Multi-task CNN を用いた 食材および調理手順情報を利用した食事画像カロリー量推定

會下　拓実[1,a)]　柳井　啓司[1,b)]

1. はじめに

近年, 健康志向の高まりにより様々な食事管理アプリケーションがリリースされ, 栄養学の知識のない人がカロリー量を記録することが可能となっている. しかしそれらは複数の操作をユーザに要求し, リアルタイム性に欠けるものが多いため, より簡便な食事記録の方法が求められている. 一方, 画像認識分野においては Deep Convolutional Neural Network(CNN) の登場以来, 精度が飛躍的に向上しており, CNN による食事画像の認識に関する研究も盛んに行われている. そこで本研究では CNN による食事画像からのカロリー量推定を行う.

カロリー量は食事カテゴリによって大きく異なり, さらにサイズや食材, 調理手順に依存しており, それらの情報は図 1 のように完成した料理の見た目に現れる. 人はそうした情報を認識することで, カロリー量の正確な値を推定することはできないまでも, “カロリー量が高いのはどちらの画像の料理か” という問いには答えられる場合が多いと考えられる. 一方, 画像認識分野では CNN により飛躍的に認識精度が向上し, ILSVRC の 1000 種類分類タスクでは人の認識精度に匹敵するほどの性能を示している. そこで本研究では CNN を用いて食事画像からのカロリー量推定を行う. 入力は食事画像のみとし, 料理の見た目を直接反映したカロリー量推定を行う. さらに Multi-task CNN [1] によりカロリー量に加えて食事カテゴリや食材, 調理手順の情報を同時に学習することで精度の向上を図る. カロリー量と食事カテゴリのような相関のある情報を同時に学習した場合, 同時に行うタスクに共通する特徴が学習され, 各タスクの精度が向上することが知られている. さらに本研究では, Web 上のレシピ情報サイトからカロリー量情報付き食事画像データを収集することで, データセット作成のコストを削減する. なお，本研究では，MVA2017 で発表予定の [2] で提案するカロリー情報推定手法を拡張し，食材情報および調理手順情報を追加して利用することによってさらなる精度向上が実現することを示すことを目的とする.

1　電気通信大学情報理工学部総合情報学
a)　ege-t@mm.inf.uec.ac.jp
b)　yanai@cs.uec.ac.jp

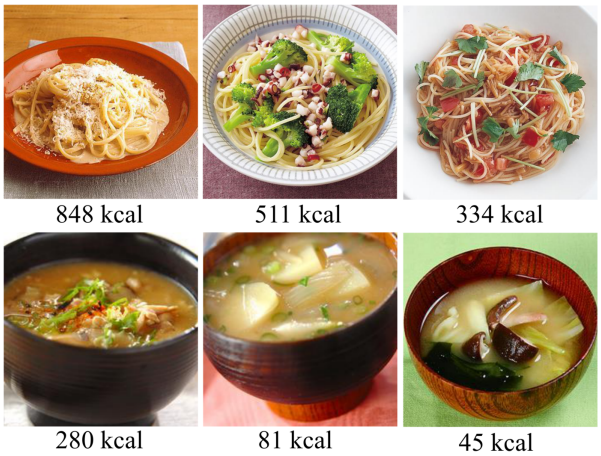

図 1　食事画像とカロリー量. 上段:“スパゲッティ”, 下段:“味噌汁”.

2. 関連研究

2.1 食事画像からのカロリー量推定に関する研究

これまで食事画像からのカロリー量推定においては, 主に料理の外見から直接推定する方法と, 料理のサイズから推定する方法の 2 つのアプローチで研究が行われてきた. 料理のサイズに注目した研究の一つに Myers らの研究 [3] がある. Myers らは, 食事/非食事の認識, 複数品目の認識, 深度推定, 食事領域分割などの複数のタスクをすべて CNN により行っている. この研究では, タスクごとに必要な学習データを独自に作成しているため, かなりのコストがかかると考えられる. また, カロリー量情報付きのデータセットが不足し, 十分に評価が行われていない問題点がある.

