

AI を用いた学内のコンビニ店における 弁当の残数を提供するシステムの開発

池本健太郎^{*1}・熊澤典良^{*1}・奈良大作^{*1}・上谷俊平^{*1}

Email: kumazawa@mech.kagoshima-u.ac.jp

*1: 鹿児島大学大学院 理工学研究科

©Key Words AI, IoT, ニューラルネットワーク, 可視化, 弁当の残数

1. はじめに

鹿児島大学郡元キャンパスでは複数の食堂とコンビニ店舗により飲食が提供されている。昼休みの時間帯になると複数ある店舗のうち食堂に利用者が集中する傾向があり非常に混雑する。先行研究^{1,2}において食堂及びコンビニ店の待ち時間を利用者に提供するシステムが開発されている。特に(2)の研究において、複数の店舗の待ち時間を一目で把握できることで利用者の店舗選択は容易となったが、待ち時間が短いからとコンビニ店に訪れても弁当の在庫がない場合があることが分かった。

本研究ではこれまでの店舗の待ち時間に加えて弁当の在庫状況を利用者に提供することにより、利用者の店舗選択に寄与することを目的とする。売り場の弁当の個数をリアルタイムに導出し、待ち時間と併せて弁当の残数を提供することで利便性は高まるはずである。本論文では陳列された弁当の残数をAIによってリアルタイムに導出し、その情報をインターネット及び店舗前に設置したモニターにて提供するシステムについて述べる。

2. 弁当の残数の導出

先行研究によって提供される情報は食堂及びコンビニ店(ミニショップ)の待ち時間であり、図1に示すように2店舗の情報が同時に提供される。図の左側には食堂において利用者が食事を受け取るまでの待ち時間が提供され、右側にはコンビニ店において利用者が食事を手に取り会計レジにたどりつくまで待ち時間が表示される。これらの待ち時間はAIによって1分ごとに導出され、両店舗前に設置しているモニターで提供されている。提供される待ち時間の情報をもとに利用者は店舗を選択するが、弁当の完売によって無駄足になることも見受けられた。本研究では、この待ち時間の情報に加え、コンビニ店の弁当の残数も表示することで利用者の店舗選択の利便性の向上を図る。

本研究の対象店舗であるコンビニ店では図2に示すように、店舗中央のワゴンで弁当が提供される。図中の実線で囲う位置には弁当が積み上げられて陳列され、四つのワゴンには最大100個以上の弁当が並ぶ。弁当の配置は決まっておらず、日や時間によってレイアウトは変わる。四つのワゴンにある弁当の数を導出するためには、弁当の種類や重なりを考慮する必要がある。

2.1 物体認識による弁当の検出

本研究では、はじめに画像内から特定の特徴をもった範囲を導出することができる物体認識のAIによって、積み上げられた一番上の弁当の数を導出する。弁当の数



図1 提供している待ち時間の情報



図2 店内の様子



図3 アノテーションされた弁当の例

を物体認識によって導出するためにCNNを使用する。CNNを使用する上で必要となるデータセットは、図3のようにアノテーションを行うことで作成した。枠で囲われている部分が弁当の特徴として学習するデータの範囲である。陳列されている弁当の種類は複数でありその弁当のサイズや色も違うが、弁当の蓋はどの弁当も同じような特徴をもつため本研究の物体認識では蓋の部分をアノテーションする。データセットの作成に用いた画像の枚数は500枚であり、約2500個の弁当の蓋を弁当としてアノ

テーションした。作成したデータセットをもとに、TensorFlow[®] ライブラリを使用して学習用データに対して学習を行った。30,000 回の学習による学習用データの認識精度はほぼ 100%であり、検証用に別途用意したテストデータに対する認識精度も 98%だったため、適切な学習がなされたと判断した。

