

CNN を用いた弱教師学習による画像領域分割

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 Convolutional Neural Network (CNN) は現在画像認識分野においてもっとも注目されている手法の一つである。CNN はこれまで主に画像認識に用いられてきた局所特徴量とは異なり、階層的な認識を行うという側面がある。また、この階層それぞれにおける認識の結果から、物体の位置の推定を行うことが可能である。本研究では、誤差逆伝搬法 (バックプロパゲーション) を用いて、この CNN の階層的な認識から得られる情報を得て、物体の位置の推定をし、画像の領域分割を行う。バックプロパゲーションによる位置の推定はピクセル単位の学習画像を必要とせず、認識と誤差の伝搬を一度行うのみでよいので、容易に領域分割を行うことができるという特色がある。

キーワード Convolutional Neural Network, 領域分割

Region Segmentation with Convolutional Neural Network by Weakly-Supervised Training

Wataru SHIMODA[†] and Keiji YANAI[†]

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

1. はじめに

画像の物体検出、領域分割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スマップとし、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ス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 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 [6] 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 [7] 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 Network [8] の考え方に基づいた研究が盛んにされている。[8] 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] [8] 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、非常に労力がかかる作業である。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 [9] がある。[9] 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 [7] 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。[9] は Overfeat [10] 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、サリエンスマップを得る
- (4) サリエンス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

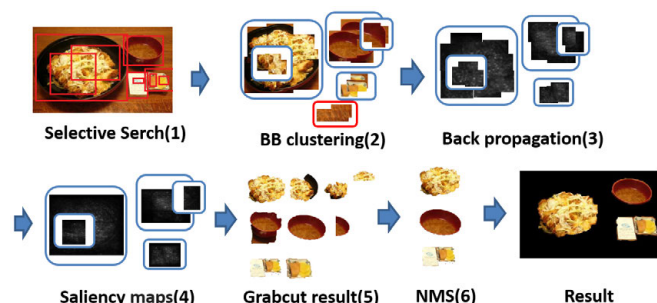


図 1 提案手法の流れ

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。

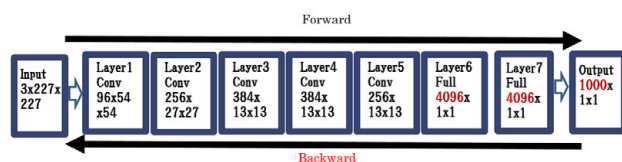


図 2 CNN のネットワーク

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

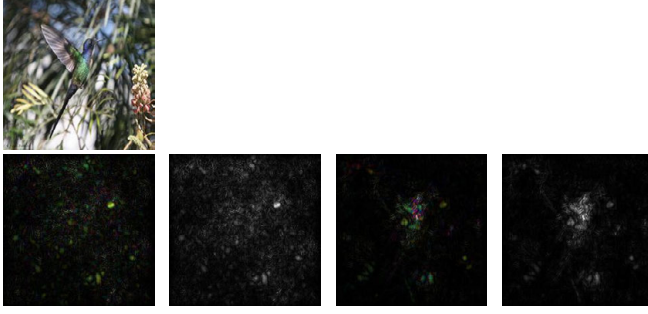


図3 (1)BP, (2) サリエンスマップ (BP), (3)GBP, (4) サリエンスマップ (GBP)

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、誤差は画像と同じ入力の次元 ($271 \times 271 \times 3$) 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、

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

とすることで、グレースケールに変換し、サリエンスマップとした。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、このサリエンスマップは CNN の認識に影響を与えた位置を示していると考えられることができる。

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) [11] 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

$$y = dz * (x > 0) \quad (2)$$

一方、GBP における relu での伝搬の式は

$$y = dz * (x > 0) * (dz > 0) \quad (3)$$

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 のみのサリエンスマップによる領域分割を行うケースと、BP と GBP のサリエンス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3は BP, GBP それぞれから得られるカラーマップとサリエンスマップの例である。

4.3 グラブカット

サリエンスマップを元に領域分割を行うために [2] と同様に、GrabCut [12] を用いた。今回は、サリエンスマップをグラブカットに渡す手法を二通り試した。[2] ではサリエンスマップの上位の値を fix 値としてこれを GMM に与え、アルファカットにより領域を拡張している。しかし、上位の値を fix 値で与えると、サリエンスマップの情報量は大量に失われてしまう。サリエンス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 の確立場に fix 値ではなく、そのままバックプロパゲーションにより得られた map 値を与える手法を用いた。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。

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

サリエンスマップを M_i ,

$$\delta_i = \begin{cases} \max(GMM) & (M_i > 0.95 * \max(M)) \\ GMM_i & (\text{上記以外}) \end{cases} \quad \text{として以下}$$

