AR DeepCalorieCam: An iOS App for Food Calorie Estimation with Augmented Reality

Ryosuke Tanno, Takumi Ege, and Keiji Yanai

Department of Informatics, The University of Electro-Communications, Tokyo
1-5-1 Chofugaoka, Chofu-shi, Tokyo 182-8585 JAPAN
{tanno-r, ege-t, yanai}@mm.inf.uec.ac.jp

Abstract. A food photo generally includes several kinds of food dishes. In order to recognize multiple dishes in a food photo, we need to detect each dish in a food image. Meanwhile, in recent years, the accuracy of object detection has improved drastically by the appearance of Convolutional Neural Network (CNN). In this demo, we present two automatic calorie estimation apps, *DeepCalorieCam* and *AR DeepCalorieCam*, running on iOS. *DeepCalorieCam* can estimate food calories by detecting dishes from the video stream captured from the built-in camera of an iPhone. We use YOLOv2 [1] which is the state-of-the-art object detector using CNN, as a dish detector to detect each dish in a food image, and the food calorie of each detected dish is estimated by image-based food calorie estimation [2, 3]. *AR DeepCalorieCam* is a combination of calorie estimation and augmented reality (AR) which is an AR version of *DeepCalorieCam*.

Keywords: Food calorie estimation, Object detection, Food image recognition, Convolutional Neural Network (CNN), Augmented reality (AR)

1 Introduction

In recent years, due to growing of health consciousness, various food photo recognition applications for recording everyday meals have been proposed. Although some applications use image-based classification for estimating food categories, most of them do not leverage object detection. This means that human assistance is necessary to detect dishes individually for the case that multiple dishes are included in a food photo. In fact, we often encounter the situation with multiple food dishes.

On the other hand, in the field of image recognition, a lot of methods using CNN have achieved various improvements. Currently, its applied technology is used for various applications. Therefore, in this demo, we use YOLOv2 [1] which is the state-of-the-art object detection system using CNN, to detect food dishes from a food photo. YOLOv2 is the latest system on object detection using CNN, and achieved high-speed and highly accurate detection. Since YOLOv2 outputs rectangular bounding boxes of objects with class labels, it is possible to aware individual objects of the same category. In case of a food photo of multiple dishes, this means it is possible to estimate each dish area in the whole image. Therefore, if we combine a system that works for a single-label image and object detection, it is possible to recognize more detail from a detected rectangular area of each object.

In this demo, we present two automatic calorie estimation app, *DeepCalorieCam* and *AR DeepCalorieCam*, running on iOS. Ege et al. [2] proposed image-based food calorie estimation which is simultaneous estimation of food categories and calories for food photos. However, this is limited to food photos which contains only one dish, so it is effective to combine with this dish detector for estimating food calories from food photos of multiple dishes. Figure 1 shows an example usage of *DeepCalorieCam* which is running on an iPhone 7 Plus. In addition, *AR DeepCalorieCam* is a combination of calorie estimation and augmented reality (AR) instead of detecting dishes as shown in Figure 2. By annotating calories of foods on the AR space, the system can recognize all the foods in the AR space by moving a smartphone.

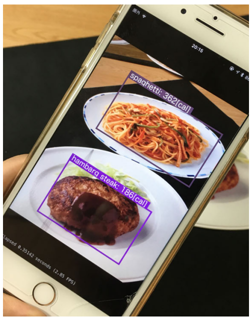

Fig. 1. *DeepCalorieCam*

Fig. 2. *AR DeepCalorieCam*

2 Proposed System

2.1 DeepCalorieCam

We estimate food calories from food photos according to [2, 3]. They proposed image-based food calorie estimation which is simultaneous estimation of food categories and calories for food photos. They collected calorie annotated recipe data from the online cooking recipe sites, and trained CNN that output food calories directly from a food photo that contained only one dish.

The network they use for food calorie estimation is based on VGG16 [4]. The fc6 layer is shared by both tasks, and the fc7 layer is branched to each task, so that each task has the fc7 layer and the output layer independently. The food calorie estimation task has the fc7 layer with 4096 dimension and an output-layer composed of one unit which outputs food calorie. The food categorization task has the fc7 layer with 4096 dimension and an output layer composed of units corresponding to each category.

However, considering the mobile implementation, we think that VGG16 is not suitable for the following reasons. (Please also see Figure 4 which shows the relationship between Inference time on iPhone 8 Plus, ImageNet Top-1 Accuracy and the size of the model weights expressed in the size of circles.)

