

# 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 総合情報学専攻

# Convolutional Neural Network

- CNN

- クラス分類タスクにおいてトップの精度



- CNNの応用(物体位置の認識)

- 物体検出

- 物体にBBを付与

- 領域分割

- ピクセル単位の認識

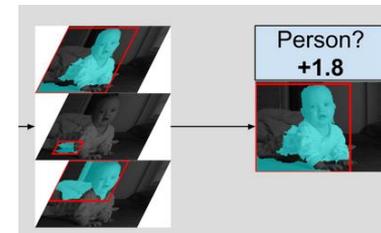


# CNNを用いた領域分割

- CNNによる完全教師ありのセグメンテーション

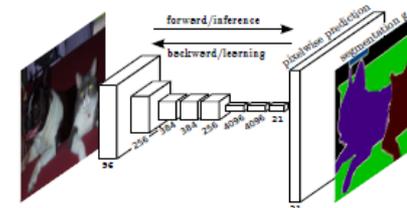
- SDS [Hariharan et al. ECCV2014]

- 領域候補のプロポージャー



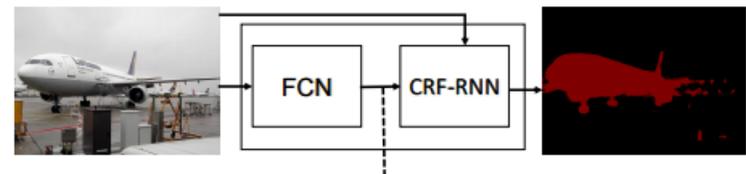
- Fully Convolutional Neural network [Long et al. CVPR2015]

- Upsampling + Deconvolution



- DeepLab [Chieh et al. ICLR2015], CRF as RNN [Zhen et al. ICCV 2015]

