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を用いた複数品食事画像の領域分割とカロリー推定

下田　　和†　柳井　啓司†

† 電気通信大学 大学院情報理工学研究科 総合情報学科

あらまし　食事認識によるカロリー推定は多くの場合，クラス分類のみを行い，標準カロリーを提示している。しかし，食事のカテゴリだけでなく，食事の量もカロリーに大きく影響するため，この推定結果には不十分な側面がある。そこで，本研究では Convolutional Neural Network(CNN) による領域分割結果を用いて，食事領域の面積の違いから，食事量を考慮したカロリー推定を行った。

キーワード　Convolutional Neural Network，領域分割

Wataru SHIMODA† and Keiji YANAI†

† Department of Informatics,　The University of Electro-Communications,　Tokyo

1. はじめに

近年，健康促進のために食事記録が行われるようになったが，手動による食事の記録には手間がかかるため，食事記録を助けるサービスが必要とされている。食事記録の手間を減らす方法として，食事の前に撮影をした写真を用いて画像認識を行うことで，自動的にカロリーを推定するものがある。ただし，この手法の多くはクラス分類タスクに基づいており，食事のカテゴリのみを認識し，認識した食事の標準カロリーを提示している。従って，提示されたカロリーには食事の量が反映されていない。食事量は食事の種類と同様に，食事から摂取するカロリーに大きく影響する。より正しい食事記録をつけるためには，食事のカテゴリの認識だけでなく，食事量の推定が必要である。画像認識における重要なタスクとして，クラス分類の他に，物体検出タスク，領域分割タスクが知られている。クラス分類タスクにおいては，物体のカテゴリのみを認識するが，物体検出タスク，領域分割タスクにおいては物体の位置を認識する必要がある。特に，領域分割タスクにおいては，ピクセル単位での認識を行うので，画像における物体の占める面積を知ることができる。この物体の面積から，物体の大きさを推定することができるため，領域分割を食事画像の認識に応用することで食事量の推定が可能である。

画像認識における物体検出タスク，領域分割タスクは画像認識における重要なタスクである。物体検出とは，画像における物体に対して位置を推定しバウンディングボックスを付与するタスクであり，一般的なクラス分類と比べて難しいとされている。領域分割は，画像における物体のオブジェクトに対してピクセル単位のラベリングを行うタスクであり，さらに発展的なタスクであると考えることができる。物体検出，領域分割の精度向上はコンピュータによる物体の位置の理解，形状の理解に繋がり，様々な分野における研究の貢献に期待できる。

近年，Convolutional Neural Network（CNN）がクラス分類タスクにおいて最も精度がよいとされ，注目されている。2014 年に ImageNet Large Scale Visual Recognition Challenge(ILSVRC) のコンペティション，1000 種類のクラス分類タスクにおいて [1] が，これまでの手法に大差をつけて 1 位となり，画像認識分野において広く知られるようになった。また，CNN はクラス分類以外のタスクにおいても有効であることがわかっている。特に，近年の領域分割の性能は CNN によって大幅に向上し，20 %以上の精度が改善された。しかし，これらの高精度な領域分割手法の多くはピクセル単位のアノテーションの情報を必要としている。ピクセル単位のアノテーションを画像に付与するのはかなりの労力が必要となりコストがかかる。もし，ピクセル単位のアノテーションを必要としない領域分割を実現することができれば，領域分割対象のカテゴリを容易に増やすことができるようになるはずである。一方で，物体検出タスクにおいてはピクセル単位のアノテーションの情報を用いずに CNN の認識精度のみで性能を大幅に向上させている。領域分割においてもピクセル単位のアノテーションを用いない弱教師有学習による精度が改善される可能性がある。

CNN は単純なクラス分類精度のよさのみでなく，階層的な認識を行うという側面で，これまでの SURF や HOG といった局所的な特徴量と異なっている。CNN は認識を行う際に一方向に一層ずつ階層を進むが，階層を進む際に非線形処理により有用な情報の取捨選択を行っている。この情報の取捨選択には粗い位置情報が含まれており，これを分析することにより，CNN が認識をする際に反応の強かった位置を推定することができる。これは，Deconvolution や Back Propagation といった CNN の認識結果を可視化する手法として用いられ，広く知られるようになった。特に，Back Propagation [2] は領域分割に応用できることが知られている。[2] の手法は弱教師有学習による領域分割が可能であるが，精度はあまり高くない。領域分割の精度を向上させる手法として, [3], [4] などの領域のプロポーザルを用いる手法がある。そこで，本研究では Back Propagation と

