社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

CNN による複数料理写真からの同時カロリー量推定

會下　拓実†　　柳井　啓司†

† 電気通信大学大学院情報理工学研究科　　〒 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　食に関する健康志向の高まりにより, 日々の食事のカロリー量を記録することができる食事管理アプリケーションが数多く登場している. 中には画像認識技術により料理の写真から, その料理名の候補を自動で提案するものまで存在する. 料理のカロリー量計算に関しては, ユーザーが選択画面に従って入力した情報からカロリー量を推定するものや, 料理の画像を送ると栄養士の方がその画像からカロリー量を推定するものがある. このようなアプリケーションにより栄養学の知識のない一般ユーザーがカロリー量を記録することが可能になったが, これらのカロリー量推定方法は人手による操作に手間がかかることや, 主観的評価であることなどの問題がある. 一方, 画像認識分野では Convolutional Neural Network(以下 CNN) の登場により精度が飛躍的に向上しており, 画像のクラス分類や, 画像に写る物体の検出や領域分割などを高精度に行うことが可能となっている. そこで本研究では, CNN を用い複数料理写真からの同時カロリー量推定を行う. 我々はまず 物体検出により複数料理写真から個々の料理を検出し, さらに検出された料理に対してカロリー量推定を行う. 物体検出には Faster R-CNN [7] を用い, これを料理画像で学習することで, 複数料理画像からの料理の検出を行う. また, カロリー量推定では食事画像からのカロリー量推定 [2] を用い, 検出された料理のクロップ画像からカロリー量を推定する. 実験では合計カロリー量がアノテーションされた複数料理画像を用意し, 複数食品画像からの同時カロリー量推定の評価を行う.

キーワード　カロリー量推定, CNN, Faster R-CNN, 食事画像認識, 物体検出

1. はじめに

近年, 食に関する健康志向の高まりにより, 日々の食事のカロリー量を記録することの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はカロリー量の推定に, ユーザー入力による情報が必要であったり, 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.

一方, 画像認識分野では CNN により飛躍的に精度が向上し, 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画像からの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ると考えられる. 実際, 料理カテゴリ分類に関する研究は盛んに行われおり, 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同時カロリー量推定を行う. 複数の料理が写る画像から個々の料理を検出し, 検出された料理のカロリー量を推定する. これにより 1 枚の複数料理画像から個々の料理のカロリー量を自動的に推定する.

會下ら [2] は CNN を用い料理画像からカロリー量を推定するネットワークを構築し, これをカロリー量情報付き食事画像データセット [2] で学習することで食事画像からのカロリー量推定を実現した. 本研究ではこのネットワークを新たに学習し, カロリー量推定に用いる. 會下らの手法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行われることが期待でき, 例えば同一カテゴリの料理だとしても食材の違いによる見た目の違いに応じて異なるカロリー量を出力する. 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, 物体検出により検出された単品料理の矩形領域をカロリー量推定ネットワークの入力として使用する. また, このネットワークの出力のカロリー量の値は, 1 人分のカロリー量の値となっている. したがって画像中の料理の量が何人前であっても, その料理の 1 人分の量

図 1　複数料理画像の例

に対応するカロリー量を出力する．

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り，これにより画像中の個々の料理のカテゴリとバウンディングボックスを推定する．同一カテゴリの物体を個々に検出することができるため，例えば，画像中に同じ料理が 2 皿写っていた場合にも個々に検出される．本研究では CNN を用いた物体検出ネットワークである Faster R-CNN [7] により複数料理画像からの単品料理の検出を行う．Faster R-CNN は提案された当時，従来研究と比較して高精度かつ高速な物体検出を実現し，後に多くの研究に応用されている．本研究ではこの Faster R-CNN を料理画像により学習することで，複数料理画像からの料理の検出を行う．さらに検出された単品料理領域に対して食事画像からのカロリー量推定 [2] を適用することで複数料理画像からの同時カロリー量推定を実現する．

