

Multi-task Learning of Dish Detection and Calorie Estimation

Takumi Ege and Keiji Yanai

The University of Electro-Communications, Tokyo

Introduction

Foodlog

Volume selection required

- Crop a dish
- Food category recognition
- Select volume **manually**, etc...

CaloNavi

Dietary advices by nutrition professional.
Human cost ➡ **Pay service**

Fully-automatic food calorie estimation from a food photo has still remained as an unsolved problem.

Introduction

Food image recognition for multiple-dish photos.

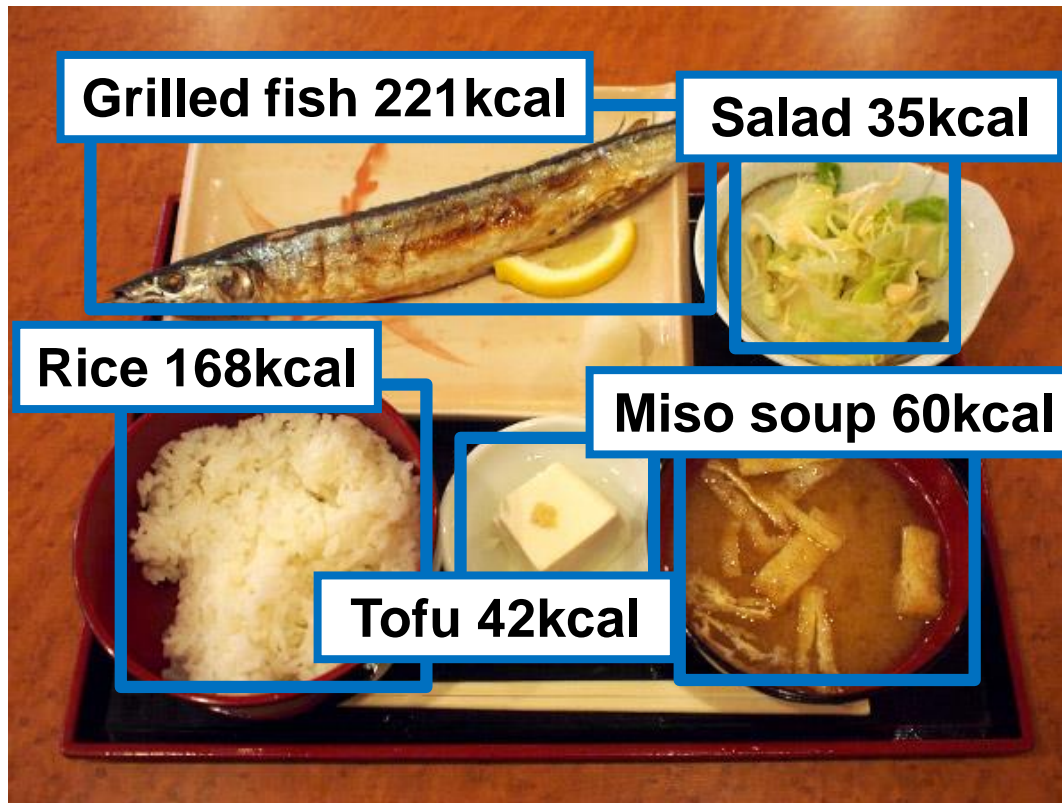


Detection of single-dish from multiple-dish photos.

➡ It is possible to record meal easily in a short time.

