

文章编号:1006-3080(2018)02-0270-07

DOI:10.14135/j.cnki.1006-3080.20170313006

基于深度学习的食物卡路里估算方法

梁炎超, 李建华

(华东理工大学信息科学与工程学院, 上海 200237)

摘要:现代智能移动设备可利用计算机视觉技术从食物照片中估算卡路里。针对现有食物卡路里估算方法识别准确率不高、估算结果误差较大的问题,本文提出了一种基于深度学习的食物卡路里估算方法。该方法需以食物俯视图和侧视图作为输入,先进行目标检测,再用 GrabCut 算法获得目标轮廓。通过对不同类别的食物采用不同的体积估算方法,以提高体积估算准确性。实验结果表明,本文方法估算结果较为准确,能为用户控制卡路里摄入提供正确的参考。

关键词:深度学习; 体积估算; 卡路里

中图分类号:TP391

文献标志码:A

Food Calorie Estimation Method Based on Deep Learning

LIANG Yan-chao, LI Jian-hua

(School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China)

Abstract: Obesity is associated with the increased risk of diseases. For obesity treatment, it is necessary to record all food intakes per day. However, in most cases, patients do have troubles in estimating the amount of food intake because they are unwillingness to record or lack of related nutritional information. The calorie in food can be estimated via computer vision methods, whose estimated accuracy is determined by two main factors: object detection algorithm and volume estimation method. In order to increase the accuracy of detection and reduce the error of volume estimation in food calorie estimation, this paper proposes a calorie estimation method based on deep learning. This proposed method takes two food images as its inputs: a top view and a side view. Each image includes a calibration object that is used to estimate image's scale factor. Food(s) and calibration object are detected by object detection method called faster region-based convolutional neural networks (Faster R-CNN) and each food's contour is obtained by applying GrabCut algorithm. The calibration object judged by Faster R-CNN is used to calculate the scale factor of each view. Each food's volume can be estimated according to its contour in top view, contour in side view, and scale factors. For improving the volume estimation accuracy, this paper divides different types of food shape into four types, for which the corresponding volume estimation formula is adopted. And then, each food's mass and calorie are estimated by means of density table and nutrition table. In the proposed volume estimation experiments, the error between a estimation result and its corresponding true value does not exceed $\pm 20\%$ for most food. The experimental results show that the estimation results are

收稿日期:2017-03-17

基金项目:国家自然科学基金(61402174)

作者简介:梁炎超(1992-),男,硕士生,研究领域为计算机视觉。

通信联系人:李建华, E-mail: jhli@ecust.edu.cn

accurate. Hence, the proposed method in this paper is helpful for those patients who want to control calorie intake. In future research, we will keep on improving our method and develop mobile application.

Key words: deep learning; volume estimation; calorie

摄入过多的卡路里会导致肥胖,进而会诱发各种疾病,如心脏疾病、II型糖尿病、阻塞性睡眠呼吸暂停以及骨关节炎等^[1]。目前,全球的肥胖形势日益严峻,到2025年,将会有18%的肥胖男性和超过21%的肥胖女性^[2]。这就要求肥胖人群减少卡路里的摄入,控制自身体重,对每个食物进行称重并查找该食物的营养成分信息,然后计算并记录该食物的卡路里。该过程非常繁琐,但随着现代智能设备以及可穿戴设备的迅速发展,研究人员已经开始研究采用计算机视觉从食物照片估算卡路里,从而简化该过程。食物卡路里估算既可以为肥胖症病人提供便利,也可以为健康人群控制卡路里摄入提供参考。

从食物照片估算卡路里一般有以下步骤:获取照片、食物与标定物的检测、以及体积与卡路里估算。照片的获取设备一般为相机、手机等智能移动设备,亦可通过可穿戴设备^[3]获取。拍摄的照片中必须包含标定物,如校准卡^[4]、盘子^[5]、餐桌背景纸^[6]、手指^[7]等。标定物用于计算图像中一个像素对应的真实大小。通过算法检测出图像中的食物和标定物后,可以估算食物的体积,再根据食物的类别查找对应的食物密度表和营养成分表来估算卡路里。在此过程中,卡路里估算的准确性主要由两个因素决定:目标检测算法和体积估算方法。

