

Multi-task Learning of Dish Detection and Calorie Estimation

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

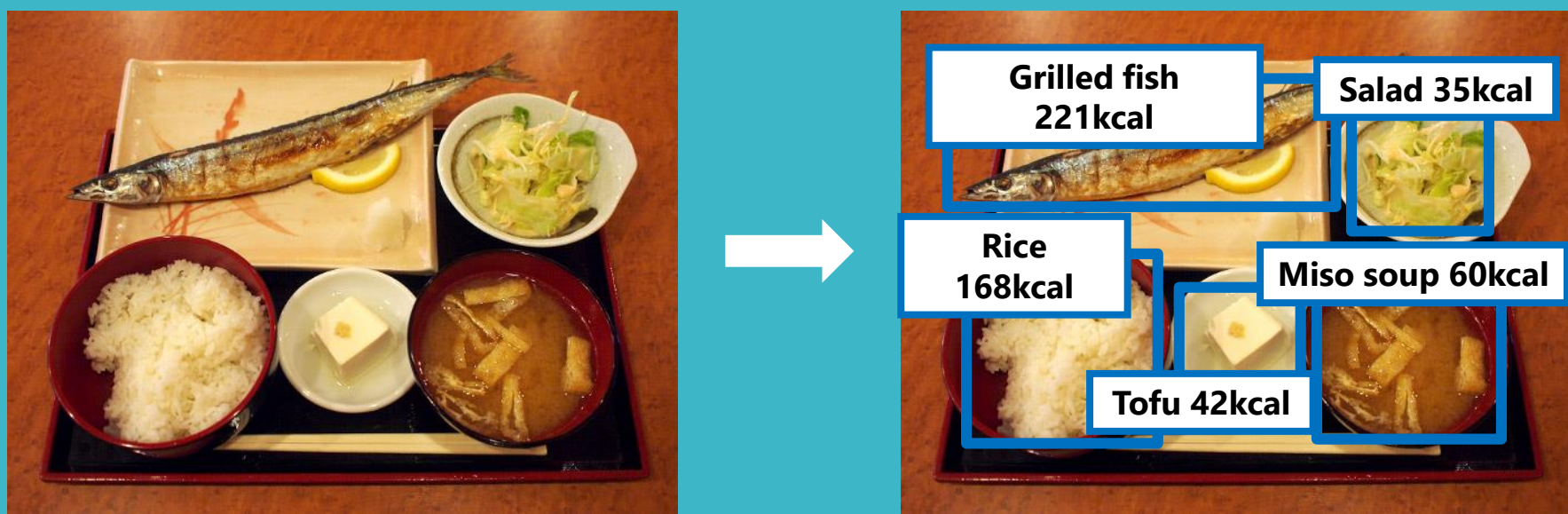
Background

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



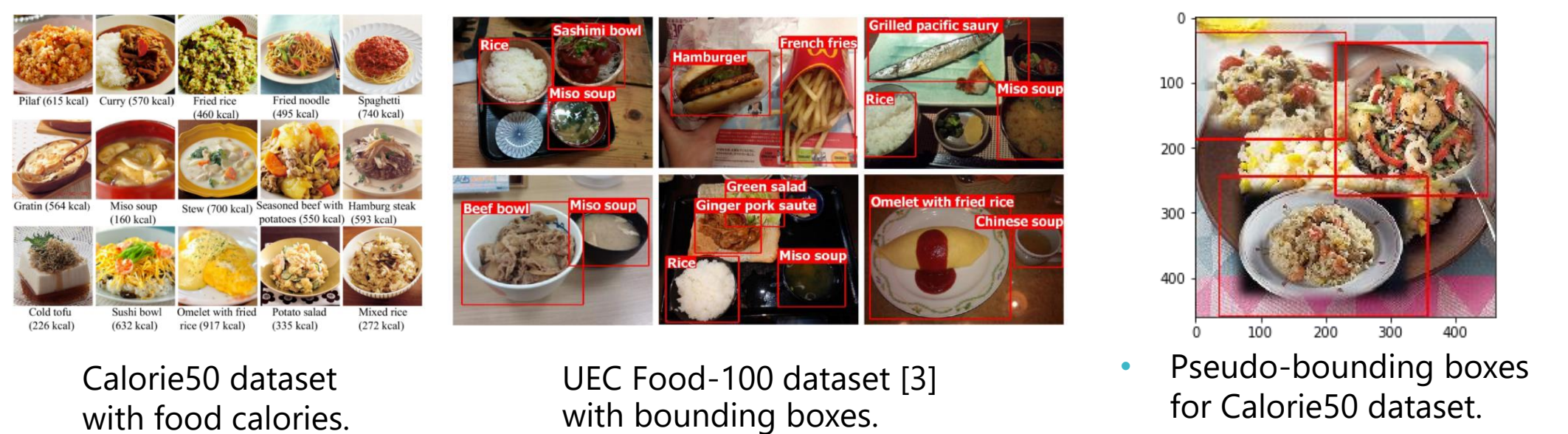
Experiment

• Baseline (Single-task learning)

