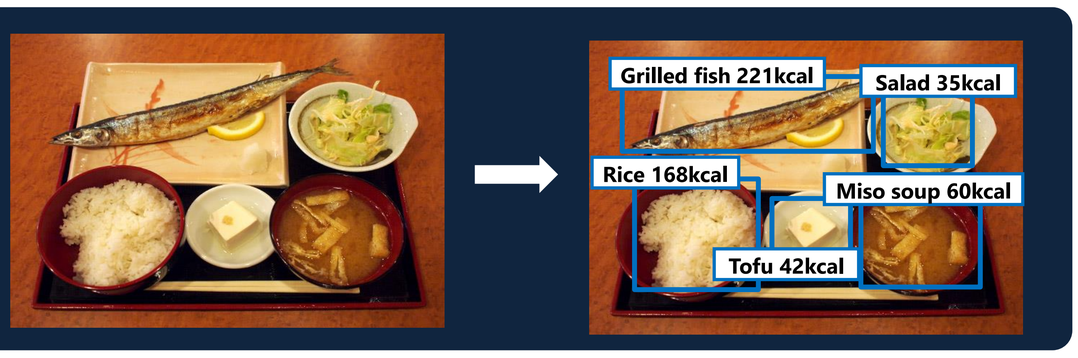
Image-Based Food Calorie Estimation Using Knowledge on Food Categories, Ingredients and Cooking Directions

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

Background & Objective

Meal management apps enable us to record food calories.
Some of them need human help for calorie estimation.

Image-based food calorie estimation using recipe information



Method: Multi-task CNN with recipe info.

Related work ① : Miyazaki et al.[1] 2011

Search-based food calorie estimation with **conventional features**.
Similar image retrieval with SURF and color histograms and so on.
Calculate food calories from retrieved images' calories.

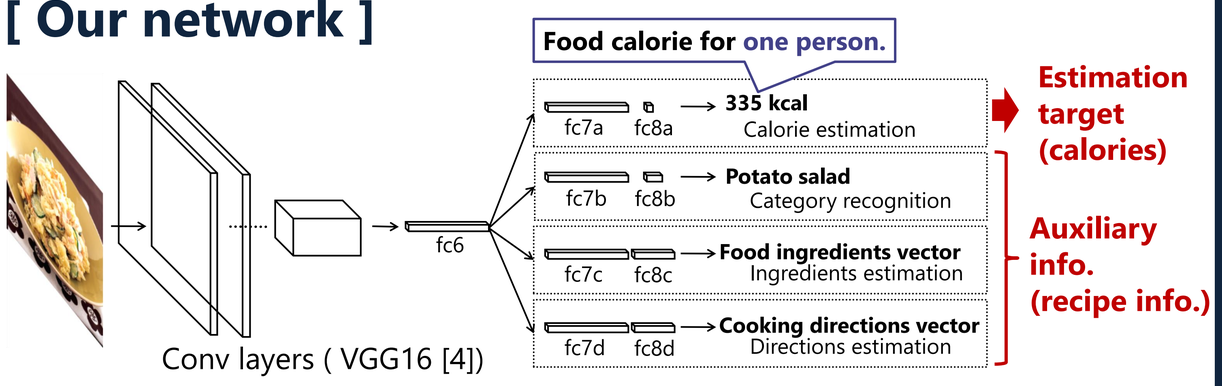
➡ We propose regression-based method using **CNN**.

Related work② : Chen and Ngo[2] 2016

Multi-task estimation of food categories and food ingredients.
Multi-task CNN[3] of food categories and food ingredients
Multi-task learning improve both task's performance.

➡We use **multi-task CNN** with **recipe info. for calorie estimation**.

[Our network]



➡ **It is expected to improve the accuracy of each task.**

[Loss function]

Total loss (combining four tasks) :

$$L = \frac{1}{N}\sum_{n=0}^{N}\left(\underline{L_{cal}} + \lambda_{cat}\underline{L_{cat}} + \lambda_{ing}\underline{L_{ing}} + \lambda_{dir}\underline{L_{dir}}\right)$$

(λ is the weight on the loss.)

Calorie estimation loss / Category recognition loss / Ingredients estimation loss / Direction estimation loss

(1) Calorie estimation loss:

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

relative err. loss $L_{re} = \frac{|y-g|}{g}$ absolute err. loss $L_{ab} = |y-g|$ (y is an estimated food calorie. g is ground-truth.)