プロポーザルを組み合わせた弱教師有学習による画像の領域分割を行った。

2. 関連研究

本研究では CNN が階層的な認識を行っている点に着目し，CNN が認識を行う際に用いている特徴量の粗い位置の情報から，物体の領域分割を行う。CNN が認識を行う際に反応した位置を可視化する方法として，逆畳み込みを行い，畳み込みから元の値を復元する Deconvolution [5] が広く知られている。また，CNN の認識結果の可視化の類似手法として，誤差逆伝搬法，Back Propagation(BP) を応用したものがある。誤差逆伝搬法は，CNN が階層的なパラメータの学習を行う際に一般的に用いられている手法である。CNN は出力における認識結果と真値との誤差を，各階層に伝搬することでそれぞれの階層におけるパラメータの最適化を行う。可視化の際には，この誤差を最大に設定し，逆伝搬を行う。この誤差は出力の時点ではスカラーであるが，各階層に伝搬するごとに次元が変化し，入力と同じ次元になる。CNN はこの誤差を最小化するために，誤差を伝搬させるはずである。よって，この画像レベルの伝搬値は CNN の反応した箇所の値が大きくなると考えることができる。

Simonyan らはバックプロパゲーションにより得られた結果をサリエンシーマップとし，伝搬値の値の大きい部分をグラブカットのポジティブ要素として渡すことで領域分割を行った [2]。この手法による領域分割は，認識をしてから画像に映っている物体のカテゴリを推測し，一度逆伝搬をするだけで容易に領域分割を行うことが可能である。しかし，グラブカットに渡す領域は固定値であるので，シーン画像のようなサリエンシーマップに複数ピークのある画像では精度が下がってしまうという欠点がある。

シーン画像の領域分割の精度をあげる手法としては領域候補をプロポーザルする手法がある。CNN とプロポーザルを組み合わせたものとしては，Rich feature Convolutional Neural Network(RCNN) [3]，Simultaneous Detection and Segmentation(SDS) [4] が代表的である。RCNN は SelectiveSerch [6] によりバウンディングボックスの領域候補をプロポーザルし，CPMC を用いて [7] それぞれの領域を分割し，CNN で認識を行いピクセル単位のラベリングを行った。SDS は，MCG [8] により，領域分割された領域候補をプロポーザルし，それぞれの領域を CNN で認識している。RCNN，SDS にはすでに分割された領域を CNN で認識し，オブジェクトのクラスのラベリングをするという共通点があり，これが以前の手法と比べて高精度となった要因であった。しかし，これらの手法では，オブジェクトの領域候補をプロポーザルする際に，CNN は用いられていないので，CNN の階層的な認識における情報が付与されていない。CNN の階層的な認識による情報が領域分割において効果的であることは先行研究で明らかになっている。特に，近年の領域分割は Fully Convolutional Netwark [9] の考え方に基づいた研究が盛んにされている。[9] は通常の CNN のネットワークに Deconvolution 層を加えることで高精度な領域分割を実現している。

[4] [9] などの高精度な領域分割は完全教師有学習であるので，ピクセル単位のアノテーション情報を必要としている。ピクセル単位のアノテーション情報の付与は，オブジェクトのラベル付け，バウンディングボックスの付与と比べて，非常に労力のかかる作業である。大規模なデータセットを作るのは困難であり，ビッグデータを利用した研究などとも相性が悪いという欠点がある。

CNN を用いた弱教師有学習による領域分割としては Pedro らの研究 [10] がある。[10] は CNN の畳み込み層のユニットが共通の情報を保持していることに着目し，少ない試行回数の CNN の認識結果からヒートマップを作成し，位置推定を行った。また，MCG [8] によるプロポーザルを用いて，領域の平滑化を行い精度を改善した。[10] は Overfeat [11] を用いて，大きめの画像を入力とし出力をマップ形式にしているため，CNN の階層的な認識による位置情報とは異なる手法で物体の位置推定を行っている。

