CONFIDENTIAL EXTENDED ABSTRACT.
DO NOT DISTRIBUTE ANYWHERE.

CNNによる料理検出とカロリー量推定のマルチタスク学習

會下 拓実[1,a)]　柳井 啓司[1,b)]

概要

近年, 画像認識技術を用いた食事記録支援アプリケーションが数多く登場している. そうしたアプリは料理写真から料理名を推定することが可能であるが、今のところ複数料理画像に対応したものは少なく, また, 料理画像からの全自動カロリー量推定が可能なものは存在しない. そこで本研究では CNN を用い, 複数料理画像からのカロリー量推定を行う. 単一の CNN での料理検出とカロリー量推定のマルチタスク学習を行うことで, 複数料理画像から各単品料理を検出すると共に各単品料理のカロリー量の推定を行う. 単一ネットワークでの同時推定による高速化と省メモリ化が期待できる.

1. はじめに

近年, 食に関する健康志向が高まる中で, 日々の食事管理を目的としてカロリー量を記録することができるアプリケーションが登場している. しかしこれらのアプリケーションのカロリー量の推定には, ユーザー入力が必要であったり, 栄養士を雇ったりと, 人手のかかるものとなっている. 最近では画像認識により料理画像から料理名の候補を自動で提案するものも存在するが, 単品料理の画像にのみ対応している場合が多く, 図 1 のように複数の料理が並ぶ場合, ユーザーは 1 品ずつ写真を撮るか, 画像から手作業で単品料理の画像を切り出す必要があり, 手間がかかるものとなっている.

一方, 画像認識分野では CNN を用いた手法により, 画像のクラス分類や物体検出を高精度に行うことが可能となっている. これを料理画像に用いることで, 料理カテゴリ分類や複数料理画像からの単品料理の検出が可能であり, 一部のアプリケーションでは実用化されている.

そこで本研究では CNN を用い, 複数料理画像からのカロリー量推定を行う. 単一の CNN での料理検出とカロリー量推定のマルチタスク学習を行うことで, 複数料理画像から各単品料理を検出すると同時に各単品料理のカロリー量の推定を行う.

會下ら [2] は CNN での回帰学習により, 料理画像からカロリー量を直接推定するネットワークを構築した. また, カロリー量情報付き料理画像データセット [2] を作成し, これを用いてネットワークを学習することで, 料理画像からのカロリー量推定を実現した. この手法では食材や調理手順の違いによる見た目の違いを考慮したカロリー量推定が期待できる. ただし, このネットワークの入力は単品料理の画像にのみ対応しており, 複数料理画像から個々の料理のカロリー量を推定することはできない. したがって本研究では, 物体検出技術を用いることで複数料理画像からのカロリー量推定を実現する. なお, 本研究のネットワークが出力するカロリー量の値は, 會下ら [2] と同様に 1 人分のカロリー量の値となっている.

図 1　複数料理画像の例

物体検出は画像に含まれる各対象物体の矩形領域とカテゴリを推定する技術である. この物体検出に関しても, CNN を用いた手法により高精度かつ高速な検出が可能となっており, 本研究では CNN に基づく物体検出手法を用い, 複数料理画像からの単品料理の検出を行う. また, 単品料理を検出すると同時にカロリー量を推定するネットワークを構築する. 一般的な物体検出を用いて複数料理画像から単品料理を検出する場合, 画像に含まれる各単品料理の矩形領域とカテゴリが推定されるが, 本研究ではこれに加えて各単品料理のカロリー量を推定することで, 料理検出とカロリー量推定を同時に行う.

まとめると, 本研究では複数料理画像からのカロリー量推定を行う. 単一 CNN での料理検出とカロリー量推定のマルチタスク学習を行うことで, 複数料理画像からのカロリー量の推定を実現する. 現状, 物体検出のためのバウンディングボックスとカロリー量の両方が各単品料理に付与された複数料理画像のデータセットは存在しないため, 本研究では, バウンディングボックスが付与された複数料理

[1] 電気通信大学大学院情報理工学研究科
[a)] ege-t@mm.inf.uec.ac.jp
[b)] yanai@cs.uec.ac.jp

画像とカロリー量が付与された単品料理画像を CNN の学習に用いる.

2. 関連研究

現在では様々な自動カロリー量推定システムが提案されている. 宮崎ら [4] は料理画像からカロリー量を直接推定している. 色ヒストグラムや SURF などの低レベル特徴量に基づいて, データベース上の類似画像を検索し, 特徴量ごとに類似度の高い上位 n 枚のカロリー量の平均値を計算し, それらの値から最終的にカロリー量を推定している. データセットには Web サービスである FoodLog に投稿された料理画像 6512 枚を使用している. このデータセットには複数料理画像も含まれており, 各料理画像には 1 人分のカロリー量がアノテーションされている. したがってこの手法では料理の量に応じたカロリー量を推定することはできない. 本研究のカロリー量もこれと同様で料理画像からカロリー量を直接推定するものである.

