Multi-task CNN による食事画像からのカロリー量推定

會下　拓実† 　柳井　啓司†

† 電気通信大学情報理工学部総合情報学科　〒 182–8585 東京都調布市調布ヶ丘 1–5–1
E-mail: †ege-t@mm.inf.uec.ac.jp, ††yanai@cs.uec.ac.jp

あらまし　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外出先でカロリー量を記録できるようになった. しかしそれらはユーザに多くの操作を要求し, リアルタイム性に欠けるものが多く, 全自動でカロリー量を推定することができるアプリケーションはまだ存在しない. 一方, 画像認識分野では Deep Convolutional Neural Network(CNN) の登場により大幅に精度が向上し, 主なタスクの最高精度を独占している. そして 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が, 高精度の食事画像からのカロリー量推定は実現されていない. 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. 複数のタスクを同時に行う Multi-task CNN [1] により, カロリー量に加えて食事カテゴリを同時に学習することで, カロリー量推定の性能の向上を図る. 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を収集し, データセットを構築する. カロリー量推定の実験を行った結果, シングルタスクとマルチタスクの相対誤差がそれぞれ 28.956%と 27.664%となり, マルチタスクにおいて精度の向上が見られた. また, 推定値と正解値の相関係数が 0.802 となり, 高い相関が得られた.

キーワード　カロリー量推定, Multi-task CNN, 食事画像認識

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. そうした中で画像認識技術による食事画像からの自動認識が盛んに行われるようになっている. 画像認識分野においては Deep Convolutional Neural Network(CNN) の登場以来, 画像認識の精度が飛躍的に向上している. そして 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おり, 実際にアプリケーションに利用された例も多い. 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. 現状では, 食事画像からのカロリー量推定は未解決の問題であり, 様々なアプローチから研究が行われている. カロリー量は健康管理の面で重要な指標であるが, 栄養学の知識のない人にとって日々の食事のカロリー量を計算することは非常に困難である. したがって食事画像からの自動カロリー量推定の実現の意義は大きいと考えられる.

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, それらの情報は完成した料理の見た目に現れる. 人はこうした料理の見た目に現れる情報を認識することで, カロリー量の値を直接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. 図 1 に食事画像とカロリー量の例を示す. 一方, 画像認識分野では Deep Convolutional Neural Network(CNN) の登場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. そして CNN を用いた食事画像の認識に関する研究も盛んに行われているが, 食事画像からのカロリー量推定はまだ未解決の問題である. 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. また, 本研究では, 一人分のカロリー量を学習することで, 料理の量の影響を排除したカロリー量の推定を行う. さらに Multi-task CNN によりカロリー量に加えて食事カテゴリを同時に学習することで, カロリー量推定の精度の向上を図る. 食事カテゴリのようなカロリー量と相関のある情報を同時に学習した場合, 同時に学習するタスクに共通する特徴が学習され, 精度が向上することが知られている. さらに本研究では, カロリー量情報付き食事画像データセットを構築する. 現在, 利用可能なカロリー量情報付き食事画像データセットは公開されておらず, 手動での作成はコストがかかる. そこで本研究では, 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.

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 Meyer らの研究 [2] がある. Meyer 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. 手法としてはまず, 食事/非食事認識により画像中に料理が存在するかを判定し, その後マルチラベル食事認識により画像中の複数の料理の有無を認識する. 次に深度推定と食事領域分割を行い, オブジェクトの三次元構造と食事領域を抽出し, これらの情報を統合して料理のサイズを推定する. 最後に食事カテゴリや

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

サイズの情報から料理のカロリー量を計算している. 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ようである.