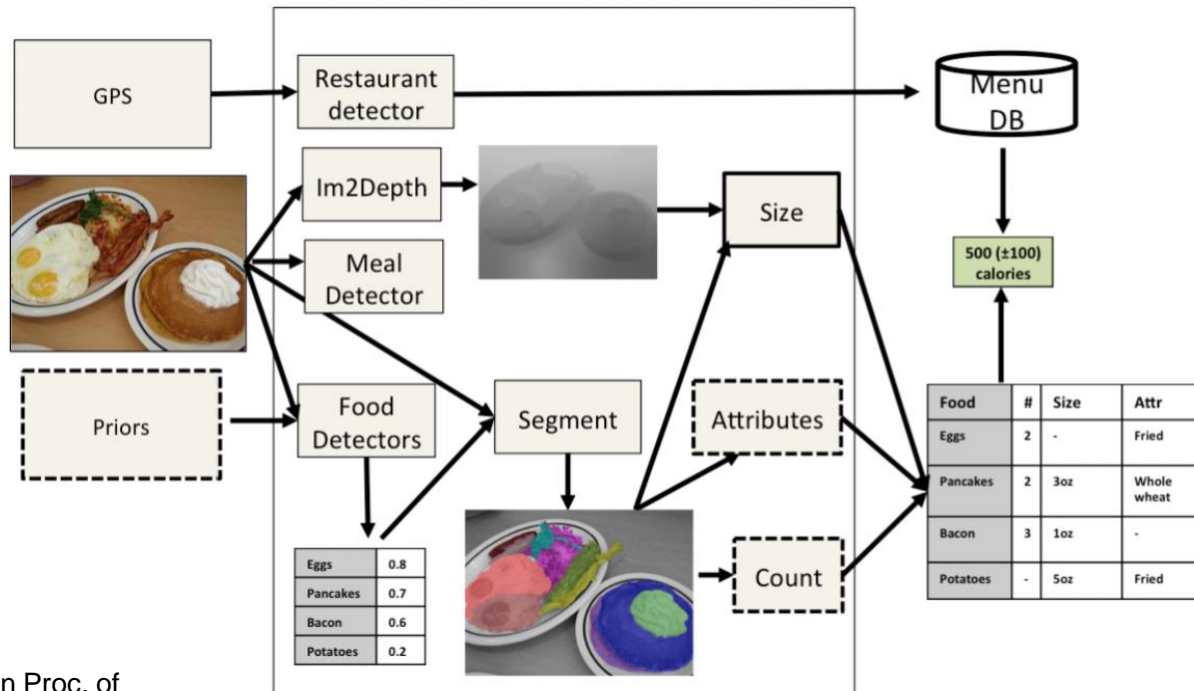
Objective

Image-based food calorie estimation



Related works : Calorie estimation

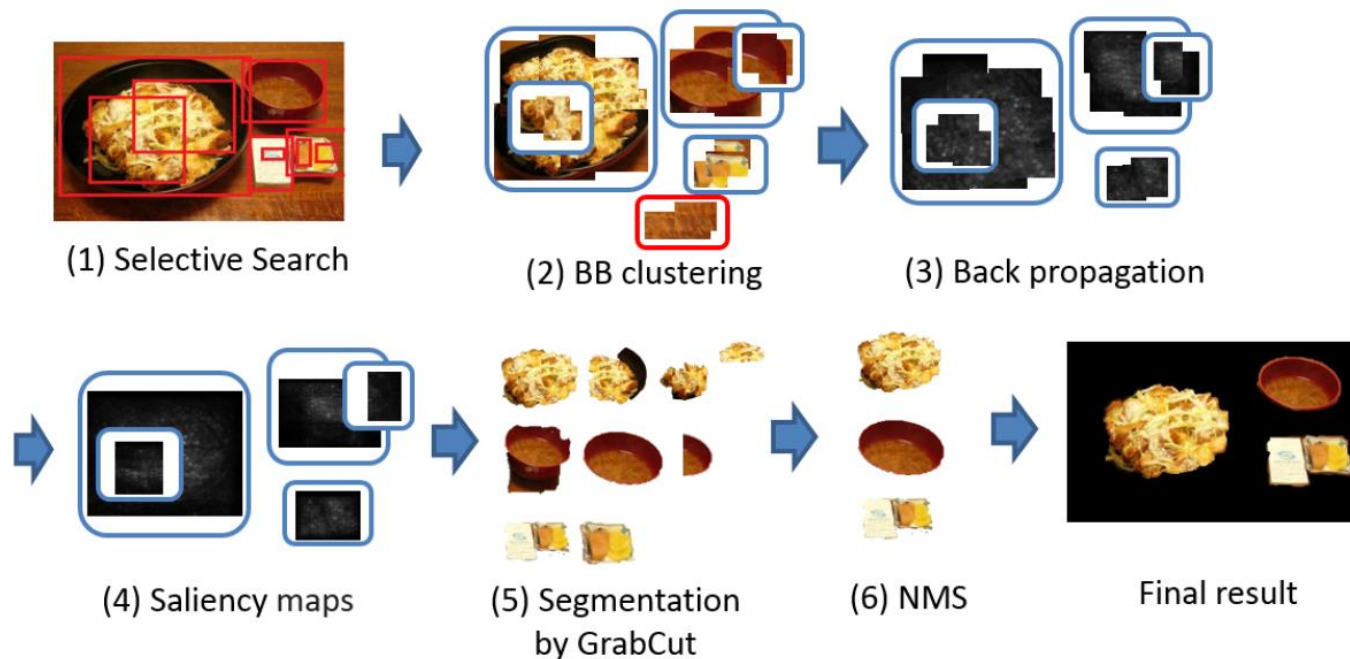
- Im2Calories [Myers et al. 2015]
 - CNN-based categorization
 - CNN-based 3D sized estimation etc...