- FCN + グラフィカルモデル

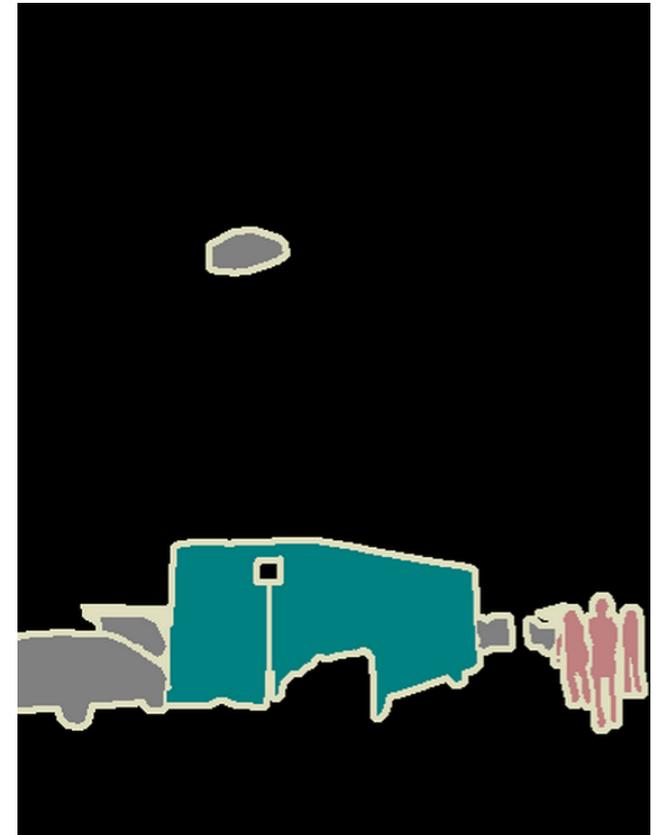


# 完全教師あり学習

- ピクセル単位のアノテーションが必要 (マスク画像)
  - かなりの労力 (コスト) が必要、ビッグデータとの相性が悪い

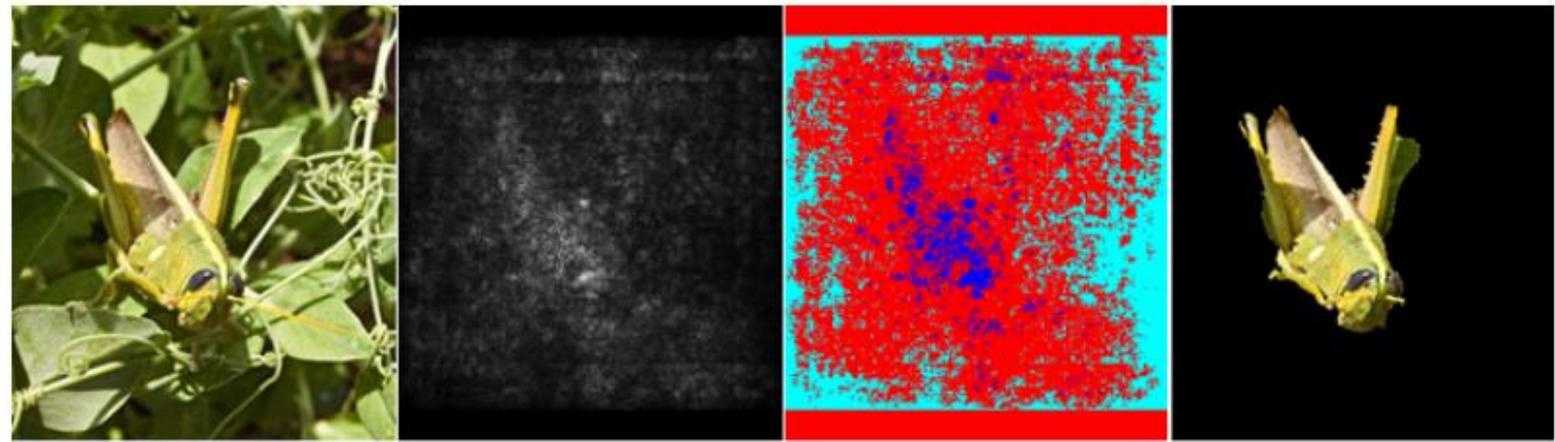


アノテーション



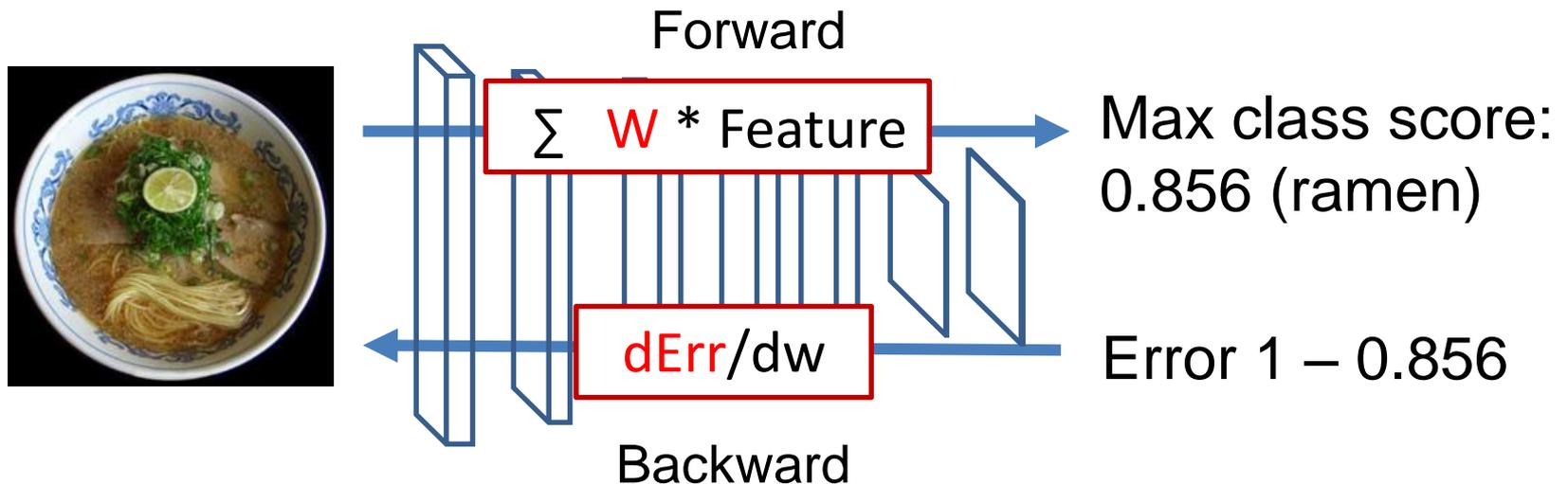
# CNNを用いた弱教師有り学習 による領域分割

- Back Propagation(BP)+グラブカットによる領域分割
  - [Simonyan et al. ICLR 2014]
  - 入力は学習済みCNNのモデルと画像のみ
  - ピクセル単位の学習画像が不要(弱教師有り学習)



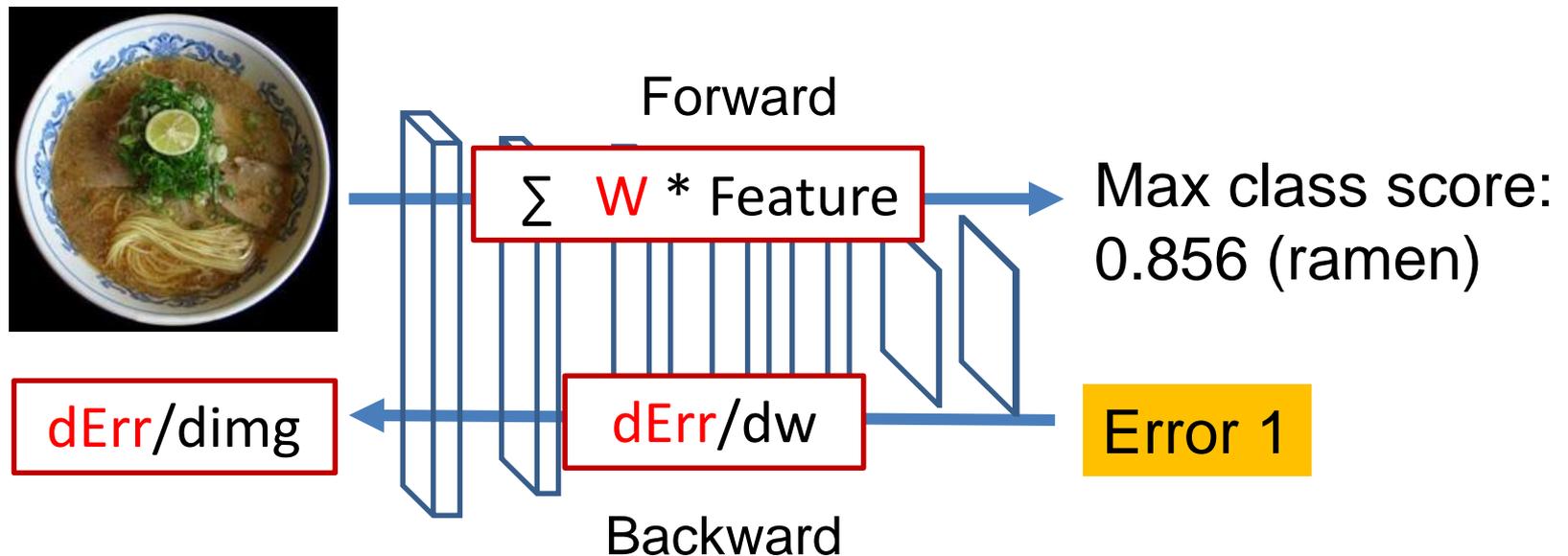
# CNNの学習プロセス

- 認識結果のエラーを最小化するためのWeightを学習
  - SGD(確率的勾配降下法)とBP(誤差逆伝搬法)を用いる
- SGD →  $-dE/dw$ の方向に沿って重みを調整
- BP →  $dErr/dw$  を各階層に伝搬



