

IoT デバイスのための深層学習を用いた自動販売機の購入商品の識別

Vending Machine Drink Recognition with Deep Learning for IoT Device.

新屋勝啓*¹
Katsuhiro Araya

長隆之*²
Takayuki Osa

山岡成光*²
Shigemitsu Yamaoka

西和彦*²
Kazuhiko Nishi

中尾政之*²
Masayuki Nakao

*¹ 株式会社 Arkth
Arkth, Inc.

*² 東京大学大学院工学系研究科
University of Tokyo, School of Engineering

In recent years, Internet of Things (IoT) with deep learning is used for various purposes. The balance of edge computing, data traffic, cloud computing and data volume is important for IoT. In this study, we present a framework that automatically recognizes bottles purchased from a vending machine in camera images with a convolutional neural network (CNN).

1. はじめに

近年 IoT と深層学習の連携が様々な用途や分野で活用され始めている。しかし現状ではデバイス、ネットワーク、深層学習による認識率など様々な面で課題があり、多くの研究者によって多岐にわたり研究されている。例えば一般的に IoT デバイスは大量に利用されるため、なるべく安価に構築する必要がある。しかし、安価なデバイスではスペックが低くなるため、計算量の大きい分析については制限がかかってしまう。そのため、深層学習による計算量を削減するための前処理やアルゴリズムの調整、ハードウェアの構成の改善などを行う必要がある。さらに、IoT デバイスが増えるとネットワークへの負荷も高くなるため、なるべく少ない通信コストで動作させる必要がある。サーバーに蓄積するデータも IoT デバイスの量に比例して増加するためデバイス側でデータ圧縮や分析を行いたい。このように IoT と深層学習の連携では、エッジ AI コンピューティング、ネットワークの転送量、クラウドでのデータ蓄積と分析の強調を、認識率やデバイスの価格などとの兼ね合いを鑑みてどのように設定するかが重要である。

IoT と深層学習を連携したシステムの中でも、本研究では近年増加している自動販売機での購入者や購入商品の分析に着目してシステムを構築する。

現状の自動販売機の購入動向の分析システムでは店舗や機器に改修を加えたり入れ替えたりする必要があるため、導入コストが高い課題がある。本研究で構築するシステムは既存の機器の内部に改修を加えることなく、安価で外付けできることを特徴として開発を進めている。

本稿ではそのシステムの中でも特にエッジ AI コンピューティングによる飲料の分類に着目し、安価で低スペックなデバイスと深層学習を用いて飲料自動販売機に設置したカメラから撮影された動画像中の購入商品を認識することを目的としている。

2. 本システムの全容

本システムでは購入者の年齢・性別と購入商品を認識すると共に、気象センサによる気象情報を取得してクラウドに保存する。ネットワークには LoRaWAN を利用することを前提としているため、通信するデータは飲料 ID や年齢、性別フラグ、気温など最低限となるようにエッジで分類を行う。

また、本システムでは購入者にリアルタイムに広告表示などのインタラクションを行うため、ディスプレイや表示用のデバイス

連絡先: 新屋勝啓, 株式会社 Arkth, info@arkth.co.jp

を設置し、クラウドから最新のデータを取得できるようにする。図 1 に本システムのシステム構成を示す。

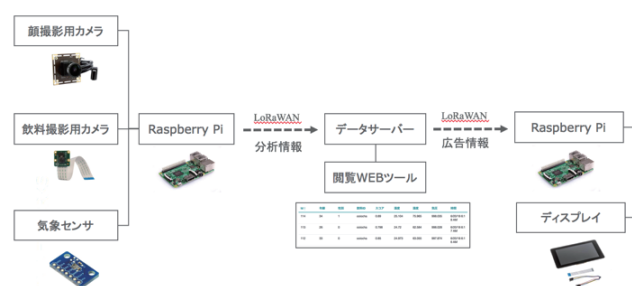


図 1 システム全容の構成図

3. 飲料の認識

3.1 デバイスについて

図 1 左側のセンサー情報の取得や、飲料の分類、サーバーへのデータ転送には Raspberry Pi 3 Model B+ を利用した。気象センサに BME280 を利用し、温度・湿度・気温の 3 つの環境情報を取得できる。カメラは購入者の撮影と飲料の撮影用にそれぞれ設置し、購入者撮影用には広角に撮影できるカメラを利用した。

3.2 飲料の認識

本研究では、深層学習を用いた物体検出技術を用いる。ニューラルネットワークは Single Shot Multibox Detector (SSD) を採用し [Liu 16], 深層学習のフレームワークとして TensorFlow を利用した。SSD はニューラルネットワークの構造に VGG16 [Simonyan 14] を用いた手法であり、入力画像が小さくても高い精度の物体検出が可能である。