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

Related works : Dish detection

- CNN-based Food Image Segmentation [Shimoda et.al. 2015]
 - Region proposals are generated by selective search.
 - The saliency map for each region are unified.

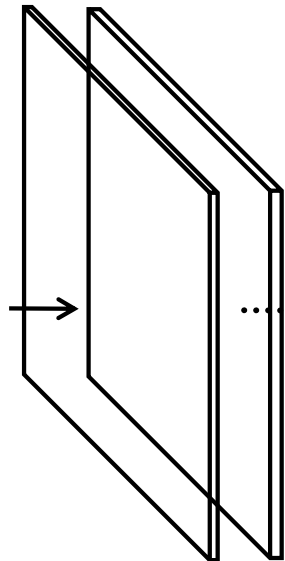


W. Shimoda and K. Yanai. CNN-based food image segmentation without pixel-wise annotation.
 In *Proc. of IAPR International Conference on Image Analysis and Processing*, 2015.

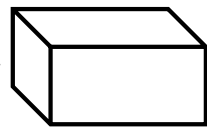
Method : Overview of our network

Dish detection

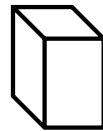
Input



Conv layers

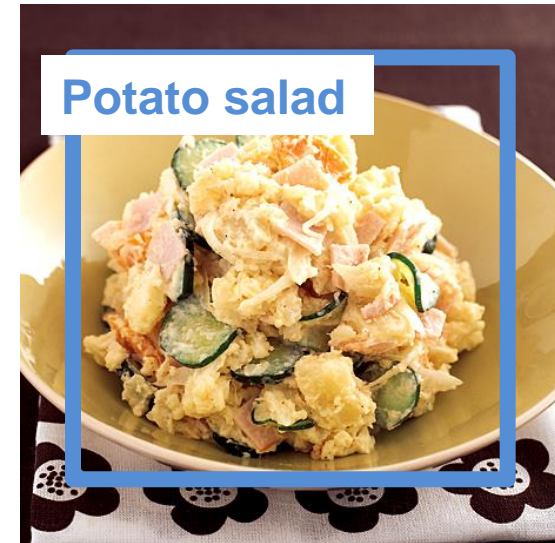


Network output



Non-Maximum suppression
+
Threshold processing

Final output

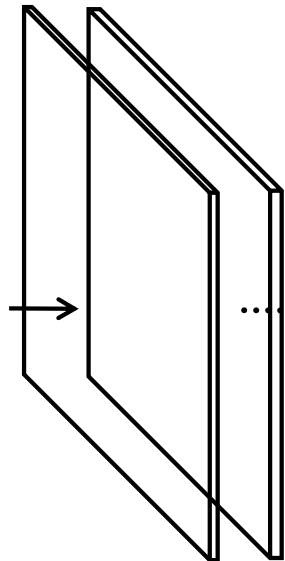


General CNN-based detection network

Method : Overview of our network