# BPによる認識の可視化

- エラーを最大に設定
- $dErr/dw$ を画像レベルまで伝搬

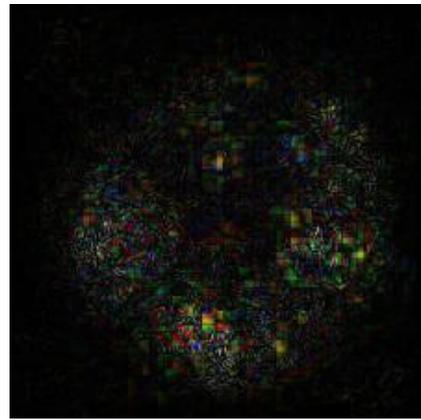


# BPから得られるサリエンスマップ

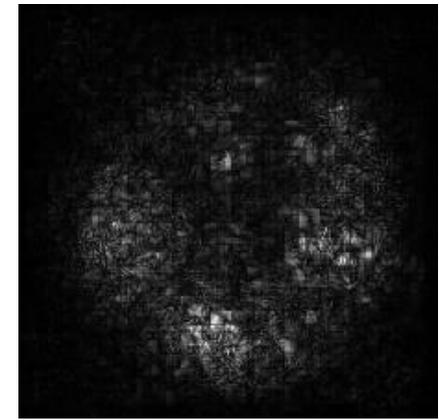
- BPにより得られた結果 ( $dE/dI_{img}$ )
  - 画像の認識結果を正解に近づけるための結果
  - 強く反応した位置 → 学習に用いた画像と共通している部分
  - 物体の位置の推定に繋がる
- RGBの最大値でグレースケールに変換



Original



BP result (im Iv)



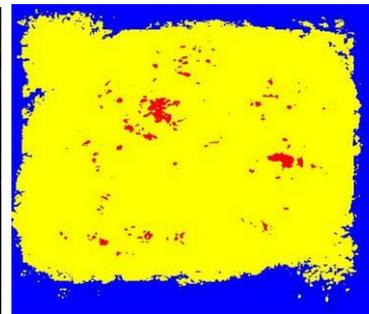
Object saliency map

# Saliency map を用いたGrabカット

- グラブカットによる2クラス領域の拡張
- サリエンシーマップからシードを作成
  - Positive 領域 (upper 5% : 赤)
  - Negative 領域 (lower 10% : 青)
  - Other (yellow)



Original



Seed



Grab cut result

# プロポーザルによる改善

- Weakポイント

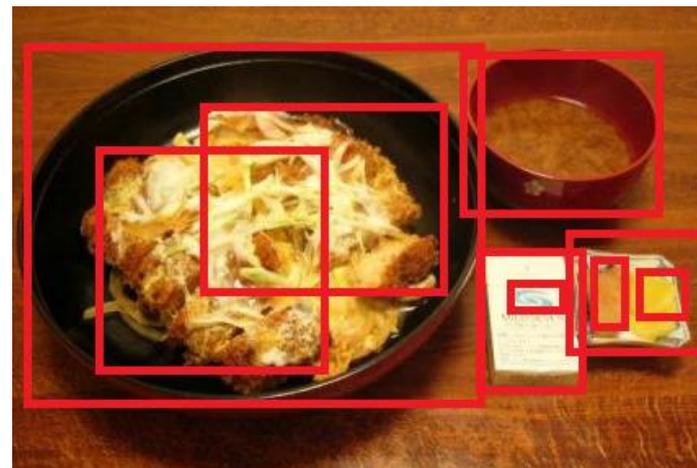
- 接触物体の分離 → ×
- 小物体の検出 → ×

- 改良案

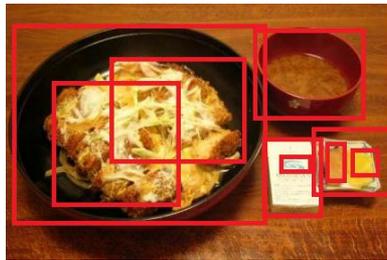
- Rich feature CNN(RCNN) [Girshk et al 2014]
- 大量の領域を提案し、全てCNNで認識
- 大幅に精度が向上



- 大量の領域をプロポーザルし、全て領域分割



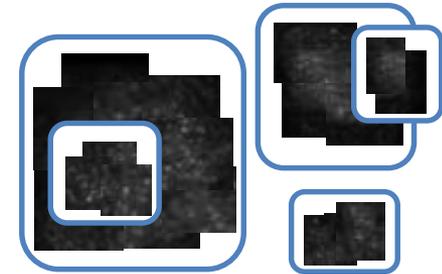
# 提案手法の流れ



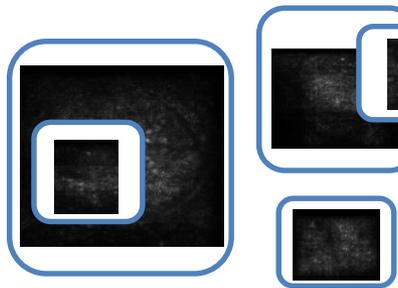
Selective Search(1)



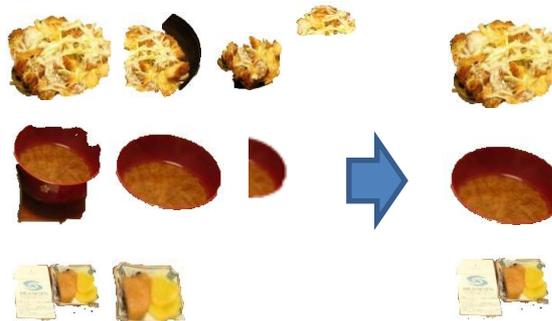
BB grouping(2)



Back propagation(3)



Saliency maps(4)



Grab cut (5)