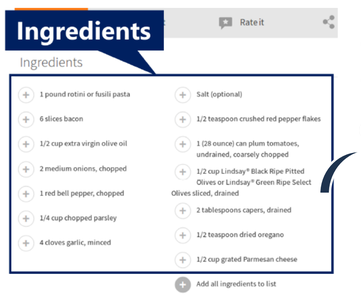
(2) Food Category loss:

$$L_{cat} = -\sum_{k=1}^{n} g_k \log y_k$$ (softmax cross entropy)

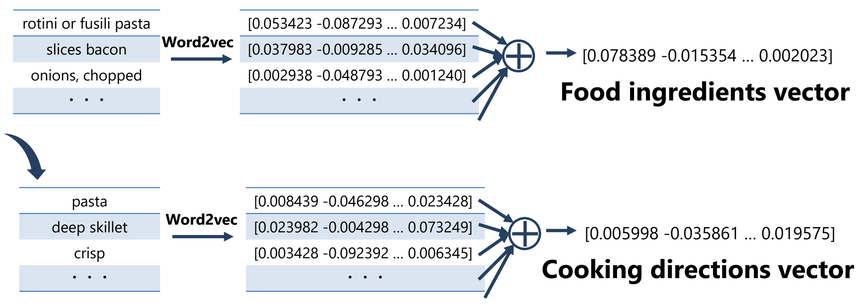
(3) Food Ingredient loss, (4) Cooking Direction loss: (mean squared error)

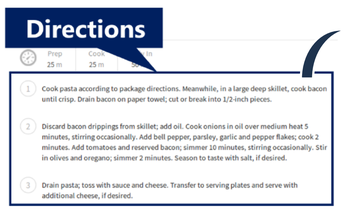
$$L_{ing,dir} = \frac{1}{2}\sum_{k=1}^{n}(y_k - g_k)^2$$

(g is a ingredient/cooking direction vector.)



We collected recipe data from the online cooking recipe sites. We converted each word of food ingredient names and cooking directions into a real-value vector by Word2Vec, and averaged them.

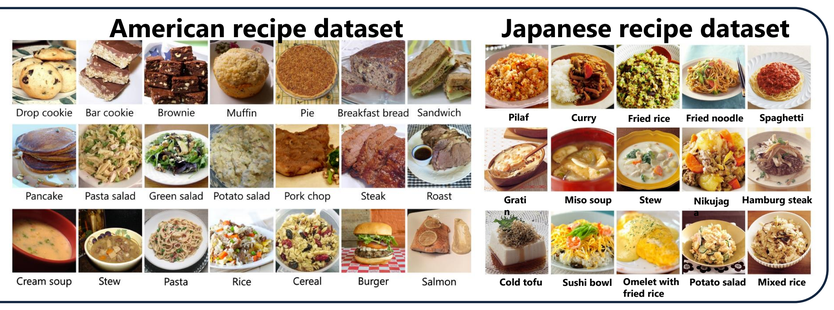




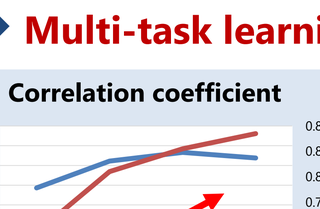
Experiment : Comparison of single & multi-task

DATASET

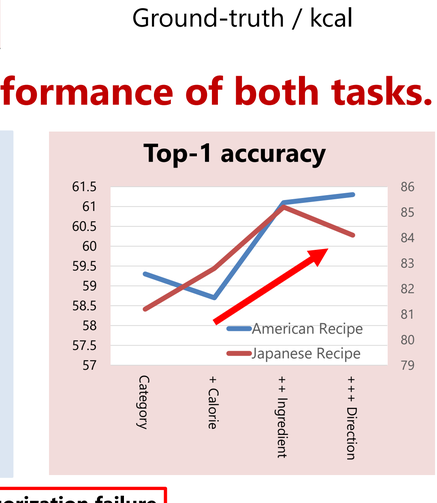
Two kinds of calorie-annotated food photo datasets.
American recipe datasets
Japanese recipe datasets
We collected recipe data from the online cooking recipe sites.