- Single-task learning of **dish detection task** with **UECFood-100 dataset** [3]. (a)
- Single-task learning of **calorie estimation task** with **Calorie-50 dataset**. (b)
- Sequential model ((a)→(b)) that detects each dish in a food image by (a) and the food calorie of each dish image based on detected results are estimated by (b).

• Our method (Multi-task learning)

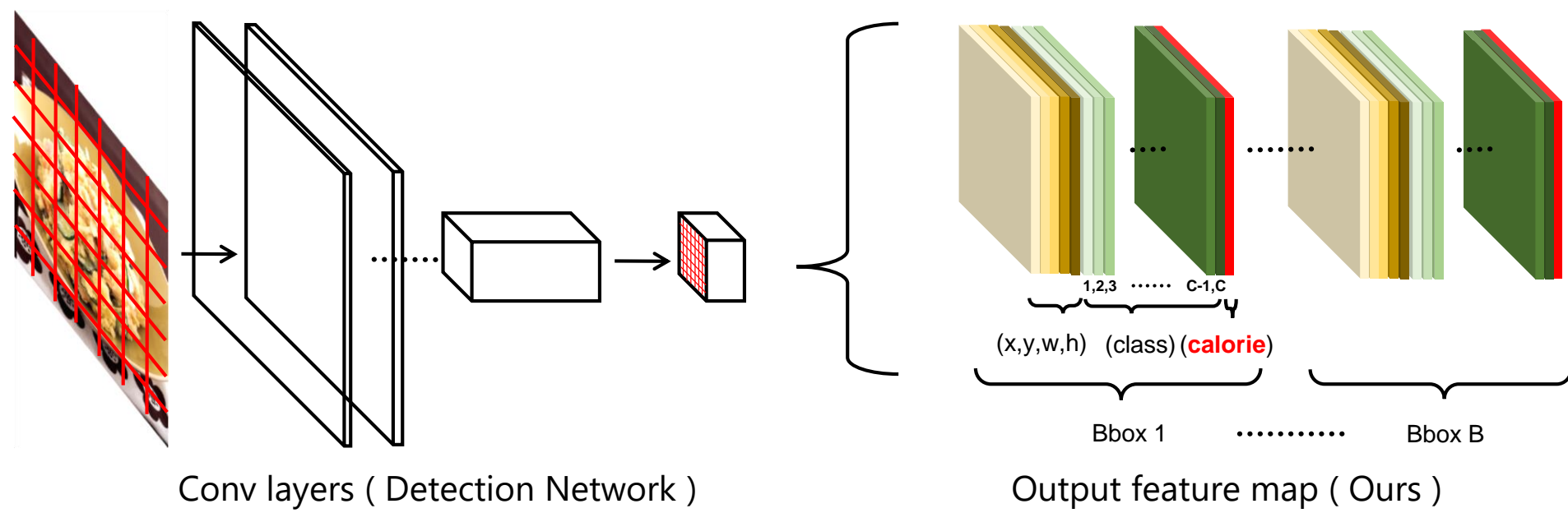
- Multi-task learning of dish detection and calorie estimation with both datasets. In training of **UECFood-100 dataset**, we use **the detection loss**. In training of **Calorie-50 dataset**, we use **the calorie loss and detection loss**.