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 [3] 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, 高精度のカロリー量推定を実現した. 手法としてはまず, 基準物体と食品を一緒に撮影する (図 2). 次に基準物体と食品のそれぞれの領域を抽出する. 最後に基準物体と食品の領域を比較して算出した食品の大きさからカロリー量を計算する. 食品領域の抽出では、まずエッジにより背景から皿領域を検出し、その皿領域に対して k-means により色情報に基づく領域分割を行い、最終的に GrabCut により皿領域から食品領域を推定する. 図 3 に食事領域の抽出の流れを示す. このシステムにより高精度の食事領域の推定を実現し, カロリー量推定では相対誤差 21%という精度を達成した. 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 [4] が存在する. 手法としてはまず, histogram 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する. そして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. 図 4 にシステムの概要を示す. データセットには Web サービスである FoodLog(注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.

(注1) : http://www.foodlog.jp/

図 2　基準物体としてプリペイドカードを使用する場合の撮影風景. ([3] より引用.)

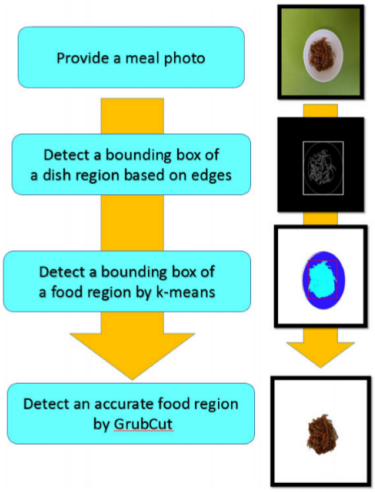

図 3　食品領域抽出の概要. ([3] より引用.)

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 [5] 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. 食事カテゴリと食材情報を推定する問題は, それぞれマルチクラス分類問題とマルチラベル認識問題として扱われている. 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.

3.　手 法 概 要

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. CNN の学習には国産の深層学習フレームワークである

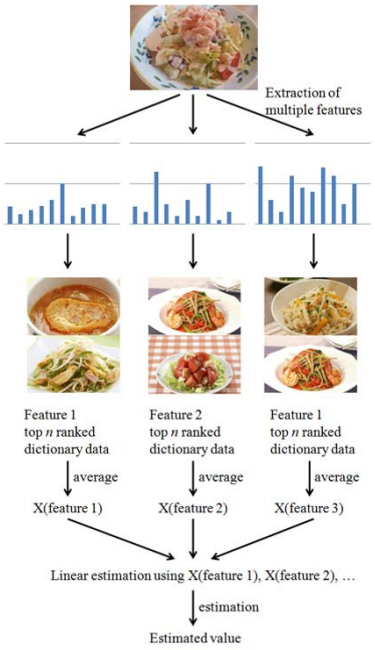

図 4　カロリー量推定システムの概要.（[4] より引用.）

Chainer(注2) を使用する. Chainer では, 既存のレイヤーを使用すれば, 簡単にネットワークを拡張することが可能である. 一般的なカテゴリ認識と異なり, カロリー量を推定する問題は回帰問題として扱われる. 本研究では, 1 種類の食品を写したシングルラベルの食事画像を入力として, 1 人分のカロリー量を出力する CNN を学習する. さらに本研究では, Multi-task CNN によりカロリー量に加えて食事カテゴリを同時に学習する. 食事カテゴリを推定する問題はマルチクラス分類問題として扱われる.

4.　手法の詳細

本章ではネットワーク構造および各タスクについて説明する.

4.1　Multi-task CNN の概要

本研究で用いる Multi-task CNN のアーキテクチャは VGG16 [6] に基づく. VGG16 は畳み込み層が 13 層, 全結合層が 2 層, 最終出力層の合計 16 層の深いネットワークであり, その性能と汎用性の高さから様々な研究に適用されている. 本研究ではこの VGG16 を拡張し, カロリー量と食事カテゴリを同時に学習する Multi-task CNN を実装する. 本研究のネットワークは図 5 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. したがって各タスクは独立した fc7 層と出力層を有する. Multi-task CNN により食事カテゴリと食材情報を同時に学習した Chen ら [5] は, いくつかのアーキテクチャで実験を行った結果, それぞれのタスクが独自の中間層を持っているとき最も性能が向上したと述べており, 本研究はそれに従う. 次にカロリー量推定タスクと食事カテゴリ分類タスクを同時に行う Multi-task CNN の全体損失関数について説明する. カロリー量推定タスクの損失関数を L_{cal}, 食事カテゴリ分類タスクの損失関数を L_{cat}, 学習データ総数 N をとすると全体損失関数 L は次のように示される.