下の式により、前景の確率場に対してサリエンスマップの値を与えた。

$$U_{fore_i}(fix) = \delta_i$$

$$U_{fore_i}(map) = GMM_i + \max(GMM) * M_i$$

背景の確率場に対しては

$$\delta_i = \begin{cases} \max(GMM) & (M_i < 0.10 * \max(M)) \\ GMM_i & (\text{上記以外}) \end{cases} \quad \text{として以下}$$

下の式により、サリエンスマップの値を与えた。

$$U_{back_i}(fix) = \delta_i$$

$$U_{back_i}(map) = GMM_i + \max(GMM) * (1 - M_i)$$

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。

$$F = CNN_{box} + CNN_{region} \quad (5)$$

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

本研究では RCNN との比較実験の為に一般に公開されている既存の学習済みモデルを用いた。ImageNet 1000 種類のカテゴリにより pretraining されたモデルを Pascal VOC 2012 データセットの背景を含む 21 カテゴリでファインチューニングされたものを使った。誤差の逆伝搬にはピクセル単位のアノテーション情報とバウンディングボックス情報は必要ない。しかし、今回の実験で用いた RCNN の学習済みモデルは Pascal データセットの BoundingBox 情報を使って finetuning が行われている。

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。サリエンス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、セレクトティブサーチ [13] を用いた。セレクトティブサーチは、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ス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 Search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. 実験結果

6.1 PASCAL VOC 2012における領域分割

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。

$$t_i = \sum_j n_{i,j} \quad (6)$$

$$acc = \sum_i n_{ii} / \sum_i t_i \quad (7)$$

この精度はピクセル単位のずれの評価になっている。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。

以上の評価方法を用いて、サリエンスマップの変化を二通り、グラブカットへのサリエンスマップの渡し方を二通り、計4パターンの手法について実験を行った。一つ目はBack Propagationにより得られたサリエンスマップの値の大きい箇所をポジティブとして固定値でグラブカットに渡す手法、二つ目はBack Propagationにより得られたサリエンスマップをそのままグラブカットに渡す手法、三つめはBack PropagationとGuided Back Propagationにより得られたサリエンスマップの値の大きい箇所をポジティブとして固定値でグラブカットに渡す手法、四つ目はBack PropagationとGuided Back Propagationにより得られたサリエンスマップをそのままグラブカットに渡す手法となっている。

- (1) Back Propagation(BP), fix value
- (2) Back Propagation(BP), map value
- (3) Back Propagation(BP) + Guided Back Propagation(GBP), fix value
- (4) Back Propagation(BP) + Guided Back Propagation(GBP), map value

PASCAL VOC 2012におけるこれら4パターンの結果を表1にまとめた。実験結果の例を図6に示した。

表1 4パターンの領域分割の精度比較

class	(1)	(2)	(3)	(4)
bg	77.2	59.4	73.5	61.8
aeroplane	42.0	22.8	40.1	22.4
bicycle	19.2	11.7	20.9	15.4
bird	27.5	22.5	28.2	21.8
boat	21.2	15.0	16.2	13.8
bottle	33.9	30.3	28.9	31.8
bus	44.9	41.8	39.0	40.7
car	47.2	39.5	47.2	42.1
cat	41.3	36.6	46.3	35.0
chair	12.4	10.5	10.5	10.2
cow	35.7	28.2	31.1	28.3
diningtable	18.8	12.6	8.5	12.8
dog	42.3	32.5	36.3	34.0
horse	29.7	24.8	28.1	24.6
motorbike	43.6	29.9	39.1	30.1
person	40.6	35.0	35.2	34.2
pottedplant	27.4	17.5	19.5	17.4
sheep	50.5	31.3	42.6	32.0
sofa	19.7	17.0	19.5	15.6
train	46.9	35.1	38.1	35.8
tvmonitor	41.7	29.8	34.9	28.3
mAP	36.4	27.8	32.6	28.0

6.2 最新手法との比較

また、表2は、PASCAL VOC 2012における最新手法との比較である。

表2 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 [8]	62.2
weakly supervised	
ours	36.4
Pedro-seg [9]	40.6

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mAP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. 考察

最新の領域分割手法における領域分割の精度比較と、RCNNとの物体検出の精度比較を行った。本手法はピクセル単位のアンノテーション情報を用いていないので、完全教師有学習の精度と比較すると劣っている。また、弱教師有学習である[9]の精度にも届かなかった。一方で、物体検出タスクにおける精度の

比較では、本手法は RCNN の精度を上回った。このことから、本手法は大まかな位置の推定の精度はよいものの、ピクセル単位の物体の位置の評価をうまく行えていないことがわかる。図 7 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 の結果はテキストの認識結果の影響を受けている関係が、バウンディングボックスが小さめに出現しているのに対して、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ため、よい精度で物体の検出が行うことができる。また、図 8 は複数品食事画像における領域分割の結果である。