3.3 認識の前処理

自動販売機の取り出し口に向けて設置したカメラから SSD によって飲料を認識する際に、背景を飲料に誤認識するフレームが多く発生した (図 2)。

そこで本手法では、認識範囲のトリミングを行うと共に、背景差分によって取り出し動作の検出を行った。

具体的には 320x240 の画像を用いた場合、取り出し口付近の背景差分を用いて取得した動きのあったピクセル数が一定のしきい値を超えた場合に飲料の認識を行うようにした。これにより認識を軽減すると共に、認識を行うフレーム数を削減できる。

背景差分を取得する範囲も図 2 に矩形で示す。

3.4 学習データのデータオーギュメンテーション

自動販売機に設置したカメラから取得される画像は、設置場所や時間、気象条件によって明るさが様々に変化する(図 3)。そこで、学習データに対してデータ拡張を行った。データの拡張はノイズ追加、コントラスト調整、ガンマ変換による明るさ調整、

拡大縮小、反転を行った。データオーギュメンテーションを行うことで、限られた数の教師データからでも明るさなどの変化にロバストな認識器を構築することができる。



図 2 背景の誤認識と背景差分取得範囲

4. 実験

4.1 自動販売機とカメラの設置

本実験に利用する自動販売機は、大学構内の屋根のある屋外に設置した。飲料認識用のカメラは図 3 のように、自動販売機前面に取り出し口付近を撮影するように設置した。カメラは Raspberry Pi 用のカメラモジュールを用いた。



図 3 飲料認識用カメラの設置

動画撮影時のフレームレートは設定やプログラム上では 20FPS と設定しているが、デバイスのスペックによる制限によって実際は 7-8FPS となった。

4.2 学習データの取得

19種類の飲料を対象に自動販売機に取り出し口に向けたカメラから、手に持った飲料の画像を収集した。

背景による影響を考慮し、過学習を避けるために屋内の白い背景で取得した学習データも追加し学習を行った。



図 4 学習データ例

学習データは各飲料で 100 枚ずつ程度取得した。

その後データオーギュメンテーションを用いて各画像を 13 枚変換し合計 26600 枚を学習データとして利用した。

4.3 実験結果

飲料が写っている画像を 19 種類の飲料ごとに 100 枚ずつ検証に利用した。その結果を下の表にまとめる。

合計画像数	正認識数	誤認識数	未検出数
1900	1651(87%)	137(7%)	112(6%)

飲料ごとの認識率では、緑茶や烏龍茶などのお茶系の認識率が低く、一方で缶コーヒーなどの缶飲料の認識率が高い傾向があった。これはお茶系の飲料は外観の色味が似ている物が多いのに対し、缶コーヒーなどはパッケージの色彩などの差異が大きいためと考えられる。

また、1枚の画像から飲料を認識するのに、Raspberry Pi 3B+では 10 秒程度を要した。

5. 結果の考察

上記の通り本実験では 85%を超える認識率を達成した。しかし、本システムの狙いは様々な環境や種類の自動販売機に設置して、エッジ AI コンピューティングにより分析を行うことである。

本実験では1箇所に設置した自動販売機で同じ季節に撮影したデータが多いため、様々な環境に設置した自動販売機で認識を行った場合よりも高い認識率が出ていると予想される。

また、自動販売機の飲料は無数に存在すると共に、商品ラインナップの入れ替えや商品ラベルの変更も数多く発生するため環境の変化だけでなく商品の多様性にも対応する必要がある。本実験は深層学習やデバイスの実装に加え、学習データの取得に多くの手間がかかった点も今後の課題である。

市販されている飲料はメーカーにラベルやボトルのデータがあるため、今後は Zero shot Learning や Few Shot Learning[Sachin 17]の模索も重要となる。

6. まとめ

本稿では IoT デバイスによる飲料認識の手法を提案し、実験を行った。その結果、限定的な環境であれば安価なエッジデバイスで高い認識率を実現できた。今後、更に多くの環境や飲料で検証を進めると共に学習データを集め、ハイパーパラメータのチューニングや本手法で用いた以外のデータオーギュメンテーション、Few Shot Learning などの学習手法の研究を進める。

また、深層学習による画像認識は顔や動物などの認識についての先行研究が多く、飲料に特化した文献はあまり見られなかった。しかし、今後の IoT の拡大につれてあらゆるものやことを認識する需要が増えてくると予想できるため、様々な物体検出に応用できる手法を目指していく。

謝辞

本研究に利用した自動販売機は株式会社伊藤園の協力により設置いただきました。心より感謝申し上げます。

参考文献

[Simonyan 14] Karen Simonyan and Andrew Zisserman: Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556*, 2014.

[Liu 16] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A. C.: Ssd: Single shot multibox detector, in European conference on computer vision, pp. 21-37Springer (2016)