(注2)：http://chainer.org/

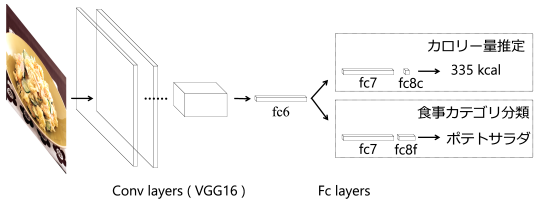

図 5　本研究で使用する Multi-task CNN のアーキテクチャ.

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

ただし λ は各タスクの損失にかかる重みであり, $\lambda_{cal} = 1$ と固定する. λ の値はすべての損失項が学習の過程で近い値に収束するように決定されるが, 最高性能を引き出すためには, 微調整が必要になる場合が多い. 式 (1) のように各タスクで計算された損失を加算したものが全体損失関数となる. 同時に学習するタスク数が増えた場合, 増えたタスクの損失を含めたすべてのタスクの損失を加算したものが全体損失関数となる. 以下で各タスクの学習方法と損失関数などについて説明する.

4.1.1　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する. このタスクは入力として 1 種類の食品が写ったシングルラベルの食事画像を受け取り, 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.

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

ただし λ は各損失にかかる重みであり, $\lambda_{re} = 1$ と固定する. 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. 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.

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

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

4.1.2　食事カテゴリの学習

カロリー量は本来的に, 使用する食材の種類と量, そして調理手順が分かればほぼ正確に推定することが可能であると考えられる. 同じ食事カテゴリの料理は使用される主な食材や調理方法が類似するため, カロリー量の値も類似する傾向があり, カロリー量の推定において有用な情報であると考えられる. 食事カテゴリを推定する問題はマルチタスク分類問題として扱われる. 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．出力層に活性化関数として Softmax 関数が適用され，出力層の各ユニットの出力値は非負となり，合計が 1 となるような確率分布が出力される．また，損失関数として交差エントロピー誤差が使用される．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．

$$L_{cat} = -\sum_{k=1}^{n} g_k \log y_k \tag{5}$$

ただし g_k はバイナリ値であり，ユニット k が正解のユニットであれば $g_k = 1$ となり，ユニット k が正解のユニットではなければ $g_k = 0$ となる．本研究では食事 15 カテゴリについて分類を行うため $n = 15$ となる．

5. 実　　験

5.1 カロリー量情報付き食事画像データセットの構築

現時点では，カロリー量情報を含む大規模な食事画像データセットは公開されておらず，手作業での収集やクラウドソーシングの使用はコストがかかるため，本研究では Web 上のカロリー量情報付きレシピ情報サイトから，カロリー量情報付き食事画像を収集する．本研究では以下の 6 つのレシピ情報サイトから食事画像とカロリー量を収集した．

（1） レシピ大百科 (注3)
（2） E・レシピ (注4)
（3） ホームクッキング (注5)
（4） みんなのきょうの料理 (注6)
（5） オレンジページ net (注7)
（6） レタスクラブニュース (注8)

これらのサイトで公開されているレシピ情報は，調理師や料理研究家などの専門家が提供したものとなっている．図 6 にレシピ情報ページの例を示す．収集した食事画像とカロリー量の値を観察すると，多くのレシピデータのカロリー量が 1 人分の値であるのに対して，図 7 のように，食事画像中の料理の量は 1 人分ではない場合が多いことがわかった．よって食事の量を考慮したカロリー量推定を行うためには，食事の量を考慮したカロリー量をアノテーションし直す必要があるが，これにはかなりのコストがかかると考えられる．したがって本研究では，1 人分のカロリー量の値を扱うものとする．