对于食物的检测,一般先采用图像分割算法,再利用分类算法完成,但现有方法对食物的检测结果精度不高。目前,深度学习已经在图像识别、语音识别等各个领域广泛应用,近期深度学习算法已用于诊断皮肤癌^[8]。因而,利用高效的深度学习算法代替现有的目标检测算法可以有效地提高检测准确性。

对于体积估算,一般利用标定物和估算公式完成。标定物用于计算像素真实大小,可根据自身特性进行检测。目前,食物的体积估算公式分为两种:(1)将所有食物利用单一体积公式(如长方体计算公式)进行计算^[9],这种方式实际上无法准确估算其他形状物体;(2)由用户手动选定模型公式进行计算^[10],这种方法需要用户手动选择,使用极其不便。

针对食物卡路里估算方法在目标检测和体积估

算两个方面存在的问题,本文提出了一种基于深度学习的食物卡路里估算方法。该方法以食物的俯视图和侧视图作为输入,用深度学习算法 Faster R-CNN^[11]进行目标检测,以提高食物识别准确率,并通过对不同食物采用不同的体积估算方法以提高体积估算准确性。

1 食物卡路里估算系统框架

卡路里估算系统框架如图1所示,其流程包含以下步骤:照片获取、检测、分割、体积估算、卡路里估算。首先对获取的照片进行目标检测,从而检测出食物以及参照物,再利用图像分割算法获取准确的食物轮廓;然后根据参照物以及食物轮廓计算食物体积;最后通过查找食物相关信息估算食物的卡路里。

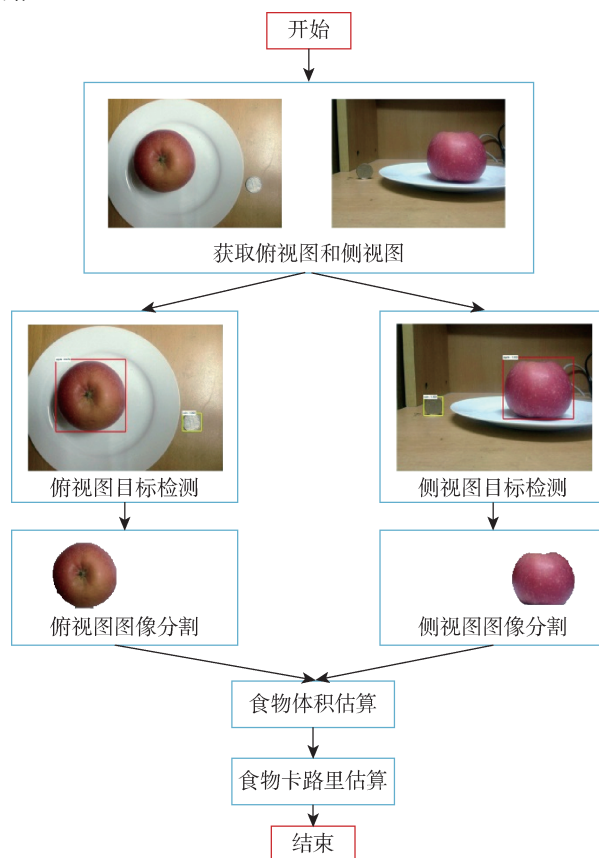


图1 卡路里估算系统框架图

Fig. 1 Architecture of calorie estimation system

为了估算食物的卡路里,本文方法需要用户提供用照相设备拍摄的待估算卡路里食物的俯视图和侧视图各一张,而且这两张图中需要有标定物。尽管采用单张食物照片^[10]亦能估算卡路里,但是对于某些拍摄角度,很难获得食物的形状、厚度(高度)等关键信息,因而体积估算误差会较大,如拍摄俯视图时会丢失食物的厚度信息。

