Estimating Food Calories for Multiple-dish Food Photos

Takumi Ege and Keiji Yanai (The University of Electro-Communications, Tokyo)

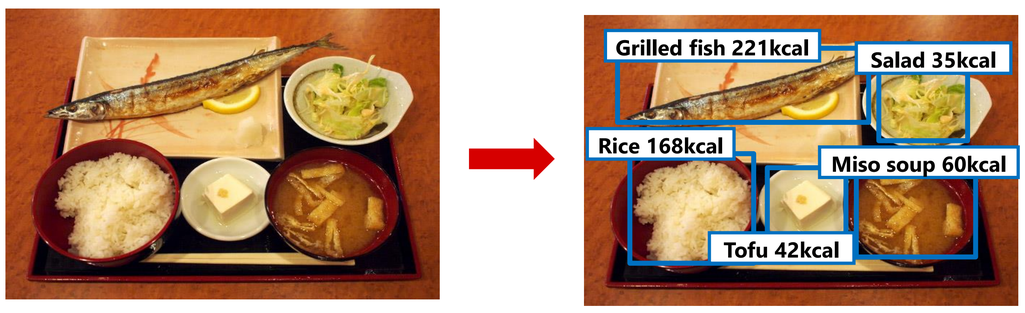
Background & Objective

Some meal recording app can estimate food calories.

But they …

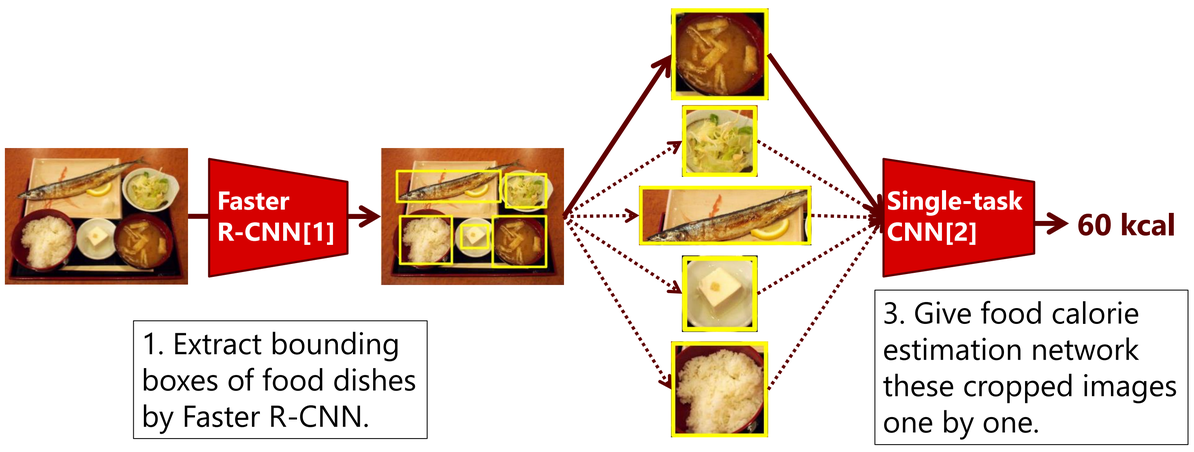
- Need user's manual input of food categories and volumes.
- Estimate food calories for each dish one by one.
- Are paid service to hire nutritionists who estimate food calories.

Purpose : Image-based food calorie estimation



Method

[Flow of Dish Detection and Calorie Estimation]



1. Extract bounding boxes of food dishes by Faster R-CNN.

2. Crop image corresponding to each bounding box.

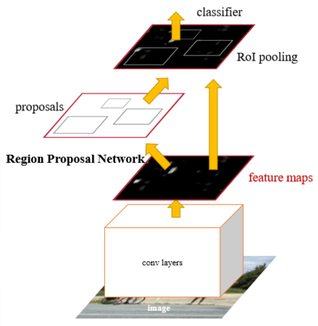
3. Give food calorie estimation network these cropped images one by one.

1. Faster R-CNN : S. Ren et al. [1] 2015

- High-speed and highly accurate detection system.
- End-to-end learning of the whole system is possible.
- Consists of two modules.

Fast R-CNN detector
Simultaneous estimation of classification and bounding boxes.
Convolution once for the entire image.

