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”, a calorie estimation system based on actual size using image recognition and AR / MR glasses. In addition to the user experience with AR / MR glasses, we have built a system that takes into account the actual size of meals, which cannot be understood from images alone, using the environment recognition function. Therefore, unlike existing methods, the calorie amount can be estimated in consideration of the actual size without any restrictions on the reference object or the camera position, so the user can easily manage meals without stress by using this a system.

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 the calories taken from meals with such devices. Therefore, there are AR DeepCalorieCam V2 [4] and DepthCalorieCam [1] as smartphone apps for managing meals. In AR DeepCalorieCam V2, you had to specify the meal area yourself. DepthCalorieCam had limitations such as having to take a meal from directly above. However, it is difficult and long-lasting to keep records by doing the above work every time you eat. So we focused on AR / MR glasses. If such a wearable device can be used to automatically manage dietary calories from images, the user's hurdle will be 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 system which estimate meal calorie based on actual size using AR / MR glasses and deep learning.

2 PROPOSED SYSTEM

The system consists of the following four parts.

1. Meal area detaction, classification, and segmantation
2. Actual size measurement
3. Calorie estimation
4. User selection and accounting of meals

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

Figure 1 shows the processing flow of the application in this system.

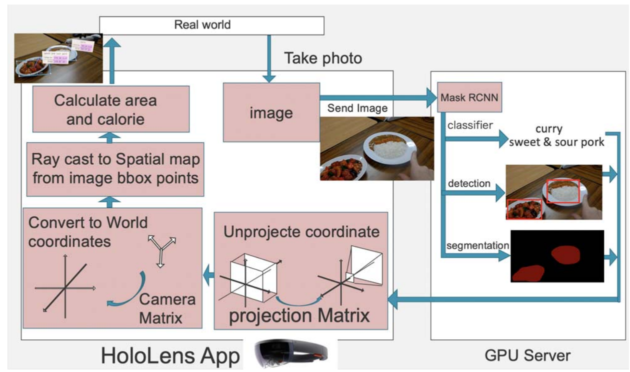

Figure 1: application over view

2.1 Meal detection, classification, segmentation

Use Deep Convolutional Neural Network to detect, classify, and segment food. In this system, we used Mask R-CNN [3] as a model that handles these three tasks. UECFoodPix [2] was used for learning Mask R-CNN. An instance label is required to learn Mask R-CNN, but UECFoodPix does not have an instance label. Therefore, we cut out a mask with the same class label from the inside of the bounding box and created an instance mask. UECFoodPix has 100 classes, but 100 classes were not used, but only 46 classes were used. The 46 classes were selected from the classes that exist in the dataset that have calorie content and area correspondence. 4297/485 images were used for training and testing. The accuracy is as shown in the 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 detection and segmantation results are shown in Fig.2.

Table 1: Mask-RCNN accuracy

	AP_{50}
Detection	42.8
Segmantation	39.1

2.2 Actual size measurement by 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. The depth of a point can be calculated by associating it with an environment map, so the application knows where the point on the image is in 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

978-1-7281-6532-5/20/$31.00 ©2020 IEEE
DOI 10.1109/VRW50115.2020.00260

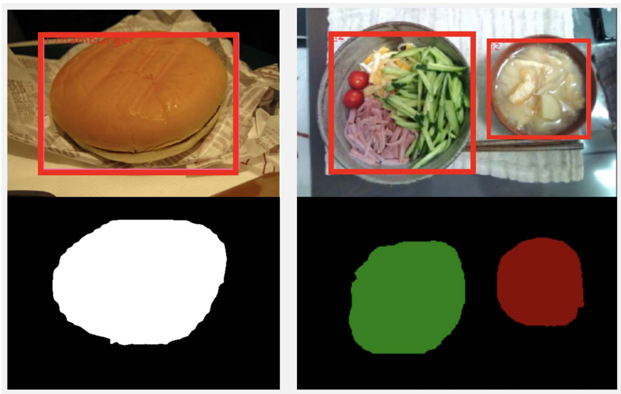

Figure 2: Detection and segmantation result by Mask R-CNN

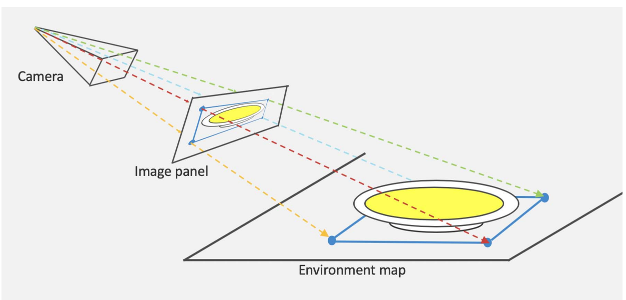

Figure 3: Correspondence between images and 3D world.

the image plane in front of the camera as shown in Figure 4. Find the intersection of the line and the environment map. By dividing the area surrounded by these four intersections into triangles, system can easily calculate the area of the object. As a result, the area of the object can be calculated. Then, calculate the area where the segmentation mask occupies the bounding box and the area of the meal area only.

2.3 Calorie estimation

We prepared a regression equation to calculate calories from the area. The regression equation was created using an image data set with calorie content. All images of the dataset used were taken with the reference object as shown on the left in Figure 4. Since the area 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 known, and a regression equation from the area to the calorie amount can be obtained. The number of pixels of the meal area taken with the reference object was calculated by segmenging the area as shown on the right in Figure 4.

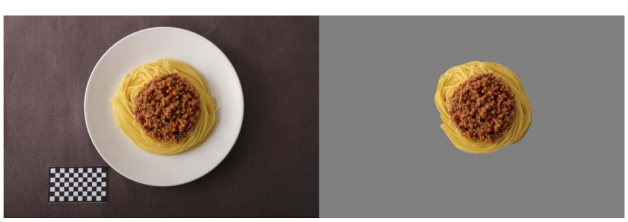

Figure 4: A meal image taken with a reference object.

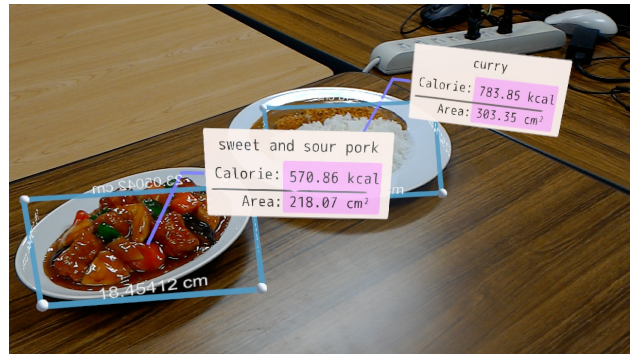

Figure 5: application view

2.4 Implementation

The client side was implemented as HoloLens, and 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, measure the actual size as shown in section 2.2, estimate the calorie from the area, and display it on the meal as shown in the figure 5. By touching the displayed panel, the user can mark the meal they ate and know the total calories they have eaten.

3 Demonstration

In this demonstration, participants will wear HoloLens. Several meals will be prepared in front of participants wearing HoloLens. By recognizing those meals, you will experience calorie estimation based on actual size using AR / MR glasses. In addition, by interacting with meals in the application, you will experience the feeling of use when such an app actually spread in the future.

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, which used UECFoodPix [2] for learning, was used for meal recognition, and the environment recognition function of AR / MR glasses was used for actual measurement. As the AR / MR glass market for general consumers expands in the future, we have shown that such meal management support applications are feasible.

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.