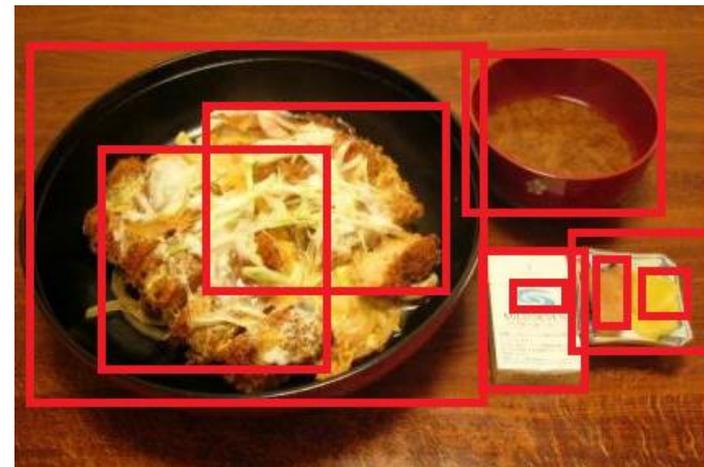
NMS(6)



Result

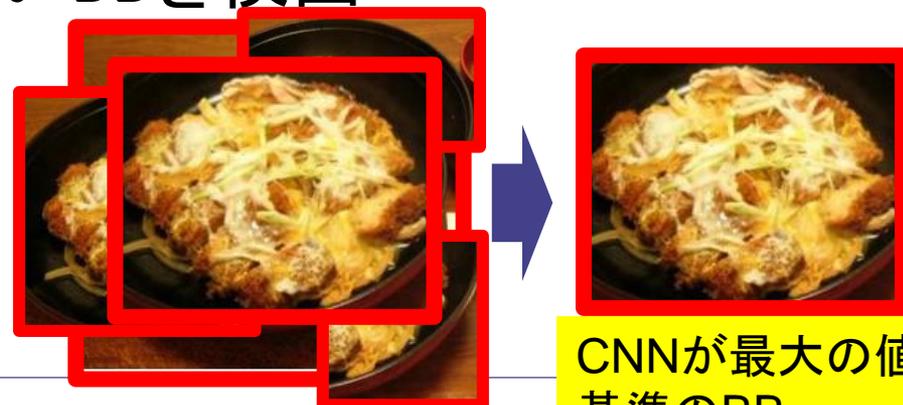
# 領域候補のプロポーザル

- Selective Search
  - [Uijlings et al. IJCV]
  - エッジ情報などから領域候補をプロポーザルする
  - 約2000のBB候補



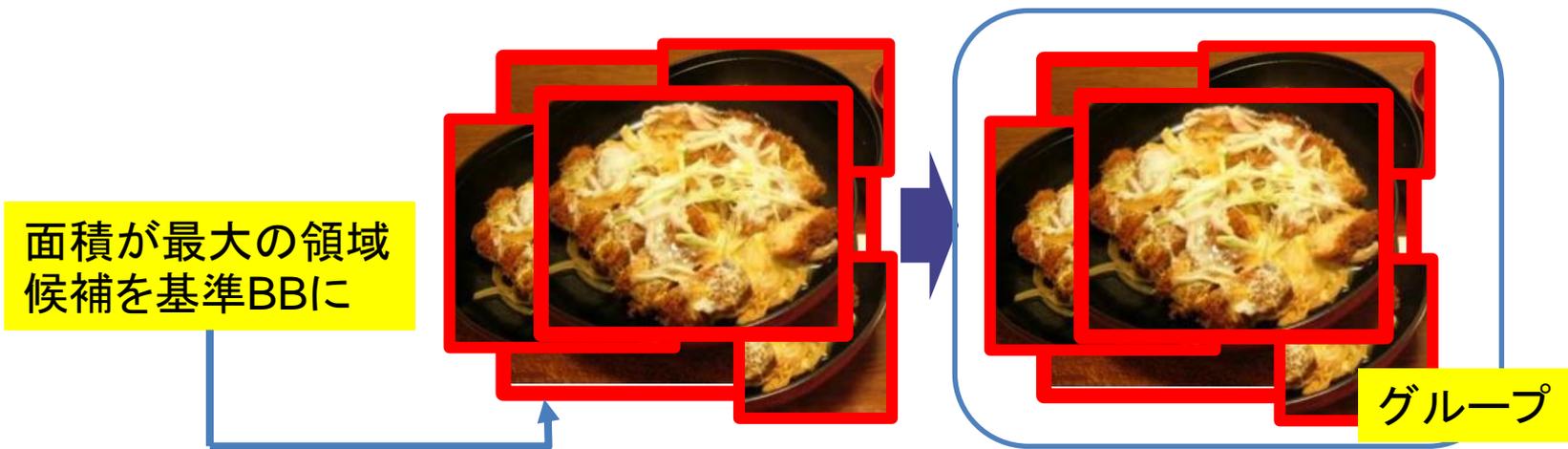
# Non Maximum Suppression(NMS)

- NMSアルゴリズム
  - Cannyエッジの検出などに使われる
  - 基準値の高いエッジを元に先鋭化
- BBの統合に応用
  - BBのCNNの認識結果を基準値に
  - オーバーラップ率の高いBBを検出
  - 基準のBB以外を除去
  - 繰り返し