まとめると，本研究では複数料理画像に対して Faster R-CNN [7] を用い，単品料理の検出を行う．また，Web 上の学校給食ブログから複数料理画像を収集し，バウンディングボックスと料理カテゴリをアノテーションした物体検出用データセットを作成し学習と評価に用いる．実験では Faster R-CNN を料理検出器として使用し，さらにカロリー量推定 [2] を組み合わせることで複数料理画像からの同時カロリー量推定を行う．また，合計カロリー量がアノテーションされた複数料理画像を Web 上の学校給食ブログから収集し，評価に用いる．

2. 関 連 研 究

近年，画像認識分野では，CNN を用いた手法が主なタスクの最高精度を独占している．そして食事画像認識においても料理カテゴリ分類や領域分割など，CNN を用いた様々な手法が提案されている．

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 [9] 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．このようにこの手法は料理領域の領域分割を行っている．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と同じになる．これに加え下田らは，はじめの候補領域の生成にも CNN を使用する手法 [10] 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が直接生成する．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．本研究では高精度かつ高速な物体検出を達成している Faster R-CNN [7] を用い，複数料理画像からの料理の検出を行う．

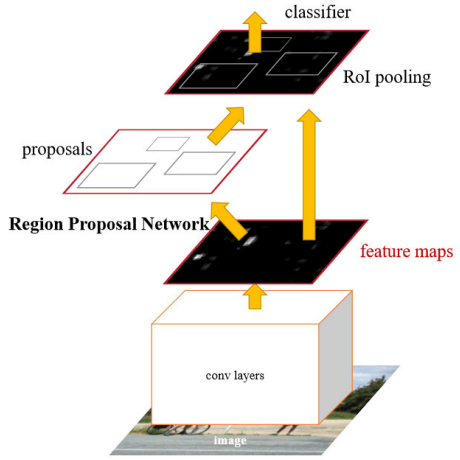

図 2　Faster R-CNN のネットワークの概略図 ([8]).

食事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 [6] 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．本研究の実験では合計カロリー量がアノテーションされた複数料理画像を収集し，複数料理画像からの同時カロリー量推定の評価を行う．

3. 手　　法

ここでは複数料理画像からの同時カロリー量推定手法について説明する．まず，構成要素である Faster R-CNN [7] と食事画像からのカロリー量推定 [2] についてそれぞれ述べ，最後にこれらを組み合わせたシステム全体について説明する．

3.1 Faster R-CNN

近年，物体検出において CNN を用いた手法が数多く提案され，飛躍的に精度が向上している．例えば，CNN を用いた物体検出の初期の研究として R-CNN [4] がある．この R-CNN では selective search により入力画像から大量の候補領域を生成し，各候補領域ついて CNN を用いて認識を行っている．また，R-CNN を改良した Fast R-CNN [3] では，画像全体の特徴マップ上から候補領域を生成するので，画像全体に一度 CNN を用いるだけで各候補領域からバウンディングボックスとカテゴリを推定することが可能になり，高速な検出を実現した．しかしこれらの手法は候補領域の生成において selective search を使用しており，この部分がシステム全体にとってボトルネックになっていた．その後提案された Faster R-CNN [7] では，候補領域の生成を CNN により行い，高精度かつ高速な検出を実現した．本研究では，この Faster R-CNN を用い，複数食事画像からの料理の検出を行う．

ここからは [7] に従い，本研究で使用する Faster R-CNN について説明する．Faster R-CNN は最新研究の基盤となっている手法であり，二つのモジュールで構成される．

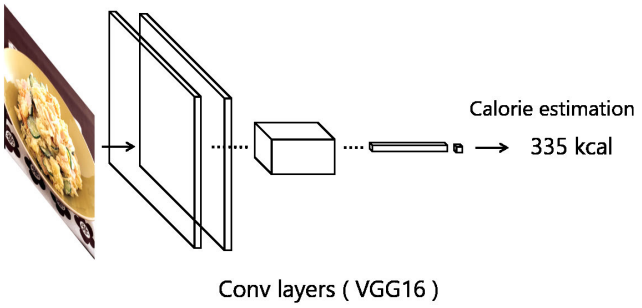

図 3　カロリー量推定のためのネットワークの概略図 ([2]).

