CS 725 - Foundations of Machine Learning

Calorie Estimation of Food Items from Images

CV Maverics

Indian Institute of Technology Bombay

November 28, 2023

Outline

Problem Statement

Object Detection

Estimating calorie content from food images poses a significant challenge. In addressing this, the project strategically employs the You Only Look Once (YOLO) algorithm. This choice is pivotal in enhancing the accuracy of food item detection and classification.

Calorie Estimation

Furthermore, the project incorporates the GrabCut algorithm to attain meticulous segmentation of food items in images. This precision is crucial for both volume estimation and, subsequently, accurate calorie estimation.

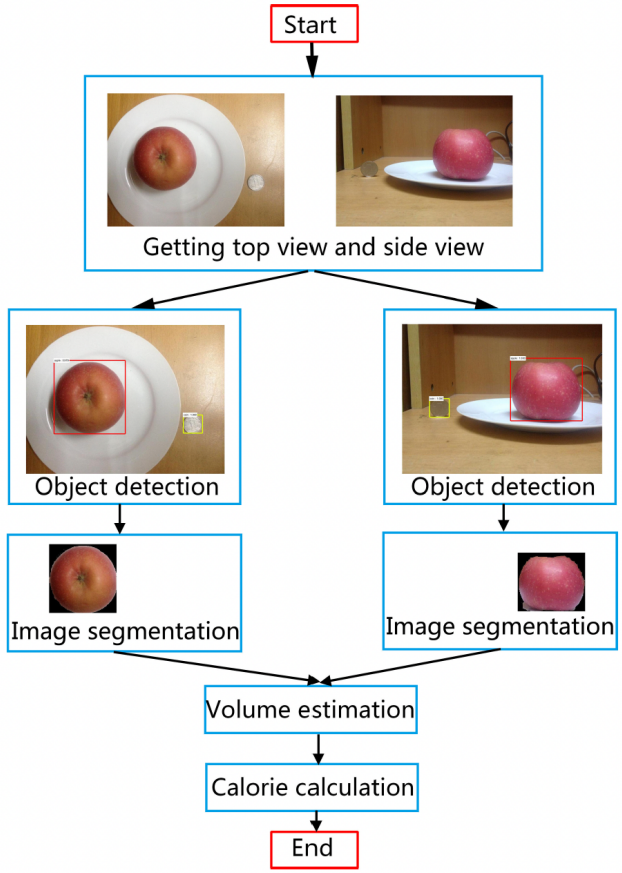

Figure: Calorie estimation of food items from images[1]

Challenges Addressed

Dataset Acquisition: Obtain and familiarize with the ECUST dataset. Understand the dataset's nuances, including its characteristics and potential challenges.

Data Preprocessing: Prepare the dataset for model training. Resize images, normalize pixel values, and organize annotations. Ensure the dataset is well-structured and ready for training.

Challenges Addressed

Dataset Acquisition: Obtain and familiarize with the ECUST dataset. Understand the dataset's nuances, including its characteristics and potential challenges.

Data Preprocessing: Prepare the dataset for model training. Resize images, normalize pixel values, and organize annotations. Ensure the dataset is well-structured and ready for training.

Object Detection with YOLO: Implement an effective object detection model. Deploy YOLO for food items detection and classification. Achieve accurate identification of food items in images.

Volume Estimation with GrabCut: Accurately estimate the volume of segmented food items. Apply GrabCut algorithm for precise contour delineation. Obtain reliable volume information for further analysis.

Challenges Addressed

Dataset Acquisition: Obtain and familiarize with the ECUST dataset. Understand the dataset's nuances, including its characteristics and potential challenges.

Data Preprocessing: Prepare the dataset for model training. Resize images, normalize pixel values, and organize annotations. Ensure the dataset is well-structured and ready for training.

Object Detection with YOLO: Implement an effective object detection model. Deploy YOLO for food items detection and classification. Achieve accurate identification of food items in images.