3. 実験の概要

3.1 領域分割の手順

以下の手順に従って，図 1 のようにして領域分割を行った。

（1） Selective Serch により 2000 程度の領域をプロポーザルする

（2） バウンディングボックスの重なり率からグルーピングを行う

（3） 各領域においてバックプロパゲーションを行い，領域グループで平均をとり，サリエンシーマップを得る

（4） サリエンシーマップを用いて各領域グループでグラブカットを行う

（5） 分割された領域を CNN で認識し，ラベリングを行う

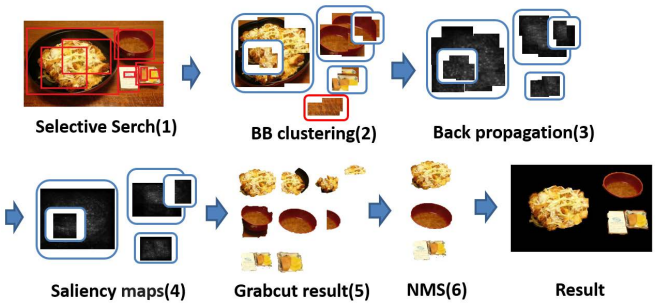

図 1 提案手法の流れ

4. 領域分割

誤差逆伝搬法を用いて，CNN の反応している位置を推定し，領域分割を行った。

4.1 バックプロパゲーション（誤差逆伝搬法）

CNN は学習を行う際に，真値との誤差を変化量として微分を行い，最適なパラメータになるように調節をする。CNN は階層構造になっているので，下の階層のパラメータを変化させるためには，誤差を伝搬させる必要がある。バックプロパゲーションはこの誤差の伝搬に用いられる手法である。図 2 は CNN の階層的なネットワークの図である。一般に，認識の際に入力から出力へと向かう処理を forward，出力から入力へと向かう逆伝搬の処理は backward と呼ばれている。

4.2 バックプロパゲーションによる位置推定

CNN の反応した位置の可視化の際には，誤差を最大に設定し，逆伝搬を行う。バックプロパゲーションはスカラー値の誤差を元に伝搬を行うが，この誤差の伝搬は各レイヤーにおいてそれぞれのレイヤーの入力の次元と同じになる。従って，畳み込み層における伝搬では，誤差は位置情報を得ることになる。この誤差を入力まで伝搬させることにより，誤差は画像と同じ

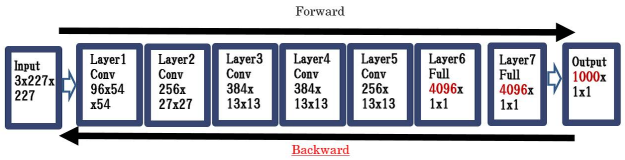

図 2　CNN のネットワーク

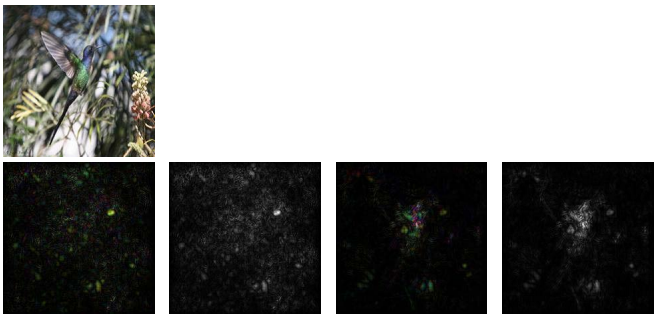

図 3　(1)BP，(2) サリエンシーマップ (BP)，(3)GBP，(4) サリエンシーマップ (GBP)

入力の次元（271*271*3）となる。そこで，この伝搬値を w，i，j を画像中におけるピクセル，c をカラーチャンネルとし，

$$M_{i,\,j} = max_c|w_{(i,\,j,\,c)}| \tag{1}$$

とすることで，グレースケールに変換し，サリエンシーマップとした。伝搬値は CNN が学習を行う際にパラメータを変化させたい部分の伝搬値を大きくしようとするはずであるので，このサリエンシーマップは CNN の認識に影響を与えた位置を示していると考えることができる。

また，バックプロパゲーションは非線形活性化関数である ReLU(Rectifier Linear Unit) での伝搬を変更することで異なる結果が得られることが知られている。Guided Back Propagation(GBP) [12] は，正の伝搬値を強調することによって通常のバックプロパゲーションとは異なる結果を得る手法である。入力を x，伝搬値を dz，出力を y とすると，バックプロパゲーションにおける relu での伝搬の式は

$$y = dz * (x > 0) \tag{2}$$

一方，GBP における relu での伝搬の式は

$$y = dz * (x > 0) * (dz > 0) \tag{3}$$

となる。GBP はバックプロパゲーションと比較して，反応した箇所がより強調される結果となる。しかし，疎な要素が多くなり，エッジが強調される傾向にある。密な結果を返す BP と疎な結果を返す GBP を組み合わせることで精度の改善が期待できる。そこで，BP のみのサリエンシーマップによる領域分割を行うケースと，BP と GBP のサリエンシーマップを足し合わせたものの二つの場合における領域分割を試した。図 4 は BP，GBP それぞれから得られるカラーマップとサリエンシーマップの例である。