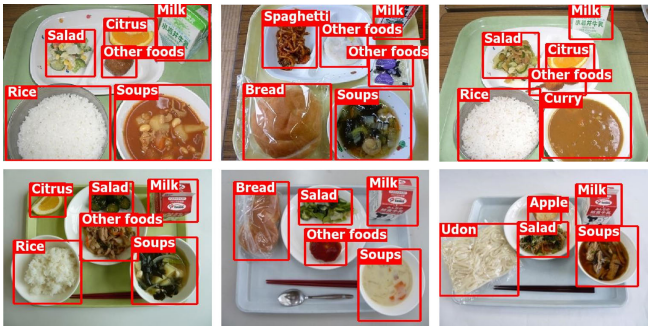

図 4　バウンディングボックス付き学校給食画像の例.

1 つ目のモジュールは, 候補領域を生成する Region Proposal Network(RPN) である. この RPN は 1 層の全結合層 (VGG16 [11] では 512 次元) と, それに続いて枝分かれする reg layer と cls layer で構成される. RPN の入力は, 特徴マップ (VGG16 では最後の畳み込み層の出力) 上の 3×3 の spatial window であり, 出力は後述する anchor の個数分のバウンディングボックスとそれに対応するスコアである. この出力のスコアは各バウンディングボックスにおいて物体が存在する確率を表している. 実際には RPN は, 3×3 のカーネルを持つ畳み込み層と, それに続いて枝分かれする 1×1 のカーネルを持つ畳み込み層により実装される. また, RPN では物体の形状に柔軟に対応するために 9 個の anchor を導入しており, この anchor は 3 スケール (128^2, 256^2, 512^2 ピクセル) と 3 アスペクト比 (2: 1, 1: 1, 1: 2) の形状をもつ. そして RPN では各 spatial window において, anchor を基準とする 9 個の候補領域を生成する.

2 つ目のモジュールは, RPN により生成された候補領域を洗練する Fast R-CNN detector [3] である. この Fast R-CNN detector の入力は RPN により生成された候補領域であり, 出力はバウンディングボックスとクラス確率である. このバウンディングボックスはクラスごとに推定される. このように Faster R-CNN は 2 つのモジュールにより構成され, 畳み込み層により実装される. 重要なことは RPN と Fast R-CNN detector の両方のネットワークは畳み込み層を共有しており, システム全体は単一のネットワークであるということである (図 2). 入力画像は一度だけ畳み込み層を通過し, RPN が生成する候補領域が Fast R-CNN detector により洗練される. 結果的に Faster R-CNN は高精度かつ高速な物体検出を実現した.

3.2　食事画像からのカロリー量推定

本研究では [2] に従い, 食事画像からのカロリー量推定を行う. [2] 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し, それにより食事画像からカロリー量を推定するネットワークを学習している. このネットワークの入力は単品料理の画像に限られ, また, 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. ここからは [2] に従い, 本研究で使用するネットワークについて説明する. 我々は [2] のカロリー量付き食事画像データセットを用い, 図 3 のネットワークを学習する. このネットワークは VGG16 [11] に基づいており, 出力層がカロリー量を出力する単一のユニットで構成され, 1 人分の量に対するカロリー量を出力する. 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, [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.

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

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

3.3　複数食事画像からの同時カロリー量推定

本研究では Faster R-CNN [7] とカロリー量推定 [2] を組み合わせることで複数料理画像からの同時カロリー量推定を行う. まず料理画像で学習した Faster R-CNN により複数料理画像中の単品料理のバウンディングボックスと料理カテゴリを推定する. 次に各単品料理のバウンディングボックスに基づき単品料理画像を切り出し, これをカロリー量推定ネットワークに入力として与えることで, 各単品料理のカロリー量が得られる. CNN の学習の際には複数枚の入力画像を一括で処理するようなバッチ処理が行われていることが多く, 今回のように切り出された単品料理画像が複数枚ある場合にも, それらをひとまとまりとして一括処理することが可能である. これにより各単品料理に関してバウンディングボックス, 料理カテゴリ, カロリー量が 1 枚の複数料理画像から推定される.

