

# 単一の畳み込みネットワークによる 料理検出とカロリー量推定の マルチタスク学習

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 情報学専攻

# 背景：料理画像からのカロリー量推定

## Foodlog

料理をクロップ  
料理の種類や量を**手入力選択**

1/4 人前	205 kcal
1/3 人前	279 kcal
1/2 人前	411 kcal
2/3 人前	551 kcal
3/4 人前	617 kcal
1 人前	823 kcal ✓
1.5 人前	1234 kcal
2 人前	1646 kcal

量の選択

## カロナビ

食事写真を撮るだけで  
カロリーがわかります

栄養士が実際にあなたの食事写真を  
見てカロリーをチェックします

傾向が分かりやすい  
体重グラフもあります

2015/3/1 Sun  
1300 kcal

2014/10/8  
54.0 kg

53.8 kg

1ヶ月 3ヶ月 12ヶ月

栄養士による食事画像からの  
カロリー量推定  
**有料サービス**

料理画像からのカロリー量推定は未解決の問題

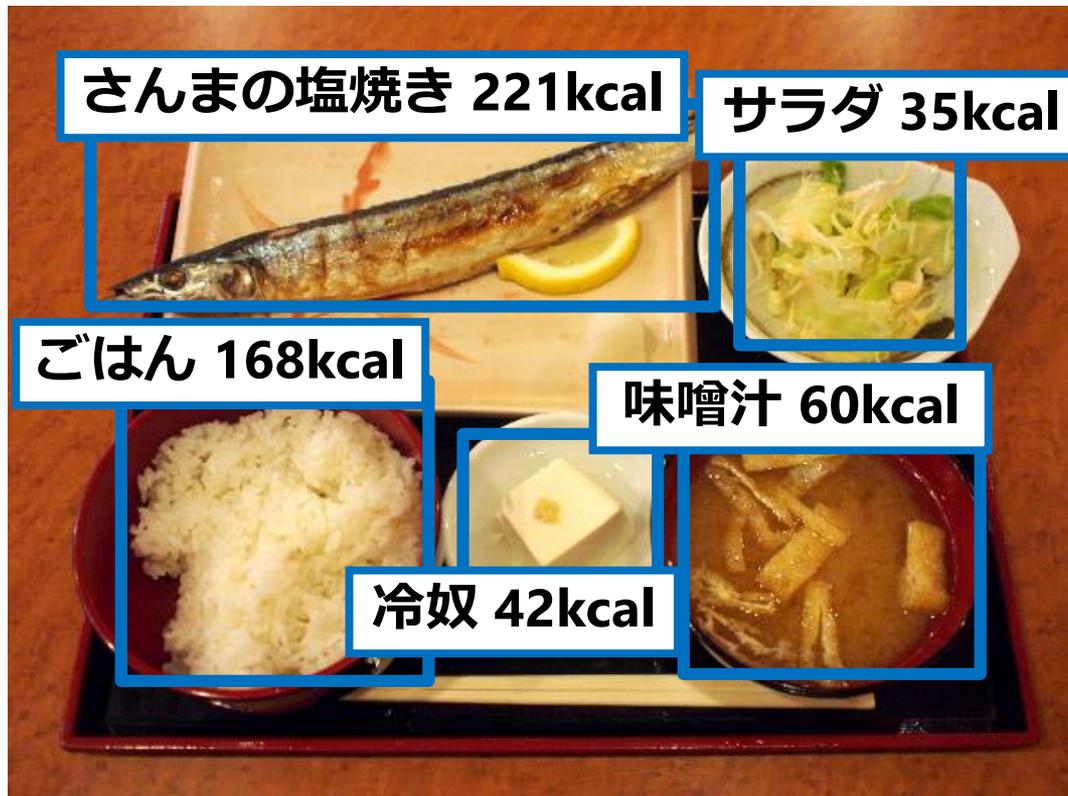
# 背景：複数品目に対する料理画像認識



物体検出により複数品目を同時に認識  
 → 少ない手間で記録することが可能

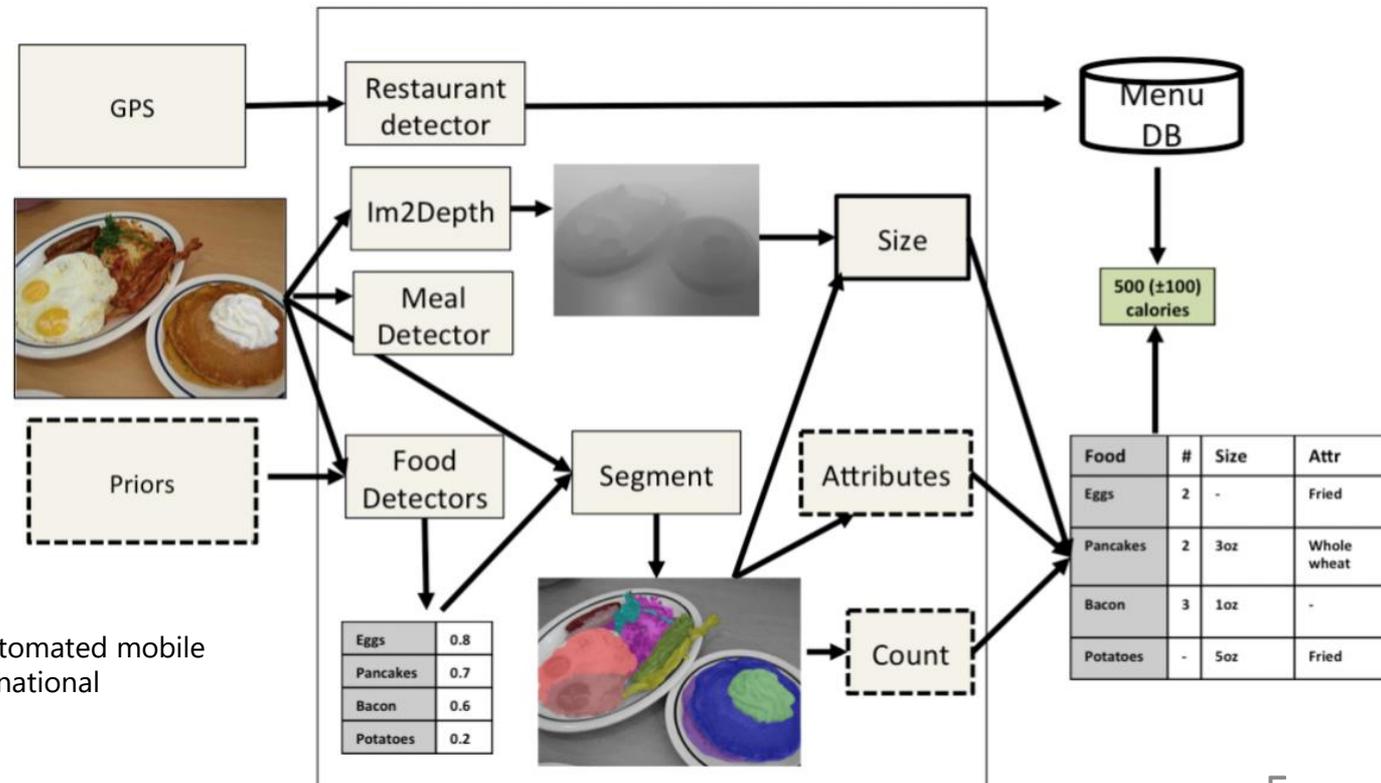
# 研究目的

## 複数料理写真からの同時カロリー一量推定



# 関連研究：カロリー量推定

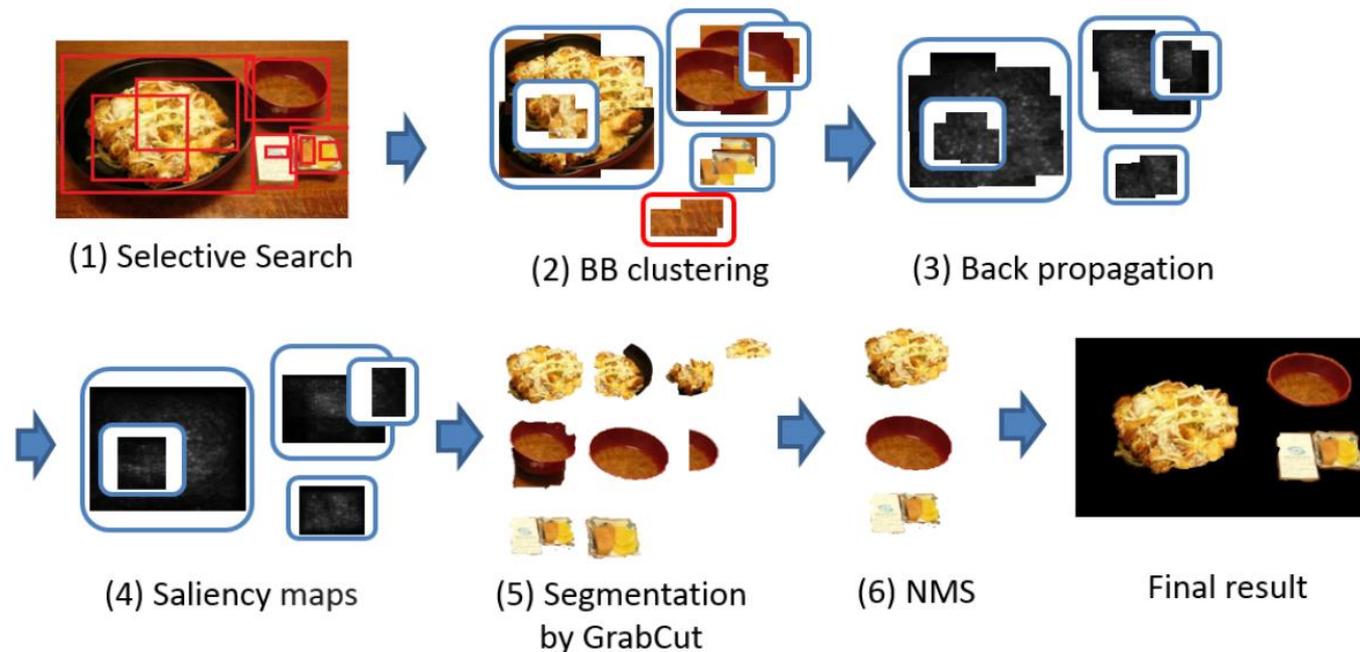
- Im2Calories [Myersら 2015]
  - CNNによるクラス分類, セグメンテーション, 深度推定
  - 食材情報とボリュームからカロリー量を推定



Myers et al. Im2calories: towards an automated mobile vision food diary. In Proc. of IEEE International Conference on Computer Vision, 2015.

# 関連研究：料理検出

- CNN-based Food Image Segmentation [下田ら 2015]
  - 各候補領域のサリエンスマップを生成
  - 領域情報からバウンディングボックスを推定

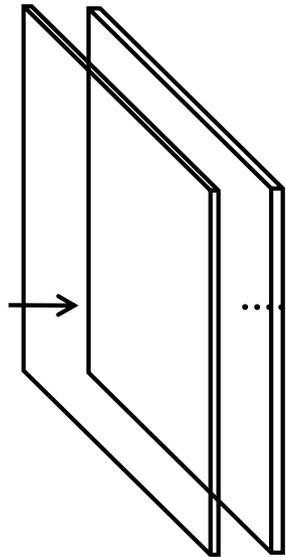


W. Shimoda and K. Yanai. CNN-based food image segmentation without pixel-wise annotation. In *Proc. of IAPR International Conference on Image Analysis and Processing*, 2015.

