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

Madima 2016

The University of Electro-Communications in Japan

Wataru Shimoda, Keiji Yanai

国立大学法人 電気通信大学

Objective

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

Training image

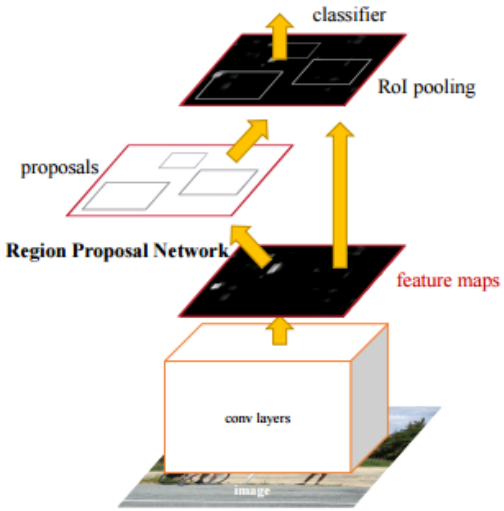
Test image

Contribution

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

Fully supervised method

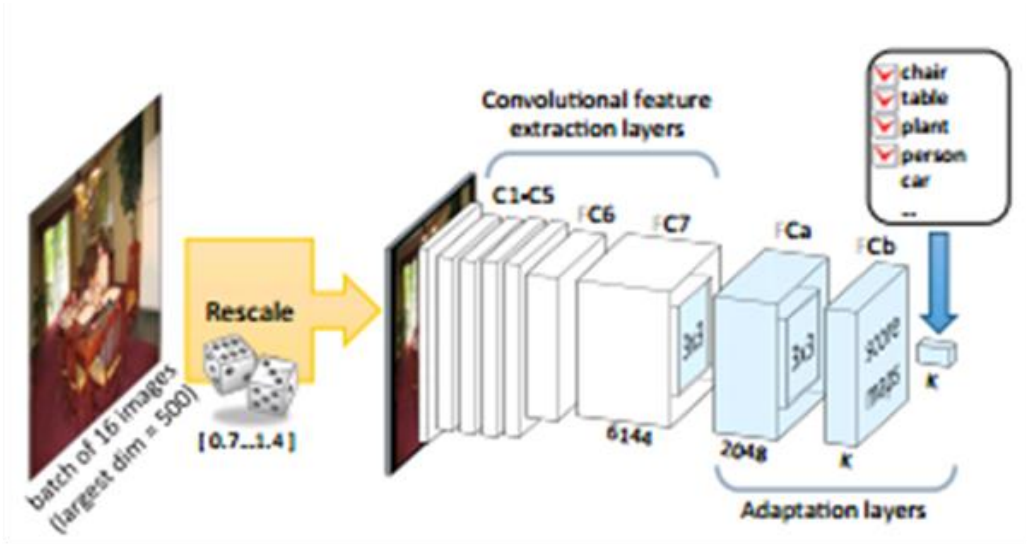
- Faster RCNN
  - Use bounding box annotation
  - Large annotation cost



[Ren et al. NIPS 2015]

Weakly supervised localization

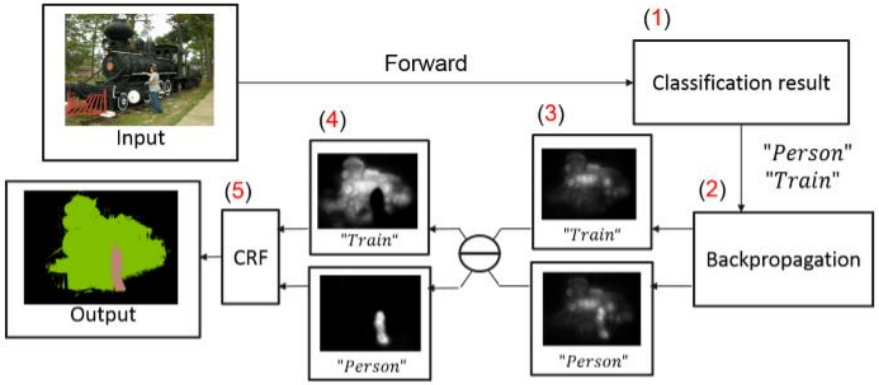
- Fully Convolutional Network + Global Max Pooling
  - Train without bounding box