4.　データセット

本研究では複数料理画像からの単品料理の検出の評価のためにバウンディグボックスがアノテーションされた学校給食画像データセットを作成し, 実験に用いる. それに加え, バウンディングボックス付きの複数料理画像を含む UEC Food-100 [5] も使用する. また, 複数料理画像からの同時カロリー量推定の評価のために合計カロリー量がアノテーションされた学校給食画

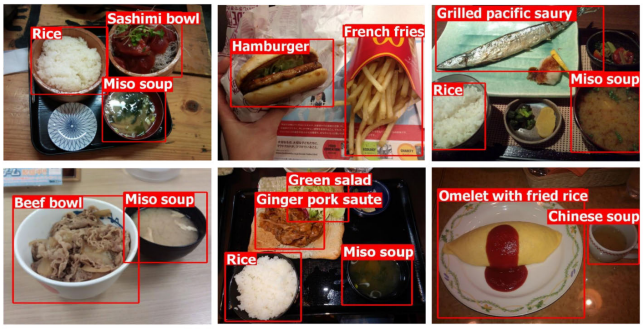

図 5　UEC FOOD-100 [5] に含まれる複数料理画像の例.

像データセットも作成する.

4.1　バウンディングボックス付き学校給食画像データセット

このバウンディングボックス付き学校給食画像データセットは本研究で作成した日本の給食の料理 21 カテゴリに関する食事画像データセットである. 合計 4877 枚の学校給食画像はすべて複数料理画像であり, バウンディングボックスが付与されている. これらの画像は, Web 上の給食センターが更新している給食ブログ(注1) (注2)から収集された. この給食ブログでは一日ごとに給食の画像と料理名, コメントなどが掲載される. このデータセットは Faster R-CNN [7] の学習と評価に使用される. 図 4 に学校給食画像データセットに含まれる画像の例を示す.

4.2　UEC Food-100

UEC FOOD-100(注3) [5] は料理 100 カテゴリに関する食事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. このデータセットは Faster R-CNN の学習と評価に使用される. 図 5 に UEC FOOD-100 に含まれる複数料理画像の例を示す.

4.3　合計カロリー量付き学校給食画像データセット

このデータセットの学校給食画像は Web 上の給食ブログ(注4) から収集された. この給食ブログでは一日ごとに給食画像と共に合計カロリー量の値が掲載される. 合計カロリー量とは, 個々の料理のカロリー量の合計値であり, 個々の料理のカロリー量は不明である. このデータセットは合計 690 枚の合計カロリー量付き学校給食画像を有し, 複数料理画像からの同時カロリー量推定の評価に使用される.

5.　実　　験

まず複数料理画像からの単品料理の検出の実験を行う. データセットとして 4.1 のバウンディングボックス付き学校給食画像データセットと 4.2 の UEC Food-100 [5] を使用する. 次に複数食事画像からの同時カロリー量推定の実験を行う. 評価には 4.3 の合計カロリー量付き学校給食画像データセットを使用する.

表 1　学校給食画像からの料理検出結果. 各カテゴリの AP(%) を示す.

Milk	99.6	Soups	92.2
Drinkable yogurt	90.6	Curry	95.1
Rice	99.7	Spicy chili-flavored tofu	99.8
Mixed rice	82.7	Bibimbap	72.9
Bread	95.5	Fried noodles	79.9
White bread	83.7	Spaghetti	90.7
Udon	98.0	Citrus	99.6
Fish	78.3	Apple	98.5
Meat	70.8	Cup desserts	93.1
Salad	94.0	Other foods	90.4
Cherry tomatoes	100.0	mean Average Precision	90.7