在获得俯视图和侧视图后,需要检测图中的食物和标定物。本文采用 Faster R-CNN 目标检测算法,它能标记目标的位置和类别。

经过 Faster R-CNN 处理后,食物和标定物的位置已经获得,虽然也能直接进行体积估算,但考虑到获取更精确的食物边缘有助于体积估算结果的准确性,用图像分割方法获得食物的准确轮廓是有必要的,本文采用 GrabCut 图像分割方法^[12]。GrabCut 是基于图论的图像分割算法,它需要用户画矩形框标记前景区域和背景区域,但由于 Faster R-CNN 已经获得了食物、标定物的所在区域,因此并不需要用户参与就可以直接进行图像分割。在完成分割后,可以得到准确轮廓。

根据食物的俯视图和侧视图,通过查询食物形状表(该表用于定义食物的形状)选择对应的计算公式进而估算得到食物体积 V ,通过查询对应食物的密度表获得密度 ρ ,根据式(1)获得食物质量 m 。

$$m = \rho \times V \quad (1)$$

通过查询对应食物的营养成分表获得该食物单位质量的卡路里 c ,由式(2)可估算出该食物的卡路里 C 。

$$C = c \times m \quad (2)$$

2 基于深度学习的目标检测算法

由于照片中食物和标定物的数目、位置不定,若采用分类算法^[13],不仅无法应对照片中有多个食物的状况,也不能确定食物在照片中的位置,因此需要采用目标检测算法。

传统的目标检测算法包含以下 3 个步骤:区域提取、特征提取、分类。输入图像中,候选窗的生成采用滑动窗口的方式,将一张图像分解成若干个候选区域(即可能为前景的区域),对每个子窗口提取特征,使用分类器进行判别。传统的目标检测算法速度慢、检测精度低。近年来,基于深度学习的目标检测算法不断被提出并取得了很好的检测效果。

Faster R-CNN 是一种基于深度学习的目标检测算法。如图 2 所示,它包含一个区域生成网络(Regional Proposal Network, RPN)和一个 Fast R-CNN^[14]。其中 RPN 用于生成候选框(即标记出图像中可能是目标的区域),它的输入是 RGB 图像,输出为一系列候选框以及每个候选框属于前景和背景的概率。利用非极大化抑制(Non-Maximum Suppression, NMS)^[15]方法,只提取 300 个候选框作为检测的对象。而 Fast R-CNN 用于检测。

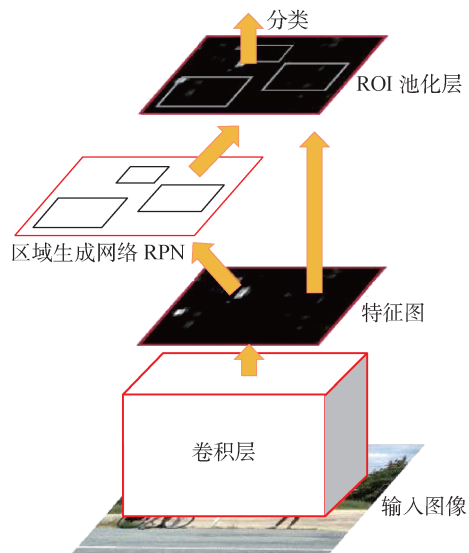


图 2 Faster R-CNN

Fig. 2 Faster R-CNN

Faster R-CNN 的训练过程如下:

(1) 用 ImageNet 模型初始化,独立训练一个 RPN 网络。

(2) 用 ImageNet 模型初始化 Fast R-CNN,并用 RPN 网络的输出作为 Fast R-CNN 的输入,训练一个 Fast-R-CNN 网络。此时,两个网络每一层的参数完全不共享。

(3) 对于 RPN 与 Fast R-CNN 共享的网络层,使用步骤(2)生成的 Fast-R-CNN 网络参数替换;对于两个网络不共有的网络层,则采用默认的方法初始化,然后再次训练 RPN,但只更新 RPN 特有的那些网络层。

(4) 固定 Fast R-CNN 与 RPN 共享的网络层,利用步骤(3)训练得到的 RPN 的输出继续训练 Fast R-CNN,训练过程中也只更新 Fast R-CNN 特有的网络层参数。

Faster R-CNN 对于图片的大小没有限制,相较于目前大多数目标检测算法,其平均精度也较高。

对于一幅 RGB 图像, Faster R-CNN 的检测步骤如下:

算法 1 基于 Faster R-CNN 的食物及标定物检测

输入 包含食物和标定物的 RGB 图像

(1) 输入经过 RPN 后, 获得若干候选框; 对于每个候选框, RPN 亦输出该候选框属于前景(即可能的目标)的概率。

(2) 将 RPN 输出的候选框按照输出的前景概率进行排列, 仅保留大于阈值(0.8)的候选框。

(3) 用非极大化抑制方法对步骤(2)获得的候选框进行处理, 以减少重叠面积过大的那些候选框。

(4) 将经过步骤(3)保留下来的候选框作为 Fast R-CNN 的输入, 判定每个候选框内图像类别, 并对那些判定为目标(即非背景)的候选框进行位置调整。

输出 Fast R-CNN 判定为目标的候选框序列以及每个候选框对应的类别

俯视图和侧视图分别经过 Faster R-CNN 处理后, 可以得到照片中食物、标定物的类别和位置, 以便于进行后续的操作。

3 体积估算方法

3.1 体积估算流程

经过检测和分割后, 食物和标定物的精确轮廓已经获得, 利用标定物计算比例系数, 然后可以进行体积估算。通过对不同食物按照形状进行体积估算可以减少误差, 提高卡路里估算准确性, 该算法的流程如下:

算法 2 基于形状的食物体积估算

输入 食物俯视图、食物侧视图、标定物俯视图、标定物侧视图

(1) 依据标定物侧视图计算侧视图比例系数 P_B 。

(2) 依据标定物俯视图和 P_B 计算俯视图比例系数 P_A 。

(3) 根据食物类别判定食物形状。

(4) 根据食物形状选取体积计算公式, 结合 P_A 、 P_B 、食物俯视图和侧视图计算食物体积。

输出 食物的体积

3.2 比例系数的计算

为了估算体积, 必须获得照片中一个像素对应

的真实大小。标定物的检测可利用目标检测算法检测, 亦可根据标定物的特性进行检测, 例如对于圆形的标定物可以利用霍夫变换检测图像中的椭圆获取。

图 3 所示为侧视图示意图。其中, O' 为镜头所处位置。 MN 为食物长度, $M'N'$ 为食物长度在照片中对应的长度; MP 为食物高度, $M'P'$ 为食物高度在照片中对应的长度; $C'D'$ 为标定物中已知的某段长度, $A'B'$ 为 $C'D'$ 在照片中对应的长度。记侧视图中一个像素对应的实际大小为 P_B , 通过构造相似三角形, 采用式(4) 计算 P_B 。