# BBグループピング

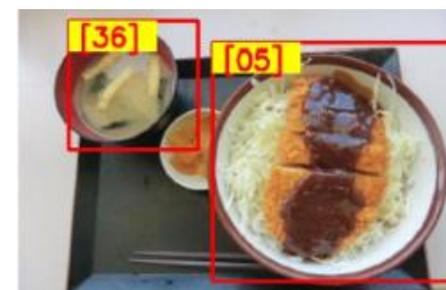
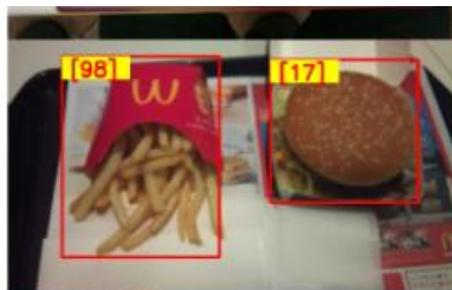
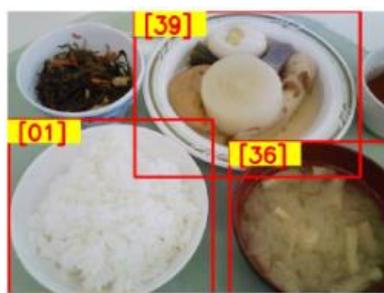
- BBグループピング
  - NMSアルゴリズムを応用
  - BBの面積を基準値に
  - オーバーラップ率の高いBBを検出
  - 基準値のBBを含めて検出されたBBを1つのグループに



# 実験

## (1) 複数食品画像 (UEC-Food 100)

- 1175 枚
- バウンディングボックスのみのアノテーション



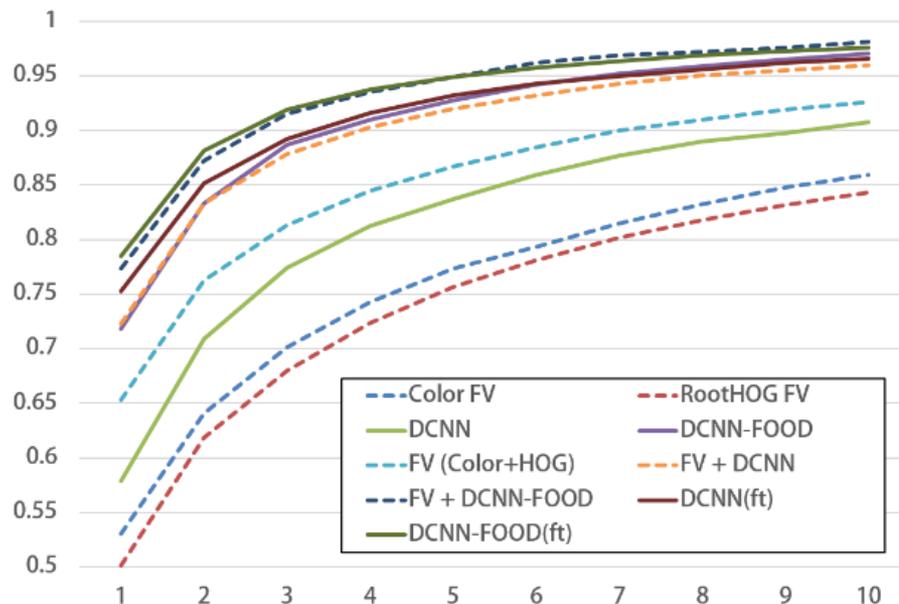
## (2) Pascal VOC 2012領域のアノテーション

- 領域のアノテーション



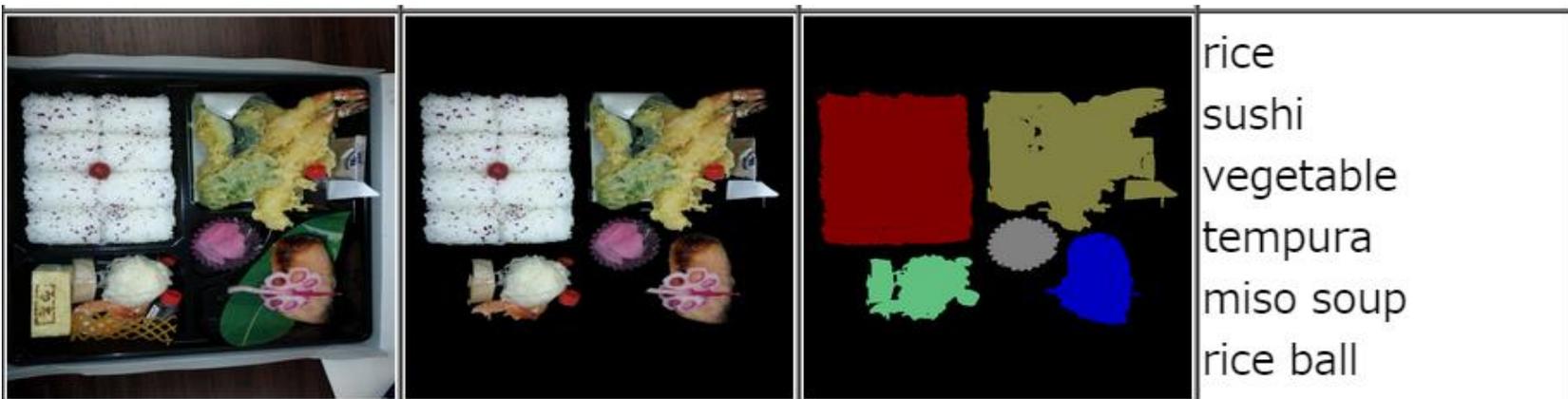
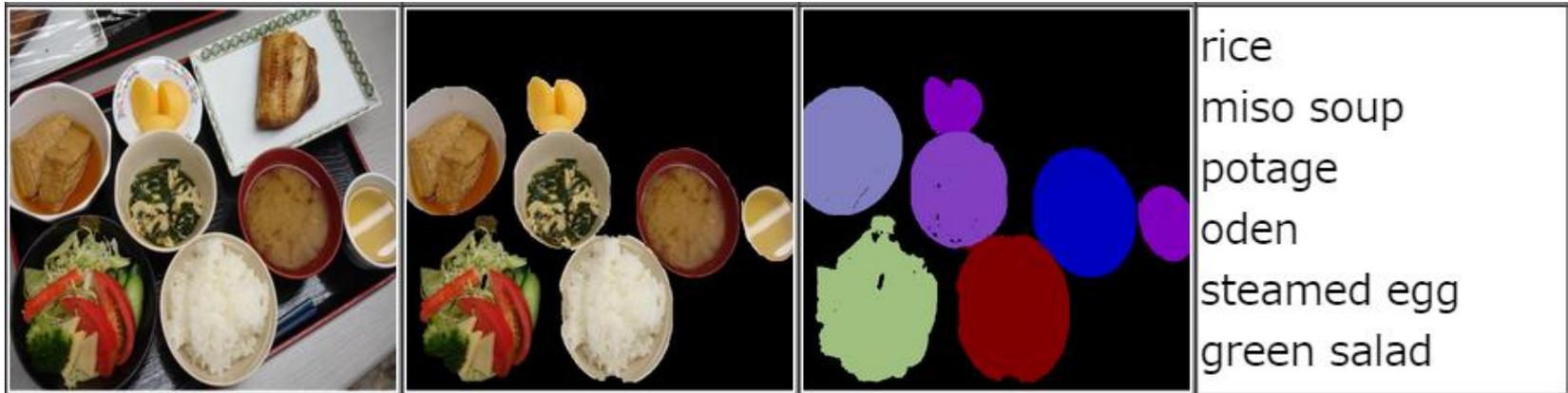
# CNNの学習(食事画像)