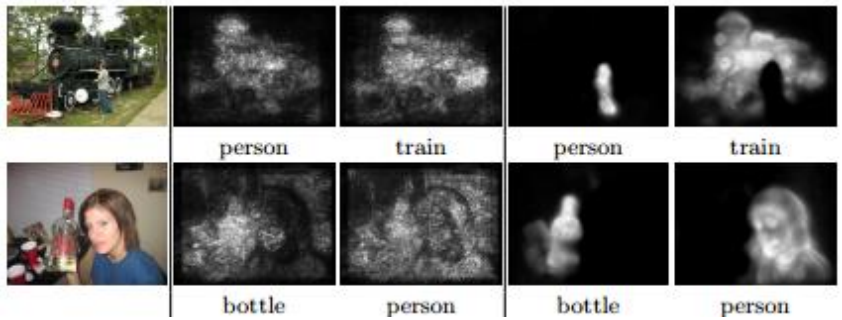


[Oquab et al. CVPR 2015]

Weakly supervised segmentation

- Distinct class specific saliency maps
  - Also use FCN and GMP
  - Pixel-wise prediction
  - Train with single label and multi label





[Shimoda et al. ECCV 2016]

Our method

- Train with only single label
  - 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
  - Test for multi object images

Training images
-single label

Test images
-multi label

Background

- Weakly supervised method by training with only single label
  - Causes significant performance drop

Result of Shimoda et al. ECCV 2016 for food images

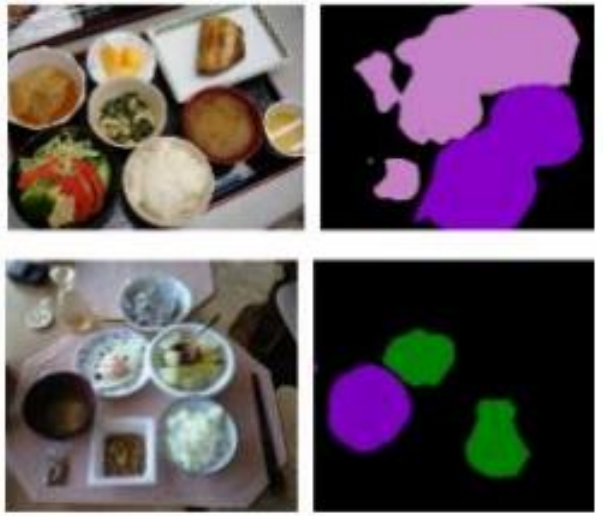
Traditional bottom up approach

- Proposal
  - previous works: RCNN, SDS
  - generates around 2000 candidates
  - Large computational cost

Key idea

- Previous weakly supervised results showed low performance
  - 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.