Method

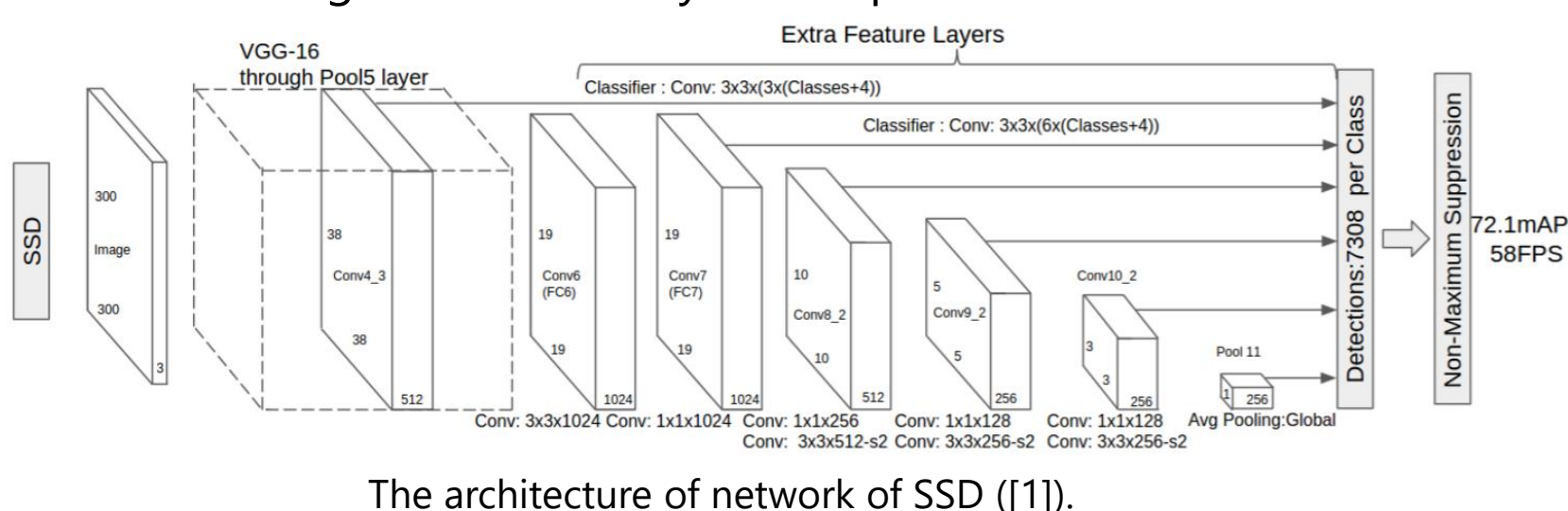
Multi-task learning of dish detection and food calorie estimation

- Simultaneous estimation of bounding boxes of food dishes and their calories.
- The output channels of estimated calories are added to the output features.



1. SSD : Wei Liu et. al. [1] 2016 (Detection Net)

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

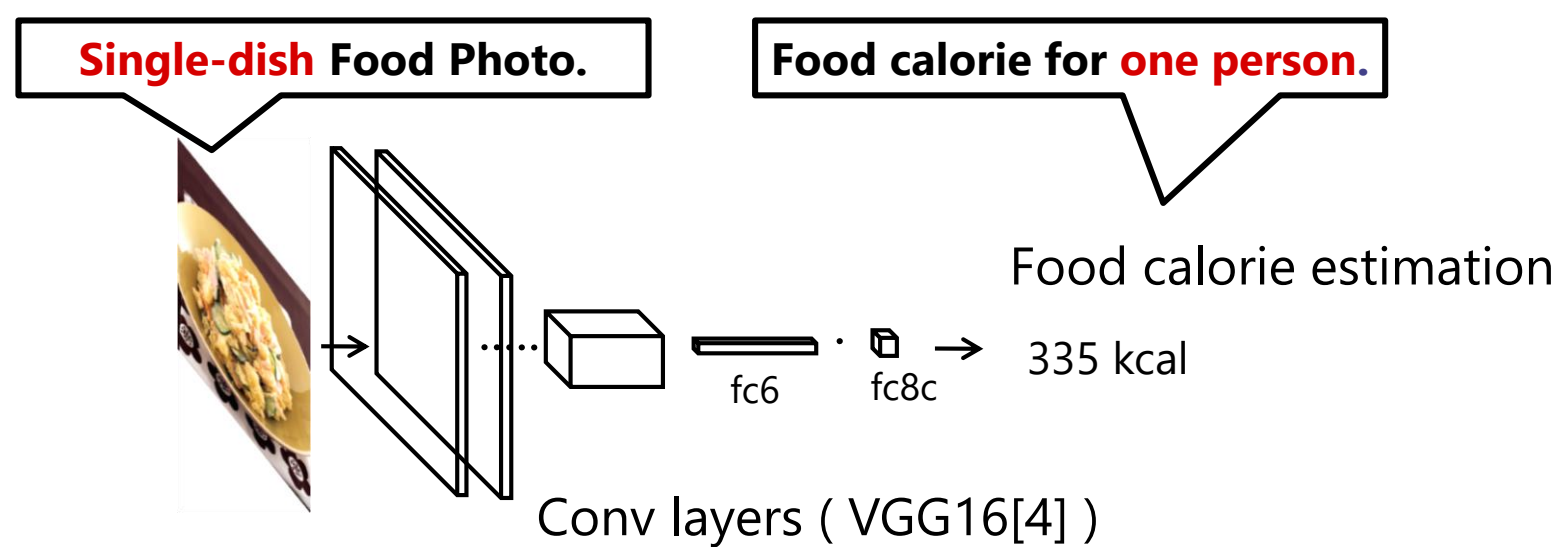


The architecture of network of SSD ([1]).

