2N-06

天井のカメラを用いた空きジョッキ検出の精度向上

七五三掛　豊和†　　中島　克人‡

東京電機大学†‡

1 はじめに

国内の飲食店業界では人手不足が深刻な問題となっている．コロナ禍の影響による景気悪化のため，少ない人員のままで効率よく店舗運営する必要もある．そうした中で利益を保つには，サービスレベルを維持したまま回転率を向上する必要がある．例えば居酒屋では食器類の回収と飲料等の追加受注の効率化が望まれる．そこで我々は居酒屋等の店舗の天井カメラを用いて，テーブル上の空いた食器類を検出し，店員に通知する支援システムを提案している．

かつて我々は一般物体検出器と分類器を用いた 2 段階式による空きジョッキ検出を試みたが，検出精度が不十分であった[1].

今回は 1 段階式の空きジョッキ検出器で精度向上を図ると共に，1 台のカメラでより多くのテーブルをカバーするために広角カメラへの対応も行う．

2 先行研究

2 段階方式の[1]では 1 段階目の検出器で検出したジョッキ画像を 2 段階目の分類器を使って空きジョッキかどうか分類した．2 段階目の分類精度は十分に高かったが，1 段階目のジョッキ検出部は他の食器とのオクルージョンや広角撮像時のシーン両端に見られるジョッキの傾きが原因で，ジョッキ検出の再現率が低く，結果として，空きジョッキの検出精度は満足の行くものではなかった．

3 提案手法

3.1 1 段階式の空きジョッキ検出

今回は 1 段階式による検出器をジョッキ画像により End-to-end の学習で作成し，オクルージョンと広角レンズによるジョッキの傾きにも頑健な空きジョッキ検出器の実現を目指す．

なお，検出ミスの原因解析を容易にするために，空きジョッキと中身在りジョッキの 2 クラス検出器として実現するが，評価では空きジョッキだけを対象とする．

3.2 YOLOv5 を使用した空きジョッキ検出器

今回は YOLOv5[2]と EfficientDet[3]の 2 種類を使用して学習し精度比較を行う．

YOLOv5 は検出と識別を同時に行う検出器で，比較的高速で，検出精度も高い．YOLOv5 には COCO dataset で学習済みの x, l, m, s の 4 つのモデルが用意されている．今回は空きジョッキの検出精度向上を図るため，速度は遅いが精度を最も重視したモデルである YOLOv5x に対して，独自のデータセットで追加学習を行う．

EfficientDet は複数の解像度での特徴マップを合わせる BiFPN によって高い検出精度を誇りつつ，パラメータ数を抑えたネットワークである．今回は EfficientDet-D4 と D5 を使用して精度比較を行う．

3.3 ジョッキデータセットの元画像

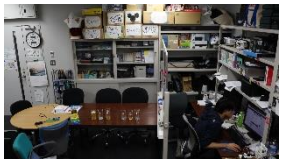
空きジョッキと中身在りジョッキが 1 個以上映るシーン画像を 352 枚用意して学習データの元画像とする．元画像には図 1 に示すような画像に加え，実際の居酒屋で GoPro HERO7 の SuperView モードで撮影した広角のシーン画像も含めている．後者の画像に映るジョッキは，複雑なオクルージョンと大きな傾きが発生している．その際のアノテーションでは，人間の目で判断してジョッキが空いているか否かがはっきりとわかる物だけを対象としている．オクルージョン発生時のアノテーション例を図 2 に示す．



(a) 白い背景の画像



(b) 図 1(a)の別角度画像



(c) 茶色い机での撮影画像



(d)別の卓上撮影画像

図 1　学習に使用した画像例



(a) 手前に物体がある状態の画像例



(b) 図 2(a)のアノテーション例

図 2　重なったジョッキのアノテーションの例

Accuracy Improvement of Empty Mug Detection using Ceiling Cameras
†Toyokazu Shimekake ・Tokyo Denki University
‡Katsuto Nakajima・Tokyo Denki University

Copyright ©2021 Information Processing Society of Japan.
All Rights Reserved.