$$P_B = \frac{MN}{M'N'} = \frac{MP}{M'P'} = \frac{C'D'}{A'B'} \quad (3)$$

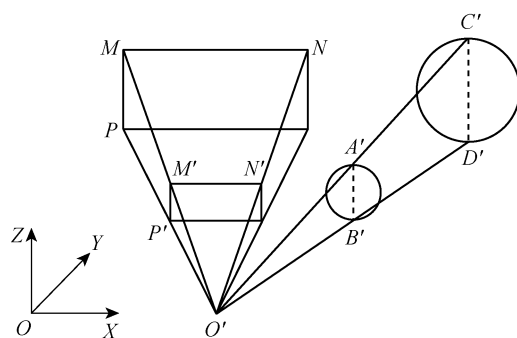


图 3 侧视图拍摄示意图

Fig. 3 Schematic diagram of shooting side view

图 4 所示为俯视图示意图。其中 O 为镜头所处的位置; CD 为标定物中某段已知长度; AB 为该标定物这段长度在照片中对应的长度(即像素的个

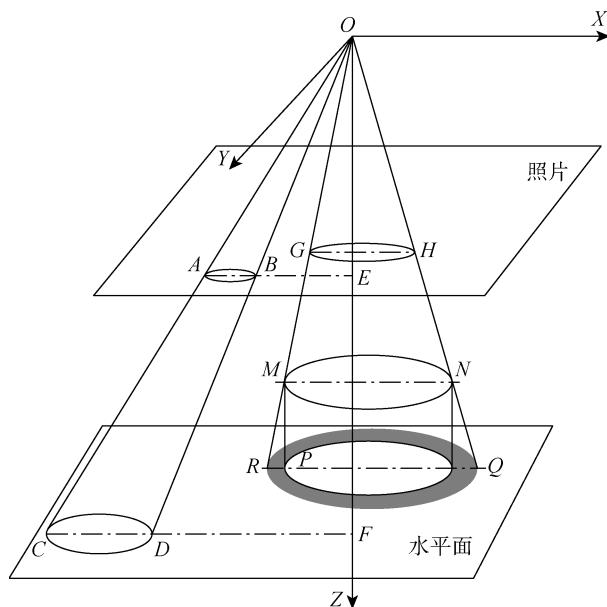


图 4 俯视图拍摄示意图

Fig. 4 Schematic diagram of shooting top view

数); MN 为待测食物的(最大)宽度; MP 为待测食物的高度; GH 为食物的宽度在照片中的长度。记俯视图中一个像素对应的实际大小为 P_A , 通过构造相似三角形, 采用式(5) 计算 P_A , 式(4) 计算得到的 P_B 用于校正 P_A 。

$$P_A = \min\left(\frac{MN}{GH}, \frac{RQ}{GH}\right) = \min\left(\frac{P_B \times M'N'}{GH}, \frac{CD}{AB}\right) \quad (4)$$

3.3 体积估算方法

将食物形状分为4种类型: 椭球体、柱体、圆环体、其他形状。根据食物的形状, 利用该食物的俯视图和侧视图构建三维模型, 从而进行体积估算。

假设经过检测和分割后, 得到了某个食物的俯视图 A(如图 5(b) 所示, 图中深色的方块代表食物所占用的像素, 白色的方块代表背景像素) 和侧视图 B(如图 5(c) 所示)。设在俯视图 A 中, 该食物占 m 行, 其中第 i ($i = 1, 2, \dots, m$) 行该食物占 L_A^i 个像素。记 $L_{A_{\max}}^i = \max(L_A^1, L_A^2, \dots, L_A^m)$, 食物所占的像素个数 $S_A = \sum_{i=1}^m L_A^i$ 。

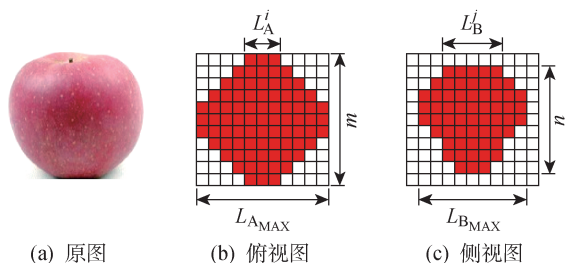


图5 体积估算示意图

Fig. 5 Schematic diagram of volume estimation

设在侧视图 B 中, 该食物占 n 行, 第 j ($j = 1, 2, \dots, n$) 行该食物占 L_B^j 个像素。记最大长度 $L_{B_{\max}}^j = \max(L_B^1, L_B^2, \dots, L_B^n)$, 食物所占的像素个数 $S_B = \sum_{j=1}^n L_B^j$ 。

对于图像分割后的子图像, 根据算法判定的食物类别, 分别采用以下公式进行体积估算:

食物形状为椭球体

$$V = \frac{\pi}{4} \times \sum_{j=1}^n (L_B^j)^2 \times P_B^3 \quad (5)$$

食物形状为柱体

$$V = (S_A \times P_A^2) \times (n \times P_B) \quad (6)$$

食物形状为圆环体(亦称胎体)时, 记圆环体食物俯视图中的中空部分(参见图 6(b) 中黄色的区域, 该图中红色部分表示食物占用的像素, 而蓝色部分和黄色部分均为背景)所占用像素数量为 s_{AE} , 用

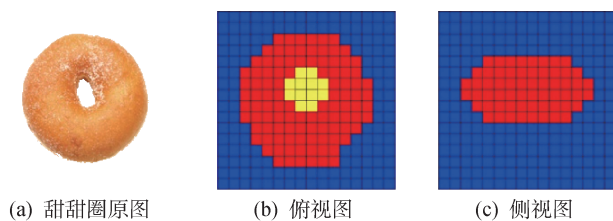


图6 甜甜圈体积估算示意图

Fig. 6 Schematic diagrams of volume estimation of doughnut

式(7) 估算体积。

$$V = \frac{\pi^2}{4} \left(\sqrt{\frac{s_A + s_{AE}}{\pi} P_A^2} + \sqrt{\frac{s_{AE}}{\pi} P_A^2} \right) (n \times P_B)^2 \quad (7)$$