Volume Estimation with GrabCut: Accurately estimate the volume of segmented food items. Apply GrabCut algorithm for precise contour delineation. Obtain reliable volume information for further analysis.

Calorie Content Estimation: Derive calorie content from volume and density values. Utilize obtained volume information to estimate the calorie content of each food item. Provide accurate assessments of the nutritional content.

Validation and Evaluation: Assess the model's performance comprehensively. Include metrics such as mean absolute error. Ensure the model's effectiveness in calorie estimation.

Outline of Method

Object Detection

- The users capture top and side views of their food, including a One Yuan coin in each picture.
- The YOLO algorithm[2] is employed for recognizing and outlining food types with precise bounding boxes.

Outline of Method

Object Detection

- The users capture top and side views of their food, including a One Yuan coin in each picture.
- The YOLO algorithm[2] is employed for recognizing and outlining food types with precise bounding boxes.

Image Segmentation

- For image segmentation, each bounding box would be broken down into individual parts before volume estimation.
- The GrabCut algorithm[3] is utilized for image processing to obtain a precise outline of each food item.
- We then obtain segmented food images with background pixels replaced by zeros, focusing solely on foreground pixels.

Outline of Method

Object Detection

- The users capture top and side views of their food, including a One Yuan coin in each picture.
- The YOLO algorithm[2] is employed for recognizing and outlining food types with precise bounding boxes.

Image Segmentation

- For image segmentation, each bounding box would be broken down into individual parts before volume estimation.
- The GrabCut algorithm[3] is utilized for image processing to obtain a precise outline of each food item.
- We then obtain segmented food images with background pixels replaced by zeros, focusing solely on foreground pixels.

Calorie Estimation

- Users provide images from both side and top views, including a One Yuan coin for scale.
- We then apply different formulas to estimate the volume of each food item.
- Using related calorie tables, we determine the calorie content of each food item.

Outline of Method

Volume Calculation

To estimate the volume[4], we calculate the scale factors (α_S and α_T) based on calibration object — One Yuan coin.

$$\alpha_S = \frac{2.5}{(W_S + H_S)/2}; \quad \alpha_T = \frac{2.5}{(W_T + H_T)/2} \tag{1}$$

Where W_S, W_T and H_S, H_T are the width and height of bounding box of coin in side and top view.

$$volume = \begin{cases} \beta \cdot \frac{\pi}{4} \cdot \alpha_S^3 \cdot \sum_{k=1}^{H_S} (L_S^k)^2, & \text{if shape is ellipsoid} \\ \beta \cdot s_T \cdot \alpha_T^2 \cdot H_S \cdot \alpha_S, & \text{if shape is column} \\ \beta \cdot s_T \cdot \alpha_T^2 \cdot \alpha_S \cdot \sum_{k=1}^{H_S} \left(L_S^k / L_S^{max} \right)^2, & \text{if shape is irregular} \end{cases} \tag{2}$$

Where L_S^k is the number of foreground pixels in side view of row k ($k \in 1, 2, 3, \ldots, H_S$). $L_S^{max} = \max_k L^k$. β is a compensation factor (default value = 1.0). s_T is the surface area of the top view.

Outline of Method

Volume Calculation

To estimate the volume[4], we calculate the scale factors (α_S and α_T) based on calibration object — One Yuan coin.

$$\alpha_S = \frac{2.5}{(W_S + H_S)/2}; \quad \alpha_T = \frac{2.5}{(W_T + H_T)/2} \tag{1}$$

Where W_S, W_T and H_S, H_T are the width and height of bounding box of coin in side and top view.

$$volume = \begin{cases} \beta \cdot \frac{\pi}{4} \cdot \alpha_S^3 \cdot \sum_{k=1}^{H_S} (L_S^k)^2, & \text{if shape is ellipsoid} \\ \beta \cdot s_T \cdot \alpha_T^2 \cdot H_S \cdot \alpha_S, & \text{if shape is column} \\ \beta \cdot s_T \cdot \alpha_T^2 \cdot \alpha_S \cdot \sum_{k=1}^{H_S} \left(L_S^k / L_S^{max} \right)^2, & \text{if shape is irregular} \end{cases} \tag{2}$$

Where L_S^k is the number of foreground pixels in side view of row k ($k \in 1, 2, 3, \ldots, H_S$). $L_S^{max} = \max_k L^k$. β is a compensation factor (default value = 1.0). s_T is the surface area of the top view.