CNN を用い複数料理画像から料理を検出した研究の一つに下田らの研究 [8] がある. 下田らはまず, selective search により大量に候補領域を生成し, 次に, CNN を用いて得られる各候補領域のサリエンシーマップに基づき, 料理領域の領域分割を行っている. 最後に non- maximum suppression(NMS) により重複している候補領域を統合し, 単品料理を検出している. この手法は料理領域の領域分割を行っているが, 領域分割はピクセル単位のカテゴリ分類であるため, ピクセル単位の領域を矩形領域で表すことで物体検出を行っている. これに加え下田らは, 候補領域の生成にも CNN を使用する手法 [9] を提案している. この手法では CNN を用いて得られるサリエンシーマップにより候補領域を直接生成する. 次に各候補領域のクラス分類を行い, 最後に NMS により重複している候補領域を統合し, 料理を検出している.

Dehais ら [1] は料理領域の領域分割の手法を提案している. 彼らはまず, CNN を用いて, おおまかな料理領域の境界線を表す Border Map を生成する. そして region growing/merging algorithm を用いて Border Map の境界線を洗練することで料理領域を推定する. 本研究では CNN に基づく物体検出手法を用い, 複数料理画像からの単品料理の検出を行う.

料理画像からのカロリー量推定を行った研究の一つに Myers らの Im2Calories [5] がある. 彼らはカロリー量を推定するために, 料理写真に含まれる食材の種類やその領域などを推定し, これらの情報から推定された体積と, 推定された料理カテゴリに対応するカロリー密度から最終的なカロリー量を計算している. ただしこの研究の実験では, カロリー量がアノテーションされたデータセットが不足し, 十分な性能評価が行われていない.

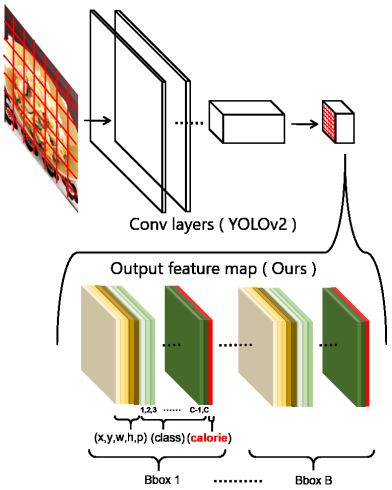

図 2 YOLOv2 [7] と本手法の出力特徴マップの概略図.

3. 手法

ここではまず本研究で提案する料理検出とカロリー量推定を同時に行うネットワークについて説明する.

3.1 料理検出とカロリー量推定のマルチタスク学習

本研究では単一 CNN での料理検出とカロリー量推定のマルチタスク学習を行い, 料理検出とカロリー量推定を同時に行うネットワークを構築する. すなわち, 複数料理画像中の各単品料理の矩形領域とカテゴリ, カロリー量の同時推定を行う. 本手法ではカロリー量推定手法として, 會下らの CNN での回帰学習による料理画像からのカロリー量の直接推定手法 [2] を用い, 物体検出手法として, Redmon らが提案した CNN に基づく物体検出手法である YOLOv2 [7] を使用する. YOLOv2 は以前に提案された YOLO [6] を改善することで, 高速かつ高精度な物体検出を達成している。

図 2 に YOLOv2 のネットワークを示す. YOLOv2 は画像を入力として, 出力は特徴マップであり, 全層が畳み込み層で構成されているため, 出力まで位置情報が保持されている. そのため出力特徴マップの各ピクセルは入力画像上のある領域に対応しており, ピクセル毎に対象物体の矩形領域とカテゴリが推定される. 出力特徴マップの幅と高さを S とすると, 入力画像上の $S \times S$ のグリッド毎に対象物体の矩形領域とカテゴリが推定される. 矩形領域とカテゴリの推定時には, 矩形領域を表す中心座標と幅・高さと, グリッド内に対象物体が存在する確率, そして各カテゴリに対応するクラス確率が出力される. したがってグリッド毎に B 個の物体を検出するとして, カテゴリ数を C とすると, 出力特徴マップのチャネル数は $B \times (5 + C)$ となる.

本研究では料理検出とカロリー量推定のマルチタスク学習を行うネットワークを提案する. 本手法では図 2 のように YOLOv2 の出力特徴マップにカロリー量の出力に対応するチャネルを追加することで, 矩形領域とカテゴリに加えてカロリー量の推定を行う. したがって, 本手法の出力

図 3 擬似的なバウンディングが付与された料理画像の例. 赤枠が正解値として擬似的に付与されたバウンディングである. (左上: ピラフ 375kcal, 右上: 肉じゃが 262kcal, 左下: スパゲッティ 391kcal, 右下: ハンバーグ 440kcal)

特徴マップのチャネル数は $B \times (1+5+C)$ となる.

