立した別クラスとして検出する 1 段階方式の検出手法の試作を行い、2 段階方式とも比較評価を行ったので報告する。

2 関連研究

2.1 天井カメラによる空きジョッキ検出 [1]

七五三掛らは、1 段階方式による空きジョッキ検出手法を提案している。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手法を実装している。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の形状は、国内で一般的にビール等に用いられる透明で、個体差が大きく無いため、1 クラス検出であった。

2.2 天井カメラによる空き皿検出 [3]

飲食店で使われる皿は多種多様であるため、空き皿検出においては、我々は 1 クラスの検出手法では困難と考えた。そこで、物体検出手法 YOLOv5x に基づく皿の検出と、その皿が空き皿かの判定を行う Siamese Network [4] による空き皿判別の 2 段階方式とした。処理の流れを図 1 に示す。COCO データセットで事前学習済の YOLOv5x に対する皿検出のための追加学習には、居酒屋で撮影した動画とインターネットから収集した皿が写る画像の合計 531 枚分の画像を用いた。空き皿判別の学習に用いた

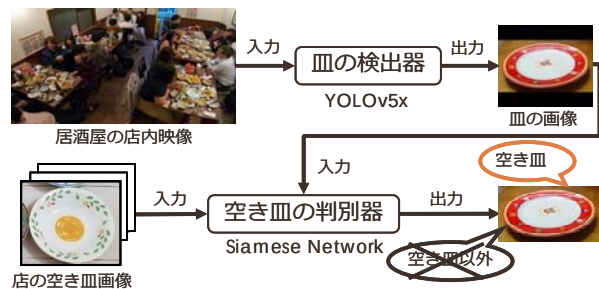


図 1 2 段階方式空き皿検出手法の処理の流れ

データセットは、17 種類の皿に限定し、皿 115 枚分の画像とした。

皿検出手法の適合率は 98.1%、再現率は 74.6%であった。空き皿判別器の適合率は 16.8%、再現率は 79.5%、正解率は 72.0%となった。皿検出手法は一定の精度を得られたが、空き皿判別器は提案システムの実現が困難な精度であった。

3 空き皿検出手法の評価

先行研究[3]の空き皿判別精度の問題は、天井カメラを想定して評価に使用したカメラが皿の柄が判別できる解像度を有していなかった事もあるが、学習に用いた画像が少ない点が大きいと考え、本研究では新たなデータセットを作成した。これにより、図 2 のような個々の空き皿を別クラスとして検出する 1 段階方式の学習も可能となった。そこで、検出対象の空き皿を同一にして、1 段階方式の空き皿検出手法と [3] の 2 段階方式の空き皿検出手法の精度比較を行う。なお今回は、前者および後者の 1 段階目は YOLOv7x[5] を検出手法として用いる。また、カトラリーが皿の上に載っている場合や、皿の内側全体に汚れが着色した場合は、空き皿と見なさないこととする。



図 2 1 段階方式空き皿検出手法の処理の流れ

3.1 1 段階方式用データセット

データセットのためのシーン画像撮影時の位置関係を図 3 に示す。一般的な飲食店の天井高の平均は 250 cm 以上とされており、テーブルの高さを

Evaluation of Empty Plate Detection
using a Ceiling Camera in Pub and Restaurant

[†] Mai Hosono · Tokyo Denki University

[‡] Katsuto Nakajima · Tokyo Denki University

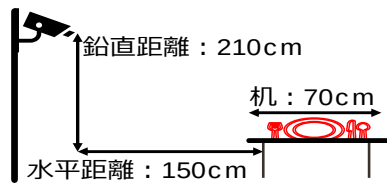


図3 撮影位置の略図

考慮して鉛直距離 210 cm, 水平距離 150 cm の位置から撮影を行った。1 画像当たり空き皿は 10~20 枚である。加えて、柄が鮮明に写り皿 1 枚のサイズが 500×500 px となる近距離からの撮影も行った。画像サイズは全て 1920×1080 px で、1 画像当たりの空き皿は 1~6 枚である。照明は白昼色と電球色の 2 パターンで、16.4~642 Lux である。撮影した 20 種類の皿は、全て空の状態、それぞれ形や色、大きさが異なる。