[Single-task CNN (baseline)]



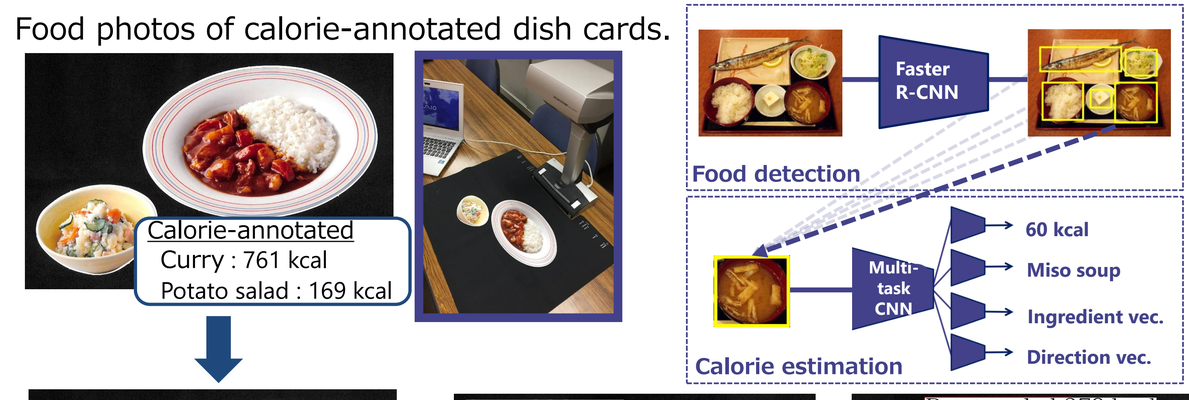
[Multi-task CNN (proposal method)]

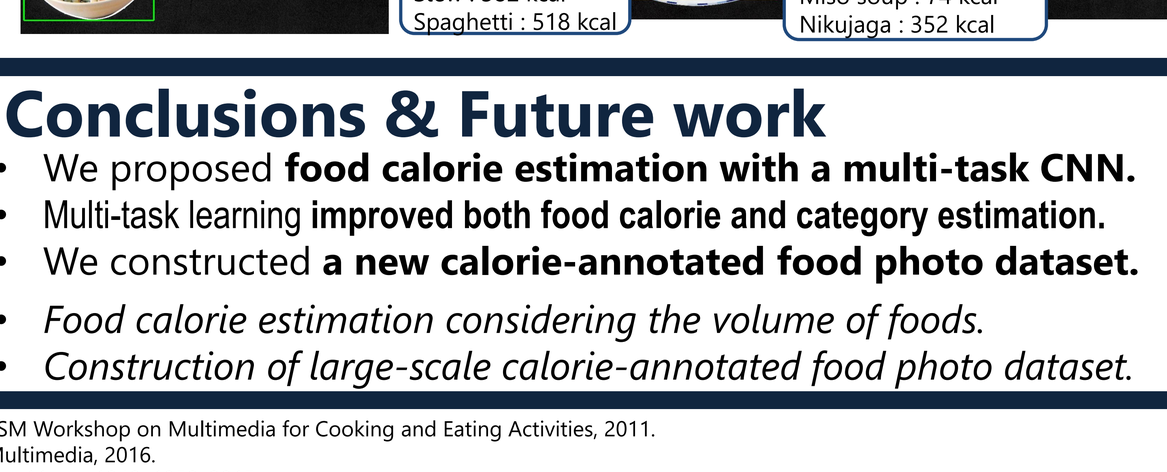


Japanese Recipe	Single	Multi	
Correlation coefficient	0.78	**0.82**	+0.04
Within 20% error	46 %	**50 %**	+4 %
Top-1 accuracy (categorization)	81 %	**84 %**	+3 %

American Recipe	Single	Multi	
Correlation coefficient	0.29	**0.37**	+0.08
Within 20% error	32 %	**34 %**	+2 %
Top-1 accuracy (categorization)	54 %	**61 %**	+7 %

➡ **Multi-task learning improves the performance of both tasks.**





Extension : Combining with Food Detector

Food photos of calorie-annotated dish cards.

Conclusions & Future work

- We proposed **food calorie estimation with a multi-task CNN.**
- Multi-task learning **improved both food calorie and category estimation.**
- We constructed **a new calorie-annotated food photo dataset.**
- *Food calorie estimation considering the volume of foods.*
- *Construction of large-scale calorie-annotated food photo dataset.*

[1] T. Miyazaki, G. Chaminda, D. Silva, and K. Aizawa. Image - based calorie content estimation for dietary assessment. In Proc. of IEEE ISM Workshop on Multimedia for Cooking and Eating Activities, 2011.
[2] J. J. Chen and C. W. Ngo. Deep-based ingredient recognition for cooking recipe retrival. In Proc. of ACM International Conference Multimedia, 2016.
[3] H. A. Abrar, W. Gang, L. Jiwen, and J. Kui. Multi-task CNN model for attribute prediction. IEEE Transactions on Multimedia, Vol. 17, No. 11, pp. 1949–1959, 2015.
[4] K. Simonyan and A Zisserman. Very deep convolutional networks for large-scale image recognition. In arXiv preprint arXiv:1409.1556, 2014.
[5] 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