Foodness Proposal for Multiple Food Detection by Training with Single Food Images

Wataru Shimoda  Keiji Yanai
The University of Electro-Communications, Tokyo, Japan

Objective

- Weakly supervised detection
- Use only image level annotation
- Use only single label for training
- Target is multi-food detection

Contribution

- Combine weakly supervised segmentation method and proposal base approach
- Improve accuracy from weakly supervised segmentation results.
- Improve computational cost from proposal base method.

Food region proposal

- We regard estimated regions of upper rank classes as proposals.
- If there are no target foods category in fact, the estimated food regions are belong to any food region.

Related work

Fully supervised method

- Faster RCNN
  - Use bounding box annotation

Weakly supervised method

- Fully convolutional network
  + Global max pooling
  - Train with single label and multi-label

  Distinct class specific saliency maps
  - Improved visualization of Simonyan et al.
  - Train with single label and multi-label
  - Achieved state-of-the-art in weakly supervised segmentation.

Our method

- Train with only single label
- Test for multi object images.
- Existence methods assume to train with Pascal VOC or MSCOCO which has multi label annotation.
- Most of existence datasets and web images have only single label.

Recognition of food region candidates

Difference in object detection and food detection
- Small region recognized as food
- Similar to texture recognition

We generate food patches by random cropping.-separate food patches class from general food.

Experiments

Training:
- UECFOOD 100 + Web images
- Food 100 class: 1000 images for each category + non-food: 10000 images
- Training without bounding box and multi label.

Test:
- UECFOOD 100 multiple food dataset
  - include at least one category of UECFOOD100
  - Each class image number vary.
  - We separate evaluation by image number.

Comparison of global pooling methods

Comparison of other proposal methods

Detection results with different conditions

Evaluation metric

Average precision for detection.
"100 class" is average of all class AP.
"53 class" is average of 53 class AP which class has at least 10 images.
"11 class" is average of 11 class AP which class has at least 50 images.

Background

Problems

- Weakly supervised segmentation for containing multi label in training data
  - Distinct class specific saliency maps
  - Caused significant performance drop by training with only single label.
- Detection and segmentation by proposal
  - We can achieve detection in bottom-up approach by proposal base method.
- Previous works:
  - RCNN, SDS
  - Generates around 2000 candidates.
  - Large computational cost.

Idea

Weakly supervised segmentation results are low confidence.
However regions respond only food regions

We consider CNN could transfer only food concept.

Regard low confidence segmentation results as proposal candidates.
Combine weakly supervised segmentation and proposal base detection method.