Dish detection + Calorie estimation

Input



Conv layers



Network output



Non-Maximum suppression
+
Threshold processing

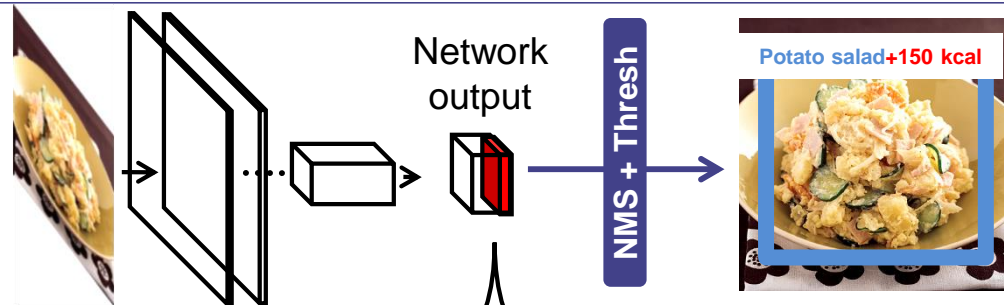
Final output



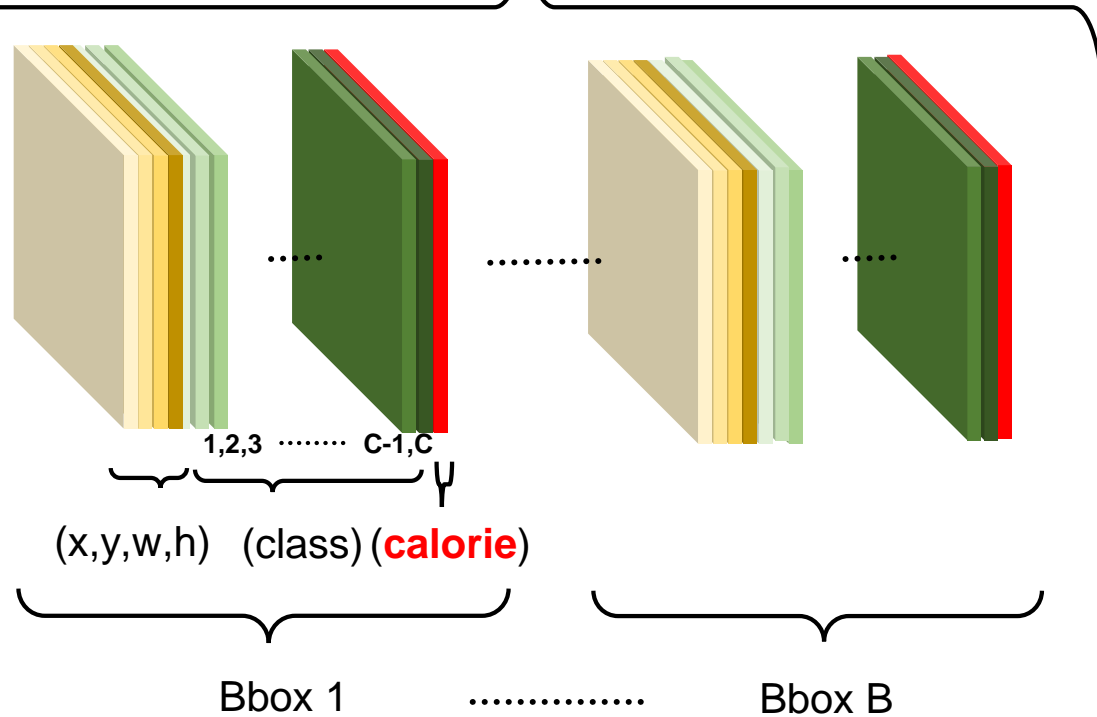
Overview of our network

Method : Overview of our network

Detection Network



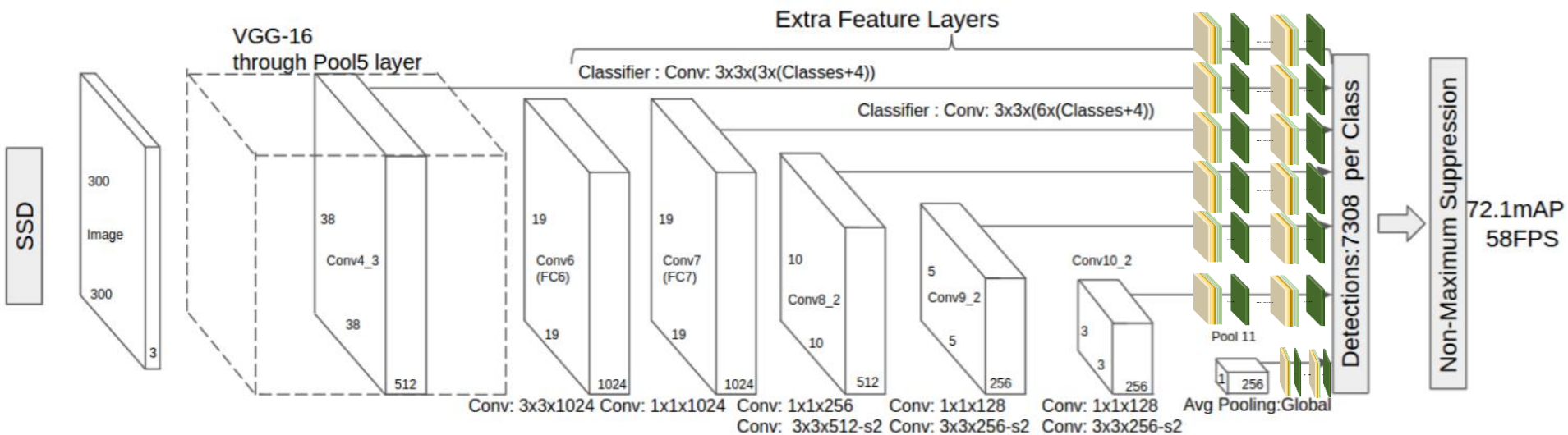
**Output feature map
(Ours)**



Method : Dish detection

Introduction of SSD (Single Shot MultiBox Detector)

- High-speed and highly accurate CNN-based detection system.
- End-to-end learning of the whole system is possible.



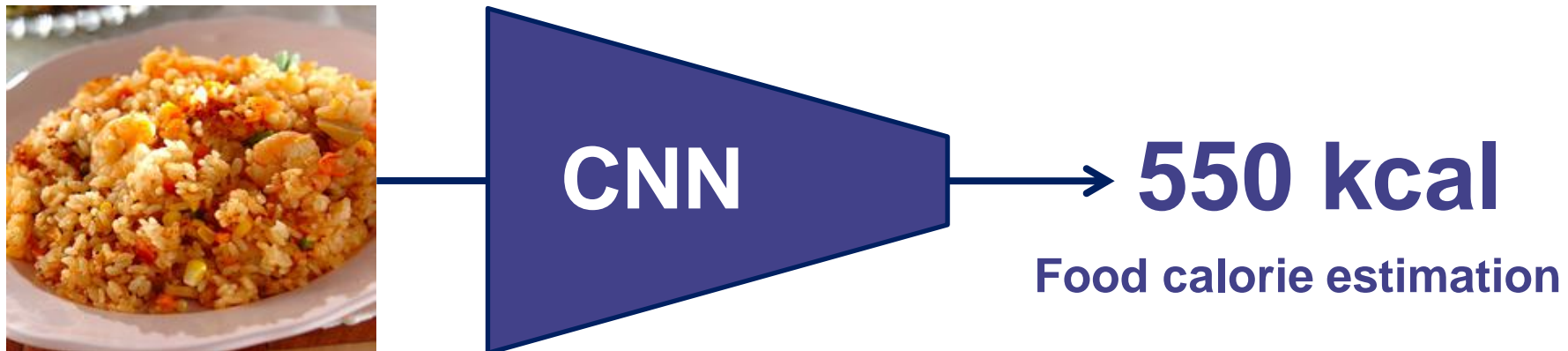
The architecture of network of SSD (quoted from [1]).

[1] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, and S. E. Reed. SSD: single shot multibox detector. CoRR, abs/1512.02325, 2015.

Method : food calorie estimation

Regression based method

- Direct calorie estimation by regression based.
- CNN-based method for single-dish food photos.



Ege and Yanai. Simultaneous estimation of food categories and calories with multi-task cnn. In Proc. of IAPR International Conference on Machine Vision Applications(MVA), 2017.

Method : Multi-task learning

Training of network with two types of datasets

- **Dataset for dish detection**
 - Bounding boxes
 - Classes

- **Dataset for calorie estimation**
 - Food calories
 - Classes



Miso soup
(160 kcal)

Multi-task learning of Dish detection and Calorie estimation with both dataset

Dataset : two types of datasets

Training of network with two types of datasets

- **UECFood-100[1]**
 - With **bounding boxes** and **classes**.
 - Include multiple-dish food photos.
- **Calorie50**
 - With **food calories** and **classes**.
 - Only single-dish food photos.