別のアプローチから料理のサイズを推定した研究として岡元らの研究 [4] がある. 岡元らは, 大きさが既知の基準物体と一緒に料理を撮影することで料理のサイズを推定し, 高精度のカロリー量推定を実現した. 実験には基準物体と料理が一緒に写った画像が必要であり, データセットは手作業で作成された.

料理のサイズ推定を行わずに直接食事画像からカロリー量を推定した研究として宮崎らの研究 [5] が存在する. 宮崎らは色ヒストグラムや 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[*1] に投稿された食事画像 6512 枚を使用し, 栄養学の知識を持った複数の専門家がカロリー量をアノテーションしている. この手法は低レベル特徴量のみを用いているため, 高精度の推定を行うことは困難であると考えられる.

2.2 Multi-task CNN と食事画像に関する研究

複数のタスクを同時に学習するために, これまでに Multi-task CNN[1] が提案されている. この研究では顔属性の認識を行っており, 複数の属性を同時に学習するために Multi-task CNN が提案されている.

Multi-task CNN に食事画像を適用した研究として Chen らの研究 [6] が存在する. Chen らは, Multi-task CNN により食事カテゴリと食材情報を同時に学習することで, 両方のタスクの精度が向上することを示した. これらは異なるタスクであるが, 食事カテゴリと食材には高い相関があるため, 両タスクに共通する特徴が学習され, 性能が向上したと考えられる. 本研究は Chen らの研究に刺激され, より困難なタスクと考えられるカロリー量推定に対して Multi-task CNN を用いることを考える. なお，本研究では，MVA2017 で発表予定の [2] で提案する食事カテゴリとカロリー量情報の同時学習によるカロリー量情報推定手法を拡張し，食材情報および調理手順情報を追加して利用することによってさらなる精度向上が実現することを示すことを目的とする.

3. 手法

3.1 Multi-task CNN の概要

本研究では VGG16 [7] を拡張し, Multi-task CNN を実装する. ネットワークは図 2 のように, fc6 層までの層が両タスクで共有され, fc7 層以降の層が各タスクで独自に学習される.

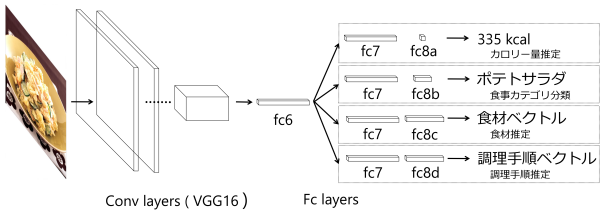

図 2 本研究で使用する Multi-task CNN のアーキテクチャ

本研究ではカロリー量に加えて食事カテゴリ, 食材, 調理手順の情報を同時に学習する. 各タスクの損失関数を L_{cal}, L_{cat}, L_{ing}, L_{met} とし学習データの総数を N とすると, 全体損失関数 L は次のように定義される.

$$L = -\frac{1}{N}\sum_{n=0}^{N}(L_{\mathrm{cal}} + \lambda_{\mathrm{cat}}L_{\mathrm{cat}} + \lambda_{\mathrm{ing}}L_{\mathrm{ing}} + \lambda_{\mathrm{met}}L_{\mathrm{met}}) \tag{1}$$

ただし λ は各タスクの損失関数にかかる重みであり, 各 λ の値はすべての損失項が同程度の値に収束するように決定される.