学習により作成した物体認識モデルは適切に弁当の数を検出するが、それは最も上部の弁当の数であり積み上げられた弁当の下部の弁当までは把握できない。さらに 4 つの陳列ワゴンのうち奥側のワゴン上の弁当も、画像解像度及び店内ポップの影響で適切な検出が行えないことがある。物体認識では把握できない箇所を補うために、本研究では他の検出方法を併用することとし、画像分類を行う CNN および背景差分法による画像処理の結果をニューラルネットワークによって物体認識の推定結果に補正をかけることを考えた。

2.2 画像分類による在庫状況の分類

カメラの画像からおおよその弁当の数を把握するために本研究では弁当の在庫状況を 5 段階に分類する。その 5 段階は「在庫なし、10 個未満、10 個以上 50 個未満、50 個以上 90 個未満、90 個以上」である。在庫なしもしくは 10 個未満の弁当が少ないという情報は利用者にとって有益な情報である。10 個以上の残数は細かく分類してもそれほど需要はないと判断したため 40 個ずつの分類とした。5 段階の在庫状況の画像に加えて、利用者の重なりの影響で弁当の在庫状況を把握できない画像もデータセットに加え、計 6 パターンの状況を学習させる。6 パターンの分類を図 4 に示しており、学習する画像は図 2 の画像をトリミングした画像となっている。図 4 に示すような 6 パターンの画像を複数用意し、それらの特徴を学習することで入力画像がどのパターンの画像に近い状況であるかを推定する。学習にはそれぞれのパターンの画像と画像に応じた情報を付随したデータセットを必要とする。作成したデータセットの内訳を表 1 に示している。在庫なしの状況は教師用のデータを 100 枚、テスト用のデータを 25 枚用意し、その他の 5 パターンの在庫状況では教師用のデータを 400 枚、テスト用のデータを 100 枚用意した。学習に使用したデータ数は合計で教師用が 2100 枚、テスト用が 525 枚である。弁当の在庫がある状況に比べて全くない状況はほぼ同じ映像であったためデータ数を少なくした。

本研究の画像分類に使用した 3 層の CNN (畳み込みニューラルネットワーク)⁴⁾ の構造を図 5 に示す。入力層には使用する画像データが当てはまり、入力した画像データを畳み込み層に入れることで特徴を際立たせることができる。プーリング層では画像の詳細な情報を消し、画像内の細かいずれやノイズに対処しやすくなる。局所領域を一つのピクセルとして計算するため元の画像よりデータが小さくなり計算量を大幅に減らすことができ、畳み込み層とプーリング層を複数回繰り返すことで精度が向上する仕組みである。繰り返しを抜けた全結合層では隣接する層のニューロンをすべて連結し出力の数を定めることができる。今回は画像の種類を 6 つに分類するため出力の数は 6 となる。畳み込み層や全結合層の後にある活性化関数とはあるニューロンから次のニューロン



図 4 在庫状況の分類

表 1 データセットの内訳

在庫状況	データ数	
	教師用	テスト用
在庫なし	100	25
10 個未満	400	100
10 個以上 50 個未満	400	100
50 個以上 90 個未満	400	100
90 個以上	400	100
判定不可	400	100
合計	2100	525

