

# 食事画像カロリー量推定における 回帰による手法と検索による手法の比較

會下 拓実<sup>1,a)</sup> 柳井 啓司<sup>1,b)</sup>

**概要:** 我々は食事画像からの直接カロリー量推定において、回帰による手法と検索による手法の比較を行う。これまでに回帰による手法と検索による手法として、それぞれ CNN を用いた回帰による食事画像からのカロリー量直接推定手法 [6] と従来の画像特徴量に基づく類似画像検索によるカロリー量推定手法 [11] が提案されている。これら二つは代表的な手法であると考えられるが、いままで厳密に比較されたことはなかった。CNN を用いた回帰による手法 [6] では料理 15 カテゴリに関して合計 4877 枚のカロリー量がアノテーションされた食事画像データセットが使用され、検索による手法 [11] では料理カテゴリを無視した合計 6522 枚のデータセットが使用されており、我々の知る限り、これらの回帰による手法と検索による手法は同一データセットによる比較実験は未だになされていない。そこで本研究では、[6] のデータセットを用いて、CNN を用いた回帰による手法と類似画像検索による手法の比較を行う。また、料理カテゴリを無視した合計 68,774 枚の大規模なカロリー量付き食事画像データセットを構築し、同様に比較実験を行う。なお本研究では、類似画像検索による手法 [11] では、BoF やカラーヒストグラムのような従来の画像特徴量の代わりに CNN の中間層から抽出した特徴量などを使用する。

## 1. はじめに

近年、食事に関する健康志向の高まりにより、多くの食事管理を目的とするモバイルアプリケーションがリリースされ、毎日の食事のバランスやカロリー量などを記録することが以前より容易になった。いくつかのモバイルアプリでは、画像認識技術を用いることで、料理写真のみから料理の名称を自動で推定することが可能となっている。しかしカロリー量の記録に関しては、多くのアプリでは、料理カテゴリや料理の量などの情報についてユーザーによる手入力が必要であり、このような手間のかかる作業や主観的な判断が要求される。こうした食事管理アプリにおけるユーザーの負担を軽減させるために、モバイルデバイス上での画像認識技術による食事画像からの自動認識は有効であると考えられ、これまでにいくつかの研究が存在する [1], [5], [8], [12], [13], [20]。

しかしそうしたものの多くは、推定された食品の種類、あるいはユーザーによって与えられた標準サイズと比較した相対的なサイズから食品の種類ごとにカロリー量を計算しており、現状では、全自動での食事画像からのカロリー量推定が可能なアプリケーションは存在しない。食事画像から

の料理カテゴリ分類などの食事画像認識タスクでは、CNN を用いた画像認識手法により精度が大幅に向上しているが、食事画像からのカロリー量推定については未解決の問題となっている。

食事画像からのカロリー量推定については、これまでにいくつかのアプローチが提案されている。主要なアプローチとして、食事画像から推定された料理カテゴリと料理のサイズもしくは体積に基づいてカロリー量を推定するものがある [4], [5], [8], [12], [13], [14]。この手法には、料理カテゴリ内の差異を考慮できないという問題がある。また、食事画像からの料理の量の推定において、CNN を使用する場合、使用する食事画像に対して高コストなアノテーションなどが必要になることが多い。しかし、カロリー量は料理カテゴリと量に強く依存していると考えられるため、このアプローチは有効かつ重要である。別のアプローチとして、料理カテゴリや量の推定なしに食事画像から直接カロリー量を推定する手法が存在する。これまでに [6], [11] がこのアプローチを適用している。宮崎ら [11] は、料理カテゴリや量を推定せず、食事画像から直接カロリー量を推定した。食事画像からの直接カロリー量推定には、カロリー量がアノテーションされた食事画像が必要であり、宮崎らは、食事管理サービス Food-Log<sup>\*1</sup> に投稿された 6512 枚の食事画