そこで、今回の実験では、サリエンスマップを 2 通りと、グラブカットへのサリエンスマップの渡し方の 2 通り、計 4 通りの実験を行い、精度の改善を試みた。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ス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が 4 パターンの中で最も良い精度となった。グラブカットに fix 値でなく map 値を与えると大きく精度がさがってしまったが、fix 値でグラブカットを行った場合に比べて、map 値でグラブカットを行うと図 4 のように背景と前景を分離できずに領域が拡張されてしまう。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、境界付近のポジティブ要素を強めてしまったのが精度低下が原因のようであった。期待していたマップ値の強弱による前景と背景の分離の精度改善は見込めなかった。また、図 5 のように、一見バックプロパゲーションと GBP のサリエンスマップを組み合わせたものはより精度の高いサリエンスマップに見えるが、結果としては精度の改善には至らなかった。これもおそらく同様の原因により、エッジ付近の確立場を強調することで、境界がむしろ曖昧になってしまい正常にポジティブ領域の拡張が行えなかったからではないかと考えられる。グラブカットではサリエンス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。

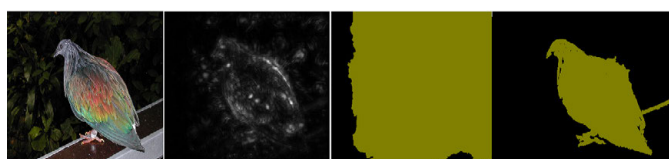


図 4 (1) 元画像, (2) サリエンスマップ, (3) マップ値, (4) フィックス値

8. まとめと今後の課題

バックプロパゲーションを用いることにより、画像中の物体の位置の推定を行い、グラブカットにより、領域分割を行った。この際に、プロポーザルを利用することで、精度の改善を図った。また、グラブカットの際の処理を工夫し、精度の改善を試みたがよい結果は得られなかった。グラブカットでは、サリエンスマップの数値の細かい変化を反映させることは難しかった。グラブカットとバックプロパゲーションはあまり相性がよくないのではないかと考えられる。Superpixel と CRF を利用することで精度の改善が得られる可能性がある。今後、様々な手法を試し、最適な方法を探したい。

また、現在はプロポーザルを用いて精度の改善を試みている

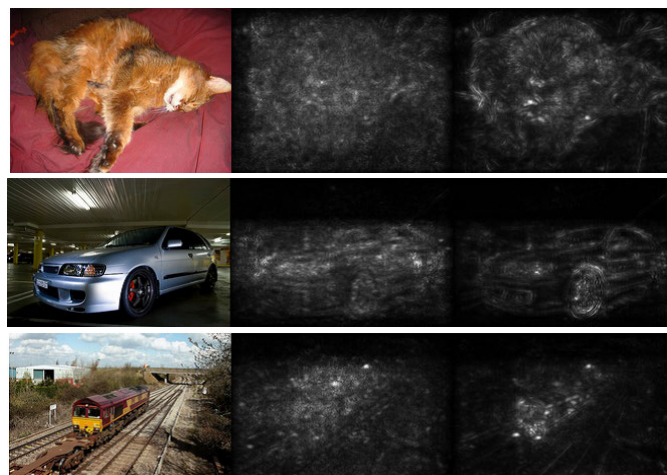


図 5 サリエンスマップの比較 (1)original image, (2) BP map, (3)BP + GBP map

が、プロポーザルを行い約 2000 の領域について認識を行い、バックプロパゲーションを行うと膨大な計算量になる。今後の方針として、スマートフォンでの食事画像の領域分割を考えている (図 8)。今後は精度を改善しつつ、計算量の削減も考えていきたい。

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.
- [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. Carreira and C. Sminchisescu. Constrained parametric min-cuts for automatic object segmentation. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2010.
- [7] A. Pablo, Jonathan T. Jordi, P., M. Ferran, and M. Jitendra. Multiscale combinatorial grouping. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2014.
- [8] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2015.
- [9] P. Pedro and C. Ronan. From image-level to pixel-level labeling with convolutional networks. In *arXiv:1411.6228*, 2014.
- [10] 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.
- [11] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller. Striving for simplicity: The all convolutional net. In *ICLR 2015 Workshop Track*, 2015.
- [12] Y. Boykov 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*