→ We use the network for single-dish detection from multiple-dish photos.

2. Single-task CNN : Ege and Yanai [2] 2017 (Calorie estimation Net)

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



- We denote L_{re} as an relative error and L_{ab} as a absolute error, L_{cal} is defined as follows:

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

$$L_{re} = \frac{|y - g|}{g} \quad L_{ab} = |y - g| \quad \left\{ \begin{array}{l} y \text{ is an estimated food calorie.} \\ g \text{ is ground-truth.} \end{array} \right.$$

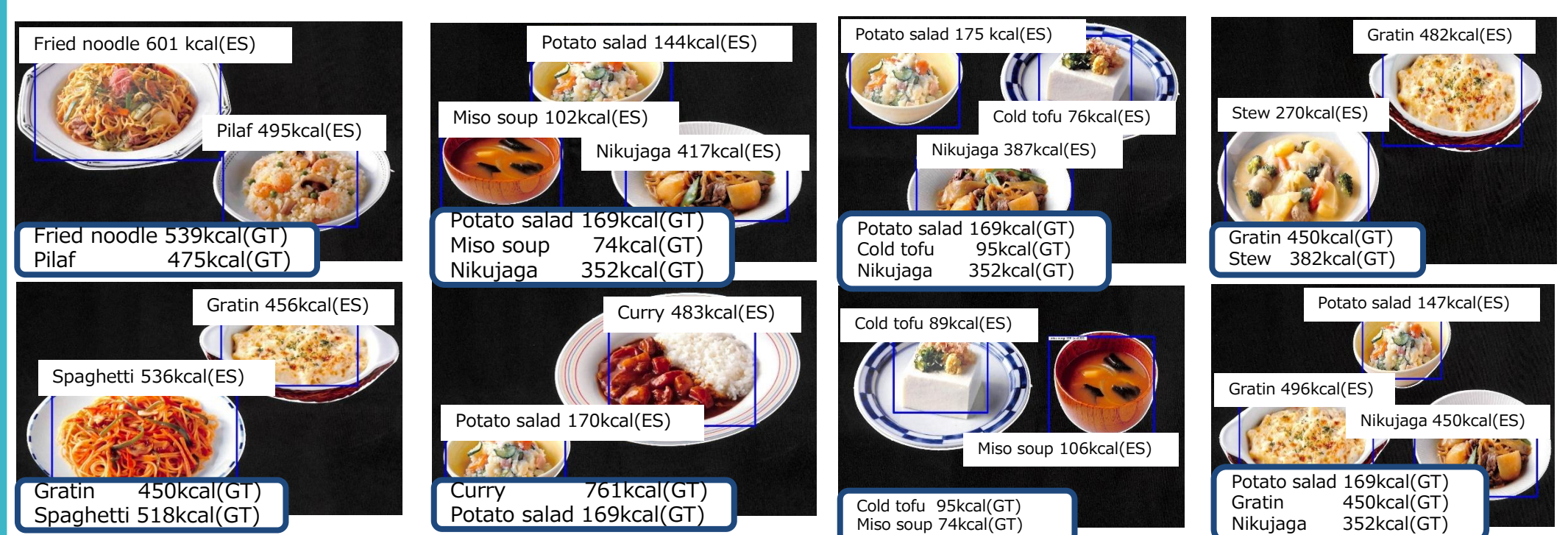
→ We use the network for food calorie estimation.

Result

Comparison of single-task and multi-task learning

Model	mAP (%)	Rel. err.	Abs. err.	Corr.	≤ 20%err.
Detection (uecfood100)	34.1	---	---	---	---
Detection (uecfood100+calorie50) (a)	37.8	---	---	---	---
Calorie estimation (calorie50) (b)	---	27.1	91.8	80.7	50.9
Sequential model ((a)→(b))	---	27.3	92.5	80.5	50.6
Detection + Calorie estimation (Ours)	37.7	26.6	89.4	81.0	50.7

- Our method achieves the calorie estimation from multiple-dish photos without degrading dish detection accuracy.
- Our method achieves **high speed** and **save memory** by simultaneous estimation in a single network.



Examples of dish detection and food calorie estimation from multiple-dish food photos. (based on YOLOv2)

Conclusion

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

Future work

- Image-based food calorie estimation based on amount of food.
- Construction of large-scale food photos dataset.

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

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

[3] 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.

[4] K. Simonyan and A Zisserman. Very deep convolutional networks for large-scale image recognition. In arXiv preprint arXiv:1409.1556, 2014.