4.3　グラブカット

サリエンシーマップを元に領域分割を行うために [2] と同様に，GrabCut [13] を用いた. [2] などの従来の手法では，色情報から得られる Gaussian Mixture Model(GMM) の値をさり園しーマップの上位のピクセルを fix 値 (GMM の最大値) で置き換え，これを Graph cut の seed としている。しかし，上位の値を fix 値で与えると，サリエンシーマップの情報量は減少してしまう。サリエンシーマップによる情報を最大限に活用し，結果に影響を与えることができればよりよい結果になりうる可能性がある。そこで，今回は GMM の値を fix 値に置き換える手法だけでなく，サリエンシーマップを map 値として GMM に加えることで，精度の改善を試みた。また，GMM に加えるサリエンシーマップについては，いくつかの関数 (上に凸や下に凸な関数など) による変換を試し結果の違いについて検証した。

グラブカットの確率モデルは，単項と共通項からなり，エネルギー関数は以下のようになる。y_i を画像におけるピクセル i，y を全ての y_i ピクセルにおけるベクトルとすると，エネルギー関数は，以下の式で書ける。

$$E(y) = \sum U_i(y_i) + \sum V_{i,\,j}(y_i, y_j) \tag{4}$$

サリエンシーマップを M_i，サリエンシーマップに適用する関数を F とすると，

$$\delta_i = \begin{cases} max(GMM) & (M_i > 0.95 * max(M)) \\ GMM_i & (\text{上記以外}) \end{cases}$$ として以下の式により，前景の確率場に対してサリエンシーマップの値を与えた。

$Ufore_i(fix) = \delta_i$

$Ufore_i(map) = GMM_i + max(GMM) * F(M_i)$

背景の確率場に対しては

$$\delta_i = \begin{cases} max(GMM) & (M_i < 0.10 * max(M)) \\ GMM_i & (\text{上記以外}) \end{cases}$$ として以下の式により，サリエンシーマップの値を与えた。

$Uback_i(fix) = \delta_i$

$Uback_i(map) = GMM_i + max(GMM) * F(1 - M_i)$

4.4　領域候補のラベリング

グラブカットによる結果は，二値分類であるので，得られる領域には前景と背景のラベルの情報のみが付与されている。しかし，領域分割においては，各カテゴリの情報を付与する必要がある。そこで，本研究では，グラブカットにより得られた領域を再度 CNN で認識することによって各領域にラベル付けを行った。グラブカットにより得られた領域を $region$，この領域を CNN で認識することで得られる特徴を CNN_{region}，前景の領域に接する直線による最小のバウンディングボックスによる領域を box，この領域を CNN で認識することで得られる特徴を CNN_{box} とする。

$$F = CNN_{box} + CNN_{region} \tag{5}$$

分割された領域とバウンディングボックスから得られる CNN の値を足し合わせてこれをラベリングの基準値とした。

4.5　CNN の学習

CNN の学習は，Image net のカテゴリ 1000 種類と Image net における food 系列のカテゴリ 1000 種類の計 2000 種類の画像を用いて pre-traing した。そして，pre-traing したモデルについて，食事画像 100 カテゴリ (各約 100 枚)+非食事画像 1 カテゴリ (約 2 万枚)，計 101 カテゴリからなる UECFOOD101 食事画像データセットを用いて fine-tuning を行った。また，caffe ツールを用いて，一般的な alex net [1] と同様のネットワークによる学習を行った。

5.　領域候補のプロポーザルと統合

グラブカットは強力な領域分割の手法であるが，シーン画像

についてはあまり有効ではない。サリエンシーマップのピークが複数のオブジェクトにまたがって検出されると，オブジェクトが繋がって領域分割されてしまう。そこで，このグラブカットの結果を改善するために，本研究では，セレクティブサーチ [6] を用いた。セレクティブサーチは，画像から約 2000 のオブジェクトの領域候補を提案する。領域を制限し，領域分割を行うことで，複数のオブジェクトについてサリエンシーマップのピークが出てくる可能性を抑えることが期待できる。最終的には，2000 の領域候補から，約 25 ほどの領域に絞り，これを統合した。