<sup>1</sup> 電気通信大学大学院 情報理工学研究所 情報学専攻所属

<sup>a)</sup> ege-t@mm.inf.uec.ac.jp

<sup>b)</sup> yanai@cs.uec.ac.jp

<sup>\*1</sup> <http://www.foodlog.jp/>

像に対して、栄養士の方にカロリー量をアノテーションしてもらうことでデータセットを構築した。宮崎らは、このデータセットを用いて、類似画像検索によるカロリー量推定手法を提案した。この手法では、与えられたクエリ画像に対して、カロリー量付き食事画像データベースから類似画像を上位  $k$  枚まで検索し、その類似画像にアノテーションされているカロリー量からカロリー量を計算する。類似画像検索には、BoF やカラーヒストグラムなどの従来の画像特徴量が用いられた。この宮崎らの手法は、料理カテゴリに依存していないため、あらゆる料理に対応することが可能であるが、検索による手法としてはデータベースの食事画像数や、使用した画像特徴量の表現力が十分ではなかったことなどが考えられる。

これに対して我々は [6] において、CNN を用いた回帰によるカロリー量推定を提案しており、食事画像からの料理カテゴリとカロリー量の同時推定のために multi-task CNN を用いた。[6] では、Web 上のレシピ情報サイトからカロリー量がアノテーションされた食事画像を収集することでデータセットを構築し、これを用いて multi-task CNN を学習した。この multi-task CNN は、一品の料理が写った写真 1 枚を入力として、料理カテゴリとカロリー量を同時に出力する。料理カテゴリとカロリー量の間には強い相関が存在するため、我々はこれらを同時に学習することで、別々に学習した場合の精度を上回ると考えた。Chen and Ngo [3] は料理カテゴリと食材の有無を同時に推定する multi-task CNN を提案し、これらの同時学習により精度が向上することを示している。[6] では、使用されるデータセットは、料理カテゴリとカロリー量の同時学習のために料理 15 カテゴリに限定され、また、類似画像検索による手法 [11] との厳密な比較実験がなされていない。

本論文では、食事画像からの直接カロリー量推定における代表的な手法である、回帰による手法 [6] と検索による手法 [11] について、同一のデータセットを用い、比較実験を行う。本実験ではデータセットとして、[6] の料理 15 カテゴリのデータセットと本研究で構築する料理カテゴリを無視した 68,774 枚の大規模データセットをそれぞれ用いる。なお、我々は検索ベースの手法 [11] に関して、従来の画像特徴量の代わりに CNN の中間層から抽出される特徴量などを使用する。

## 2. 手法

本研究では、食事画像からのカロリー量直接推定において、回帰による手法と検索による手法の比較を行う。本章では本研究で使用するこれら二つの手法について述べる。回帰による手法では [6] に従い、図 1 のような食事画像から料理カテゴリとカロリー量を同時に推定する multi-task CNN を用い、カロリー量を推定する。検索による手法では [11] に従い、画像特徴量に基づき類似画像検索を行うこ

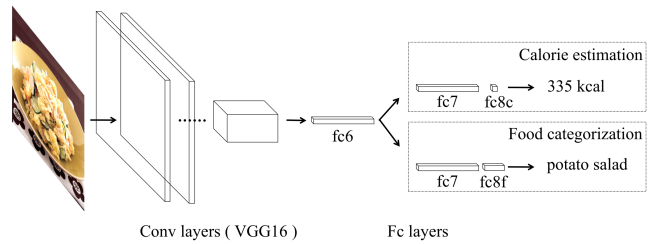


図 1 回帰によるカロリー量推定に使用する multi-task CNN のアーキテクチャ ([6]).