- 食事画像のCNNモデル(alex net)
  - ImageNetによるpre-train
    - 一般画像1000+ 食事系列画像1000クラス
  - UECFOOD101によるfine-tuning(BBアノテーション有り)
    - 食事画像100クラス
    - Top-1 78.5 %
    - Top-5 94.9 %



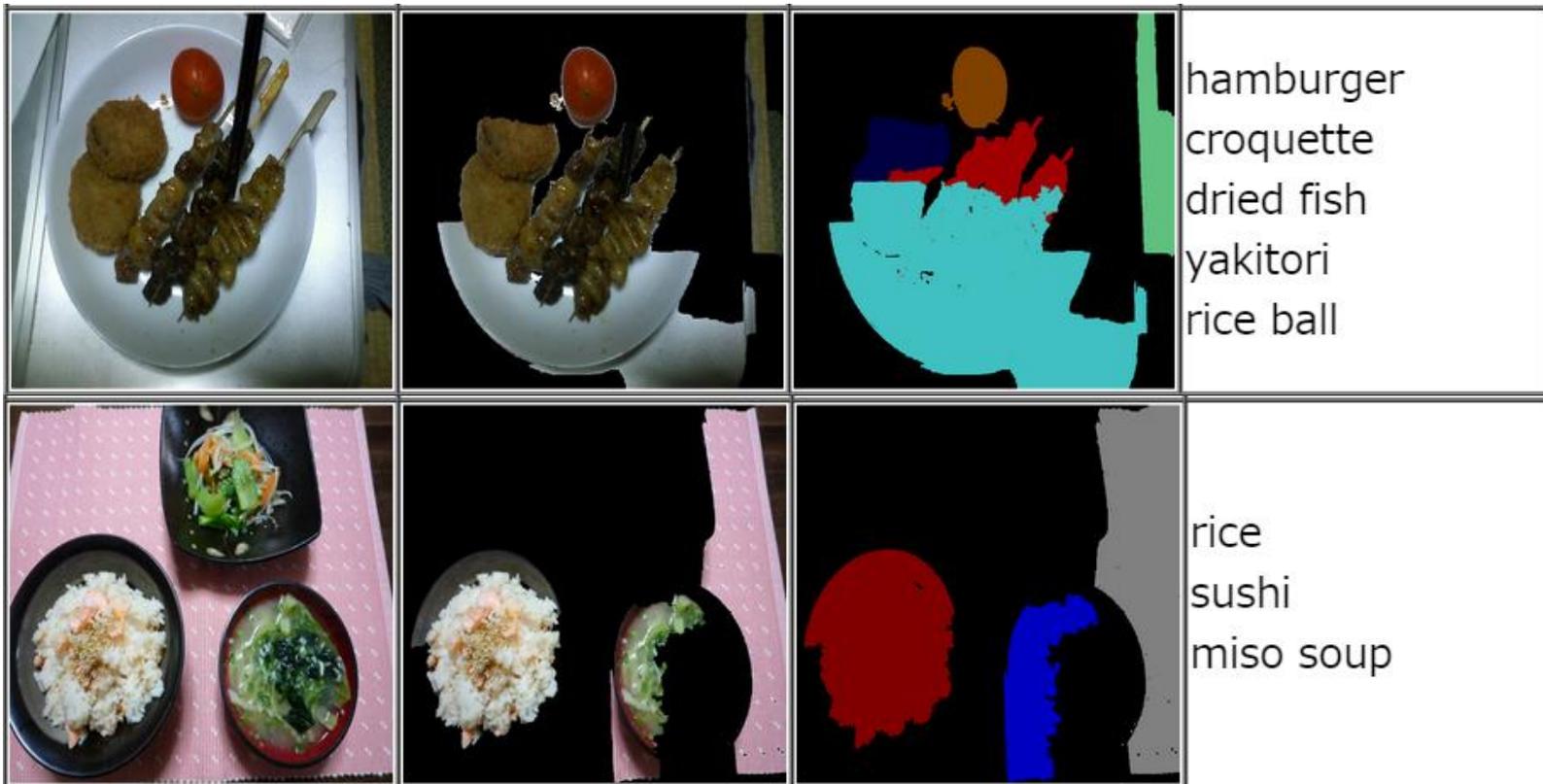
# 領域分割結果1

- 成功例



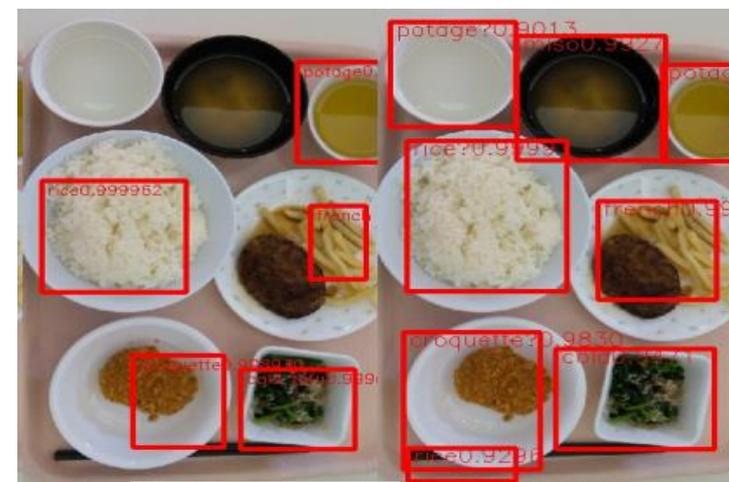
# 領域分割結果2

- 失敗例



# UEC FOODにおける物体検出の評価

- 物体検出の精度を評価 (BB level)
- RCNNとの比較
  - RCNNは小領域を検出