5.1　複数料理画像から料理の検出

本実験では Faster R-CNN [7] を料理画像で学習し評価を行う. 実装は主に [7] に従う. スクラッチ学習は行わずに VGG16 [11] の事前学習済みモデルを使用する. 最適化手法として SGD を使用し, momentum 値を 0.9, weight decay 値を 0.0005 とする. また, バッチサイズを 1 として学習率 10^{-3} において 50k 回反復し, その後, 学習率 10^{-4} において 20k 回反復する. RPN と Fast R-CNN detector の 2 つのモジュールを学習するためにネットワーク全体を Approximate joint training [8] により学習する. この Approximate joint training では, RPN により生成される候補領域は事前計算されたものとして扱われ, RPN の損失と Fast R-CNN の損失が独立に誤差逆伝搬され, 共有された畳み込み層において 2 つの損失が結合されて誤差逆伝搬される. 評価指標として PASCAL VOC detection task(注5) の標準的な評価指標である mean Average Precision(mAP) を使用する.

5.1.1　バウンディングボックス付き学校給食画像

4.1 のバウンディングボックス付き学校給食画像データセットを用いて Faster R-CNN の学習と評価を行う. 学習にはデータセットの 80%を使用し, 残りの 20%を評価に使用する. 表 1 に各料理カテゴリの Average Precision(AP) を示す. 表 1 に示す通り, mAP は 90.7%となり, 高精度に検出された. 図 6 に検出結果の例を示す. 学校給食画像の場合, 図 6 のように料理の量や盛り付け方, 背景などが統一されており, また, 学習データとテストデータに同じ給食センターで作られた給食が含まれていることなどが高精度の検出の要因であると考えられる.

5.1.2　UEC Food-100

4.2 の料理 100 カテゴリの食事画像データセットである UEC Food-100 [5] を用いて Faster R-CNN の学習と評価を行う. 学習には単品料理画像を使用し, テストには複数料理画像を使用する. また, 下田ら [9] の手法との比較を行う. 下田らは誤差逆伝搬により得られるサリエンシーマップに基づき領域分割を行い料理の検出を行った. [9] に従い, 100 カテゴリ, 53 カテゴリ (評価用サンプル数 ≧10 のカテゴリ), 11 カテゴリ (評価用サンプル数 ≧50 のカテゴリ) それぞれについての mAP を表 2 に示

(注1) : http://tate-cook.seesaa.net/
(注2) : http://blog.canpan.info/takizawa
(注3) : http://foodcam.mobi/dataset100.html
(注4) : http://inzai.ed.jp/kyusyoku/?page_id=32
(注5) : http://host.robots.ox.ac.uk/pascal/VOC/voc2012/index.html

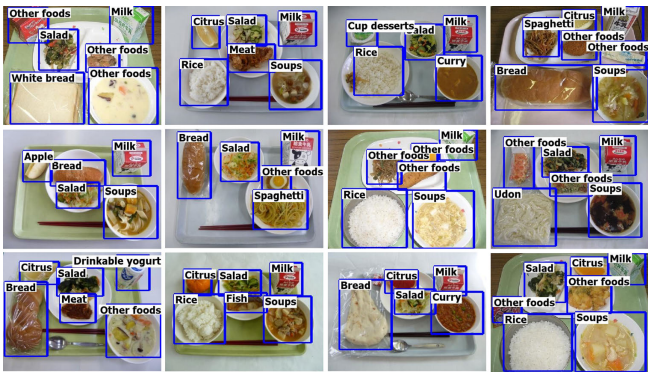

図 6　学校給食画像からの料理検出例.

表 2　UEC Food-100 [5] に含まれる複数料理画像からの料理検出結果 (料理カテゴリごとに AP を計算). 100 カテゴリ (全体), 53 カテゴリ (評価用データ数 ≧10 のカテゴリ), 11 カテゴリ (評価用データ数 ≧50 のカテゴリ).