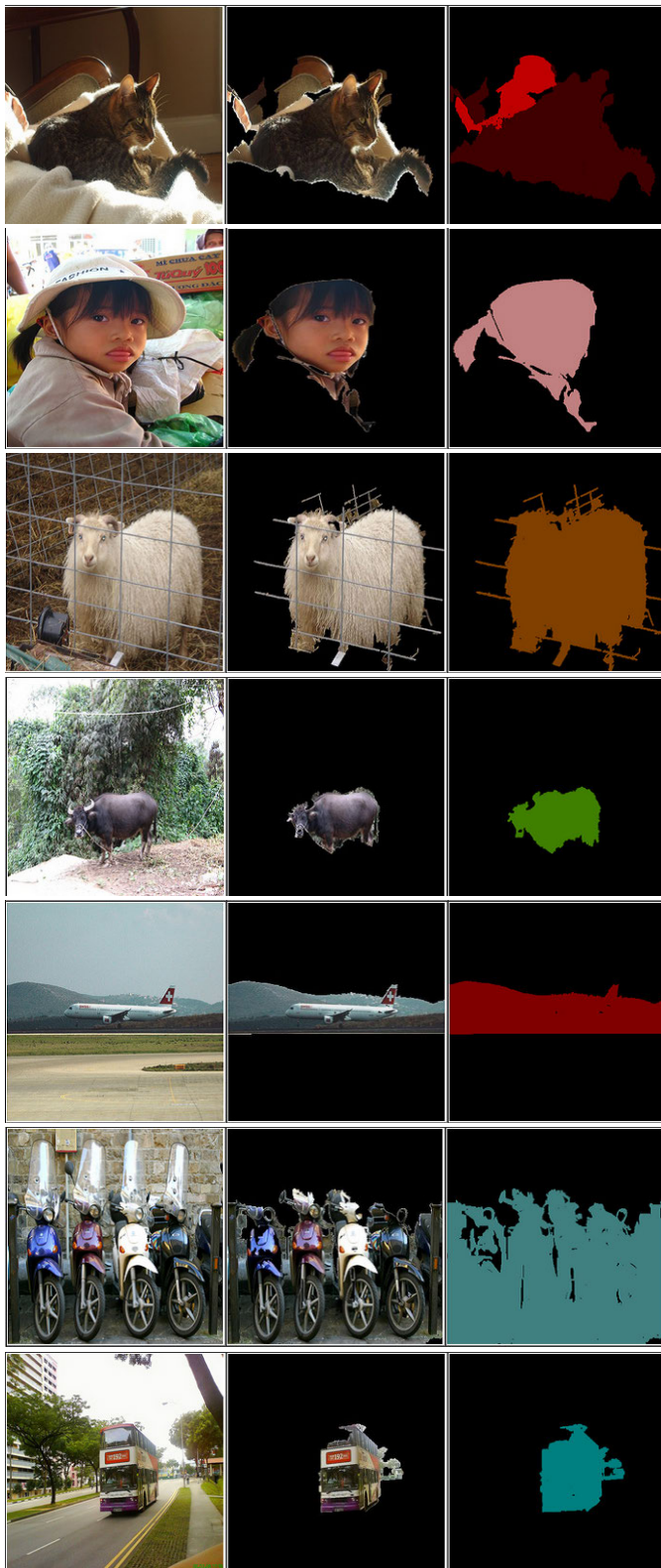


図 6 領域分割結果の例 (1)original image , (2) segmentation result , (3)label

Recognition, 2001.

- [13] 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.



図 7 食事画像における RCNN との比較 (1)original image , (2) rcnn result (3) ours (4) ground truth

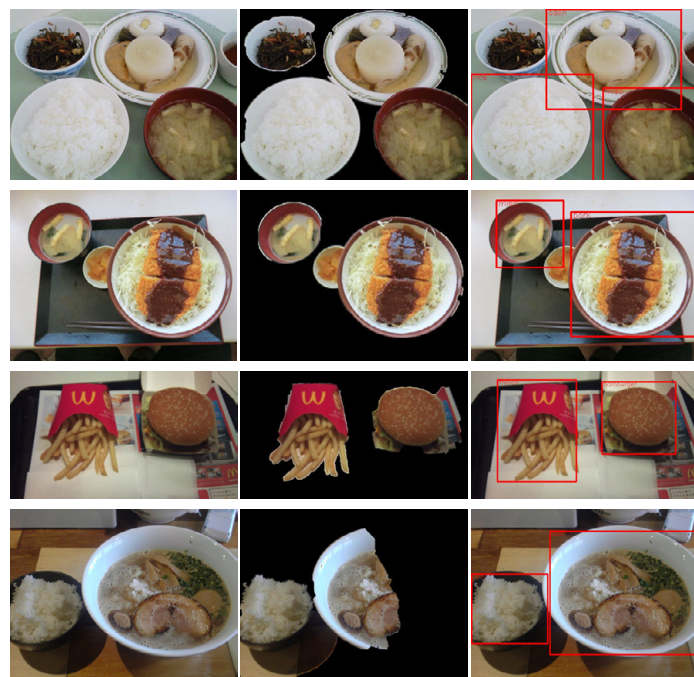


図 8 食事画像の領域分割結果 (1)original image , (2) segmentation result