とでカロリー量を推定する。ただし、BoF やカラーヒストグラムなどの従来の画像特徴量の代わりに CNN 特徴ベクトルなどを使用する。

### 2.1 回帰によるカロリー量推定

我々は [6] に従い、CNN を用いて回帰による食事画像からのカロリー量直接推定を行う。[6] では、Web 上のレシピ情報サイトからカロリー量付き食事画像を収集することでデータセットを構築し、これを用いて multi-task CNN を学習している。この multi-task CNN は図 1 に示すように、料理カテゴリ分類タスクとカロリー量推定タスクを同時に行っており、一品の食品が写った食事画像から料理カテゴリとカロリー量を同時に推定する。Multi-task CNN を用いて料理カテゴリと食材を同時に推定した Chen and Ngo [3] は、それぞれのタスクが独自の間置層を持っているとき性能が向上したと述べており、本研究で使用する multi-task CNN のアーキテクチャはこれに従う。学習には ImageNet の 1000 種類分類タスクによって事前学習済みのモデルを使用する。

[6] に従い、カロリー量推定タスクの学習に使用する損失関数  $L_{cat}$  は、絶対誤差を  $L_{ab}$ 、相対誤差を  $L_{re}$  とすると次のように定義される。

$$L_{cat} = \lambda_{re} L_{re} + \lambda_{ab} L_{ab} \quad (1)$$

$\lambda$  は各損失項にかかる重みである。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  $L_{re}$  は下のよう定義される。

$$L_{ab} = |y - g| \quad (2)$$

$$L_{re} = \frac{|y - g|}{g} \quad (3)$$

また、料理カテゴリ分類タスクの損失関数  $L_{cat}$  として、交差エントロピー誤差を使用する。ある画像  $x$  を入力したときの出力層のユニット  $i$  の出力値を  $y_i$ 、 $y_i$  に対する教師データの値を  $g_i$  とすると、食事カテゴリ分類タスクの損失関数  $L_{cat}$  は次のように定義される。

$$L_{cat} = - \sum_{n=1}^n g_n \log y_n \quad (4)$$

これら二つのタスクを同時に学習するために, [6] では各タスクの損失関数の線形和を全体損失関数  $L$  として定義している.

$$L = \frac{1}{N} \sum_{n=0}^N (\lambda_{re} L_{re} + \lambda_{ab} L_{ab} + \lambda_{cat} L_{cat}) \quad (5)$$

$\lambda$  は各損失項にかかる重みであり, 各  $\lambda$  の値が近い値に収束するように決定されることが多い. 本実験ではすべての  $\lambda$  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,  $L_{re} = 1$  と固定した.

## 2.2 検索によるカロリー量推定

宮崎ら [11] は, BoF やカラーヒストグラムなどの従来の画像特徴量に基づき, カロリー量付き食事画像データベースから類似画像上位  $k$  枚を検索し, それらの画像にアノテーションされたカロリー量の平均値を計算することでカロリー量を推定している.

我々は宮崎ら [11] の手法に従い, 検索によるカロリー量推定を行う. ただし, 従来の画像特徴量の代わりに CNN の中間層から得られる特徴ベクトルを使用する. 我々は ImageNet1000 種類分類タスクによって事前学習済みの VGG16 [19] を画像特徴量抽出器として使用し, 全結合層の出力である 4096 次元ベクトルを画像特徴量として利用する. 実験ではまず, データベース用画像から画像特徴ベクトルを計算し, カロリー量付き食事画像データベースを構築する. カロリー量推定時には, 各テスト画像から画像特徴ベクトルを計算し, 特徴ベクトル間のユークリッド距離に基づきデータベースから類似画像を検索する.