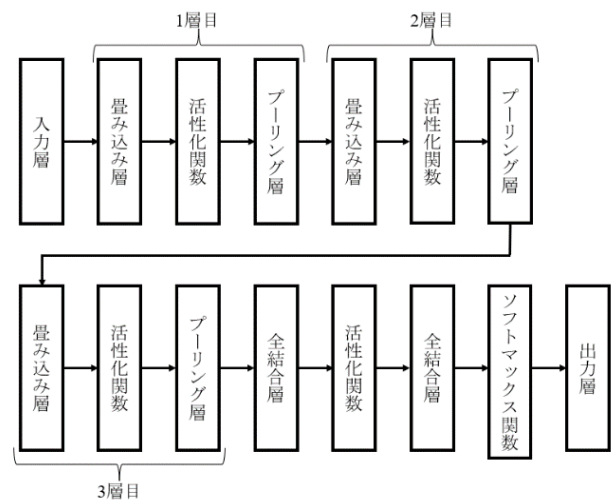


図 5 CNN の構造

へ出力する際に別の数値に変換する関数であり、活性化関数により学習効率は高まる。

表 2 に中間層は 3 層、入力画像はカラー画像、各中間層のニューロンの数は 32 に設計した CNN によって作成したモデルに対するテスト用のデータセットに対する推定結果を示す。縦の列は入力画像を示しており、横の行は推定された結果の割合を示している。0~4 の数値は 5 段階に分類した在庫状況を表し、×は判定不可であることを表している。下線で示す数値は入力画像が正しく推定された割合である。正しく推定される確率はどの種類の入力画像においても 90%以上の結果を得られた。

表2 画像分類の推定結果

		推定結果					
		0	1	2	3	4	×
入力画像	0	<u>92%</u>	8%	0%	0%	0%	0%
	1	1%	<u>96%</u>	2%	0%	0%	0%
	2	0%	5%	<u>92%</u>	3%	0%	0%
	3	0%	0%	4%	<u>94%</u>	2%	0%
	4	0%	0%	0%	10%	<u>90%</u>	0%
	×	0%	0%	2%	4%	2%	<u>92%</u>

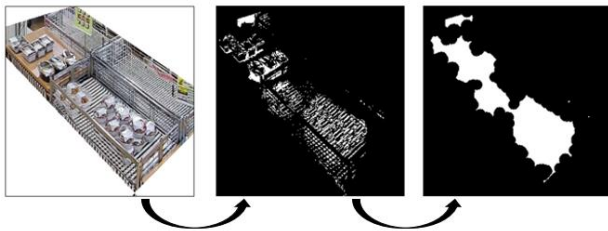


図6 画像処理の様子

2.3 画像内に占める弁当の面積抽出

ここでは入力画像内に占める弁当の映る面積について考える。弁当の占める面積を把握できることで、弁当の積みあがっている部分および奥のワゴンの物体認識では把握できない箇所を補正する。入力画像における弁当の占める面積を画像処理によって導出することでその面積と実際の弁当の数の関係を学習し、弁当の数の補正に役立てる。

画像内の弁当の面積を抽出するために背景差分法およびモルフォロジー演算を使用した。まずワゴン上に弁当が陳列されていないときの画像と現在の画像をグレースケール化し、それぞれの画素値の差分をとることで弁当の映るピクセルのみを抽出する。つぎに、弁当のみに注目し余分な情報を無視するために弁当を陳列しているワゴン以外の箇所はマスク処理を行い、背景差分処理を行った。背景差分処理では大まかに弁当のあるピクセルを把握できるが、処理の精度は粗く面積を導出するにはさらに画像処理を行わなければならないため、さらにモルフォロジー演算を使用して処理を行う。モルフォロジー演算によってあるピクセルとその近傍のピクセルを比較して変換することでノイズを除去し、弁当の面積を明確に抽出する。この画像処理全体の流れを図6に示す。左図はワゴン以外をマスク処理した画像、中央の図は背景差分処理後の画像、右図はモルフォロジー演算による処理後の画像である。左図と右図を比較すると、一連の流れによって弁当の面積を導出することを可能とすることが分かる。処理後の画像における画素値が255(白)のピクセル数を10段階の指標で表し、その結果をニューラルネットワークの入力として利用する。ピクセル数を10段階に分割することでおよそその弁当の数と紐づけることができ、ニューラルネットワークの入力として扱うことで精度の向上が期待できる。