UEC Food-100 mAP(%)	100 カテゴリ (all)	53 カテゴリ (#item≧10)	11 カテゴリ (#item≧50)
R-CNN (in [9])	26.0	21.8	25.7
[9]'s method (BP)	49.9	55.3	55.4
Faster R-CNN	42.0	46.3	57.9

表 3　UEC Food-100 [5] に含まれる複数料理画像からの料理検出結果 (料理カテゴリを無視し, 評価用データ全体で AP を計算). 100 カテゴリ (全体), 53 カテゴリ (評価用データ数 ≧10 のカテゴリ), 11 カテゴリ (評価用データ数 ≧50 のカテゴリ).

UEC Food-100 AP(%)	100 カテゴリ (all)	53 カテゴリ (#item≧10)	11 カテゴリ (#item≧50)
[9]'s method (BP)	57.3	58.0	58.8
Faster R-CNN	57.7	59.2	67.0

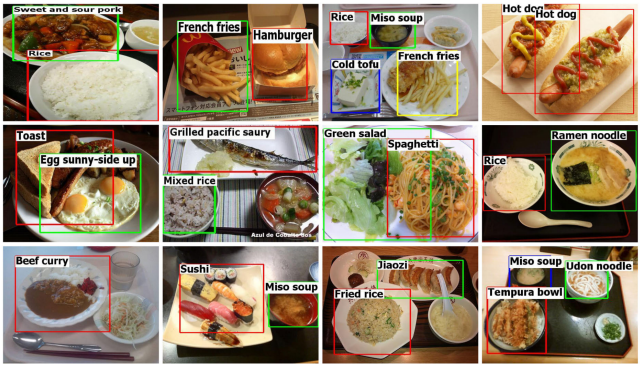

図 7　UEC Food-100 [5] に含まれる複数料理画像からの料理検出例.

す. また, 図 7 に検出結果の例を示す. 我々の手法は 11 カテゴリの mAP において [9] の精度を超え, 特に日本食に頻繁に現れるようなサンプル数の多い “ごはん” や “味噌汁” において高い精度での検出が見られた. [9] と比較すると, “ごはん” と “味噌汁” に関してそれぞれ, 60.0%から 90.2%, 68.3%から 80.2%の向上が得られた. さらに [9] に従い, 料理カテゴリを無視し, 全体で AP(%) を計算した結果を表 3 に示す. サンプル数の多い “ごはん” や “味噌汁” が強調されるため, 精度の向上が見られた.

5.2　複数料理画像からの同時カロリー量推定

本実験では Faster R-CNN [7] とカロリー量推定ネットワーク [2] を組み合わせ, 複数料理画像からの同時カロリー量推定を行う. そのため本実験では学習済みの Faster R-CNN とカロリー量推定ネットワークを個別に用意する必要がある. 複数料理画像からの料理の検出には 4.1 の学校給食画像を学習した Faster R-CNN を用いる. これは評価に用いる画像が学校給食画像であるためである. また, 単品料理画像からのカロリー量推定には, [2] において収集されたカロリー量付き食事画像で学習したカロリー量推定ネットワークを使用する.

[2] で構築されたカロリー量付き食事画像データセットは, Web 上のレシピ情報サイトから大量に収取されたものであったが, [2] では料理カテゴリ分類を行うために料理 15 カテゴリに含まれる画像のみを使用していた. この料理 15 カテゴリは本実験の評価に使用する学校給食画像中の料理にはほとんど対応しておらず, この料理 15 カテゴリのデータセットで学習したカロリー量推定ネットワークを学校給食画像に用いることは難しいと考えられる. そこで本研究では, 料理カテゴリを限定せず, [2] で収集された大量のカロリー量付き食事画像を使用し, カロリー量推定ネットワークの学習を行う. ノイズ除去として我々はまず 256×256 以下の画像を取り除いた. また, [2] のカロリー量推定ネットワークの入力は単品料理の画像にのみ対応しているため, UEC Food-100 により学習済みの Faster R-CNN により複数の料理が検出された画像の除去も行った. 最終的に我々は 55,020 枚のカロリー量付き食事画像を使用して図 3 に示すカロリー量推定ネットワークの学習を行った. 本実験ではカロリー量推定ネットワークの学習のために, 深層学習フレーム Chainer [12] を使用する. パラメータなどの値は主に [2] に従う. 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 10^{-3} において 150k 回反復し, その後 10^{-4} において 50k 回反復する.