3.1.1 カロリー量の学習

カロリー量推定タスクは 4096 次元の fc7 層と, カロリー量を出力する単一のユニットで構成される出力層を有し, 1 人分のカロリー量の値を出力する. このような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, 本研究では次のような損失関数を使用する. 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.

$$L_{\mathrm{cal}} = \lambda_{\mathrm{re}}L_{\mathrm{re}} + \lambda_{\mathrm{ab}}L_{\mathrm{ab}} \tag{2}$$

ただし λ は各損失にかかる重みである. 式 (2) のように絶対誤差と相対誤差を組み合わせた損失関数を使用することで性能が向上する. 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.

$$L_{\mathrm{ab}} = |y - g| \tag{3}$$

$$L_{\mathrm{re}} = \frac{|y - g|}{g} \tag{4}$$

3.1.2 食事カテゴリの学習

この食事カテゴリ分類タスクは 4096 次元の fc7 層と, カテゴリ数分のユニットで構成される出力層を有し, 食事カテゴリのクラス確率を出力する. 損失関数として交差エントロピー誤差を使用する.

3.2 食材情報の学習

本研究では Word2Vec [8] による単語の分散表現を用いることで, 各レシピデータの食材名の単語を低次元実数ベクトルに変換し, それを食材情報の学習のために教師データとして利用する. Word2Vec の学習にはクックパッドのレシピデータセット [*2] の約 871 万文の調理手順文章を使用し, 単語の分散ベクトルの次元は $n = 500$ とする. 本実験では各レシピデータにおいて tf-idf 値の高い食材名の単語のみを利用する. 1 レシピデータあたりの食材名の単語の平均単語数が 12 であることから, 使用する単語数の上限は 12 個とする.

こうして得られた単語の分散表現と tf-idf 値から各レシピデータの食材ベクトルを生成する. あるレシピデータ r_j の食材情報を食材名の単語 w_i とすると, レシピデータ r_j の食材ベクトル v_j は次のように表される.

$$v_j = \sum_{k=1}^{N} tfidf_{k,j} * word2vec(w_k) \tag{5}$$

ただし $N = 12$ である. $word2vec(w_k)$ は Word2Vec による w_k の分散表現であり, $tfidf_{k,j}$ はレシピデータ r_j から抽出された単語 w_k の tf-idf 値である. 食材情報の学習は, この食材ベクトルを推定するタスクとして実現される. この食材ベクトル推定タスクでは 4096 次元の fc7 層と, 食材ベクトルの次元数のユニットで構成される出力層をもつ. 本実験では食材ベクトルは 500 次元とし, 損失関数として 2 乗

[*1] http://www.foodlog.jp/

[*2] http://www.nii.ac.jp/dsc/idr/cookpad/cookpad.html

和誤差を使用する.

3.3 調理手順情報の学習

調理手順に関しても食材情報の学習と同様に, 教師データを作成する. 本実験では名詞, 動詞, 形容詞のみを使用し, tf-idf 値の高い単語を利用する. 1 レシピデータあたりの平均単語数が 44 であることから, 使用する単語数の上限は 44 個とする.

こうして得られた単語の分散表現と tf-idf 値から各レシピデータの調理手順ベクトルを生成する. あるレシピデータ r_j の調理手順の文章中の単語を w_i とすると, レシピデータ r_j の調理手順ベクトル v_j は式 5 により得られる. ただし調理手順ベクトルでは $N = 44$ である. 調理手順の学習は, この調理手順ベクトルを推定するタスクとして実現される. この調理手順ベクトル推定タスクは 4096 次元の fc7 層と, 調理ベクトルの次元数のユニットで構成される出力層をもつ. 本実験では調理手順ベクトルは 500 次元とし, 損失関数として 2 乗和誤差を使用する.

4. 実験

4.1 カロリー量情報付き食事画像データセットの構築

本研究では 6 つのレシピ情報サイト (“レシピ大百科 *3”,“E・レシピ *4”,“ホームクッキング *5”,“みんなのきょうの料理 *6”,“オレンジページ net*7”,“レタスクラブニュース *8”) からカロリー量情報付き食事画像を収集した. 図 3 のようにレシピ情報ページには食事画像, カロリー量に加えて, 必ず食材情報と調理手順情報が含まれている. 収集したデータを観察すると, 食事画像の多くは 1 種類の料理の画像であり, カロリー量情報の多くは 1 人分の値であることがわかった. したがって本研究では, 1 種類の料理が写ったシングルラベルの食事画像を入力とし, 1 人分のカロリー量の値を推定する.