# 手法概要：料理検出＋カロリー量推定

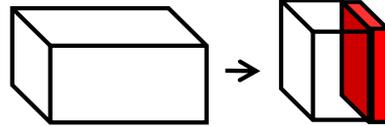
## 料理検出 + カロリー量推定

入力



Conv layers (YOLOv2)

ネットワーク  
出力



Non-Maximum suppression

+ 閾値処理

最終出力



YOLOv2 : J. Redmon, A. Farhadi , YOLO9000: Better, Faster, Stronger, CVPR 2017

# 手法概要：料理検出＋カロリー量推定

## 二つの既存データセットを用いた学習

- **物体検出用のデータセット**
  - バウンディングボックス
  - クラス
- **カロリー量推定用のデータセット**
  - カロリー量



Miso soup  
(160 kcal)

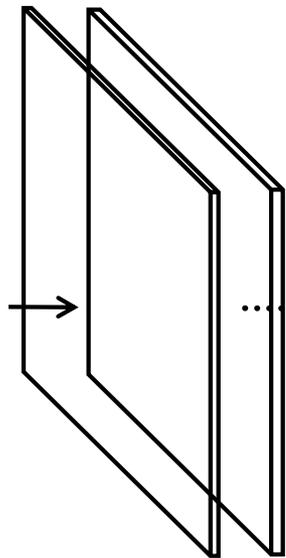
**両方用いることで料理検出とカロリー量推定のマルチタスク学習を実現**

# 手法：YOLOv2による料理検出

## YOLOv2の導入

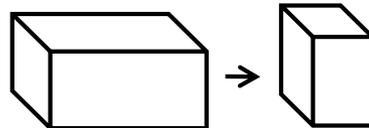
- システム全体のend-to-endな学習が可能
- 全層が畳み込み層で構成される

入力



Conv layers (YOLOv2)

ネットワーク  
出力



Non-Maximum suppression  
+  
閾値処理

最終出力



# 手法：カロリー量推定

## CNNを用いた回帰による手法

- 損失関数  $L_{cal}$ 
  - 相対誤差  $L_{re}$ と絶対誤差  $L_{ab}$ の線形和

$$L_{ab} = |y_i - g_i| \quad L_{re} = \frac{|y_i - g_i|}{g_i}$$

$y_i$  : 画像  $x_i$  から推定されたカロリー量

$g_i$  : 正解カロリー量

# 実験：データセット

## 二つの異なるデータセットを使用

- UEC Food-100[1]
  - バウンディングボックス、クラス
  - 学習画像 約 10,000 枚
- カロリー量付き料理画像データセット[2]
  - カロリー量
  - 学習画像 約 3,500 枚



Miso soup  
(160 kcal)

[1] Y. Matsuda, H. Hajime, and K. Yanai. Recognition of multiple-food images by detecting candidate regions. ICME, 2012.

[2] Ege and Yanai. Simultaneous estimation of food categories and calories with multi-task cnn. MVA 2017.