Region Proposal Network (RPN)
CNN-based region proposal method
Share a conv layers with Fast R-CNN detector.

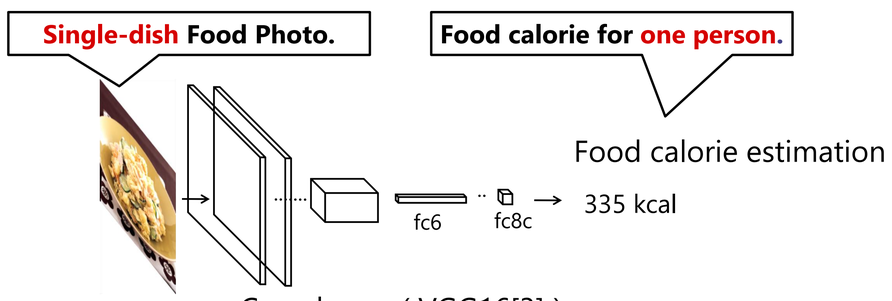


The architecture of network of Faster R-CNN ([1]).

➡ We use Faster R-CNN as a food detector to detect each dish in a food image.

2. Single-task CNN : Ege and Yanai [2] 2017

- Image-based food calorie estimation with CNN.
- Regression based-method.
- Output food calories directly from single-dish food photos.



- We denote $\boldsymbol{L\_re}$ as an relative error and $\boldsymbol{L\_ab}$ as a absolute error, $\boldsymbol{L\_cal}$ is defined as follows:

$$\boldsymbol{L_{cal}} = \lambda_{re} L_{re} + \lambda_{ab} L_{ab}$$

$$L_{re} = \frac{|y - g|}{g} \qquad L_{ab} = |y - g|$$

(y is an estimated food calorie. g is ground-truth.)

➡ The food calorie of each detected dish are estimated by image-based food calorie estimation.

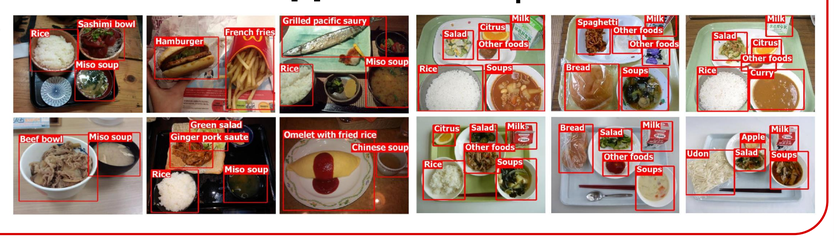
Experiment 1 : Food detection

DATASET

Two kinds of bounding box annotated food photo datasets.

- **UEC Food-100[4]**
- **Japanese school lunch**

Japanese school lunch photos are collected from online school lunch sites.



[Food Detection① (on UEC Food-100)]

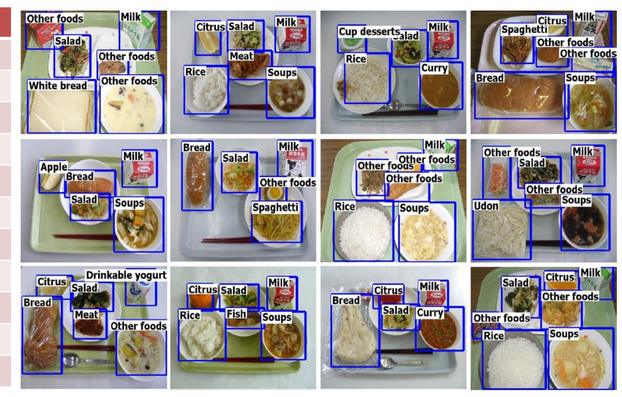
- Comparison to an exist work of Shimoda et al. [5] .
- We use mean Average Precision(mAP) of PASCAL VOC detection task for the evaluation.
- 11,566 single-dish photos for training, 1,174 multiple-dish photos for evaluation.

UEC FOOD-100 mAP(%)	100 class (all)	53 class (test images ≧10)	11 class (test images ≧50)
R-CNN	26.0	21.8	25.7
[5]'s method	**49.9**	**55.3**	55.4
Faster R-CNN	42.0	46.3	**57.9**

[Food Detection② (on Japanese school lunch)]

- 80% of food photos for training and 20% for the evaluation.