さらに我々は, より画像検索タスクに適した特徴量を得るために, Simo-Serra ら [18] が提案した joint learning の手法を使用する. 彼らは特徴量抽出ネットワークと分類ネットワークの共同学習を行うことで, 識別に有効な特徴表現を学習した. 学習時には入力として参照画像, 類似画像, 非類似画像それぞれがパラメータを共有する同一のネットワークに与えられ, 各画像から CNN 特徴ベクトルが得られる. 特徴量抽出ネットワークでは参照-類似と参照-非類似それぞれの特徴ベクトル間のユークリッド距離を最適化することにより, 有用な特徴表現を学習し, 分類ネットワークでは, 得られた特徴ベクトルから分類タスクを行うことで有用な feature map を学習する. 推定時には画像一枚を入力することで CNN 特徴ベクトルが得られ, これを画像特徴量として検索タスクに使用する. 有用な特徴量を学習する手法として他にも siamese-network [2] や triplet-network [21] などが存在するが, 下田ら [17] は, これらの手法を食事画像検索タスクに適用し Simo-Serra ら [18] の joint-learning

が検索タスクにおいて最も優れていることを示しているため, 我々は joint-learning を使用する.

Joint-learning では予め, 入力として与える参照画像, 類似画像, 非類似画像を決める. 我々はまず, 類似画像と非類似画像のそれぞれを, 参照画像と同じ料理カテゴリの画像と参照画像と, 異なる料理カテゴリの画像と定義する. さらに, より確実に学習を進めるために, 画像  $I_i$  と画像  $I_j$  の類似度を表す類似スコア  $r_{i,j} = r(I_i, I_j)$  を導入し, 参照画像  $I$  に対して  $r(I, I_+) > T_+$  を満たす画像  $I_+$  を類似画像として使用し,  $r(I, I_-) < T_-$  を満たす画像  $I_-$  を非類似画像として使用する. 学習時にはこれら三枚の画像を入力として, 各画像から CNN 特徴ベクトルが得られる.

特徴量抽出ネットワークでは, 参照-類似と参照-非類似それぞれの特徴ベクトル間のユークリッド距離が, 参照-類似間では小さく, 参照-非類似間では大きくなるように学習する. 参照画像, 類似画像, 非類似画像それぞれの CNN 特徴ベクトルを  $f, f_+, f_-$  とすると, [21] に従い, 特徴量抽出ネットワークの損失関数  $L_R$  は次のように定義される.

$$L_R = \max(0, g + \|f_+ - f\|_2 - \|f_- - f\|_2) \quad (6)$$

分類ネットワークでは, 各画像で料理カテゴリ分類タスクが行われ, 損失関数として交差エントロピー誤差が使用される. 分類ネットワークの損失関数  $L_C$  は, 各画像の損失の線形和で表される.

最終的に全体損失関数  $L$  は以下のように定義される.

$$L = L_R + \lambda L_C \quad (7)$$

$\lambda$  は損失項にかかる重みであり, 本実験では  $\lambda = 1$  とする.

## 3. データセット

本実験では, [6] の料理 15 カテゴリのカロリー量付き食事画像データセットを用い, 食事画像からのカロリー量推定について, 回帰による手法 [6] と検索による手法 [11] の比較を行う. また, 本研究では料理カテゴリを無視した大規模データセットを構築し, 同様に比較実験を行う. 本章では実験に使用するこれら二種類のデータセットについて述べる.

### 3.1 料理 15 カテゴリのカロリー量付き食事画像データセット

[6] では, Web 上のレシピ情報サイトからカロリー量付き食事画像を収集することでデータセットを構築した. また, 料理カテゴリとカロリー量の同時学習のために料理 15 カテゴリがアノテーションされており, 合計 4877 枚のデータセットとなっている. 実験では学習用画像として 70% を使用し, 残りの 30% をテストに用いる. このデータセットを用い, 回帰による手法と検索による手法の比較を行う.