其他形状采用式(8) 进行体积估算。

$$V = s_A \times P_A^2 \times \sum_{j=1}^n \left(\frac{L_B^j}{L_{B_{\max}}} \right)^2 \times P_B \quad (8)$$

4 实 验

用手机拍下部分食物的俯视图和侧视图, 并用一元人民币作为标定物, 同时记录了每个食物的体积, 食物的体积通过排水法测得。

实验所用机器型号为中科曙光 A650。实验平台为 Windows 7 (64 位), 显卡型号为英伟达 GTX1070。Faster R-CNN 代码基于 Matlab 以及 C++, GrabCut 图像分割以及体积估算代码基于 C++。Matlab 版本号为 R2014b, Visual Studio 版本号为 2013, opencv 版本号为 3.0, CUDA 版本号为 8.0。

以俯视图和侧视图作为输入, 输出卡路里估算值, 并计算与真实值的误差。限于实验条件, 无法获取卡路里真实值, 因此仅验证到体积估算这一步, 即比较估算的体积与参考体积之间的误差。

对 19 种常见食物进行了估算实验。目标检测对象包含: 苹果、香蕉、面包片、馒头、甜甜圈、鸡蛋、麻花、葡萄、柠檬、荔枝、芒果、月饼、橙子、油桃、梨、黑布林、猕猴桃、沙琪玛、番茄, 样本总数合计为 2 978 张, 其中训练样本数目为 1 245 张, 测试样本数目为 1 733 张。

在用训练集对 Faster R-CNN 进行训练后, 对测试集的样本进行了体积估算实验。在体积估算实验中, 只采用 Faster R-CNN 在测试集中正确检测的图片对(一对图片包含一张俯视图和一张侧视图), 所用到的各类图片样本数目参见图 7。剔除 Faster R-CNN 误识别或者未识别的样本后, 体积估

算实验所用的各类样本数参见图7。体积估算的评价标准为平均误差 ME 。

$$ME_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{v_j - V_j}{V_j} \quad (9)$$

其中, i 表示食物类别,用于体积估算实验的图片总数为 $2n_i$ 。由于本文方法需要俯视图和侧视图才能估算体积,所以体积估算的结果总数为 n_i 。对于第 j 次体积估算, v_j 表示本文方法得到的体积, V_j 表示真实体积。

体积估算的结果如图8所示。从图8中可以看

出,除了香蕉、葡萄、月饼这3类食物外,本文方法的估算结果与真实值之间的误差不超过 $\pm 20\%$,而且部分食物(如猕猴桃)的估算结果与真实值非常接近,这表明本文提出的体积估算是有效的。对于那些误差较大的食物,则可以通过对原来的体积计算公式附加补偿系数等方式降低体积估算误差。

除去初始化所需时间,对于一组食物图片(仅包含一个食物和一个标定物,每张图片大小为 816×612),用本文方法估算卡路里平均耗时 0.52 s ,其中对两张图片的目标检测所花费的时间为 0.17 s 。