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

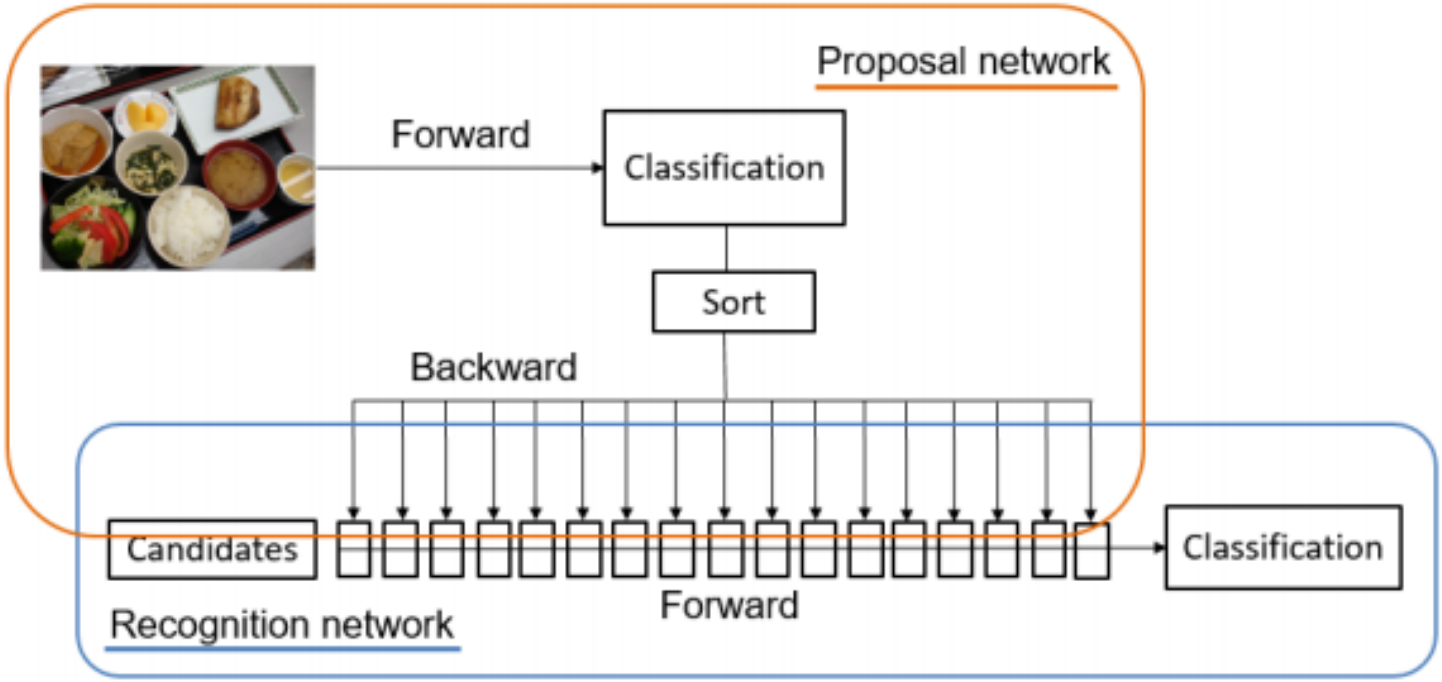
Proposals

deep-frid chiken	Ginger fried pork	Boiled beef	Beaf steak	Fried vegetable	Pork **cutlet**	Chiken rice

rice	Rice deep-frid chiken	rice	deep-frid chiken	Non food	Rice deep-frid chiken Green salada	deep-frid chiken

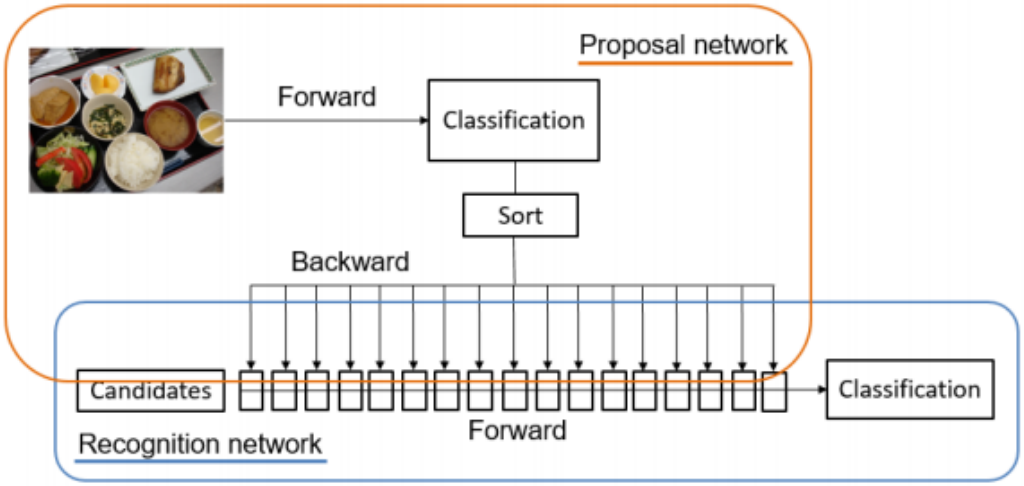
Method

- We re-recognize low confidence segmentation result



Overview

- Sort recognition result
- Estimate upper rank food region
- Re-recognize estimated region
- Unify recognition result by NMS