5.1 Selective Serch による領域候補のプロポーザル

領域候補の提案には RCNN と同じ Selective Search [6] を用いた。一枚の画像から約 2000 のバウンディングボックス候補をプロポーザルし全ての領域候補についてバックプロパゲーションを行った。

5.2 領域候補のグルーピング

Non Maximum Suppression(NMS) を応用し，バウンディングボックスのグルーピングを行った。バウンディングボックスの面積の大きさを基準値として，このバウンディングボックスのオーバーラップ率を NMS と同様のアルゴリズムで計算し，バウンディングボックスを複数のグループに分割した。

5.3 領域候補の統合

グラブカットにより得られた領域にバウンディングボックスを付与する。それぞれについて，領域とバウンディングボックスを CNN で認識し，ラベリングを行う。また，得られた CNN の値に基づき，バウンディングボックスの領域について NMS を適用し，領域の統合を行った。

6. 領域分割を用いた食事量の推定

領域分割の結果は，画像における物体の大きさの情報を含んでいる。食事画像の領域分割の結果から食事の領域の面積から食事の量を推定する。

6.1 面積比による食事量の推定

領域分割結果における物体の面積は相対的なものであるが，物体間の面積の比率は不変であり，実際の食事量を反映したものになる。そこで，本研究では，領域分割の結果の面積比から複数食事画像の食事の量を推定した。

6.2 基準物体の決定

領域分割の結果の面積の比率から食事量の推定を行うが，面積の比率から実際の食事の量を計算するためには，基準となる値が必要となる。そこで，本研究では，食事の量に変化が少ないことで知られている味噌汁などの食事の量は一定であると考え，この食事の面積を基準として，食事量の推定を行った。各食事のカテゴリについて基準の物体となる優先順位を決定しておき，この優先順位にもとづき基準物体を決定した。

6.3 凸包による領域分割結果の改善

領域分割の結果は中心部が空洞になってしまう場合や，器の一部がかけてしまうことがあった。中には領域の面積に大きな影響を与える結果もあった。こういった分割に失敗した領域は複雑な形状をとるのに対して，実際の食事領域は食器の形に影響を受けるので，円や楕円などの単調な形である場合が多い。そこで，凸包による食事領域の改善を行い，カロリー推定を行った。

7. 実 験 結 果

100 種類の食事カテゴリからなる食事画像データセット UEC-FOOD101（特に複数品を含んでいる食事画像）と 20 種類の一般画像カテゴリからなる PASCAL VOC 2007 及び，PASCAL VOC 2012 のデータセットについて実験を行った。特に，UECFOOD101，PASCAL VOC 2007 については物体検出についての評価を，PASCAL VOC 2012 については領域分割についての評価を行った。これら評価方法の違いは，データセットの正解画像に付与されている情報の違いによるものであり，UECFOOD101 と PASCAL VOC 2007 にはバウンディングボックス情報のみのアノテーションが，PASCAL VOC 2012 にはピクセル単位のアノテーションがされている。

7.1 UECFOOD101 複数品食事画像における精度評価

UECFOOD101 における物体検出の精度評価を行った。また，BP によるサリエンシーマップと BP+GBP によるサリエンシーマップの 2 パターン，サリエンシーマップをグラブカットに渡す際の方法をそれぞれ 6 パターンの計 12 パターンについての実験を行った。

また，UECFOOD101 は各カテゴリ 100 枚，計 10000 枚，そのうち複数品食事画像は約 1000 枚である。ただし，複数品食事画像に含まれる食事カテゴリの数には偏りがある。そのため，0 枚以上（全て）のカテゴリにおける精度，10 枚以上のカテゴリにおける精度，50 枚以上のカテゴリにおける精度の 3 つの場合について精度比較を行った。

表 1　UECFOOD101 における物体検出精度の評価

method	100 class $\#item \geqq 0$	53 class $\#item \geqq 10$	11 class $\#item \geqq 50$
BP map			
fix value	41.2	45.5	50.5
x	28.6	32.8	32.3
$x^{1/2}$	25.5	29.3	28.2
$tanh(x)$	28.0	31.7	31.0
x^2	32.6	35.8	37.7
$-log(x)$	32.8	34.8	36.3
BP + GBP map			
fix value	40.3	44.5	49.7
x	29.0	32.0	32.2
$x^{1/2}$	25.8	29.2	27.5
$tanh(x)$	29.0	31.8	31.2
x^2	33.4	35.2	38.4
$-log(x)$	31.7	33.2	36.6