本研究では Multi-task CNN により，カロリー量に加えて食事カテゴリを学習するため，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．本実験では UEC Food-100 食事画像データセット [7] 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

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

図 6　レシピ情報ページの例．

図 7　料理の量とカロリー量が対応していない食事画像の例．レシピページから収集したカロリー量が 1 人分であるのに対して，写真中の料理の量は 1 人分ではないと考えられる．

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．図 8 に UEC Food-100 食事画像データセットの食事 100 カテゴリの一覧を示す．

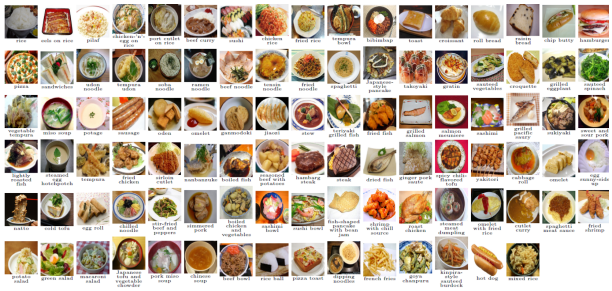

図 8　UEC Food-100 食事画像データセット [7] の食事 100 カテゴリ．

次にノイズ除去を行う．本実験では低解像度画像 (ぼやけた画像)，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た．さらにノイズ除去後に，サンプル数が 100 枚以下になった食事カテゴリを除いた．これにより，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 9 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．図 10 にデータセット全体のカロリー量の分布を示す．“味噌汁” の画像が多いために 100kcal 以下のサンプル数が多くなっているが，おおよそ 500kcal 付近の料理が多いことがわかった．また，1000kcal 以上のサンプルはほどんどなかった．図 11 に “味噌汁”，“スパゲッティ”，“カレーライス” に関するカロリー量の分布を示す．食事カテゴリとカロリー量との間に相関があり，また，同じ食事カテゴリであっても

ばらつきがあることがわかる．本実験ではこのデータセットを用いて学習と評価を行う．

図 9　本研究の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．

5.2　CNN の学習とハイパーパラメータ

本実験で用いる Multi-task CNN のアーキテクチャは VGG16 [6] に基づく．ただし本研究では Dropout [8] は用いず, fc6 層と fc7 層において Batch Normalization [9] を使用し, すべての層が畳み込み層により実装される．Batch Normalization 層と出力層以外の層は, ImageNet の 1000 種類分類タスクのデータセットで学習済みのものを使用する．5.1 章のカロリー量情報付き食事画像データセットの 70%を学習に使用し, 残り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については, λ_{ab} は 0.003 とし, λ_{cat} は 0.3 とする．λ_{ab} に関しては, L_1 の各損失項が近い値に収束するように設定された．λ_{cat} に関しては, 0.1 刻みでの実験を行い決定した．

5.3　Multi-task CNN によるカロリー量推定の評価

5.1 章のカロリー量情報付き食事画像データセットのテストデータによる評価を行った．カロリー量推定については評価指標として相対誤差, 絶対誤差, 相関係数, 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, 相対誤差は絶対