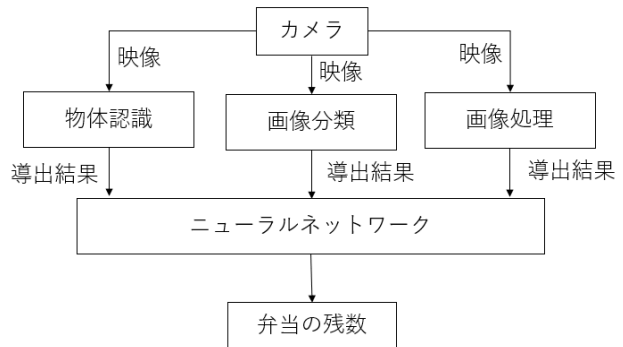


図7 弁当の残数導出のフローチャート

表3 テストデータでのモデルの精度 (単位: 個)

モデルの種類	平均誤差	誤差の標準偏差
C+P	1.98	4.87
R+C	2.85	7.21
R+P	2.50	6.68
R+C+P	0.25	3.83

3. 残数の推定実験

物体認識の結果に補正をかけて弁当の残数を推定するために、画像分類の結果および画像内に占める弁当の面積を使用し、ニューラルネットワーク作成する。本研究ではより精度の高いモデルの作成を目指して4パターンのモデルを作成し比較した。4パターンのモデルはそれぞれ3種類の入力値「物体認識の結果 (R), 画像分類の結果 (C), 画像内に占める弁当の面積 (P)」の組み合わせを変えて作成した。本研究ではC+P, R+C, R+P, R+C+Pの4種類のモデルを作成しそれぞれ精度の確認を行った。作成したモデルを使用した弁当の残数の導出のフローチャートを図7に示す。はじめにカメラ映像を取得し、取得した画像を用いてそれぞれの処理を行う。処理によって導出された結果をニューラルネットワークに入力することで弁当の残数を導出する。

3.1 予備実験

300パターンのデータを用いて学習を行い、学習したモデルを用いて100パターンのテストデータで評価を行った。作成した4パターンのモデルの平均誤差と誤差の標準偏差を示したものを表3に示す。表よりテストデータでの検証ではどのモデルも非常に誤差は小さくなったが、入力値に画像分類の結果、画像内に占める弁当の面積、物体認識の結果の全てをいれたR+C+Pのモデルの精度が最も高くなる結果を得た。

3.2 実証実験

実装して問題ないかを確認するために、ある一日の弁当の残数と作成したモデルの推定結果を比較した。比較した様子を図8に示す。上図はある一日の実際の弁当の残数の推移を示しており、横軸は時間、縦軸は弁当の残数を表している。図より10時過ぎに弁当は並べられ、14時半ごろに弁当は完売したタイミングで補充されていることが分かる。下図はR+PおよびR+C+Pのモデルによって推定された弁当の残数と実際の残数との誤差を表して

いる。横軸は時間、縦軸は誤差を表しており、実線はR+Pの誤差、点線はR+C+Pの誤差である。誤差を比較すると予備実験とは異なってR+PよりもR+C+Pの方が誤差が大きくなった。この原因はR+C+Pで使用している画像分類の実証実験に対する精度が80%程度に減少して、誤判定を起こしていることが原因である。この誤判定は使用した教師用データの数が少なく様々なパターンを網羅しきれていないことと、陳列のレイアウト変更及び想定外の弁当の残り方に対応しきれていないからである。

画像分類を使用していないR+Pでは一日の最大誤差は弁当が100個以上ある時に生じた20個であり、この結果は提供に支障はないと判断した。よって物体認識の結果と画像内に占める弁当の面積を入力としたニューラルネットワークによって導出した弁当の残数を利用者に提供する。

4. 情報提供

本研究で構築したシステムは3.2節の実験より、図7に示す画像分類を除いた物体認識(R)および画像処理(P)を使用する。はじめにカメラ映像を取得し、物体認識をAIコンピュータ(Jetson Xavier NX)によって行い、その他の処理をPC(CPU: Intel Core i7-10510U, メモリ: 16GB)によって行う。それぞれの直近の導出結果をニューラルネットワークの入力とし、毎秒弁当の残数の推定を行う。推定された結果を用いて提供する画像を作成してWebサーバ上にアップロードし、店舗前に設置されたモニタにて情報を提供する。