テスト画像は、同様の環境下で中身(料理)有を含む食事風景を撮影したものを用いる。

3.2 2段階方式用データセット

2段階方式の1段階目である皿検出器の学習用の元画像は、インターネットから収集した皿が1枚以上映る解像度 320×320~1920×1080 px のシーン画像 116 枚である。アノテーションは1画像当たり16皿程度である。ここから、検証画像15枚、テスト画像11枚を除いた90枚に対して $\pm 10^\circ$ 範囲の色相変換と $\pm 5^\circ$ 範囲の水平および垂直方向せん断でデータ増分した270枚を訓練画像とした。

2段階目である Siamese Network に基づく皿判別器の学習用の元画像は、3.1 節で説明した1段階方式データセットの元画像を、皿の矩形で切り取り、皿の種類ごとにラベルを付与したものである。

3.3 実験① 1段階方式と2段階方式の精度

2段階方式の2段階目の精度が、データセットによって改善されたか調査し、1段階方式とどちらの精度が良いか比較を行った。

空き皿として学習したのは色柄の異なる20種類の皿で、四角い皿も含む。図4はその一部を示す。その内、代表的な柄や形状として図4(a)~(e)の5



図4 検出対象の空き皿(20種類の内の一部)

表1 1段階方式と2段階方式の精度比較(%)

	適合率	再現率	正解率
1段階方式 5/20 種類	7.7	16.7	73.4
2段階方式 5/20 種類	1.5	33.4	71.6

種類に対しての検出精度を表1に示す。

なお、1段階方式において、皿の種類を無視した空き皿かどうかの判別では、適合率 38.5%, 再現率 83.4%, 正解率 85.0%であった。また、2段階方式における1段階目の皿自体の検出精度は、適合率は 92.2%, 再現率は 82.5%であった。

1段階方式も2段階方式も皿の種類を特定しての空き皿検出は十分な精度を得られておらず、優劣もつけ難い。皿の種類を無視した場合も適合率が満足いくものにはなっていない。

3.4 実験② 皿の柄の有無での1段階方式の精度

図4(d)~(i)のような皿の底の柄が、空き皿検出の精度に影響を及ぼすのかを調査するため、3.1 節のデータセットの20種類全ての空き皿と、その中の底柄無しの11種類の空き皿のみでそれぞれ1段階方式を学習し、精度を比較した。表2は、3.1 節の内、未学習の空き皿のみのシーン画像と、料理の載った皿を含むシーン画像でテストを行った結果を示す。なお、皿の種類を無視した空き皿判別では、それぞれ、適合率 47.9% / 84.2%, 再現率 64.8% / 45.7%, 正解率 77.0% / 83.8%であった。

底柄無しの皿に限定した場合は一定の精度向上があるが、実用に供する場合は、適合率が 84% ではまだ不十分であると言える。

表2 皿の底柄の有無による精度比較(%)

	適合率	再現率	正解率
1段階方式 20 種類 (空き皿のみ / 料理有含む)	98.0 / 25.0	100.0 / 38.0	- / 66.5
1段階方式 11 種類 (空き皿のみ / 料理有含む)	91.5 / 41.7	87.8 / 31.3	- / 75.0

4 まとめ

底柄無しの皿だけを対象としても、料理の有無の判定精度はまだまだ不十分な結果となった。主な理由は、空き皿判別のための学習用画像に負例として料理が載った皿を全く含めなかったことと考えられる。また、天井カメラからの俯瞰角度といった撮影条件の影響も精査が必要である。今後も学習データの充実などを図り、精度向上を目指す予定である。

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,” 情報処理学会第83回全国大会講演論文集, 2021.
- [2] G. Jocher, “YOLOv5,” <https://github.com/ultralytics/yolov5>, 2021年10月18日参照.
- [3] 細野 真衣 他, “飲食店の天井カメラによる空き皿検出,” 情報処理学会第84回全国大会講演論文集, 2022.
- [4] G. Koch, et al., “Siamese Neural Networks for One-shot Image Recognition,” ICML deep learning workshop, 2015.
- [5] Chien-Yao Wang, et al., “YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” arXiv preprint arXiv:2207.02696, 2022.