本研究では収集した画像に食事カテゴリの情報をアノテーションする必要があるため, 今回は UEC Food-100 食事画像データセット [9] の 100 種類の食事カテゴリについてラベリングを行う. UEC Food-100 食事画像データセットは主に日本の料理に関するデータセットであり, カロリー量の情報はアノテーションされていない.

また, 本実験では低解像度画像, 複数の種類の料理が写る画像, 付随するカロリー量が 1 人分の値であると断定できない画像をノイズとして除去し, その後サンプル数が 100 枚以下になった食事カテゴリを除いた. 最終的に総画像枚数 4877 枚, 食事 15 カテゴリのカロリー量情報付き食事画像データセットが完成した. 図 4 にカロリー量情報付き食事画像データセットの食事 15 カテゴリを示す. 本実験で

*3 http://park.ajinomoto.co.jp/
*4 http://erecipe.woman.excite.co.jp/
*5 https://www.kikkoman.co.jp/homecook/
*6 http://www.kyounoryouri.jp/
*7 http://www.orangepage.net/
*8 http://www.lettuceclub.net/recipe/

図 3 レシピ情報ページの例.

図 4 カロリー量情報付き食事画像データセットの食事 15 カテゴリ.

はこのデータセットを用いて学習と評価を行う.

4.2 モデルの学習とハイパーパラメータ

モデルの学習では ImageNet の 1000 種類分類タスクの pre-train モデルを初期値として利用し, 4.1 章のカロリー量情報付き食事画像データセットの 70%を使用して fine-tune する. 残りの 30%はテストに使用する. バッチサイズは 8 とし, 最適化手法として SGD を使用する. Momentum 値を 0.9 とし, 学習率 0.001 において 50k イテレーション, さらに 0.0001 において 20k イテレーション学習する. 式 1 と式 2 の損失項にかかる重みは事前に決定する必要があるが, 本実験では同時に行う全てのタスクの損失項にかかる重みを 1 に設定した状態で一度学習を行い, そのとき各イテレーションで得られる損失の値をタスクごとに保持しておき, 最終的に全イテレーションにおける損失の値の平均値の逆数を各タスクの損失項にかかる重みとして使用する. ただし本実験では, $\lambda_{\mathrm{re}} = 1$ と固定した.

4.3 Multi-task CNN によるカロリー量推定の評価

テストデータを用いてカロリー量の推定を行った結果を表 1 にまとめた. テストには, 学習時に最後の 1k イテレーションから 100 イテレーション間隔で得られた 10 個のモ

デルを使用し, 各モデルから得られた推定値の平均値を最終的な推定値とした. 全タスクでのマルチタスクの場合では, シングルタスクに比べて相対誤差が–1.967%, 絶対誤差が–9.488kcal, 相関係数が+0.039, 相対誤差 20%以内の割合が+4.189%となり, 食事カテゴリ分類においては正解分類率が+2.954%となり改善が見られた. 図 5, 図 6 にカロリー量推定の推定値と正解値の相関関係を示す. 図 6 は全タスクでのマルチタスクでの結果である. 図 5 と図 6 を比較すると, 95%信頼楕円などからマルチタスクにより精度が向上していることがわかる. 図 7, 図 8 に成功例と失敗例を示す.