  - 領域分割により一貫性のある領域に絞られた?

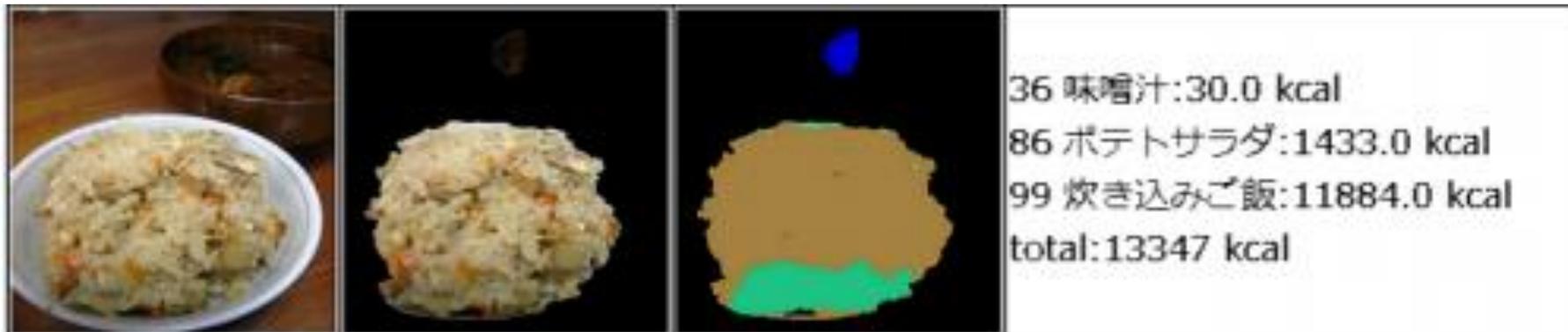
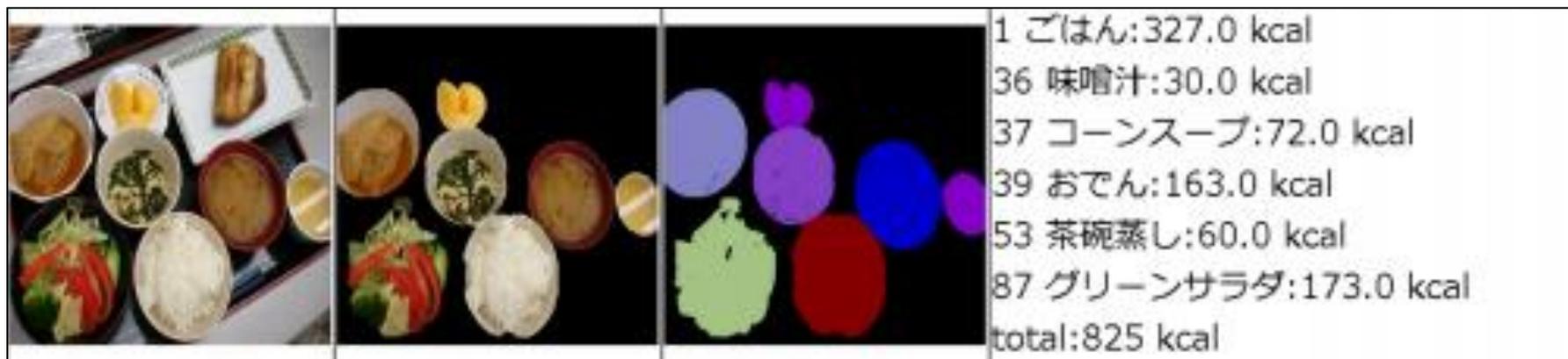


RCNN Ours

Method	100クラス Num>0	53クラス Num>10	11クラス Num>50
RCNN	26.0	21.8	25.7
Ours	49.9	55.3	55.4

# 領域分割結果を 利用したカロリー推定

- 味噌汁などの面積の変化が少ないと考えられる食事クラスを基準にして領域面積からカロリーを計算



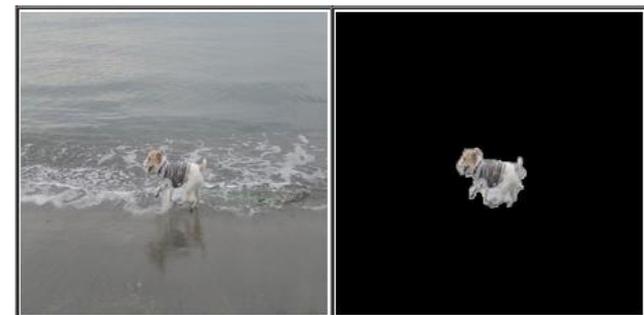
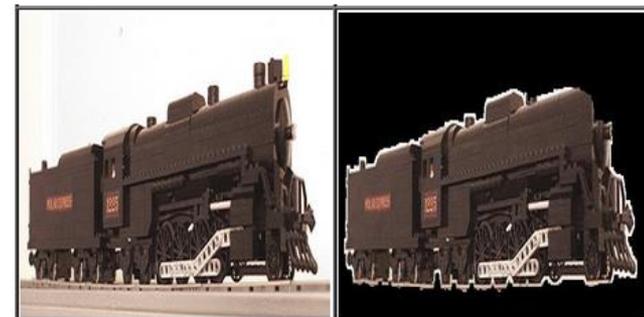
## (2) Pascal分類におけるCNNの学習