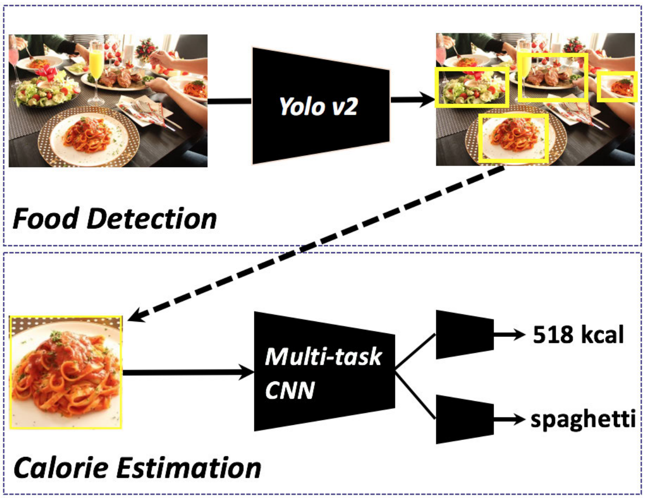

Fig. 3. Flow of Dish Detection and Calorie Estimation

1. The size of the model weights (553MB) is too large for mobile implementation.
2. The inference time is longer.

In particular, for mobile implementation, the memory capacity and processing time of the device is an important factor in implementing deep learning. Also, from the research of [2, 3], it is suggested that the result of calorie estimation tends to depend on classification accuracy. For this time, we decided to use inception-v3 [5] which has the lighter memory, faster inference and higher classification accuracy.

Figure 3 shows a flow of dish detection and calorie estimation system. We combine this food calorie estimation network and YOLOv2 [1] for food photos of multiple dishes. Firstly, we extract bounding boxes of food dishes by YOLOv2 from a food photo of multiple dishes, and obtain a cropped image corresponding to each bounding box which contains only one dish. Then, we provide the cropped dish images to the food calorie estimation network one by one. Finally, the total amount of food calories are calculated from food calories estimated for all the cropped dishes.

2.2 AR DeepCalorieCam

We propose a new food recognition method using AR technology. Since *AR DeepCalorieCam* omits YOLOv2 which performs dish detection, it recognizes only one food at the same time. However, since the recognition results remain in the 3D AR-view space, we realized sequential multiple food recognition by moving a smartphone.

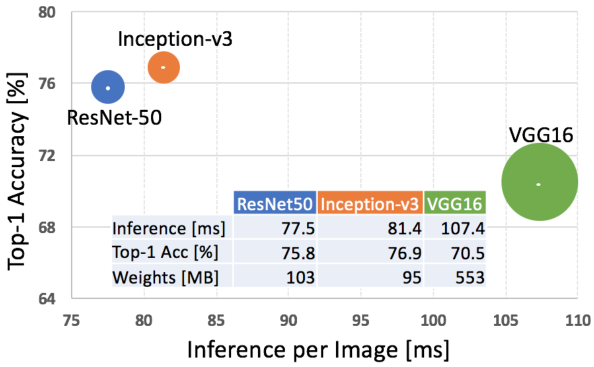

Fig. 4. Inference time on iPhone 8 Plus, ImageNet Top-1 Accuracy and the size of the model weights expressed as the size of circles.

3 Demo Video

We prepared the videos recorded that *DeepCalorieCam* and *AR DeepCalorieCam* app were running in the two kinds of the settings.

- *DeepCalorieCam*
  https://www.youtube.com/watch?v=FOho7WyntOw
- *AR DeepCalorieCam*
  https://www.dropbox.com/s/huwzhe9gwft2zcx/deepcaloriecam_ar.mp4
- Project Pages
  http://foodcam.mobi/deepcaloriecam/

We will release *DeepCalorieCam* and *AR DeepCalorieCam* at iOS App Store by the time of the MMM2018 conference.

References

1. Redmon, J., Farhadi, A.: YOLO9000: Better, Faster, Stronger. In: Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2017)
2. Ege, T., Yanai, K.: Simultaneous Estimation of Food Categories and Calories with Multi-task CNN. In: Proc. of ACPR International Conference on Machine Vision Applications (MVA). (2017)
3. Ege, T., Yanai, K.: Estimating Food Calories for Multiple-dish Food Photos. In: Proc. of Asian Conference on Pattern Recognition (ACPR). (2017)
4. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. In: Proc. of International Conference on Learning Representations. (2015)
5. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the Inception Architecture for Computer Vision. In: Proc. of arXiv:1512.00567. (2015)