Difference in object detection and food detection

- Small region recognized as food
  - Similar to texture recognition

General Object

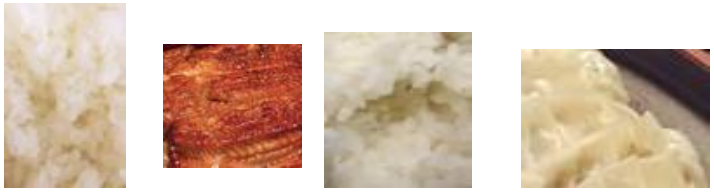
Back ground

Food

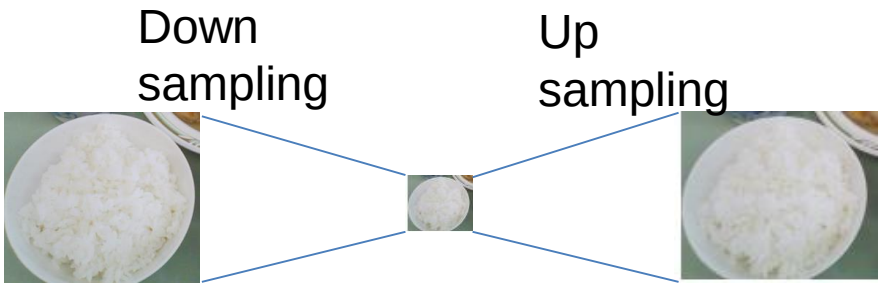
Rice

Data augmentation

- Food patch images
  - Generate by cropping
  - Separate food patches class from general food.



- Low resolution images
  - Generat by down sampling and up sampling
  - Add low resolution images to all classes



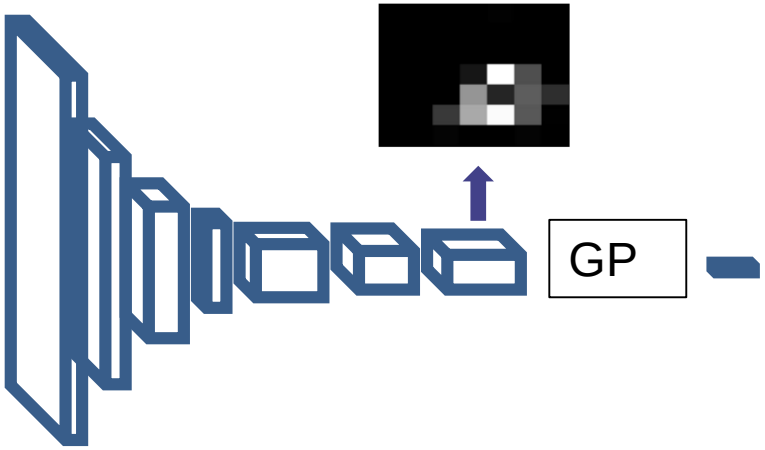
Experiments

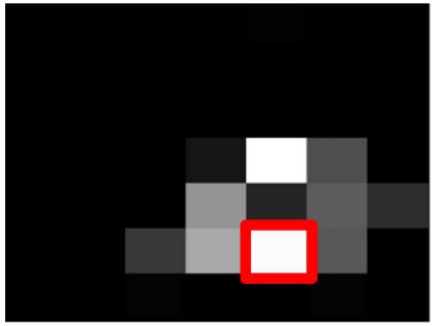
- Training
  - UECFOOD 100＋Web images
  - food 100 class:1000 images + non- food:10000 images
  - Training without bounding box and multi label.
- Test
  - UECFOOD 100 multiple food dataset
  - include at least one category of UECFOOD100
  - Each class image number vary
  - We separate evaluation set by each class image number.

Detection results with different conditions

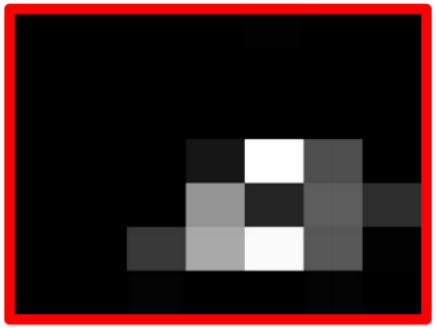
Patch images	Low resolution images	100 class	53 class	11 class
—	—	33.5	35.1	33.3
○	—	32.2	34.8	31.8
○	○	**36.4**	**39.9**	**36.3**