Evaluation Metric

We use mean absolute error as the evaluation metric for the volume estimation of each of the food item class.

$$MAE = \sum_{k=1}^{N} \frac{|v_{true}^k - v_{pred}^k|}{v_{true}^k} \times 100\% \tag{3}$$

Experiment Details and Main Results

Dataset Processing

- Utilized ECUST dataset with 19 diverse food types, each having top and side views.
- This dataset is a collection of 2978 images, each meticulously annotated and supplemented with necessary food metrics i.e. volume, density, calorie content etc.
- The images were calibrated images using a One Yuan coin (25 mm diameter) for accurate measurements.
- ECUST dataset considered important factors that affect the accuracy of estimation results: camera, lighting, shooting angle, displacement, calibration object, food type.

Experiment Details and Main Results

Dataset Processing

- Utilized ECUST dataset with 19 diverse food types, each having top and side views.
- This dataset is a collection of 2978 images, each meticulously annotated and supplemented with necessary food metrics i.e. volume, density, calorie content etc.
- The images were calibrated images using a One Yuan coin (25 mm diameter) for accurate measurements.
- ECUST dataset considered important factors that affect the accuracy of estimation results: camera, lighting, shooting angle, displacement, calibration object, food type.

Object Detection Setup

- The food and fruit images are divided into train and test sets. We used 80-20 ordered split for training and testing set.
- After YOLO version 8 is well trained, we use those pairs of test images which YOLO correctly recognizes to estimate volumes. In other words, those images YOLO cannot identity or misidentify in test sets will be discarded.
- We use mean absolute error to evaluate volume estimation results.

Experiment Details and Main Results

Volume Estimation Results

- Volume estimation results are shown in the below Figure. For most types of food in our experiment, the estimation volume are closer to reference volume.
- The mean error between estimation volume and true volume does not exceed 20% except banana, grape, mooncake. For some food types such as orange,our estimation result is close enough to the true value.