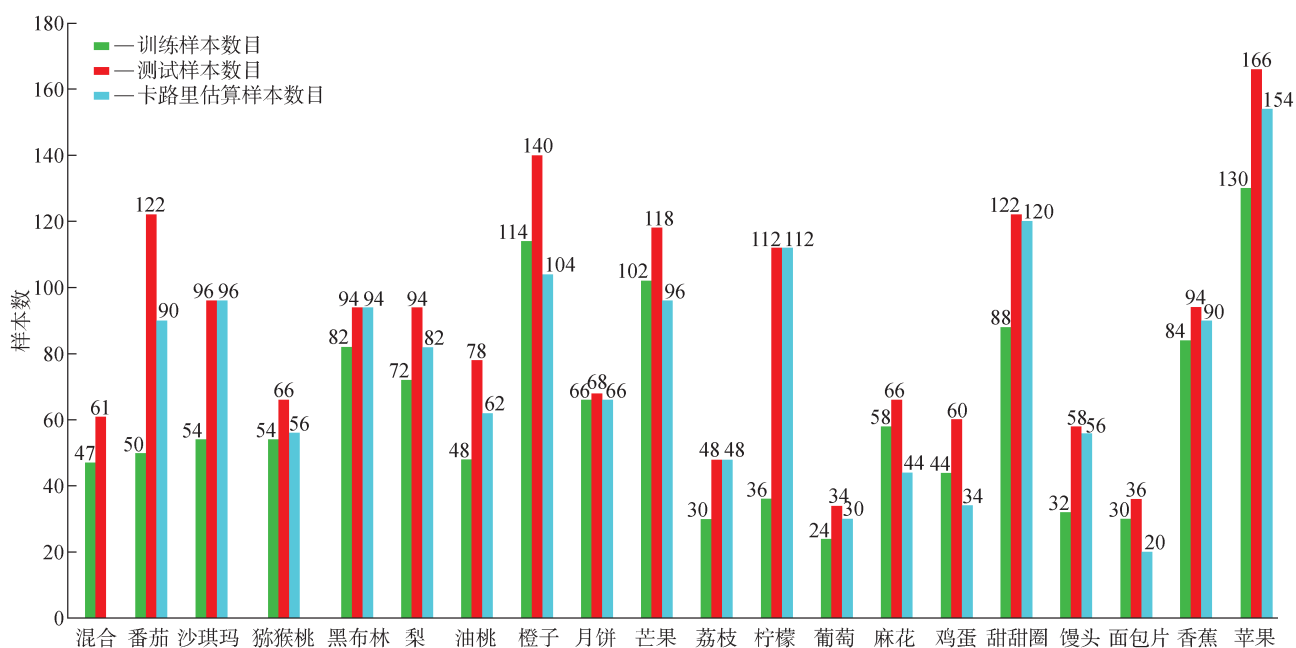


图7 各类别训练样本数同、测试样本数目以及卡路里估算所用样本数目

Fig. 7 Training image numbers, testing image numbers and image numbers involved in food calorie estimation experiment for each type