3.4 データ拡張後のデータセット

学習に使用するアノテーション済みの元画像各 352 枚に対して輝度を±30%の範囲で，画像回転を±5 度の範囲でランダムかつ同時に変動させた．適用後の画像は 1,056 枚となり，このデータセットを用いて検出精度向上を試みる．

4 評価

4.1 評価指標とハイパーパラメータ

YOLOv5x と EfficientDet-D4, -D5 を使用した空きジョッキの検出精度を比較する．評価指標を表 1 に示す．この指標から空きジョッキ検出の適合率と再現率を求め，各モデルの精度比較を行う．なお，シーン内で空きジョッキ以外を空きジョッキとして検出していないことを表す True Negative は，空きジョッキが存在しない場所が数えきれないため，評価対象とはしない．

表 1　空きジョッキ検出器の評価指標

True Positive (TP)	False Positive (FP)
入力シーン内の空きジョッキを，空きジョッキとして検出	入力シーン内の空きジョッキ以外を，空きジョッキとして検出
False Negative (FN)	True Negative (TN)
入力シーン内の空きジョッキを非検出	

4.2 学習環境とハイパーパラメータ

学習には CPU に Intel(R)Xeon(R)の 1 コアを,GPU に NVIDIA Tesla P100 を使用した．入力解像度は YOLOv5x では 960px×960px の 3ch，EfficientDet-D4 は 1024px×1024px の 3ch，EfficientDet-D5 において 1280px×1280px の 3ch とし，各モデルで学習を行う．各学習のパラメータを表 2 に示す．学習ではデータセットを 7:2:1 の比率で分割し，訓練データを 739 枚(1056×0.7)，検証データ 211 枚(1056×0.2)，テストデータ 105 枚(1056×0.1)とした．

表 2 各モデルのハイパーパラメータ

モデル名	Step	Batch	学習率	Optimizer
YOLOv5x	4,440	5	1e-2	SGD[4]
EfficientDet-D4	18,500	4	1e-4	AdamW[5]
EfficientDet-D5	6,150	4	1e-4	AdamW

4.3 広角レンズを用いた食事シーンでの評価

GoPro HERO7 の SuperView モードを使用して撮影した食事シーンを評価に使用した．動画の長さは 10.3 秒（309 フレーム）で 1 フレームずつ検出結果の目視確認を行った．

表 3 に示す結果から YOLOv5x は全体的に検出精度が高く，空きジョッキ自体の検出率が高いことがわかる．一方で，EfficientDet-D4 では再現率が先行研究手法と同等の低さであり，Efficient-D5 では再現率は高くなるものの，適合率が低いことから，空きジョッキとして他の物体を誤検出することが多いことがわかる．

表 3 1 段階式空きジョッキ検出器の適合率，再現率(%)

モデル	適合率	再現率
YOLOv5x	100.00	73.65
EfficientDet-D4	100.00	24.47
EfficientDet-D5	55.06	43.62
先行研究手法	14.94	20.39

5 考察

店員に空きジョッキを通知する支援システムという面では，完璧な再現率を必要としないため，YOLOv5x の再現率は実用に耐え得ると考えるが，EfficientDet のそれは低すぎる．一方，わずかな誤報(False Positive) も店舗運営には障害であるため，完璧な適合率が求められるが，YOLOv5x はその期待にも応える結果となった．

YOLOv5x と EfficientDet の大きな性能差は今回行った学習用のジョッキ画像に対する輝度調整および回転によるデータ拡張と，各ネットワークの特徴抽出部分の相性によると考えられる．

6 まとめと今後の課題

我々は COCO dataset で事前学習済の YOLOv5x と EfficientDet-D4, -D5 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，2 クラス検出器を実装・評価した．シーン画像 352 枚から輝度および回転のデータ拡張を行った 1,056 枚のデータセットによる学習の結果，空きジョッキ検出に関し YOLOv5x が最も高く，かつ，実用に耐えうる適合率と再現率を得る事ができた．

今回はジョッキ画像に回転を加えて学習を行い，ある程度のジョッキの傾きに対応できたが，広角映像の左右端では更に大きい傾きが生じ得るため，入力映像の両端の歪を補正して検出器に入力することも今後の検討課題である．また，ジョッキ以外の食器の空き検出も今後の課題である．

参考文献

[1] 七五三掛 豊和 他, “飲食店における天井のカメラを用いた空きジョッキ検出,” 第 82 回情処全大, 2020
[2] “YOLOv5,” https://github.com/ultralytics/yolov5
[3] M.Tan,et al.,“EfficientDet: Scalable and efficient object detection,” In CVPR, 2020
[4] Duchi, et al., "Adaptive subgradient methods for online learning and stochastic optimization," In Journal of Machine Learning Research 12.Jul: 2121-2159, 2011
[5] L. Ilya, et al.,"Decoupled weight decay regularization," In arXiv:1711.05101, 2017

Copyright ©2021 Information Processing Society of Japan.
All Rights Reserved.