CalorieCaptorGlass: Food Calorie Estimation Based on Actual Size using HoloLens and Deep Learning

Shu Naritomi*
The University of Electro-Communications, Japan

Keiji Yanai†
The University of Electro-Communications, Japan

ABSTRACT

We propose "CalorieCaptorGlass", an image-based food calorie estimation system implemented on HoloLens. We take advantage of the function of 3D environment recognition of HoloLens for estimating the actual size of foods for accurate estimation of their amounts and calories. We combine it with CNN-based food image segmentation, and estimate calorie intake of 46 kinds of meals based on estimated 2D sizes and meal-dependent quadratic-curve between meal size and calories. The estimated meal categories and calories are displayed on HoloLens over the AR space, which invites us to the new eating experience.

Keywords: Food calorie estimation, Image recognition, Head mounted display, HoloLens

Index Terms: Human-centered computing—Human computer in-teraction (HCI)—Mixed / augmented reality—; Computing methodologies—Machine learning—Machine learning approaches—Neural networks

1 INTRODUCTION

Recently, device-based healthcare has attracted attention. For example, a wristwatch-type device constantly monitors the user's heart rate, records calories burned during exercise, and manages the health status. However, it is difficult to record calorie intake from meals with such devices. To make it easier to record meal calories, smartphone apps such as AR DeepCalorieCam V2 [4] and DepthCalorieCam [1] have been proposed. In AR DeepCalorieCam V2, a user has to specify meal areas by moving a smartphone. DepthCalorieCam can estimate food volumes with builtin dual cameras on an iPhone. However, it can recognize only three kinds of foods. Moreover, it is difficult to keep recording every mealtime. So we focus on HoloLens which are one of the most common AR / MR glasses. If such a wearable device can be used to automatically manage dietary calories from images, the user's hurdles will become very low. In addition, these devices measure the environment, which is useful not only for the user but also for accurate calorie estimation. The recent rapid development of deep learning technology has made it possible to recognize meal areas more accurately. In this demo, we propose a system which estimates meal calorie intakes based on the actual size of foods using AR / MR glasses and deep learning. This is one of the applications of integration of the AR/MR technology and deep learning.

2 PROPOSED SYSTEM

The system consists of the following four parts:

(1) Meal area detection, classification, and segmentation
(2) Actual size estimation
(3) Calorie estimation
(4) Showing meal information on the AR space

Figure 1 shows the processing flow of the proposed system.

*e-mail: naritomi-s@mm.inf.uec.ac.jp
†e-mail: yanai@cs.uec.ac.jp

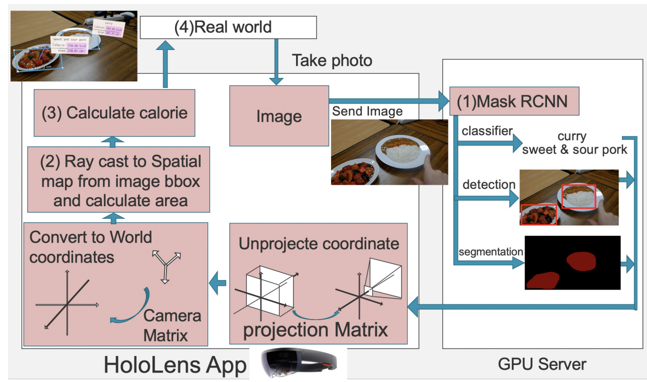

Figure 1: Overview of the proposed system.

2.1 Meal detection, classification and segmentation

In the system, we use Mask R-CNN [3] to detect, classify and segment foods in the view of HoloLens simultaneously. For training the Mask R-CNN, we use the UEC-FoodPix dataset [2] which contains about 10,000 food images annotated with 100-class food regions and food instance bounding boxes. To estimate food calories, we need to know a meal-dependent relation between a meal size and meal calories. We have collected photos of 46 kinds of meals in which the calorie values and the actual 2D size of the meal are known. Regarding each of 46 kinds of the meals, meal photos of four different sizes were taken. We estimated quadratic curves which represent the relation between the actual 2D size and the calories for each of the 46 meals. Therefore, we used only the food photos of the 46 kinds of foods contained in the UECFoodPix dataset, in which 4297 and 485 images were used for training and testing, respectively.

The accuracy of food detection and region segmentation are shown in Table 1. AP_{50} is the Average Precision when assuming that the output area of the model matches 50% or more of the ground truth area. The results of food detection and segmentation are shown in Figure 2.

Table 1: The accuracy of detection and segmentation by Mask R-CNN.

	AP_{50}
Detection	42.8
Segmentation	39.1

2.2 Actual size measurement by the correspondence between image coordinates and world coordinates

By using the camera projection matrix, the 2D image coordinate system and the 3D camera coordinate system can be converted bidirectionally. Since the depth of a point can be calculated by associating it with an environment map provided from HoloLens API, the system can estimate where the point on the image is in the world coordinates. The calculation method is as follows: For each of the four vertices of the bounding box detected on the image, a straight line is extended from the camera position to the image plane in front of the camera as shown in Figure 3. We find the

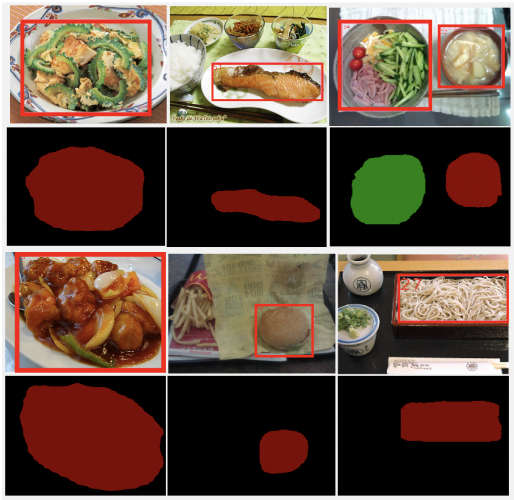

Figure 2: Detection and segmentation results by Mask R-CNN.