表 2 は UECFOOD101 の物体検出における本手法と RCNN の精度の比較である。

表 2　UECFOOD101 における RCNN との比較

method	100 class $\#item \geqq 0$	53 class $\#item \geqq 10$	11 class $\#item \geqq 50$
x	28.6	32.8	32.3
$tanh(x)$	28.0	31.7	31.0
x^2	32.6	35.8	37.7
fix	41.2	45.5	50.5
RCNN [3]	26.0	21.8	25.7

7.2 カロリー推定

領域分割の結果に基づいて，複数品の食事画像の食事量を計算し，カロリー推定を行った。図 8 はその結果である。また，図 9 は凸包により食事領域が改善された例である。なお，現時点ではテスト食事画像の正解カロリーデータがないため，推定結果の評価は行っていない。

7.3 PASCAL VOC 2007 における物体検出精度の評価

表 3 は，PASCAL VOC 2007，物体検出タスクにおける RCNN [3] との比較である。物体検出は PASCAL の基準に基づいて精度評価を行った。真のバウンディングボックスとのオーバーラップが 50 %以上であれば正解，それより下であれば不正解となる。これのm AP を計算する。

表 3　PASCAL VOC 2007 における RCNN との物体検出精度比較

method	mAP on PASCAL VOC 2007
RCNN [3]	54.2
ours	58.7

7.4 PASCAL VOC 2012 における領域分割精度の評価

PASCAL VOC 2012 データセットについて領域分割の実験を行った。精度は PASCAL の基準に基づいて，評価を行った。画像のピクセルを i，j として，以下の式により Mean IU を計算している。

$t_i = \sum_i n_{i,\,j}$

$acc = \sum_i n_{ii} / \sum_i ti$

ピクセル単位の評価を行っており，領域が大きすぎても，小さすぎても精度は低下する。また，ラベリングが間違っている場合は極端に低い精度となる。

また，表 4 は，PASCAL VOC 2012 における最新手法との比較である。

表 4　PASCAL VOC 2012 における SDS との領域分割制度の比較

method	mean IU on PASCAL VOC 2012
fully supervised	
SDS [4]	51.6
FCN [9]	62.2
weakly supervised	
ours	36.4
Pedro-seg [10]	40.6

8. 考　　察

UECFOOD101 について，fix 値でなくサリエンシーマップを GMM に加える手法を試した。結果としては，これらの工夫は精度の改善には繋がらず，バックプロパゲーションにより得られたサリエンシーマップの上位の値を fix 値としてグラブカットを行う既存の手法の精度には及ばなかった。これは，fix 値ではオブジェクトの中央付近のポジティブの要素を強くするのに対して，map 値ではオブジェクトのエッジ，境界付近における影響を強めてしまったのが精度低下の原因のようであった。図 4 は fix 値と map 値による領域分割の例である。map 値を GMM に加える際に，いくつかの関数による変換を行ったが，fix 値により近い結果となる下に凸な関数を用いたほうがよい精度になっていることがわかる。グラブカットではサリエンシーマップの情報を最大限に反映させることは難しいことがわかる。

図 6 は複数品食事画像における RCNN と本手法との比較結果である。RCNN の結果はテクスチャの認識結果の影響を受けている関係か，バウンディングボックスが小さめに出ている。一方，本手法は領域分割を行ってからバウンディングボックスを付与しているので，よい精度で物体の検出が行うことができている。また，図 7 は複数品食事画像における領域分割の結果である。

また，領域分割の結果からカロリー推定を行った。食事の量を考慮したカロリー推定を行うことができた。しかし、見きれていたり、領域分割に失敗し、本来より小さくなっている食事領域を基準としカロリー推定をした場合、極端な結果になってしまう場合があった。図 8 の最下段がその例である。味噌汁の領域分割に失敗し、小領域をカロリー計算の基準としてしまったため、炊き込みご飯のカロリー推定結果が極端に大きなものとなってしまった。今後の改善の課題としたい。