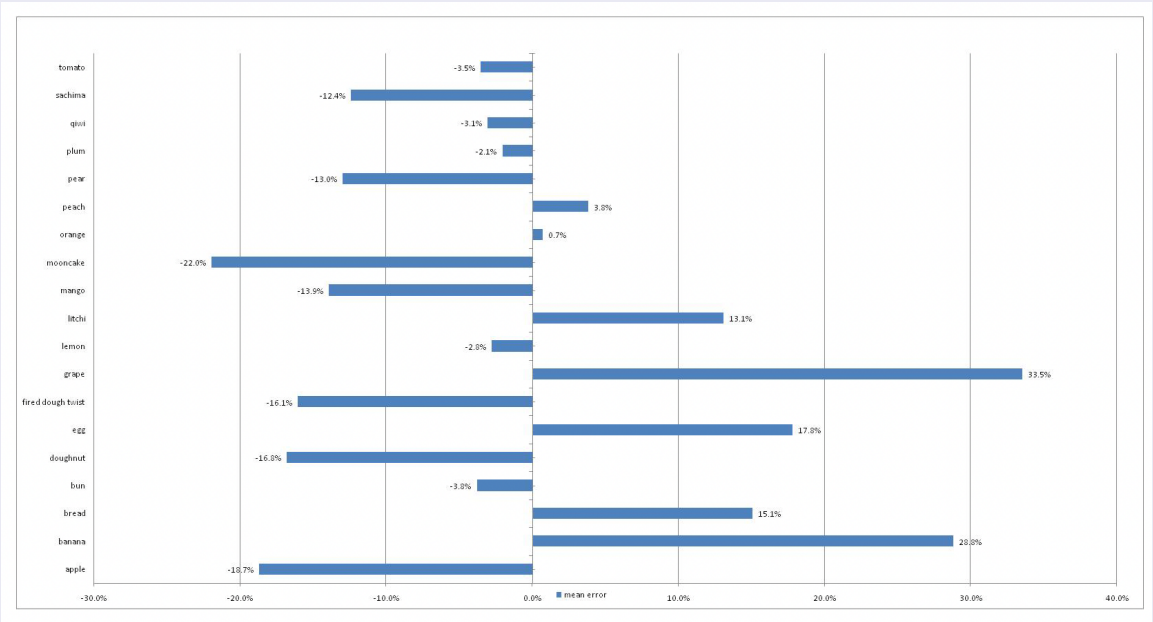

Figure: Volume estimation results of food items from images

Related Work

Previous Studies

- Surveyed existing literature related to calorie estimation from food images.
- Identified key studies addressing challenges in object detection, volume estimation, and calorie assessment.

Object Detection Approaches

- Explored methodologies in the literature for accurate detection and classification of food items.
- Noteworthy studies include Ren et al. and Shang et al.

Volume Estimation Techniques

- Reviewed literature on precise volume estimation for food items in image analysis.
- Notable works include Girshik et al. using Mask RCNN.

Conclusion

- Successfully addressed the challenge of estimating calorie content from food images.
- Implemented a robust methodology utilizing YOLO for object detection and GrabCut for image segmentation.
- Leveraged the ECUST dataset with 2978 annotated images, overcoming challenges in dietary assessment.

Conclusion

- Successfully addressed the challenge of estimating calorie content from food images.
- Implemented a robust methodology utilizing YOLO for object detection and GrabCut for image segmentation.
- Leveraged the ECUST dataset with 2978 annotated images, overcoming challenges in dietary assessment.
- Achieved key milestones in dataset utilization, preprocessing, object detection, volume estimation, and calorie content estimation.
- Validated and evaluated the model's performance, demonstrating effectiveness in accurate food item identification and calorie estimation.

Individual Contributions

Team Members

23M2157	Soumen Kumar Mondal
23M2162	Shubhranil B
23M2154	Siddhant Dnyanesh Gole
23M2156	Vaibhav Rathore
23M2158	Akash Pal

Contributions

Soumen	Led the dataset acquisition and preprocessing efforts. Implemented and fine-tuned the YOLO model for efficient object detection. Handled documentation, code-base, report and presentation.
Vaibhav, Shubhranil	Applied expertise in image segmentation, incorporating the GrabCut algorithm. Contributed to the preparation of report.
Sidhhant, Akash	Contributed to the development of the calorie estimation workflow and validation processes.

References

Codebase

Project Repo	All the works of the team can be found in this GitHub link.
Dataset Repo	The dataset can be found in this GitHub link.
Volume Repo	This GitHub link provided by Liang et al.[1] was referred to calculate the volume.
External Repo	The following libraries are used for the implementation of different algorithms — Ultralytics, OpenCV

External Papers

[1] Yanchao Liang and Jianhua Li. "Computer vision-based food calorie estimation: dataset, method, and experiment". In: *arXiv preprint arXiv:1705.07632* (2017).

[2] Joseph Redmon et al. "You Only Look Once: Unified, Real-Time Object Detection". In: (2016), pp. 779–788. DOI: 10.1109/CVPR.2016.91.

[3] Yubing Li et al. "Grab Cut Image Segmentation Based on Image Region". In: (2018), pp. 311–315. DOI: 10.1109/ICIVC.2018.8492818.

[4] V.Subapriya. "Prediction of Health Risks Using Semi-Supervised Learning". In: *International Journal of Scientific Research and Engineering Development* (2020).