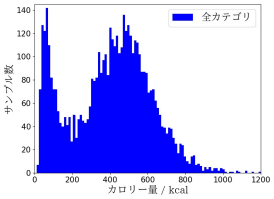

図 10　データセット全体のカロリー量の分布．

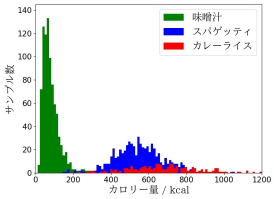

図 11　“味噌汁”, “スパゲッティ”, “カレーライス” のカロリー量の分布．

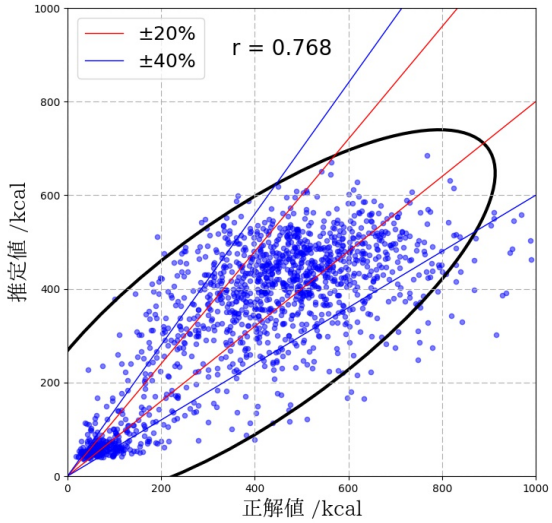

図 12　シングルタスクでのカロリー量推定の推定値と正解値の相関関係．

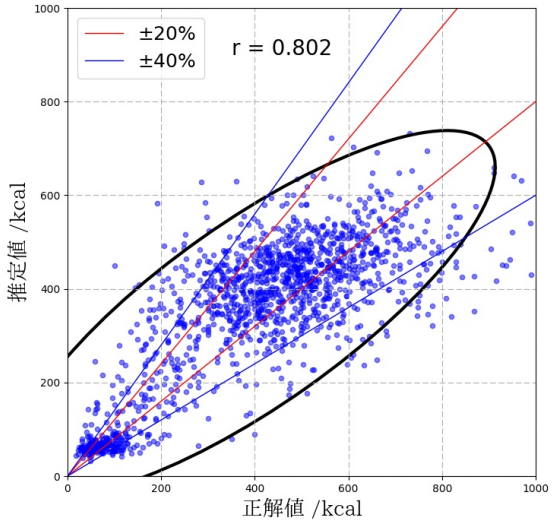

図 13　マルチタスクでのカロリー量推定の推定値と正解値の相関関係．

誤差と正解値の比である．また, 食事カテゴリ分類については正解分類率により評価を行う．カロリー量推定と食事カテゴリ分類のマルチタスクでの推定結果と, 各タスクのシングルタスクでの推定結果を表 1 にまとめる．シングルタスクとマルチタスクとを比較すると, すべての評価指標においてマルチタスクがシングルタスクを上回った．マルチタスクではシングルタスクに比べて, 相対誤差が–1.292%, 絶対誤差が–7.804kcal, 相関係数が+0.034 となり, 改善が見られた．これは食事カテゴリを同時に学習することで料理の誤認識が減り, カロリー量推定の精度が向上したと考えられる．また, 食事カテゴリ分類においては正解分類率が 2.884%向上した．カロリー量と食事カテゴリとの間には相関があるため, 両タスクが相乗効果を生み出し

表 1　シングルタスクとマルチタスクでの推定結果.

	カロリー量推定					食事カテゴリ分類
	相対誤差 (%)	絶対誤差 (kcal)	相関係数	相対誤差 20%以内 (%)	相対誤差 40%以内 (%)	Top-1(%)
カロリー量推定 (シングルタスク)	28.956	105.543	0.768	43.269	76.236	—
+食事カテゴリ (マルチタスク)	**27.664**	**97.739**	**0.802**	**47.527**	**80.151**	**83.104**
食事カテゴリ分類 (シングルタスク)	—	—	—	—	—	80.220

たと考えられる. 図 12 にシングルタスクでのカロリー量推定の推定値と正解値の相関関係を示す. 図 13 にマルチタスクでのカロリー量推定の推定値と正解値の相関関係を示す. 図 12 と図 13 とを比較すると, 95%信頼楕円からマルチタスクでの精度が向上していることがわかる. 図 14 にカロリー量推定の成功例, 図 15 にカロリー量推定の失敗例を示す. 高精度のカロリー量推定を行うには, 食事カテゴリを正確に認識する必要があると考えられる.

推定値	347 kcal 肉じゃが	566 kcal スパゲッティ	61 kcal 味噌汁	725 kcal カレーライス
正解値	386 kcal 肉じゃが	510 kcal スパゲッティ	65 kcal 味噌汁	730 kcal カレーライス
誤差	-39 kcal	+56 kcal	-4 kcal	-5 kcal

図 14　カロリー量推定成功例.