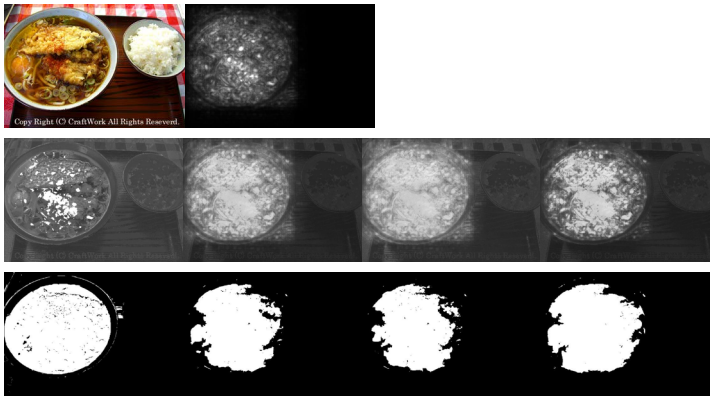

図 4　top:元画像とサリエンシーマップ，middle:GMM+サリエンシーマップ，bottom:グラフカットの結果，(1)fix value，(2) x (3) $tanh(x)$（上に凸） (4)x^2 （下に凸）

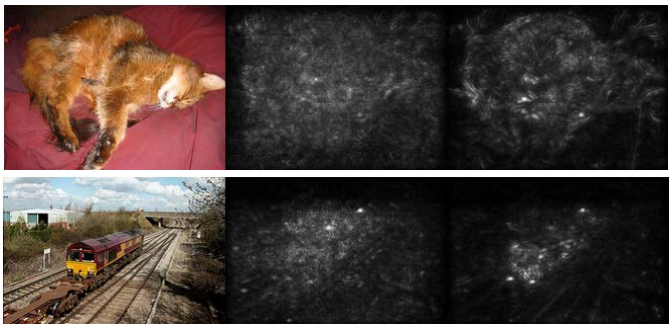

図 5　サリエンシーマップの比較 (1)original image，(2) BP map，(3)BP + GBP map

9. まとめと今後の課題

CNN の逆伝搬を利用した食事画像の領域分割を行った。既存の物体検出手法より良い精度で物体の一を推定することができた。また，領域分割の結果から，食事量を考慮したカロリー推定を行った。ただし，正解データがないため，精度評価を行うことができていない。今後は，領域分割の精度向上を図り，カロリーデータの評価方法を考えたい。

文　　献

[1] A. Krizhevsky, I. Sutskever, and G E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, 2012.

[2] K. Simonyan, A. Vedaldi, and A. Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. In *ICLR 2014 Workshop Track*, 2014.

[3] 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.

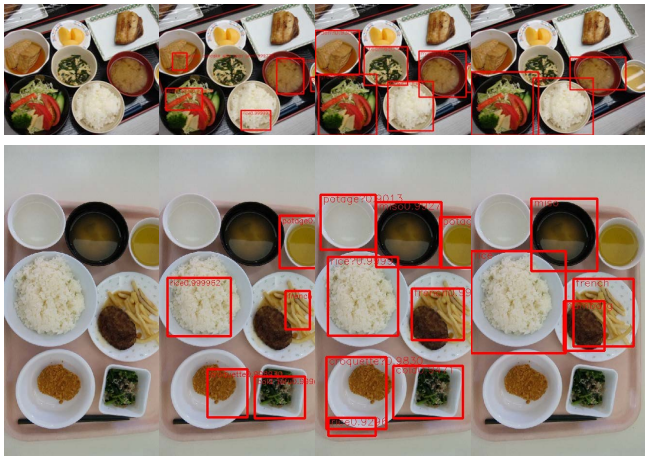

図 6　食事画像における RCNN との比較 (1)original image, (2) rcnn result (3) ours (4) ground truth