### 3.2 料理カテゴリを無視した大規模カロリー量付き食事画像データセット

[6] では料理カテゴリとカロリー量の同時学習のために料理 15 カテゴリに制限されている。現状, CNN を用いた料理 100 カテゴリに対するクラス分類の正解率が 90% [9] に近くなり, 高精度に分類することが可能になっていることを考えると, CNN を用いた手法を評価するにあたり料理 15 カテゴリは少ないと考えられる。そこで本研究では, [6] において除かれていた料理 15 カテゴリ以外のすべての食事画像も使用し, 料理カテゴリを無視した大規模カロリー量付き食事画像データセットを構築する。データセットの構築にあたり, 我々はノイズ除去を行った。まず  $256 \times 256$  より小さい画像を除去し, 次に物体検出ネットワークである Faster R-CNN [15] を用い, 食事画像に対して料理の検出を行い, 複数の料理が含まれる画像を自動で除去した。Faster R-CNN は料理の検出を行うために, バウンディングボックス付きの食事画像データセットである UECFOOD-100 [10] を用いて学習したものを使用した。最終的に我々は, 料理カテゴリを無視した 68,774 枚のカロリー量付き食事画像データセットを構築した。実験では学習用画像として 80 % を使用し, 残りの 20 % をテストに用い, 二つの手法の比較実験を行う。

## 4. 実験

本実験では, 料理 15 カテゴリのデータセットと料理カテゴリを無視した大規模データセットそれぞれを用い, 回帰による手法と検索による手法を用い, 食事画像からのカロリー量推定を行う。回帰による手法では, CNN 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する。

### 4.1 料理 15 カテゴリのデータセットでのカロリー量推定

ここでは料理 15 カテゴリのカロリー量付き食事画像データセットを用い, 食事画像からのカロリー量推定を行う。

#### 4.1.1 回帰によるカロリー量推定

回帰による手法では, 料理 15 カテゴリのカロリー量付き食事画像データセットの学習用画像を用い, CNN の学習を行う。料理カテゴリとカロリー量の同時学習を行う multi-task CNN と, カロリー量のみを学習する single-task CNN の両方について実験を行う。最適化手法として SGD を使用し, Momentum 値は 0.9 とし, バッチサイズは 8 とする。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。モデルには VGG16 [19], ResNet50/101 [7] を使用し, いずれも ImageNet の 1000 種類分類タスクにおいて事前学習済みのモデルを利用する。

表 1 に料理 15 カテゴリのデータセットでの回帰による

表 1 料理 15 カテゴリのデータセットでの回帰によるカロリー量推定結果

	相対誤差 (%)	絶対誤差 (kcal)	相関係数	誤差 20%以内 (%)	Top-1 (%)
VGG16 single-task	29.2	100.4	0.783	46.0	—
ResNet50 single-task	31.1	96.7	0.789	47.0	—
ResNet101 single-task	29.4	94.7	0.797	49.4	—
VGG16 multi-task	<b>28.0</b>	96.5	0.805	47.2	83.2
ResNet50 multi-task	30.6	93.2	0.803	49.0	<b>83.9</b>
ResNet101 multi-task	29.6	<b>91.9</b>	<b>0.807</b>	<b>50.8</b>	82.0

カロリー量推定結果を示す。カロリー量推定の評価指標として相対誤差, 絶対誤差, 推定値と正解値の相関係数, 相対誤差 20% 以内の推定値の割合を用いた。また, 料理カテゴリ分類に関しては Top-1 精度を示す。表 1 からいずれのモデルに関しても, 料理カテゴリとカロリー量の同時学習による精度の向上が見られ, ResNet101 が最も優れていた。

#### 4.1.2 画像検索によるカロリー量推定

検索による手法では, 学習用画像から画像特徴量データベースを作成する。推定時には, クエリ画像から検索された上位 k 枚の画像を使用し, それらにアノテーションされたカロリー量の平均値を計算する。ベースラインとして, ImageNet1000 種類分類タスクによって事前学習済みの VGG16 [19] から得られる fc6 層と fc7 層の出力である 4096 次元ベクトルを画像特徴量として使用する。画像特徴量データベースの作成には料理 15 カテゴリのカロリー量付き食事画像データセットの学習用画像を用いる。