推定値	299 kcal ポテトサラダ	452 kcal グラタン	163 kcal 味噌汁	476 kcal カレーライス
正解値	682 kcal シチュー	362 kcal ピラフ	44 kcal 味噌汁	944 kcal カレーライス
誤差	-383 kcal	+90 kcal	+119 kcal	-468 kcal

図 15　カロリー量推定失敗例.

次にカテゴリ依存のカロリー量推定の手法との比較を行った. カテゴリ依存の手法では, あらかじめ食事カテゴリごとにカロリー量を固定しておき, カテゴリ認識の結果からカロリー量の値を推定する. 本実験ではこの手法をカテゴリ依存のカロリー量推定とする. カテゴリ依存のカロリー量推定では, 食事カテゴリ分類タスクにおいてカロリー量を推定することが可能であるが, 図 11 に示すような, 各食事カテゴリにおけるカロリー量の分散を考慮することができない. 本実験では, 学習データから算出した各食事カテゴリの平均カロリー量を固定値として使用する. このカテゴリ依存の手法に対して本研究の手法は, 食事画像から直接カロリー量を推定するため, 図 1 のような料理の外見を反映することで, 各カテゴリにおけるカロリー量の分散を考慮することが可能であると考えられる.

表 2 では, まず表 1 のシングルタスクでの食事カテゴリ分類によるカテゴリ依存のカロリー量推定の結果を示す. さらに, 表 1 のマルチタスクでのカロリー量推定において, カロリー量推定タスクの出力は用いず, 食事カテゴリ分類タスクの出力からカテゴリ依存のカロリー量推定を行った結果を示す. また, 本研究の手法として, 表 1 のシングルタスクでのカロリー量推定の結果と, マルチタスクでのカロリー量推定の結果を示す. 図 16 にカテゴリ依存のカロリー量推定での推定値と正解値の相関関係を示す. 表 2 を見ると, カテゴリ依存の手法は本研究の手法に劣らない性能を示していることがわかる. また, 表 1 のように, マルチタスクにおいては食事カテゴリ分類の精度も向上するため, マルチタスクでのカテゴリ依存のカロリー量推定の精度も向上している. このカテゴリ依存のカロリー量推定の精度は, 各カテゴリにおけるカロリー量の分散に大きく左右されるため, カテゴリによっては本研究の優位性が得られると考えられるが, 全体としては本研究の手法の特長を十分に生かせていない結果になったと考えられる. したがって本手法では, 各食事カテゴリにおけるカロリー量の分散を考慮したカロリー量推定が, 十分に行われていないと推測される.

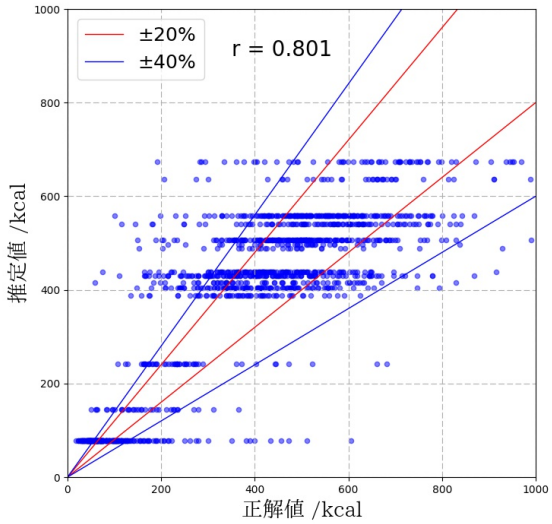

図 16　マルチタスクでのカテゴリ依存のカロリー量推定の推定値と正解値の相関関係.