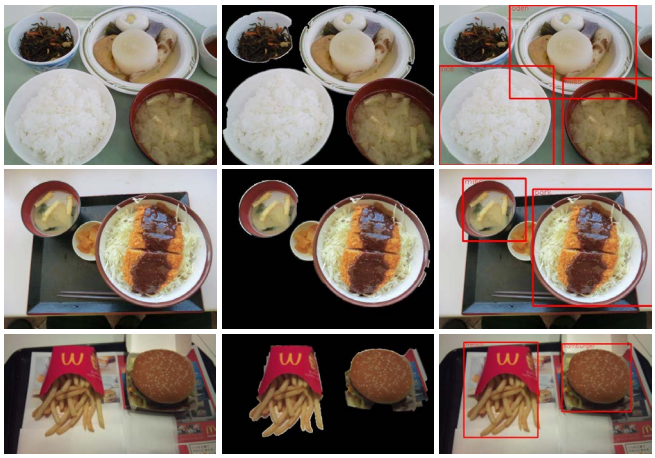

図 7　食事画像の領域分割結果 (1)original image, (2) segmentation result

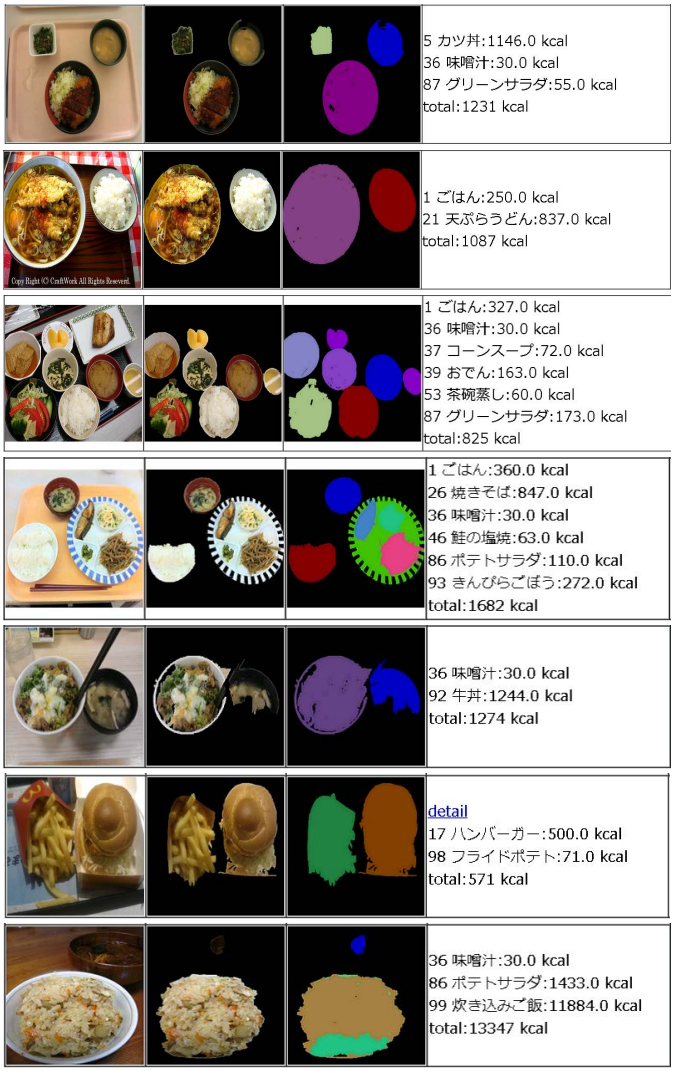

図 8　カロリー推定の結果 (1)original image, (2) segmentation result, (3)outline, (4)estimated calory

図 9　凸包による食事領域の (1)original image, (2) segmentation result, (3)outline, (4)conv hull, (5)calory estimation

[4] B. Hariharan, P. Arbeláez, R. Girshick, and J. Malik. Simultaneous detection and segmentation. In *Proc. of European Conference on Computer Vision*, 2014.

[5] M. Zeiler and R. Fergus. Adaptive deconvolutional networks for mid and high level feature learning. In *Proc. of IEEE International Conference on Computer Vision*, 2011.

[6] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders. Selective search for object recognition. Vol. 104, pp. 154–171, 2013.

[7] J. Carreira and C. Sminchisescu. Constrained parametric min-cuts for automatic object segmentation. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2010.

[8] A. Pablo, Jonathan T. Jordi, P., M. Ferran, and M. Jitendra. Multiscale combinatorial grouping. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2014.

[9] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation . In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2015.

[10] P. Pedro and C. Ronan. From image-level to pixel-level labeling with convolutional networks. In *arXiv:1411.6228*, 2014.

[11] S. Pierre, E. David, Z. Xiang, M. Michael, F. Rob, and L Yann. Overfeat: Integrated recognition, localization and detection using convolutional networks. In *Proc. of International Conference on Learning Representations*, 2014.

[12] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller. Striving for simplicity: The all convolutional net. In *ICLR 2015 Workshop Track*, 2015.

[13] Y. Boyokov and M. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2001.