こうしてそれぞれ学習された Faster R-CNN とカロリー量推定ネットワークを組み合わせることで複数料理画像からの同時カロリー量推定を行う. まず Faster R-CNN により複数料理画像中の単品料理のバウンディングボックスと料理カテゴリを推定する. 次に各単品料理のバウンディングボックスに基づき単品料理画像を切り出し, これをカロリー量推定ネットワークに入力として与えることで, 各単品料理のカロリー量が得られる. 最終的に推定された各料理のカロリー量から合計カロリー量を計算する. 評価には 4.3 の合計カロリー量付き学校給食画像データセットを使用する. 評価では, 単品料理ごとに推定されたカロリー量の正確さではなく, それらの合計カロリー量の正確さの評価を行う. ただしこの実験では Faster R-CNN により “牛乳” と分類された料理に関しては 134kcal と固定する. カロリー量推定ネットワークの学習データに学校給食の “牛乳” が含まれていることは考えづらく, また, サイズが定まっており, カロリー値に大きな変動はないと考えられるためである.

表 4 に複数料理画像からの同時カロリー量推定の結果を示す. 比較のために [2] の料理 15 カテゴリを対象とする単品料理画像に対するカロリー量推定の結果を示すが, 使用されるデー

表 4　複数料理画像からの同時カロリー量推定結果．単品料理カロリー量推定 [2] は料理 15 カテゴリのみを対象としており，使用したデータセットが異なるため参考値である．

	相対誤差 (%)	絶対誤差 (kcal)	誤差 20%(%) ≦	誤差 40%(%) ≦
単品料理カロリー量推定 [2]	30.2	105.7	43	76
複数料理合計カロリー量推定	21.4	136.8	53.0	85.1

図 8　複数料理画像からの同時カロリー量推定結果の例．バウンディングボックス中の数字は推定された各料理のカロリー量 (kcal) である．ES は推定されたカロリー量の合計値 (kcal) であり，GT は合計カロリー値の正解値 (kcal) である．

タセットが異なるため参考比較である．[2] に基づき，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た．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．表 4 からわかるように，相対誤差や相対誤差 20%/40%以内の推定値の割合に関して，[2] の料理 15 カテゴリの単品料理画像に対するカロリー量推定の結果を上回っている．本手法では合計カロリー量を計算しているため，各単品料理の推定カロリー量の誤差が緩和されているのではないかと考えられる．

図 8 に複数料理画像からの同時カロリー量推定結果の例を示す．図 8 のように各単品料理のカロリー量が推定され，それらの合計値が計算される．同一の料理を複数回検出してしまったり，料理が存在しない領域を検出してしまうような過剰な検出により，計算された合計カロリー量が正解値よりも高い値になってしまう例が多く見られた．密集する同一カテゴリのバウンディングボックスの統合や，非料理画像 (0kcal) の学習によりこれに対処することができるのではないかと考えられる．

6. おわりに

本研究ではまず Faster R-CNN [7] による複数料理画像からの単品料理の検出を行い，これとカロリー量推定 [2] を組み合わせることで，複数料理画像からの同時カロリー量推定を行った．また，Faster R-CNN の学習のためのデータセットと複数料理画像からの同時カロリー量推定の評価に必要なデータセットをそれぞれ作成した．今後の課題として，物体検出とカロリー量推定を単一のネットワークで行い同時学習することを考えている．

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] R. Girshick. Fast R-CNN. In *Proc. of IEEE International Conference on Computer Vision*, 2015.

[4] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2014.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

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

[12] S. Tokui, K. Oono, S. Hido, and J. Clayton. Chainer: a next-generation open source framework for deep learning. In *Proceedings of Workshop on Machine Learning Systems (LearningSys) in The Twenty-ninth Annual Conference on Neural Information Processing Systems (NIPS)*, 2015.