Joint-learning [18] によるネットワークの学習には, UECFOOD-100 中の本実験が対象とする料理 15 カテゴリに対応する画像を使用する。最適化手法として SGD を使用し, Momentum 値は 0.9 とし, バッチサイズは 8 とする。学習率 0.0001 において 10k イテレーション学習する。モデルには ImageNet の 1000 種類分類タスクにおいて事前学習済みの VGG16 を用い, fc6 層から得られる特徴ベクトルを学習時とテスト時に画像特徴量として使用する。画像特徴量データベースの作成には料理 15 カテゴリのカロリー量付き食事画像データセットの学習用画像を用いる。

まず, 表 2 に各特徴量での検索精度を示す。クエリ画像と同じ料理カテゴリの画像が検索された場合を正解として, 各クエリ画像に対して, 検索された上位 k 件の画像から平均適合率 AP@k が求まる。評価指標として, 料理カテゴリごとに計算される平均適合率の平均値 mAP@k の平均値を用いる。表 3 に料理 15 カテゴリのデータセットでの検索によるカロリー量推定結果を示す。Joint-learning [18] により UECFOOD-100 を学習した VGG16 の特徴量において, 大幅に精度が向上したことがわかる。

表 2 各特微量における画像検索精度

	mAP@1	mAP@5	mAP@10
VGG16(ImageNet) fc6	39.8	48.7	46.2
VGG16(ImageNet) fc7	34.0	45.4	43.8
VGG16(UECFood-100) fc6	<b>63.4</b>	<b>70.6</b>	<b>68.0</b>

表 3 料理 15 カテゴリのデータセットでの検索によるカロリー量推定結果

	相対誤差 (%)	絶対誤差 (kcal)	相関係数	誤差 20%以内 (%)
VGG16 fc6 (k=5)	48.2	117.2	0.677	42.3
VGG16 fc6 (k=10)	48.9	112.9	0.698	44.2
VGG16 fc6 (k=15)	50.1	113.2	0.699	45.0
VGG16 fc7 (k=5)	53.3	122.9	0.641	40.1
VGG16 fc7 (k=10)	54.5	118.9	0.666	42.7
VGG16 fc7 (k=15)	55.1	117.5	0.675	43.4
VGG16(UECFood-100) fc6 (k=5)	38.5	100.6	0.763	46.5
VGG16(UECFood-100) fc6 (k=10)	<b>38.4</b>	97.0	0.780	49.4
VGG16(UECFood-100) fc6 (k=15)	38.7	<b>96.0</b>	<b>0.781</b>	<b>50.3</b>

表 4 料理 15 カテゴリのデータセットでの回帰による手法と検索による手法の比較

	相対誤差 (%)	絶対誤差 (kcal)	相関係数	誤差 20%以内 (%)
Baseline	32.4	<b>93.6</b>	0.784	50.0
VGG16 single-task	29.2	100.4	0.783	46.0
VGG16 multi-task	<b>28.0</b>	96.5	<b>0.805</b>	47.2
Joint-learning (k=5)	38.5	100.6	0.763	46.5
Joint-learning (k=10)	38.4	97.0	0.780	49.4
Joint-learning (k=15)	38.7	96.0	0.781	<b>50.3</b>

#### 4.1.3 回帰による手法と検索による手法の比較

ここでは回帰による手法と検索による手法についてまとめる。

さらに、上記二つの手法に加えて、よりシンプルなベースラインとして、推定された料理カテゴリに応じて、登録されたカロリー量を推定値とする手法の評価も行う。この手法のために料理 15 カテゴリに関して CNN を用いてクラス分類を行う。モデルには ImageNet の 1000 種類分類タスクにおいて事前学習済みの VGG16 を用い、学習率 0.001 において 20k イテレーション学習したものを使用する。

表 4 に回帰による手法と検索による手法をまとめる。表 4 の手法は、いずれも VGG16 [19] を使用したものである。また、図 2 に各カロリー量推定手法での推定値と正解値の相関を示す。図 2 を見ると、回帰による手法は他の手法と比べて大幅な推定誤りが少なく、全体的に安定していることがわかる。表 4 の相対誤差、相関係数からも同様のことが言える。