- PascalのCNNモデル(一般公開済みのRCNNのモデル)
  - ImageNetによるpre-train
    - 一般画像1000クラス
  - Pascal voc によるfine-tuning (BBアノテーション有り)
    - Pascal voc 2012, 20クラス
    - プロポーザルによる学習画像の拡張

# (2) Pascal VOCにおける 領域分割の評価

- PASCAL VOC 2012
  - 一般的な物体 20 class (バス, 犬, etc)
- 領域分割の精度評価(pixel level)

Method	Mean IU
完全教師有り学習	
SDS[3]	51.6
FCN[4]	62.2
弱教師有り学習	
Ours	36.4
Pedro-seg[5]	40.6



# 精度改善についての試み

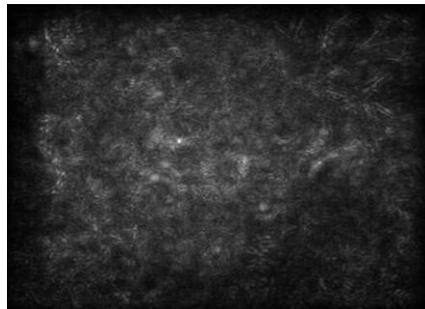
- CNN → おおよその位置を推定
- グラフィカルモデル → エッジを検出

# CNNによるおおよその 位置推定の改善

- Guided Back Propagation(GBP)→BPからReluの処理を変更
- Relu (back ward)
  - BP  $y = dz * (x > 0)$
  - GBP  $y = dz * (x > 0) * (dz > 0)$
 (x : 入力 (CNN特徴) y : 出力, dz : 伝搬値)

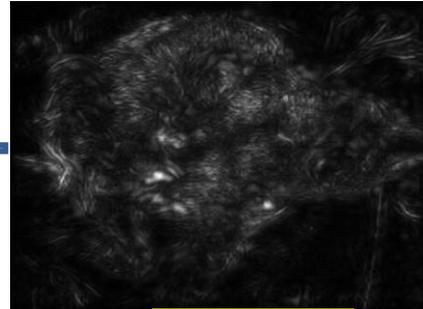


Original

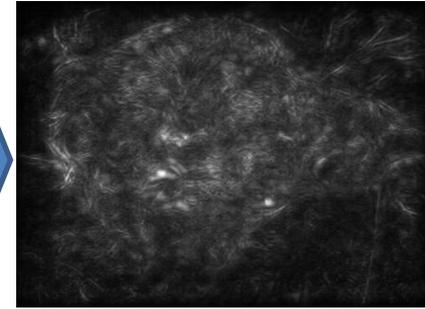


BP

+



GBP



BP+ GBP

# グラフィカルモデルの改善

- グラフィカルモデル

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

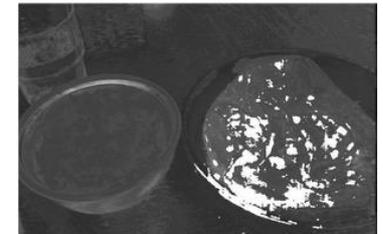
- 単項

– Gaussian mixture model(GMM)+seed領域

- 平滑化項

– ピクセル間の結びつき

– エッジを利用



GMM+seed



Y edge



X edge

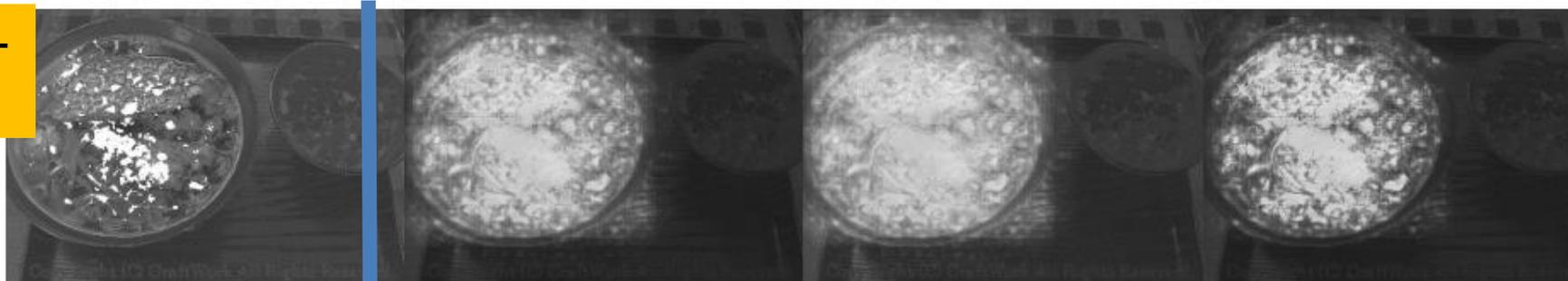
# 単項に渡すseed領域を変更

- GMM+fix値 → GMM+map値
  - サリエンスマップの情報量を活かす
  - 関数による変換



Copy Right (C) CraftWork All Rights Reserved.

GMM+  
seed



結果



mask

x

Tanh(x)

X<sup>2</sup>

# result

method	100 class	53 class	11 class
	$\#item \geq 0$	$\#item \geq 10$	$\#item \geq 50$
<u>BP map</u>			
<u>fix value</u>	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
<u>BP + GBP map</u>			
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

# まとめ

- CNNに基づいた領域分割を行った
- 食事画像と一般物体画像について手法を適用した
- Future work
  - Super pixels + グラフィカルモデル (GMMが不要)