次に食事 15 カテゴリ中の “味噌汁”, “スパゲッティ”, “カレーライス” についてカロリー量推定の実験を行い, 各カテゴリにおけるカロリー量推定の評価を行った. これらの食事カテゴリは図 11 に示すように特徴的な異なるカロリー量の分布を有している. “味噌汁”, “スパゲッティ”, “カレーライス” における推定結果を表 3, 4, 5 に示す. 表 3, 4, 5 を見ると, 表 1 の

表 2　カテゴリ依存の手法との比較.

	カロリー量推定					食事カテゴリ分類
	相対誤差 (%)	絶対誤差 (kcal)	相関係数	相対誤差 20%以内 (%)	相対誤差 40%以内 (%)	Top-1(%)
カテゴリ依存 (シングルタスク)	37.004	94.030	0.780	50.069	77.473	80.220
+カロリー量 (マルチタスク)	31.784	**90.871**	0.801	**50.618**	78.091	**83.104**
カロリー量推定 (シングルタスク)	28.956	105.408	0.768	43.269	76.236	—
+食事カテゴリ (マルチタスク)	**27.664**	97.739	**0.802**	47.527	**80.151**	**83.104**

表 3　“味噌汁” での推定結果.

	カロリー量推定					食事カテゴリ分類
	相対誤差 (%)	絶対誤差 (kcal)	相関係数	相対誤差 20%以内 (%)	相対誤差 40%以内 (%)	Top-1(%)
カロリー量推定 (シングルタスク)	**35.871**	31.604	0.442	37.692	65.385	—
+食事カテゴリ (マルチタスク)	36.439	**30.038**	**0.451**	**38.077**	**66.538**	**98.846**
食事カテゴリ分類 (シングルタスク)	—	—	—	—	—	95.000

表 4　“スパゲッティ” での推定結果.

	カロリー量推定					食事カテゴリ分類
	相対誤差 (%)	絶対誤差 (kcal)	相関係数	相対誤差 20%以内 (%)	相対誤差 40%以内 (%)	Top-1(%)
カロリー量推定 (シングルタスク)	20.722	109.742	0.201	**59.140**	90.323	—
+食事カテゴリ (マルチタスク)	**19.750**	**104.457**	**0.359**	58.602	**93.011**	**91.398**
食事カテゴリ分類 (シングルタスク)	—	—	—	—	—	88.172

表 5　“カレーライス” での推定結果.

	カロリー量推定					食事カテゴリ分類
	相対誤差 (%)	絶対誤差 (kcal)	相関係数	相対誤差 20%以内 (%)	相対誤差 40%以内 (%)	Top-1(%)
カロリー量推定 (シングルタスク)	35.638	241.133	**0.064**	25.000	60.000	—
+食事カテゴリ (マルチタスク)	**32.142**	**206.083**	0.022	**33.333**	**75.000**	**86.667**
食事カテゴリ分類 (シングルタスク)	—	—	—	—	—	73.333

結果と比較して, 相関係数の値が小さいことがわかる. これは, 図 11 に示すような, 各食事カテゴリにおけるカロリー量の分散を考慮したカロリー量推定が, 十分に行われていないことを示していると考えられる. また, 本研究で扱った食事カテゴリの情報では, 各食事カテゴリにおけるカロリー量の分散を考慮することはできないため, マルチタスクによる効果を得ることが難しいと考えられる. “味噌汁”, “スパゲッティ”, “カレーライス” におけるカロリー量推定の推定値と正解値の相関関係を, 図 17, 18, 19 に示す. 図 17, 18, 19 を見ると “味噌汁”, “スパゲッティ” では弱い相関があり, “カレーライス” では相関がないことがわかる. より高精度のカロリー量推定を実現するためには, 各食事カテゴリにおけるカロリー量の分散を考慮する必要があると考えられる.

6. 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った. 食事 15 カテゴリを対象に行ったカロリー量推定の実験では, Multi-task CNN により カロリー量に加えて食事カテゴリを同時に学習することで, カロリー量推定の精度が向上することを示した. また, Web 上のカロリー量情報付きレシピサイトからカロリー量情報付き食事画像を収集することで, データセット作成のコストを削減した.

