FoodPhone: Food Image Recognition Using SVM Classifier

Group 16: Lijun Chen, Mingyang Yan, Weipeng Wang, Yibin Zhang
CS542: Machine Learning Class Project, Boston University

Figure 1. Example of four food categories images: apple pie, baby back rib, waffle, and sushi

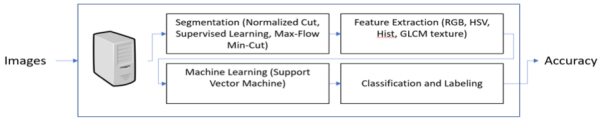

Figure 2. Project Overall System Architecture

Figure 3. Image segmentation after supervised learnings; images correspond to Fig.1

Project Task

The goal of our project is to recognize and classify the type of the food given 101 food types by analyzing images of food/meals like the ones shown in Fig 1. The input of the model we used is an image shown in Fig.1 and the output is a number corresponding to its label.

Some challenges of the project:

- How not to over segment the images
- Choose proper features and training model for the task

Related Work

- According to F. Zhu, A. Myers, Zhang, etc. [1] [2] [3], the food recognition problem usually starts with an image segmentation process, where from each segment low and high levels of the features indicating the type of food are extracted.
- Although different projects take different datasets and variant image processing methods such as active contour, SIFT, salient/region based sampling are applied, CNN and SVM are the most commonly used machine learning models to classify the type of food.

Dataset and Metric

Dataset:
75750 training images provided on Kaggle, where 80% of it will become the training set and 20% will be used as the validation set to optimize the performance of the training model. The other 25250 provided test images will be used as the test set. There are in total of 101 categories.

Evaluation Metric:
The classification accuracy will be using the simplest metric by dividing the number of correct predictions by the total number of predictions over the test set.

$$Accuracy = \frac{\text{Number of correct predictions}}{\text{Total number of prediction}}$$

Approach

As is shown in Figure 1, our team feed features extracted from the pre-processing steps [4][5] of the input dish images into the machine learning model, assign a label to each test image and calculate the accuracy. The machine learning model, based on SVM (LinearSVC), applies the one-vs-rest scheme for multiple classes and has a squared hinge loss function.

- Pre-processing Step: Image Segmentation
  - Apply SLIC (Simple Linear Iterative Clustering) to get about 400 superpixels and used normalized cut to find the most obvious boundaries in the image and combine those that are more similar in color space.
  - A pre-trained supervised learning model based on the color, histogram, position of the food over 200 labelled images of food is used to assign a probability of being food to each segment
  - Use a simplified Max-Flow Min-Cut method to keep the segments with fairly higher probability. The cut method is more aggressive on training images to make sure all the noise are removed and more conservative on test images in case of over segment.
- Training in Support Vector Machine
  - 3 cues of RGB values, 3 cues of HSV values, 5 cues of 5-bin histogram, 3 cues of 3-bin histogram, 8 cues of GLCM dissimilarity in 8 directions, 8 cues of GLCM correlation in 8 directions, totally 30 cues of features are extracted from the processed training images to feed into the learning model.

Evaluation

- The pre-processing turned out to be good on most of the images, examples shown in Figure 3. The food segments are well separated from the rest part of the image.
- One drawback of the pre-processing step is that the training step can be hugely influence by it. For example, if only part of the food is preserved, the feature of its shape can never be used. Information can be lost due to the error in segmentation.
- The normalized cut method requests quite some processing power and processing time, which might not be very applicable over test images, one possible way might be looping over all sub-regions of a test image and give possibility of the region being a certain kind of food and then combine all the regions to give a final decision.

Conclusions

- There exists method using CNN which gives about around 85% of accuracy in food classification, which can be a more efficient method. But with better features and supervised training model, our accuracy can be improved to similar or higher accuracy.
- A faster cutting method may be needed to make the approach applicable to the real world. One possible solution is to loop over blocks of pixels rather than superpixels and directly apply Max-Flow Min-Cut algorithm. We took advantage of the skimage normalized-cut library but it really should be a especially designed library for separating food from a image.

Welcome to have a look at our GitHub Page:

References

1. F. Zhu *et al.*, "The Use of Mobile Devices in Aiding Dietary Assessment and Evaluation," in *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 4, pp. 756-766, Aug. 2010.
2. A. Myers et al., "Im2Calories: Towards an Automated Mobile Vision Food Diary," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, pp. 1233-1241.
3. Zhang, W., Yu, Q., Siddiquie, B., Divakaran, A., & Sawhney, H. (2015). "Snap-n-Eat": Food Recognition and Nutrition Estimation on a Smartphone. Journal of Diabetes Science and Technology, 9(3), 525-533.
4. R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua and S. Süsstrunk, "SLIC Superpixels Compared to State-of-the-Art Superpixel Methods," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274-2282, Nov. 2012.
5. M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," Int. J. Comput. Vis., vol. 1, pp. 321–331, 1998.