利用者への情報提供は先行研究によって作成された図1の情報に弁当の残数の情報を加えた図9の画像をモニタに表示することで可能にする。本研究で導出したコンビニ店の弁当の残数の情報を待ち時間の下部に追加表示している。右側はニューラルネットワークで推定された1個単位で表示される弁当の残数となっており、左側に5段階の顔文字で弁当の残数をしている。顔文字はニューラルネットワークによって導出した弁当の残数をもとに5段階に分割した結果を表示している。一番左のマークは在庫なし、左から二番目のマークは10個未満、中央のマークは10個以上50個未満、右から二番目のマークは50個以上90個未満、一番右のマークは90個以上在庫がある状況となっており、現在の状況に当てはまるマークは大きく表示されている。顔文字を利用することで画像を見て瞬時に大体の弁当の残数を把握することが可能となっている。本研究によって利用者は昼食を摂る店舗を選択する際に待ち時間の情報に加え、コンビニ店の弁当の残数の情報も加味することが可能となった。

5. おわりに

本研究では、コンビニ店のカメラの映像を使用し物体認識のAIおよび背景差分処理によって弁当の在庫状況を把握し、それらの情報に加え画像処理によって得られた弁当の面積の情報をもとにニューラルネットワークによって弁当の残数を導出するシステムを開発した。複数のモデルを検証した結果、予想に反して物体認識の結果お

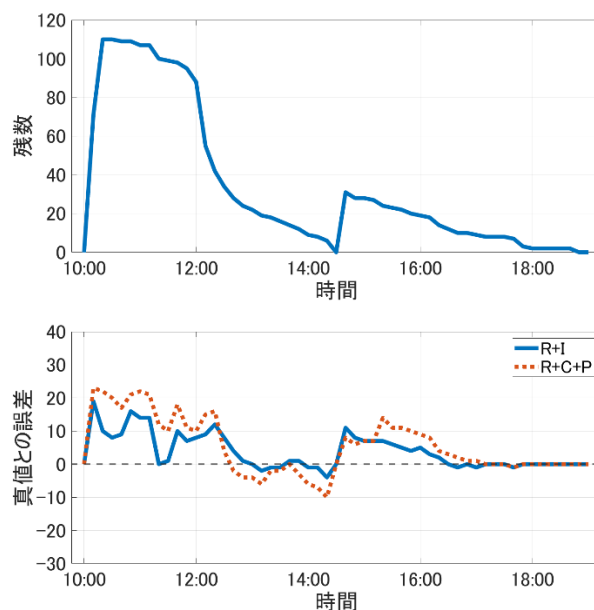


図8 実験結果



図9 提供する情報

よび弁当の面積の情報をニューラルネットワークの入力にしたモデルが最も精度の高いモデルとなった。導出した弁当の残数の情報を待ち時間と併せて店舗前のモニタに表示することで利用者にその情報を提供した。今まで、待ち時間が短いからとコンビニ店に訪れても弁当の在庫がないという状況が生じていたが、このシステムによってそのような状況を改善することができた。

参考文献

- (1) 吉野陽, 熊澤典良, 奈良大作, 上谷俊平: “AIを用いた学生食堂内の待ち時間を見える化するシステムの開発”, 2021 PCカンファレンス論文集, pp.260-261, (2021).
- (2) 池本健太郎, 吉野陽, 熊澤典良, 奈良大作, 上谷俊平: “学生食堂の店舗選択に寄与する混雑情報システムの開発”, 2021 PCカンファレンス論文集, pp.262-263, (2021).
- (3) TensorFlow: “TensorFlow”, <https://www.tensorflow.org/>, 2022年5月10日アクセス.
- (4) 斎藤康毅, “ゼロから作るDeep Learning”, pp.66-71, オライリー・ジャパン, (2016).