このようにグリッド毎に推定を行うためには, 正解値として付与されたバウンディングボックスに応じて正解グリッドを設定し, その正解グリッドに関して物体検出とカロリー量推定が適切に行われるように学習する必要がある. しかし, 本実験においてカロリー量推定の学習に使用するカロリー量付き料理画像データセット [2] にはバウンディングボックスが付与されていない. そこで本研究では, 次の手順により図 3 のように擬似的にバンディングボックスを付与することで YOLOv2 の学習を行う. まず, カロリー量付き料理画像をランダムな位置に埋め込み, 埋め込んだ画像領域を正解バウンディングボックスとする. さらに背景との境界線を目立たなくするために, 埋め込んだ画像と同じものを反転させながら背景部分に埋め込む.

3.2 料理画像からのカロリー量推定

本研究では CNN での回帰学習による料理画像からのカロリー量の直接推定手法 [2] を用い, 料理検出と同時にカロリー量推定を行う. 會下ら [2] が提案したネットワークでは, 入力は単品料理画像に限られており, 出力のカロリー量は料理画像中の料理の量に関係なく 1 人分の量に対応するカロリー量を出力する. 本手法のネットワークは複数料理画像にも対応しているが, 出力されるカロリー量の値は [2] と同様に 1 人分の量に対応する値である. また, 本手法では [2] に従い, カロリー量の学習に使用する損失関数として式 (1) を使用する.

回帰問題においては, 一般的に損失関数として 2 乗和誤差が用いられるが, [2] では次のような損失関数 L_{cal} を使用している. 絶対誤差を L_{ab}, 相対誤差を L_{re} とすると, カロリー量推定タスクの損失関数 L_{cal} は下のように定義される.

$$L_{cal} = \lambda_{re} L_{re} + \lambda_{ab} L_{ab} \tag{1}$$

ただし λ は各損失にかかる重みである. 絶対誤差は推定値と正解値の差の絶対値であり, 相対誤差は絶対誤差と正解値の比である. ある画像 x を入力したときの推定値を y, y に対する正解値を g とすると, 絶対誤差 L_{ab} と相対誤差 L_{re} は下のように定義される.

$$L_{ab} = |y - g| \tag{2}$$

$$L_{re} = \frac{|y - g|}{g} \tag{3}$$

4. データセット

現状, バウンディングボックスとカロリー量の両方が各単品料理に付与された複数料理画像のデータセットは存在しないため, 本研究では, バウンディングボックスが付与された複数料理画像と, カロリー量が付与された単品料理画像を用いて CNN の学習を行う. バウンディングボックスが付与された複数料理画像として UEC Food-100 [3] を使用し, カロリー量が付与された単品料理画像として [2] で用いられたデータセットを使用し, 単一のネットワークの学習を行う.

UEC FOOD-100 [3] は料理 100 カテゴリに関する料理画像データセットであり, 複数料理画像も含まれている. 単品料理画像は各料理カテゴリ 100 枚以上あり, 合計 11566 枚ある. それに加え, 複数料理画像は 1174 枚ある. 合わせて 12740 枚のすべての画像にバウンディングボックスが付与されている.

本研究で使用するカロリー量付き料理画像 [2] は Web 上のカロリー量情報を提供するレシピ情報サイトから収集されたものであり, 単品の料理画像に 1 人分のカロリー量が付与されている.

5. 実験

本実験ではバウンディングボックスが付与された複数料理画像として UEC Food-100 [3] を使用し, カロリー量が付与された単品料理画像として [2] で用いられたデータセットを使用し, 単一のネットワークの学習を行う. 学習時にはバッチ単位で使用するデータセットを切り替えることで, 料理検出タスクの学習とカロリー量推定タスクの学習を交互に行う. 料理検出タスクの学習には UEC Food-100 [3] を使用し, 損失関数として料理検出タスクに関連する損失項のみを用いる. カロリー量推定タスクの学習にはカロリー量付き料理画像 [2] を使用し, 損失関数としてカロリー量推定タスクに関連する損失項のみを用いる.

5.1 単品料理画像からのカロリー量推定

本実験の評価にはカロリー量が付与された単品料理画像データセット [2] のテスト用画像を使用する. 會下ら [2] に基づき, 評価指標としてカロリー量推定の評価指標として相対誤差, 絶対誤差, 相対誤差 20% 以内の推定値の割合, 相対誤差 40%以内の推定値の割合を用いる. 絶対誤差は推定値と正解値の差の絶対値であり, 相対誤差は正解値に対する絶対誤差の割合である.