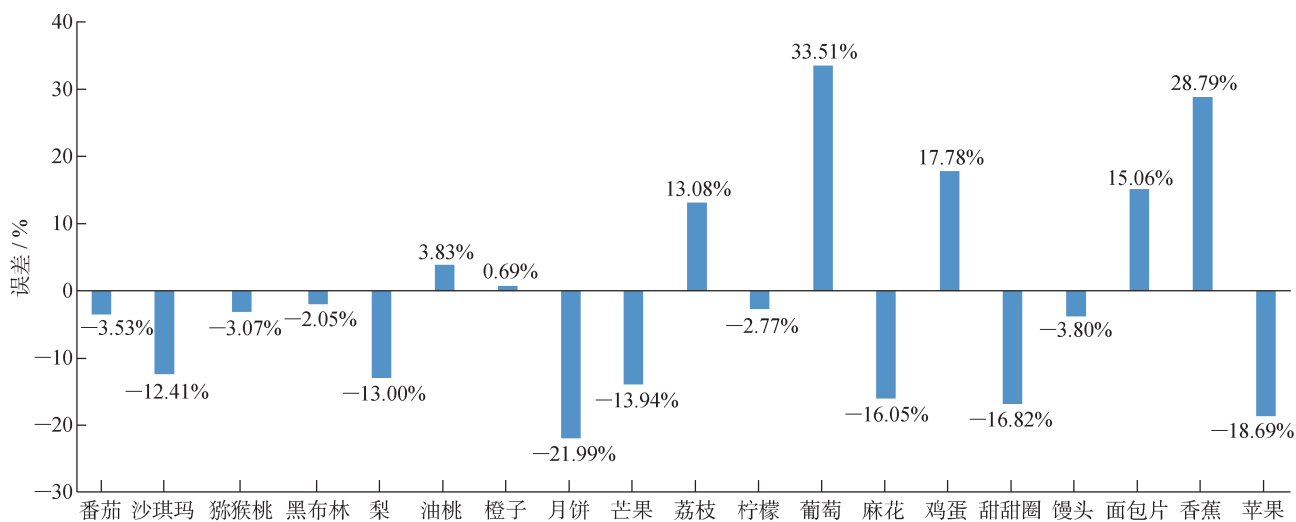


图8 体积估算结果

Fig. 8 Volume estimation results

5 结束语

本文提出了一种基于深度学习的食物卡路里估算方法。该方法需要包含食物和标定物的俯视图、侧视图各一张。深度学习算法 Faster R-CNN 的应用可以提高检测的准确性,而通过对不同食物采用不同的体积估算方法可以使体积估算结果更接近真实值。实验结果表明本文的体积估算方法能为患者控制卡路里摄入提供正确的参考。

接下来的工作主要集中在以下几个方面:

(1) 尝试对其他种类的食物进行实验,并改进本文的体积估算方法;

(2) 进行移动应用开发,为用户提供卡路里估算服务。

参考文献:

- [1] ZHENG W, MCLERRAN D F, ROLLAND B, *et al.* Association between body-mass index and risk of death in more than 1 million Asians[J]. *New England Journal of Medicine*, 2011, 364(8): 719-29.
- [2] DI C M, BENTHAM J, STEVENS G A, *et al.* Trends in adult body-mass index in 200 countries from 1975 to 2014: A pooled analysis of 1698 population-based measurement studies with 19.2 million participants [J]. *Lancet*, 2016, 387(10026): 1377-1396.
- [3] ANTHIMOPOULOS M M, GIANOLA L, SCARNATO L, *et al.* A food recognition system for diabetic patients based on an optimized bag-of-features model[J]. *IEEE Journal of Biomedical and Health Informatics*, 2014, 18(4): 1261-1271.
- [4] PURI M, ZHU Z, YU Q, *et al.* Recognition and volume estimation of food intake using a mobile device[C]//2009 Workshop on Applications of Computer Vision (WACV). USA: IEEE, 2009: 1-8.
- [5] SUN M, LIU Q, SCHMIDT K, *et al.* Determination of food portion size by image processing[C]//Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE. USA: IEEE, 2008: 871-874.
- [6] 宋成琳, 王文媛, 刘箫, 等. 即时性图像法膳食调查技术应用于学龄前儿童膳食调查的效果评价[J]. *中国儿童保健杂志*, 2015, 23(7): 684-688.
- [7] VILLALOBOS G, ALMAGHRABI R, POULADZADEH P, *et al.* An image processing approach for calorie intake measurement[C]//IEEE International Symposium on Medical Measurements and Applications Proceedings. USA: IEEE, 2012: 1-5.
- [8] ESTEVA A, KUPREL B, NOVOA R A, *et al.* Dermatologist-level classification of skin cancer with deep neural networks[J]. *Nature*, 2017, 542(7639): 115-118.
- [9] POULADZADEH P, SHIRMOHAMMADI S, AL-MAGHRABI R. Measuring calorie and nutrition from food image [J]. *IEEE Transactions on Instrumentation & Measurement*, 2014, 63(8): 1947-1956.
- [10] JIA W, CHEN H C, YUE Y, *et al.* Accuracy of food portion size estimation from digital pictures acquired by a chest-worn camera[J]. *Public Health Nutrition*, 2014, 17(8): 1671-1681.
- [11] REN S, HE K, GIRSHICK R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks[C]//Advances in Neural Information Processing Systems. Canada: ACM, 2015: 91-99.
- [12] ROTHER C, KOLMOGOROV V, BLAKE A. GrabCut: Interactive foreground extraction using iterated graph cuts[J]. *ACM Transactions on Graphics*, 2010, 8(1): 170-173.
- [13] KAGAYA H, AIZAWA K. Highly accurate food/non-food image classification based on a deep convolutional neural network[C]//International Conference on Image Analysis and Processing. Switzerland: Springer, 2015: 350-357.
- [14] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. USA: IEEE, 2015: 1440-1448.
- [15] NEUBECK A, VAN GOOL L. Efficient non-maximum suppression[C]//International Conference on Pattern Recognition. USA: IEEE, 2006: 850-855.