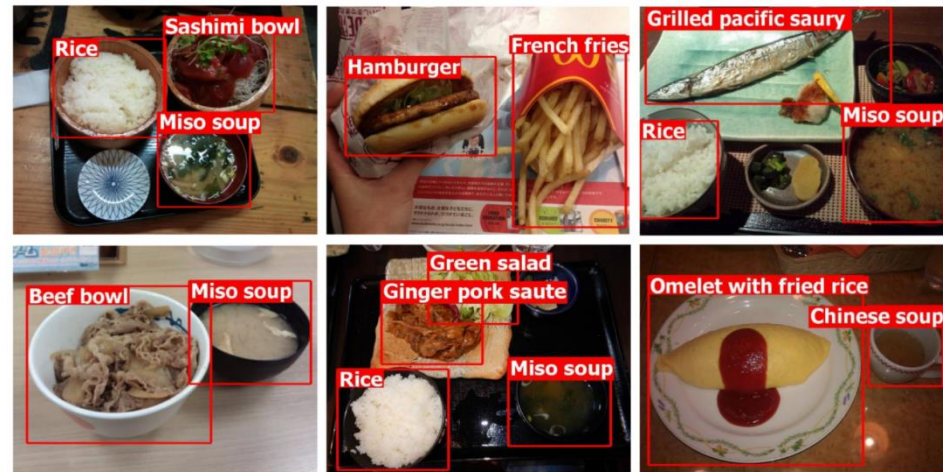


Miso soup
(160 kcal)

[1] Matsuda et al. Recognition of multiple-food images by detecting candidate regions.
In Proc. of IEEE International Conference on Multimedia and Expo, 2012.

Dataset : UECFood-100

- 100 kinds of foods.
- Train: 11566 single-dish food photos.
- Test: 1174 multiple-dish food photos.
- Annotation:
Bounding boxes,
class labels.



Some of examples of test images

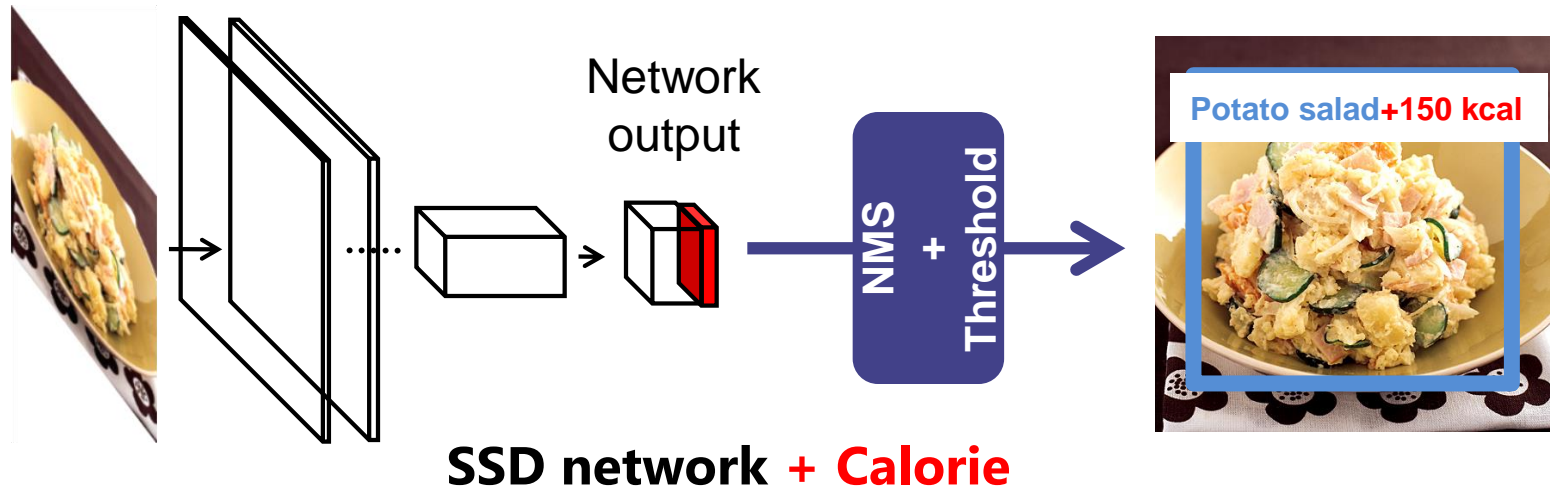
Dataset : Calorie50

- 50 kinds of foods included in UECFood-100.
- Train : 5370 single-dish photos.
- Test : 2317 single-dish photos.
- Annotation:
Food calories,
class labels.



Experiment : Our network

Detection network : SSD[1] +
Calorie estimation : Regression method[2]



[1] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, and S. E. Reed. SSD: single shot multibox detector. CoRR, abs/1512.02325, 2015.

[2] Ege and Yanai. Simultaneous estimation of food categories and calories with multi-task cnn. In Proc. of IAPR International Conference on Machine Vision Applications(MVA), 2017 © 2018 UEC Tokyo.

