

# 深層学習を用いた画像からの食品質量推定

1235073 三浦 康寛 【 知能情報学研究室 】

## Food Mass Estimation from Images Using Deep Neural Networks

1235073 Yasuhiro Miura 【 Intelligent Informatics Lab. 】

### 1 はじめに

食品加工業界ではベルトコンベア上の食品の計量の必要があるが、秤による計測速度の問題があった。そこで、画像認識を用いた計量による高速計量の研究が必要である。関連するものとして、画像中の群衆の人数を推定する Crowd Counting[1] などがあり、学習済み畳み込みニューラルネット (CNN) である VGG16, ResNet50, Inception\_v3 が用いられている。そこで本研究では、食品を対象に CNN を用いた画像からの質量推定を提案する。

### 2 提案手法

本研究では、学習済み CNN モデルを用いた食品画像からの食品質量の推定を提案する。用いる学習済みモデルは、Xception, VGG16, ResNet50, Inception\_v3 の4つで、これらは ImageNet データセットによって事前学習されている。実験では、これらを用いて推定モデルを構築する。推定モデルに食品画像を入力し、質量の正解データを与え学習させることで食品質量を推定する。

### 3 実験

#### 3.1 計量推定モデル

本研究では、学習済み CNN モデルをバックボーンとし、全結合ブロック (FC block) を再構築し、出力層のノード数を1に設定して、回帰モデルを構築する。今回回帰モデルはバックボーンの ImageNet の重みは固定し、FC block のみで学習される。VGG16 を例とし図1に推定モデルを示す。

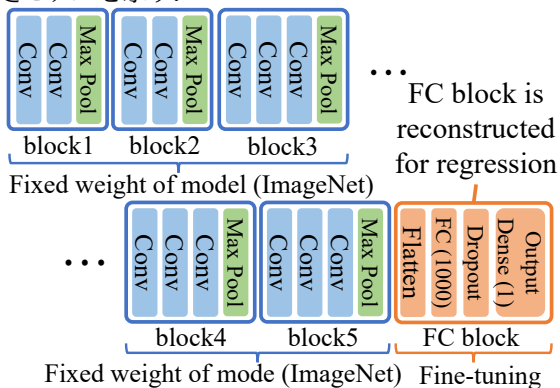


図1 推定モデルの構成

#### 3.2 食品画像データセット

データセットは、3種類あり、1つ目は、白のベルト上にカットしたニンジンで撒いたデータセットである (Carrot White データセット: CW). 2つ目は、青のベルト上にカットしたニンジンで撒いたデータセットである (Carrot Blue データセット: CB). 3つ目は、青のベルト上にカットしたタマネギで撒いたデータセットである (Onion Blue データセット: OB). 画像サイズは 570 × 330 と設定する。画像データ数はそれぞれ 3784, 2236, 2688 である。図2にデータセットを示す。これらの食品は共同研究企業との関係で選定された。

#### 3.3 実装詳細

それぞれの推定モデルの学習では、損失関数に平均二乗誤差 (MSE), 最適化手法に Adam を用いる。学習率は事前学習によって決定する。性能評価の手法は、平均絶対パーセント誤差 (MAPE) を用いる。最終的な性能評価は5分割の交差検証の平均で算出する。

#### 3.4 結果

MAPE は5%以下であれば計量システムの許容範囲とされている。全てのデータセットで VGG16 が最も性能が高いモデルとなった (表1)。CW データセットで、MAPE が 10.9, CB データセットで、15.8, OB データセットで、16.1 であった。

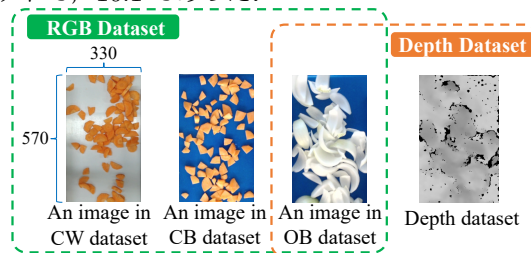


図2 食品画像データセット

表1 4つの学習済みモデルの推定質量のパーセント誤差

Model	CW dataset MAPE(%)	CB dataset MAPE(%)	OB dataset MAPE(%)
Xception	63.5	38.1	90.8
VGG16	<b>10.9</b>	<b>15.8</b>	<b>16.1</b>
ResNet50	79.0	70.3	113.2
Inception_v3	54.3	52.5	77.8

表2 Fine-tuning の適用範囲を変えた5つのモデルの推定質量のパーセント誤差

Model	CW dataset MAPE(%)	CB dataset MAPE(%)	OB dataset MAPE(%)
FT-b5	6.1	7.8	9.5
FT-b54	5.6	<b>6.0</b>	9.3
FT-b543	5.8	6.2	<b>8.4</b>
FT-b5432	<b>5.4</b>	6.5	8.7
Without-IN	5.7	6.2	8.5