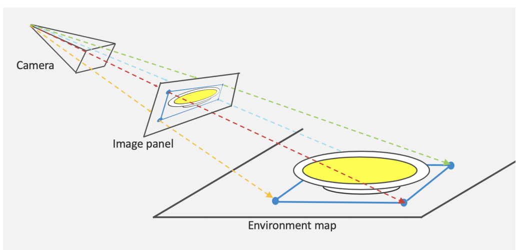

Figure 3: Correspondence between images and 3D world.

intersection of the line and the environment map. By dividing the area surrounded by these four intersections into triangles, the system can easily calculate the actual size of the target area. Then, we calculate the actual area where the food segmentation mask occupies in the bounding box.

2.3 Calorie estimation

We have prepared a regression curve for each of the 47 kinds of meals to calculate calories from the real area size in advance. The regression curve was created using an image data set annotated with calorie values. All images used for estimating of the regression curve were taken with a reference object as shown on the left in Figure 4. Since the actual area size of the reference object is known, by comparing the number of pixels of the meal area with the reference object, the area of the meal can be calculated easily. The number of pixels of the meal area was counted within the food regions obtained by the MASK R-CNN as shown on the right in Figure 4.

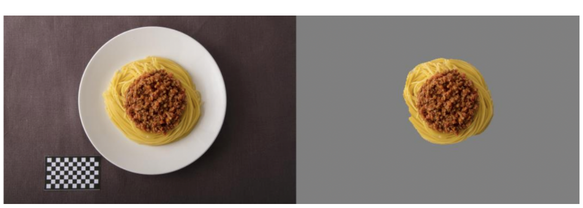

Figure 4: A meal image taken with a reference object.

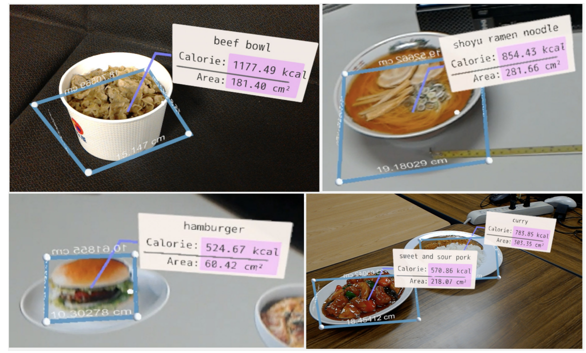

Figure 5: Meal information displayed in the AR space.

2.4 Showing estimated food information on the AR space

The client-side software was implemented on the Microsoft HoloLens, while the function of meal recognition was implemented as a REST API on the HTTP server. When HoloLens posts an image via HTTP, a JSON containing the class label, bounding box, and meal area ratio is returned as a response. Based on this information, the client-side system measures the actual size as shown in Section 2.2, estimates the calorie from the area and displays it over the meal in the AR space as shown in Figure 5. By touching the displayed panel, users can mark the meal they ate and know the total calories of the taken meals.

3 Demonstration

In the demonstration, participants can wear HoloLens. Several meals will be prepared in front of participants wearing HoloLens. By recognizing those meals, the participants will experience calorie estimation based on actual size using AR / MR glasses. In addition, by interacting with meals in the application, they will experience the feeling of a new style of eating.

4 Conclusion

In this study, we proposed a calorie estimation system that takes into account the actual size of meals using AR / MR glasses and image recognition technology based on deep learning. Mask R-CNN trained with the UECFoodPix dataset [2] was used for meal recognition, and the environment recognition function of AR / MR glasses was used for measuring of the actual meal size. As the AR / MR glass market for general consumers expands in the future, we have shown that such meal management support applications are feasible.

More information is available at the project page: `https://neno.dev/project/caloriecaptorglass/`.

Acknowledgements: This work was supported by JSPS KAKENHI Grant Number 15H05915, 17H01745, 17H06100 and 19H04929.

References

[1] Y. Ando, T. Ege, J. Cho, and K. Yanai. Depthcaloriecam: A mobile application for volume-based foodcalorie estimation using depth cameras. In *Proceedings of the 5th International Workshop on Multimedia Assisted Dietary Management*, pp. 76–81. ACM, 2019.

[2] T. Ege, W. Shimoda, and K. Yanai. A new large-scale food image segmentation dataset and its application to food calorie estimation based on grains of rice. In *Proceedings of the 5th International Workshop on Multimedia Assisted Dietary Management*, pp. 82–87. ACM, 2019.

[3] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask R-CNN. In *Proc.of IEEE International Conference on Computer Vision*, pp. 2961–2969, 2017.

[4] R. Tanno, T. Ege, and K. Yanai. Ar deepcaloriecam v2: Food calorie estimation with cnn and ar-based actual size estimation. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, p. 46. ACM, 2018.