表 5 料理カテゴリを無視した大規模データセットでの回帰による手法と検索による手法の比較

	相対誤差 (%)	絶対誤差 (kcal)	相関係数	誤差 20%以内 (%)	誤差 20%以内 (%)
VGG16 fc6 k=10	121.5	142.9	0.330	24.7	47.3
VGG16 fc7 k=10	126.9	145.6	0.315	24.6	46.5
VGG16 single-task	60.0	132.0	0.436	23.5	48.0

#### 4.2 料理カテゴリを無視した大規模データセットでのカロリー量推定

この実験では、本研究で構築した料理カテゴリを無視した大規模カロリー量付き食事画像データセットを用いてカロリー量推定を行う。

回帰による手法では、学習用画像を用いて single-task CNN を学習し、残りの 20%をテストに使用する。学習率 0.001 において 150k イテレーション、さらに 0.0001 において 50k イテレーション学習する。今回使用するデータセットには料理カテゴリがアノテーションされていないため、multi-task CNN は使用しない。

検索による手法では、ImageNet の 1000 種類分類タスクにおいて事前学習済みの VGG16 [19] をそのまま利用し、全結合層の fc6 層と fc7 層から得られる 4096 次元ベクトルを画像特徴量として使用する。画像特徴量データベースの作成には学習用画像を用いる。この実験で使用するデータセットには料理カテゴリのアノテーションがないため、joint learning [18] は使用しない。

表 5 に料理カテゴリを無視した大規模なデータセットでのカロリー量推定の結果を示す。

#### 5. おわりに

本研究では、回帰による手法と検索による手法をそれぞれ用いて食事画像からのカロリー量を推定し、手法の比較を行った。実験では、料理 15 カテゴリのデータセットと料理カテゴリを無視した大規模データセットをそれぞれ用い、各手法の性能評価を行った。

本研究で扱った回帰による手法と検索による手法は、料理の量を考慮しないカロリー量推定であった。そこで今後は、料理の量に基づくカロリー量推定に取り組むことを考えている。カロリー量は料理カテゴリと量に強く依存するためこのアプローチは有効であり、これまでに、領域分割と基準物体を用いる手法 [13], [16] や、複数視点からの画像や奥行き推定により三次元推定を行う手法 [5], [8] などが提案されているが、我々はまず、カロリー量推定と物体検出を組み合わせることを考えている。