	クラス	AP (%)		クラス	AP (%)
1	Milk	99.6	12	Soups	92.2
2	Drinkable yogurt	90.6	13	Curry	95.1
3	Rice	99.7	14	Spicy chili tofu	99.8
4	Mixed rice	82.7	15	Bibimbap	72.9
5	Bread	95.5	16	Fried noodles	79.9
6	White bread	83.7	17	Spaghetti	90.7
7	Udon	98.0	18	Citrus	99.6
8	Fish	78.3	19	Apple	98.5
9	Meat	70.8	20	Cup desserts	93.1
10	Salad	94.0	21	Other foods	90.4
11	Cherry tomatoes	100.0		mAP	**90.7**



Experiment 2 : Food calorie estimation of multiple dishes

DATASET

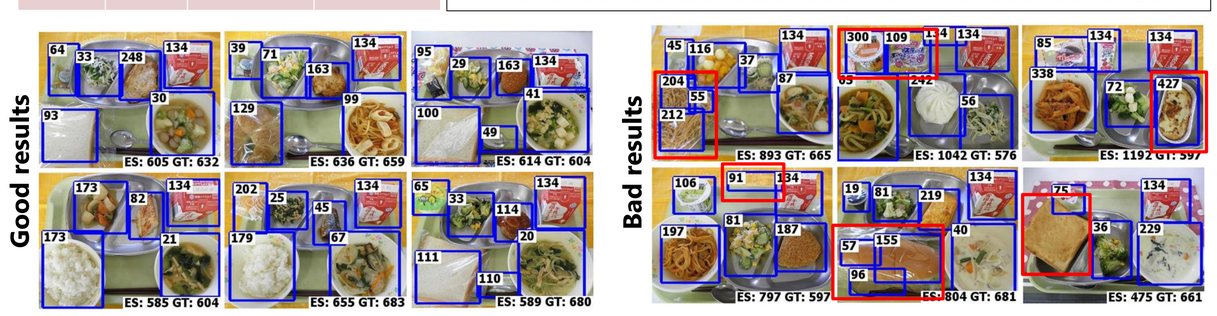
- **Bounding box annotated japanese school lunch photos** for training of Faster R-CNN [1] .
- **Calorie-annotated food photos** [2] for training of Single-task CNN [2] .
- **Total food calorie annotated Japanese school lunch photos** for the evaluating .

We also collected calorie annotated school lunch photos for the evaluation.
Each image has a total calorie value of all the dishes.

- We fixed the calorie of "Milk" detected by Faster R-CNN to 134 kcal.
- 690 total food calorie annotated Japanese school lunch photos for the evaluating.

Rel. error (%)	Abs. error (kcal)	<20% (Rel. err.) (%)	<40% (Rel. err.) (%)
21.4	136.8	53.0	85.1

Relative error (%) :
The differences between estimated values and ground-truth.
Absolute error (kcal) :
The ratio between absolute error and ground-truth.
< 20% (Relative error) (%) :
The ratio of the estimated value within the relative error 20 %.



The numbers in bounding boxes are estimated food calories of foods in each bounding box (kcal).
ES : the estimated total food calorie (kcal).
GT : the ground-truth of total food calorie (kcal).

Conclusion & Future work

- Food detection by Faster R-CNN.
- We collected school lunch photo dataset by Web image mining.
- We estimate food calories from multiple-dishes food photos.

- Multi-task learning of food detection and food calorie estimation.
- Construction of large-scale food photos dataset.

[1] S. Ren et al. Faster R-CNN: Towards realtime object detection with region proposal networks. NIPS 2015.
[2] Ege and Yanai. Simultaneous estimation of food categories and calories with multi-task cnn. MVA 2017.
[3] K. Simonyan and A Zisserman. Very deep convolutional networks for large-scale image recognition. In arXiv preprint arXiv:1409.1556, 2014.
[4] Y. Matsuda, H. Hajime, and K. Yanai. Recognition of multiple-food images by detecting candidate regions. In Proc. of IEEE International Conference on Multimedia and Expo, 2012.
[5] 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.