最適化手法として SGD を使用し, momentum 値を 0.9 する. また, バッチサイズを 8 として学習率 10^{-5} において

表 1　単品料理画像 [2] に対するカロリー量推定の結果

	相対誤差 (%)	絶対誤差 (kcal)	誤差 20%(%) ≤	誤差 40%(%) ≤
VGG16 でのカロリー量推定 ([2])	30.2	105.7	43	76
料理検出+カロリー量推定 (本手法)	36.1	121.7	34	64

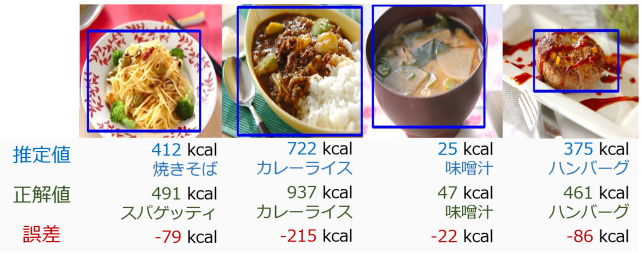

図 4　単品料理画像 [2] に対する料理検出とカロリー量推定の例．青枠が推定されたバウンディングである．

40,000 回反復し，その後，学習率 10^{-6} において 20,000 回反復する．

本実験で使用するテスト画像は単品料理画像であるため，検出されたバウンディングボックスのうちグリッド内に対象物体が存在する確率が最も高いバウンディングボックスを最終出力とする．表 1 にカロリー量推定の結果を示す．

VGG16 [10] を用いてカロリー量推定のみを行った手法 [2] と比較すると，本手法の精度は全ての評価指標に関して下回った．原因として，擬似的なアノテーションを行ったことが考えられる．図 4 に料理検出とカロリー量推定の例を示す．

また，複数料理画像に対する料理検出とカロリー量推定の例を図 5 に示す．テスト画像として「実物大・そのまんま料理カード 食事バランスガイド編[*1]」の料理カードを用いる．この「実物大・そのまんま料理カード 食事バランスガイド編」は実物大の料理カード 131 枚を有し，各料理カードには材料や作り方，カロリー量の値などの情報が付与されている．

6. おわりに

本研究では単一の CNN での料理検出とカロリー量推定のマルチタスク学習を行うことで，複数料理画像からのカロリー量の推定を行った．また、本実験ではバウンディングボックスが付与された複数料理画像として UEC Food-100 [3] を使用し，カロリー量が付与された単品料理画像として [2] で用いられたデータセットを使用し，単一のネットワークの学習を行った．

今後の課題としてカロリー量付き複数料理画像データセットの作成がある．バウンディングボックス付き料理画像とカロリー量付き料理画像を使用し，CNN を学習することで，新たに生成することなどを考えている．

[*1] http://www.gun-yosha.com/book/balanceguide.html

図 5　複数料理画像に対する料理検出とカロリー量推定の例．青枠が推定されたバウンディングである．(ES: 推定値, GT: 正解値)

謝辞: 本研究は科研費 (17H01745) の助成を受けたものである．

参考文献

[1] J. Dehais, M. Anthimopoulos, and S. Mougiakakou. Food image segmentation for dietary assessment. In *Proc. of ACM MM Workshop on Multimedia Assisted Dietary Management*, 2016.

[2] T. Ege and K. Yanai. Simultaneous estimation of food categories and calories with multi-task cnn. In *Proc. of IAPR International Conference on Machine Vision Applications(MVA)*, 2017.

[3] Y. Matsuda, H. Hajime, and K. Yanai. Recognition of multiple-food images by detecting candidate regions. In *Proc. of IEEE International Conference on Multimedia and Expo*, pages 25–30, 2012.

[4] T. Miyazaki, G. Chaminda, D. Silva, and K. Aizawa. Image - based calorie content estimation for dietary assessment. In *Proc. of IEEE ISM Workshop on Multimedia for Cooking and Eating Activities*, pages 363–368, 2011.

[5] A. Myers, N. Johnston, V. Rathod, A. Korattikara, A. Gorban, N. Silberman, S. Guadarrama, G. Papandreou, J. Huang, and P. K. Murphy. Im2calories: towards an automated mobile vision food diary. In *Proc. of IEEE International Conference on Computer Vision*, pages 1233–1241, 2015.

[6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You Only Look Once: Unified, real-time object detection. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2016.

[7] J. Redmon and A. Farhadi. YOLO9000: Better, faster, stronger. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2017.

[8] W. Shimoda and K. Yanai. CNN-based food image segmentation without pixel-wise annotation. In *Proc. of IAPR International Conference on Image Analysis and Processing*, 2015.

[9] W. Shimoda and K. Yanai. Foodness proposal for multiple food detection by training of single food images. In *Proc. of ACM MM Workshop on Multimedia Assisted Dietary Management*, 2016.

[10] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *arXiv preprint arXiv:1409.1556*, 2014.