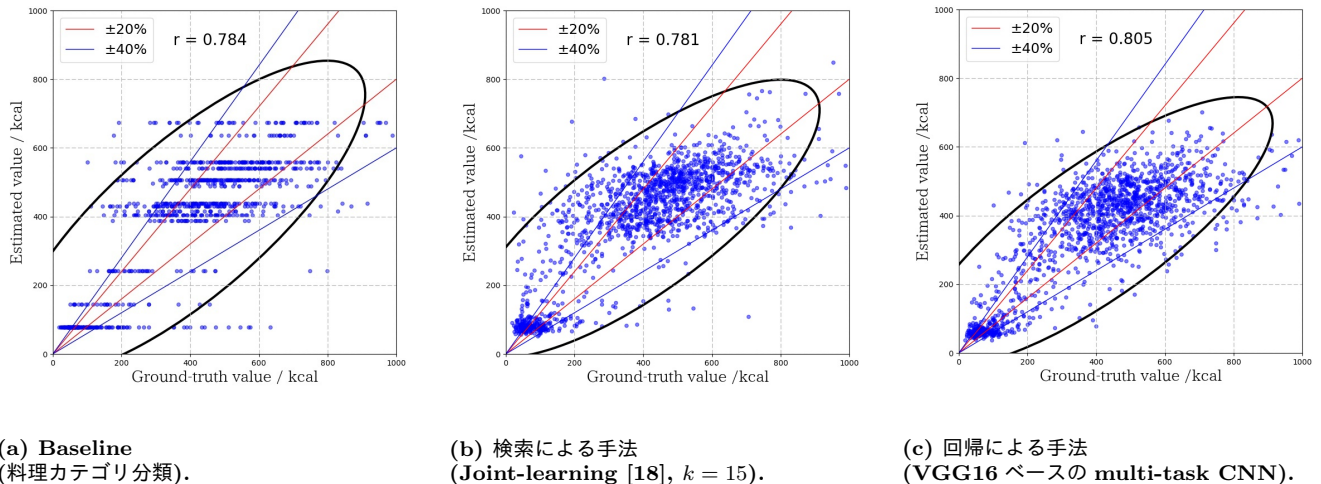


図 2 各カロリー量推定手法における推定値と正解値の相関

### 参考文献

- [1] V. Bettadapura, E. Thomaz, A. Parnami, D. G. Abowd, and A. Essa. Leveraging context to support automated food recognition in restaurant. In *Proc. of the 2015 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2015.
- [2] J. Bromley, I. Guyon, E. Sckinger, and R. Shah. Signature verification using a siamese time delay neural network. In *Advances in Neural Information Processing Systems*, 1994.
- [3] J. Chen and C. W. Ngo. Deep-based ingredient recognition for cooking recipe retrieval. In *Proc. of ACM International Conference Multimedia*, 2016.
- [4] M. Chen, Y. Yang, C. Ho, S. Wang, E. Liu, E. Chang, C. Yeh, and M. Ouhyoung. Automatic chinese food identification and quantity estimation. In *Proc. of SIGGRAPH Asia Technical Briefs*, page 29, 2012.
- [5] J. Dehais, M. Anthimopoulos, and S. Mougiakakou. Gocarb: A smartphone application for automatic assessment of carbohydrate intake. In *Proc. of ACM MM Workshop on Multimedia Assisted Dietary Management*, 2016.
- [6] 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.
- [7] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *arXiv preprint arXiv:1512.03385*, 2015.
- [8] F. Kong and J. Tan. Dietcam: Automatic dietary assessment with mobile camera phones. In *Proc. of Pervasive and Mobile Computin*, pages 147–163, 2012.
- [9] N. Martinel, G. L. Foresti, and C. Micheloni. Wide-slice residual networks for food recognition. In *arXiv preprint arXiv:1612.06543*, 2016.
- [10] 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.
- [11] 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.
- [12] A. Myers, N. Johnston, V. Rathod, A. Korattikara, A. Gorban, N. Silberman, S. Guadarrama, G. Papan-dreou, J. Huang, and P. K. Murphy. Im2calories: towards an automated mobile vision food diary. In *Proc. of IEEE International Conference on Computer Vision*, 2015.
- [13] 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.
- [14] P. Pouladzadeh, S. Shirmohammadi, and R. Almaghrabi. Measuring calorie and nutrition from food image. In *IEEE Transactions on Instrumentation and Measurement*, pages 1947–1956, 2014.
- [15] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016.
- [16] 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.
- [17] W. Shimoda and K. Yanai. Learning food image similarity for food image retrieval. In *Proc. of International Conference on Multimedia Big Data (BIGMM)*, 2017.
- [18] E. Simo-Serra and H. Ishikawa. Fashion style in 128 floats: Joint ranking and classification using weak data for feature extraction. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2016.
- [19] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *arXiv preprint arXiv:1409.1556*, 2014.
- [20] R. Tanno, K. Okamoto, and K. Yanai. Deepfoodcam: A dcnn-based real-time mobile food recognition system. In *Proc. of ACM MM Workshop on Multimedia Assisted Dietary Management*, 2016.
- [21] J. Wang, Y. Song, T. Leung, C. Rosenberg, J. Wang, J. Philbin, B. Chen, and Y. Wu. Learning fine-grained image similarity with deep ranking. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2014.