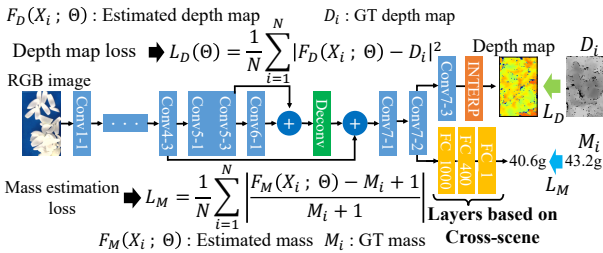


図3 Depth-model

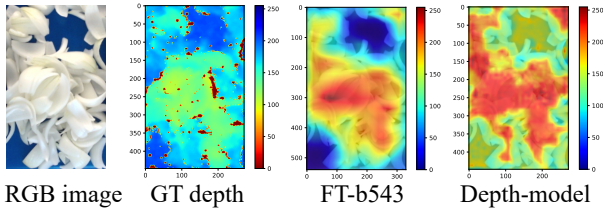


図4 Grad-CAM 適用結果

### 3.5 Fine-tuning 適用範囲変更

VGG16 モデルの構成次第では更なる性能改善の余地がある。Crowd Counting の研究では、Fine-tuning の適用範囲の変更によって性能が向上した報告がある。そこで、VGG16 の Fine-tuning の適用範囲を変えた5つのモデルを性能評価する。それぞれのモデルは、FC block, block5 を Fine-tuning するモデル (FT-b5), FC block, block5, block4 を Fine-tuning するモデル (FT-b54), FC block, block5~block3 を Fine-tuning するモデル (FT-b543), FC block, block5~block2 を Fine-tuning するモデル (FT-b5432), ImageNet の重みを用いず1から学習させるモデル (Without-IN) である。表2の実験結果から、CW データセットでは、FT-b5432 が最も高い性能で MAPE が 5.4 であった。CB データセットでは、FT-b54 が最も高い性能で MAPE が 6.0 であった。OB データセットでは、FT-b543 が最も高い性能で MAPE が 8.4 であった。全てのデータセットの実験結果の平均を取ると FT-b543 が最も高い性能となった。

## 4 深度情報を用いた推定モデル

3.5 節では、VGG16 の FC block~Block3 を Fine-tuning した推定モデル (FT-b543) が全てのモデルの中で最も高い性能となった。しかし、この推定モデルは入力が食品画像で出力が食品質量という単純なモデルであるため、新たな情報を学習する推定モデルで、更なる精度向上が考えられる。多くの Crowd Counting の研究で

は、群衆画像を CNN モデルで密度マップを推定し、密度マップ上の積分から群衆の人数を推定している。そして、このような無数の人間が写った画像からその密度を推定するアプローチは高い性能である。そのため、このアプローチは多くの物体の画像からその量を推定するタスクに有効であると考えられ、画像からの質量推定への応用に期待ができる。また、深度カメラで物体の厚みを計測することで質量推定の精度向上ができることから、密度マップ推定モデルで深度マップ推定を行うことで、更なる精度向上ができると考える。そこで、追加実験では食品深度画像データセットを用いた CNN による食品質量推定を提案し食品の質量と深度を学習する CNN モデルを構築する。図3に推定モデル (Depth-model) を示す。本推定モデルは、Cross-scene と SaCNN[1] を基に構築されたマルチタスクのモデルである。そのため、質量の誤差を計算する  $L_D$  と深度マップの誤差を計算する  $L_M$  ロス関数を用いる (図3)。データセットには、図2の深度画像データセットを用いる。また、本推定モデルは、深度情報の有効性を検証するためのものであるため、深度情報を用いない FT-b543 と性能を比較する。実験の結果、Depth-model は MAPE が 4.1 で、FT-b543 が MAPE が 3.0 となった。

## 5 考察

本研究では、精度向上を目的として深度を学習する CNN モデル (Depth-model) を構築し質量推定を行なったが、VGG16 の FC block block3 を Fine-tuning したモデル (FT-b543) が全てのモデルの中で最も高い性能を示した。そこで、Grad-CAM を用い FT-b543 と Depth-model の画像データの注目領域の可視化を行う。図4はその結果である。教師データの深度画像と FT-b543 の結果を比較すると、Depth-model と同様に、高密度の部分と低密度の部分の識別を行なっていることが分かる。全ての実験から、VGG16 やそれを Fine-tuning したモデルは、様々なモデルの中で最も性能の高い特徴抽出器であり、質量や人数といった「物の量」の推定タスクに適していると考えられる。また、VGG16 を Fine-tuning したモデルは、図4で表されるように、深度情報が教師データとしてなくとも、画像中の食品の密度や深度に着目し質量を推定していると考えられる。

## 6 まとめ

本研究では、秤を用いない高速計量・低導入コストの計量システムの実現を目的とし、CNN を用いた食品画像からの食品質量推定を行なった。実験の結果、VGG16 の FC block~block3 を Fine-tuning したモデル (FT-b543) が最も高い性能となった。また、Grad-CAM を適用した結果、FT-b543 は、画像中の食品の密度や深度に着目し質量推定をしている可能性が示唆される。

## 参考文献

[1] L. Zhang, et al.: WACV, pp. 1113–1121, 2018.