# 実験：擬似的なバウンディングボックスの付与

## YOLOv2の学習にはBBが必須

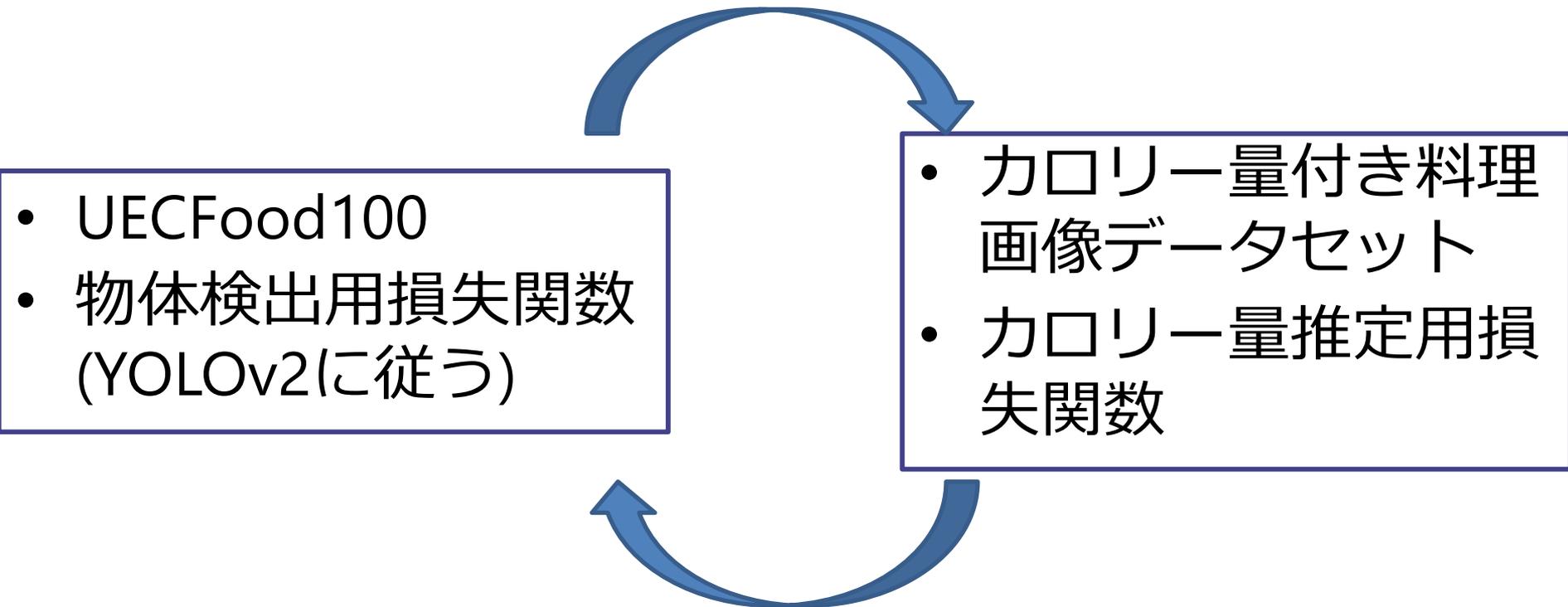
### ➡ カロリー量推定用データセットにBBを付与



- (1) ランダムな位置に画像を貼り付け, その画像領域を正解バウンディングボックスとする.
- (2) 貼り付けた画像と同じものを反転させながら背景を埋める.

# 実験：二つのデータセットを用い学習

## バッチ単位でデータセットと損失関数を入れ替えることで交互に学習



# 実験結果：料理検出＋カロリーー量推定



推定値

412 kcal  
焼きそば

722 kcal  
カレーライス

25 kcal  
味噌汁

375 kcal  
ハンバーグ

正解値

491 kcal  
スパゲッティ

937 kcal  
カレーライス

47 kcal  
味噌汁

461 kcal  
ハンバーグ

誤差

-79 kcal

-215 kcal

-22 kcal

-86 kcal

# 実験結果：料理検出＋カロリーー量推定

- カロリーー量推定評価指標

- 絶対誤差(kcal) :  $|y_i - g_i|$

- 相対誤差(%) :  $\frac{|y_i - g_i|}{g_i}$

- < 20% (相対誤差)(%) : 相対誤差20%より小さい推定値の割合

	相対 誤差 (%)	絶対 誤差 (kcal)	< 20% (相対誤 差)(%)	< 40% (相対誤 差)(%)
カロリーー量推定 (VGG16) [1]	30.2	105.7	43	76
<b>料理検出＋カロリーー量推定(本手法)</b>	36.1	121.7	34	64

[1] Ege and Yanai. Simultaneous estimation of food categories and calories with multi-task cnn. MVA 2017.

# 実験結果：料理検出＋カロリーー量推定

焼きそば 601 kcal(ES)



ピラフ 495kcal(ES)



焼きそば 539kcal(GT)  
ピラフ 475kcal(GT)

ポテトサラダ 144kcal(ES)



味噌汁 102kcal(ES)



肉じゃが 417kcal(ES)



ポテトサラダ 169kcal(GT)  
味噌汁 74kcal(GT)  
肉じゃが 352kcal(GT)

グラタン 456kcal(ES)



スパゲッティ 536kcal(ES)



グラタン 450kcal(GT)  
スパゲッティ 518kcal(GT)

カレーライス 483kcal(ES)



ポテトサラダ 170kcal(ES)



カレーライス 761kcal(GT)  
ポテトサラダ 169kcal(GT)

# まとめ

---

- 料理検出 + カロリー量推定
- 二つの異なるデータセットを使用することでマルチタスク学習を実現

# 今後

- 物体検出ネットワークの検討
  - Faster R-CNNの使用
- カロリー量付き複数料理画像データセットの作成
  - 背景画像にカロリー量付き単品料理画像を貼り付ける
  - 画像生成により食事風景らしい背景を生成