Experiment : Multi-task learning

Training of network with two types of datasets

- **UECFood-100 – Detection loss function[1]**
 - With bounding boxes and classes.
 - Dish detection
- **Calorie50 – Detection loss[1] + Calorie loss[2]**
 - With **pseudo-bounding boxes**, classes and calories.
 - Food calorie estimation

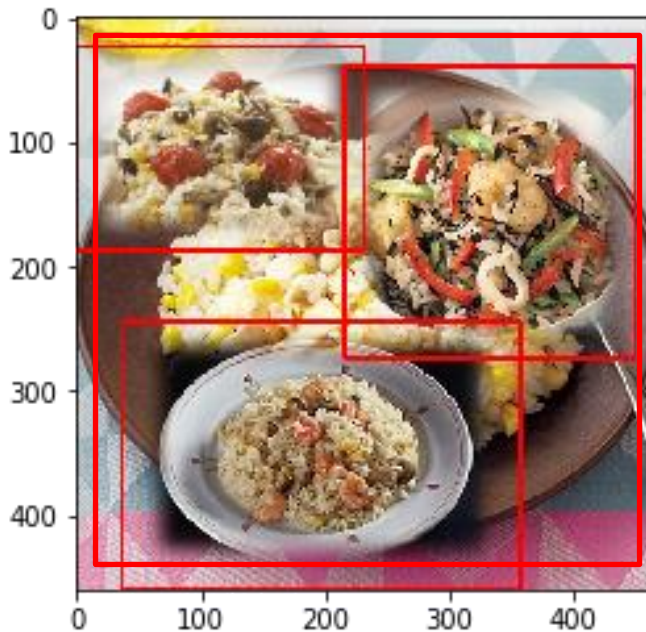
[1] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, and S. E. Reed. SSD: single shot multibox detector. CoRR, abs/1512.02325, 2015.

[2] Ege and Yanai. Simultaneous estimation of food categories and calories with multi-task cnn. In Proc. of IAPR International Conference on Machine Vision Applications(MVA), 2017© 2018 UEC Tokyo.

Experiment : Pseudo-bounding boxes

Bounding box and calorie correspondence required.

➔ pseudo-bounding-boxes to Calorie50.



Pseudo-bounding boxes
for Calorie50 dataset.

- Use Train images.
1. Prepare one background image.
 2. Paste single-dish image at a random position, and make the image area the correct bounding box.

Experiment : Training settings

- **Datasets**
 - **UECFood-100**
 - Train : 11566 images (single-dish)
 - Test : 1174 images (multiple-dish)
 - **Calorie50**
 - Train : 5370 images (single-dish)
+ 10000 images with pseudo-bboxes (multiple-dish)
 - Test : 2317 images (single-dish)
- **Training of networks**
 - Input size : 300 x 300
 - SGD (momentum of 0.9)
 - Mini-batch of 32.
 - 10^{-3} of learning rate for 40,000 iterations and then used 10^{-4} for 10,000 iterations.

Experiment

Comparison of single-task and multi-task learning

- **Baseline (Single-task learning)**
 - Single-task learning of **dish detection task** (a)
 - Single-task learning of **calorie estimation task** (b)
 - Sequential model (a)→(b)
- **Our method (Multi-task learning)**
 - Multi-task learning of dish detection and calorie estimation with both datasets.


Results : Dish detection + Calorie estimation

- Evaluation values

- Detection
 - mAP : mean Average Precision
- Calorie
 - Absolute error (kcal) : $|y_i - g_i|$
 - Relative error (%) : $\frac{|y_i - g_i|}{g_i}$ (Let y_i as the estimated value of an image x_i and g_i as the ground-truth.)
 - Correlation value between estimated calorie value and ground-truth.

| Model | mAP (%) | Rel. err. | Abs. err. | Corr. |
|--|-------------|-------------|-------------|-------------|
| Detection (uecfood100) | 34.1 | --- | --- | --- |
| Detection (uecfood100+calorie50) (a) | 37.8 | --- | --- | --- |
| Calorie estimation (calorie50) (b) | --- | 27.1 | 91.8 | 80.7 |
| Sequential model ((a)→(b)) | --- | 27.3 | 92.5 | 80.5 |
| Detection+Calorie estimation (Ours) | 37.7 | 26.6 | 89.4 | 81.0 |