表 1 カロリー量推定の結果

	相対誤差 (%)	絶対誤差 (kcal)	相関係数	誤差 20%以内 (%)	Top-1(%)
カロリー量 (single)	29.4	100.7	0.778	45.9	—
+食事カテゴリ	27.9	95.2	0.802	48.8	82.8
++食材情報	27.6	94.4	0.811	49.5	**85.2**
+++調理手順情報	27.4	**91.2**	**0.817**	**50.1**	84.1
+食材情報	29.2	96.8	0.795	46.8	—
++調理手順情報	28.0	97.9	0.806	47.2	—
+調理手順情報	28.2	95.5	0.808	48.1	—
++食事カテゴリ	**27.3**	96.0	0.808	48.8	84.8
食事カテゴリ (single)	—	—	—	—	81.2

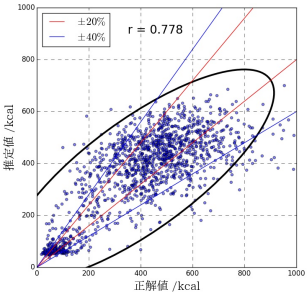

図 5 推定値と正解値の相関関係 (single-task).

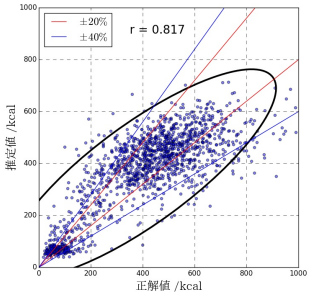

図 6 推定値と正解値の相関関係 (multi-task).

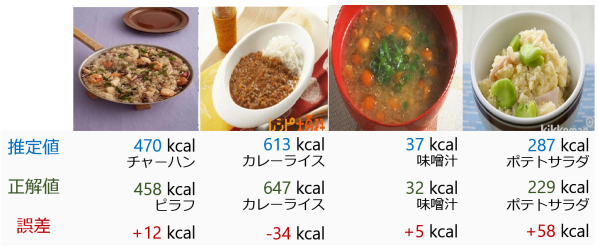

図 7 カロリー量推定成功例.

5. おわりに

本研究では CNN による食事画像からのカロリー量推定を行い, Multi-task CNN によるカロリー量と食事カテゴリの同時学習に, さらに食材情報, 調理手順情報を追加して同時学習することによって, さらなる精度向上が得られることを示した.

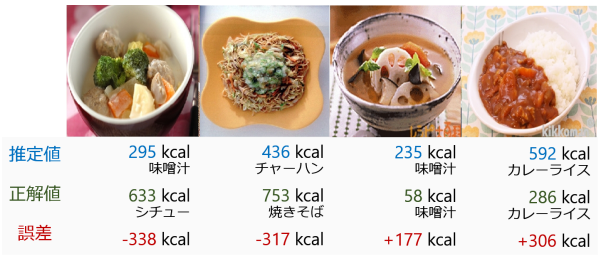

図 8 カロリー量推定失敗例.

今後の課題としては, カロリー量推定では食事の量を考慮することが不可欠であるため, 食事領域の検出や領域分割, さらに [4] や [10] のように予め基準物体を設けるなどして, 写真中の料理の実世界上でのサイズを推定することなどがある. また, 複数視点からの画像や奥行き推定を行うことで三次元情報を考慮することも考えられる.

謝辞 本研究は科研費 (17H01745,17H06026) の助成を受けたものである. 本研究では, クックパッド株式会社と国立情報学研究所が提供する「クックパッドデータ」を利用した.

参考文献

[1] H. A. Abrar, W. Gang, L. Jiwen, and J. Kui. Multi-task CNN model for attribute prediction. *IEEE Transactions on Multimedia*, Vol. 17, No. 11, pp. 1949–1959, 2015.

[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] 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*, 2015.

[4] K. Okamoto and K. Yanai. An automatic calorie estimation system of food images on a smartphone. In *Proc. of ACM MM Workshop on Multimedia Assisted Dietary Management*, 2016.

[5] 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*, 2011.

[6] J. J. Chen and C. W. Ngo. Deep-based ingredient recognition for cooking recipe retrival. *Proc. of ACM International Conference Multimedia*, 2016.

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

[8] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, 2013.

[9] 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*, 2012.

[10] 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.