今後の課題として, 食事カテゴリに加えて食材や調理手順の情報を同時に学習することで, 各食事カテゴリにおけるカロリー量の分散を十分に考慮したカロリー量推定を行うことが考えられる. また, 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 [3] や [11] 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が必要であると考えられる.

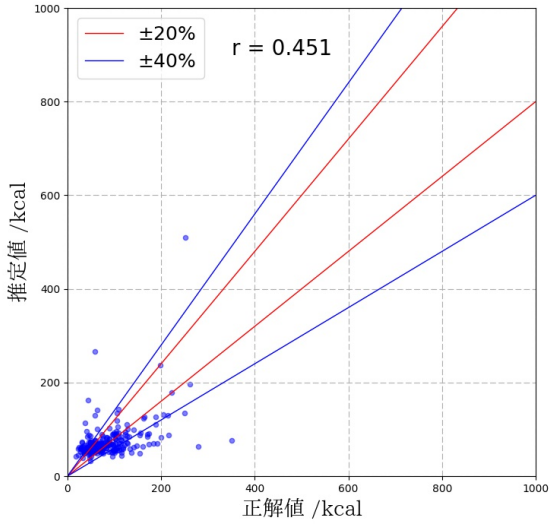

図 17　マルチタスクでのカロリー量推定の推定値と正解値の相関関係 (“味噌汁”).

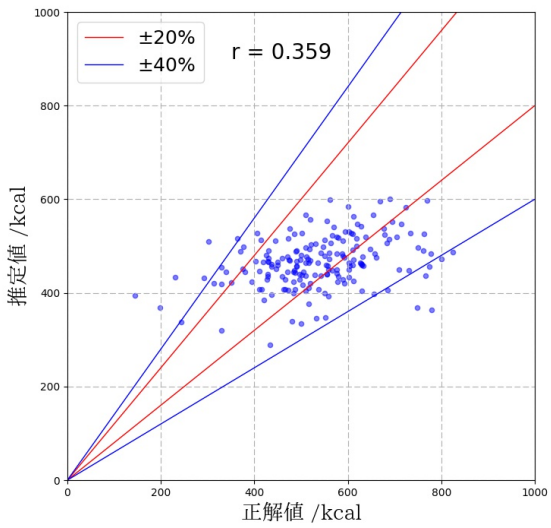

図 18　マルチタスクでのカロリー量推定の推定値と正解値の相関関係 (“スパゲッティ”).

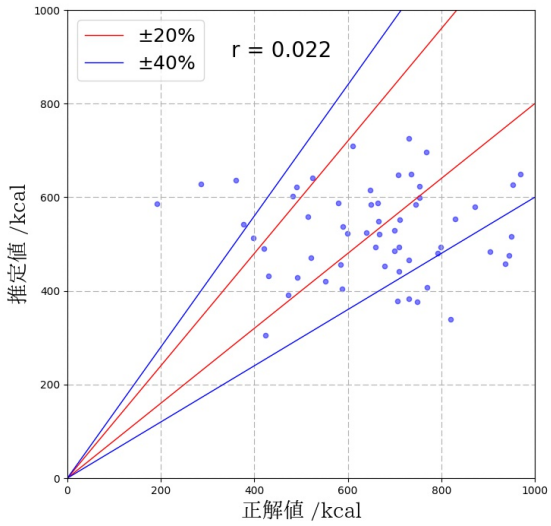

図 19　マルチタスクでのカロリー量推定の推定値と正解値の相関関係 (“カレーライス”).

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] A. Meyers, 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.

[3] 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.

[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*, 2011.

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

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

[7] 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.

[8] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, Vol. 15, pp. 1929–1958, 2014.

[9] S. Ioffe and C. Szegedy. Batch Normalization: Accelerating deep network training by reducing internal covariate shift. In *Proc. of International Conference on Machine Learning*, 2015.

[10] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*, 2015.

[11] 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.