Results : Dish detection + Calorie estimation

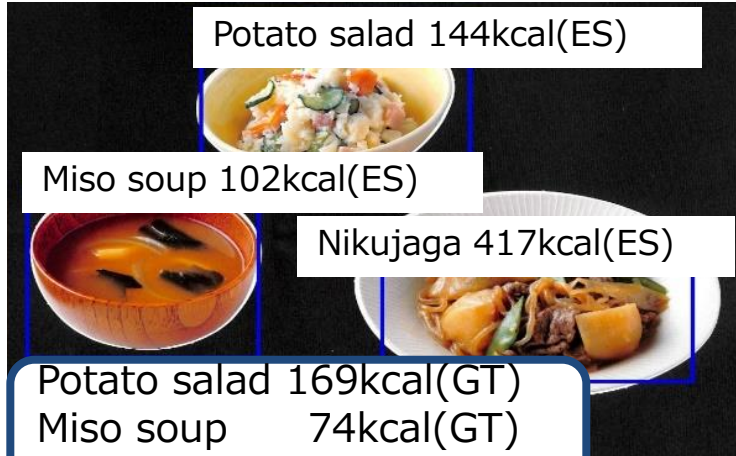


Fried noodle 601 kcal(ES)

Pilaf 495kcal(ES)

Fried noodle 539kcal(GT)

Pilaf 475kcal(GT)



Potato salad 144kcal(ES)

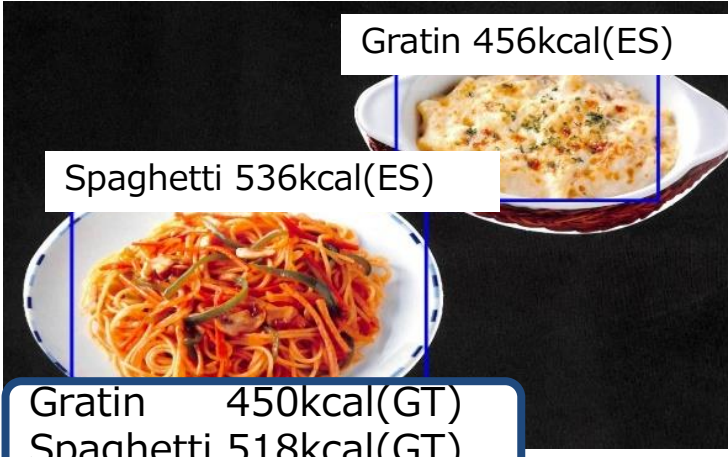
Miso soup 102kcal(ES)

Nikujaga 417kcal(ES)

Potato salad 169kcal(GT)

Miso soup 74kcal(GT)

Nikujaga 352kcal(GT)

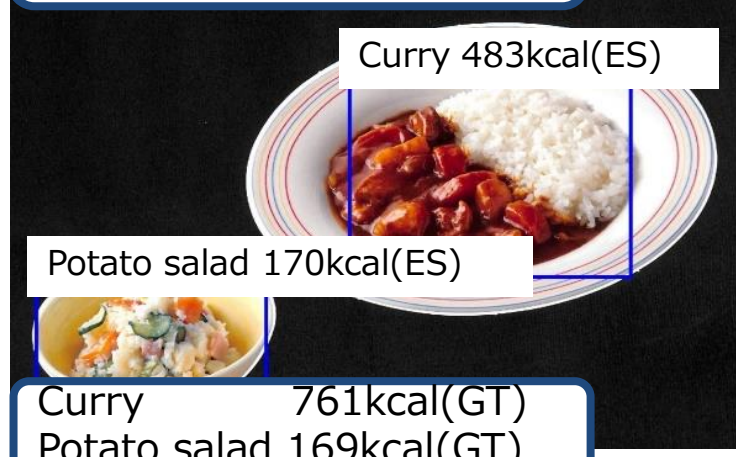


Gratin 456kcal(ES)

Spaghetti 536kcal(ES)

Gratin 450kcal(GT)

Spaghetti 518kcal(GT)



Curry 483kcal(ES)

Potato salad 170kcal(ES)

Curry 761kcal(GT)

Potato salad 169kcal(GT)

The results of dish detection from multiple-dish food photos.(Our papers model)

Conclusion

- We estimate food calories from multiple-dish food photos.
- Multi-task learning of dish detection and food calorie estimation.

Future work

- Construction of large-scale calorie-annotated food photos dataset.
- Comparative experiments with additional learning of dish detection and calorie estimation.