Comparison of global pooling methods





Average pooling



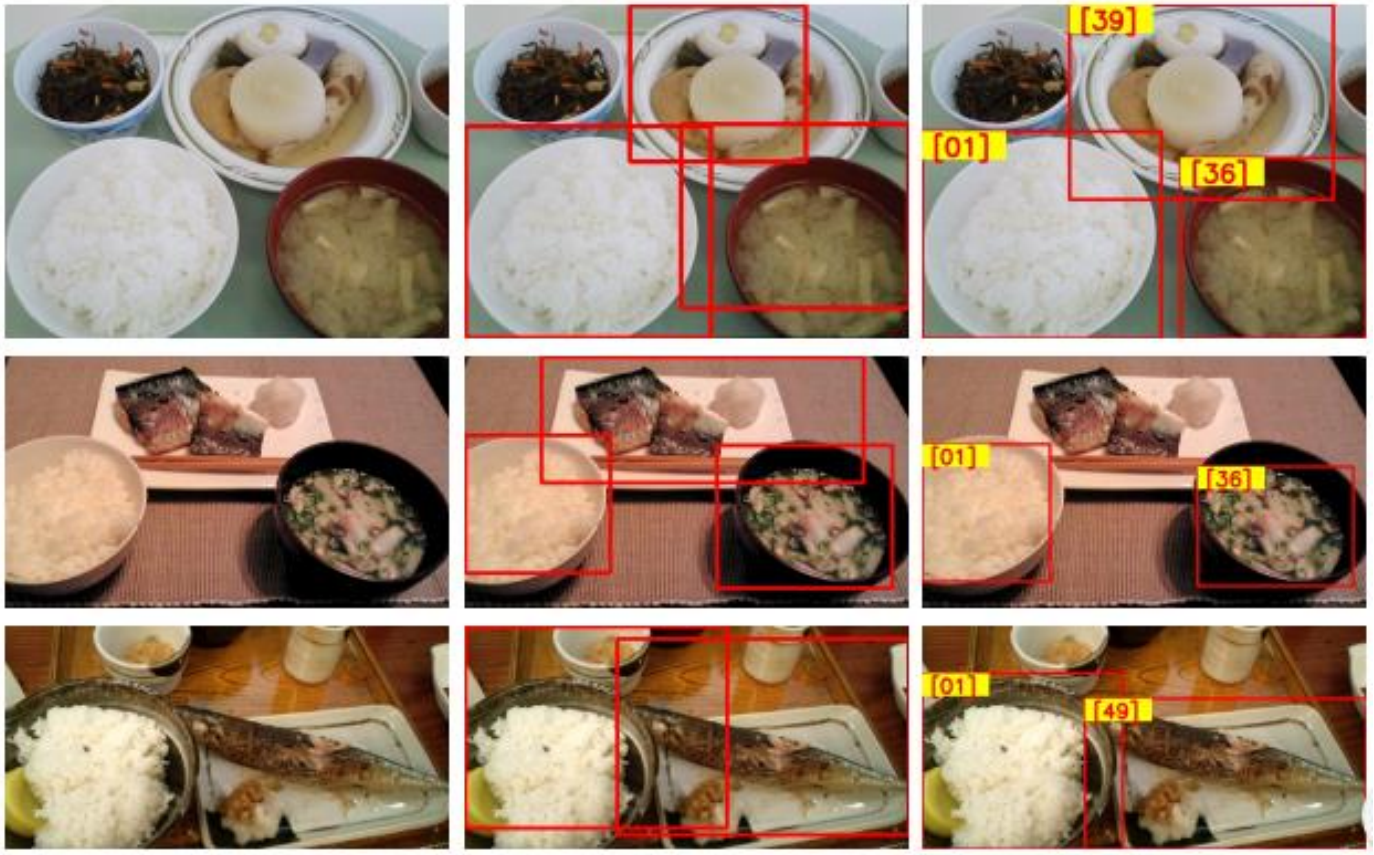
Max pooling

method	100 class	53 class	11 class
Average pooling	36.4	39.9	36.3
Max pooling	**38.9**	**42.5**	**38.1**

Comparison of other proposal methods

Method	100 class	53 class	11 class	Proposal speed [s]	recognition speed [s]
SS	38.3	39.1	35.7	7.6	35.0
MCG	33.9	**43.7**	33.4	2.5	35.0
Ours 10 class	33.1	33.0	33.2	**0.5**	**1.1**
Ours 20 class	36.5	40.1	37.7	1.0	2.6
Ours 30 class	**38.9**	42.5	**38.1**	1.4	3.8

Examples



Conclusion

- Achieved weakly supervised detection by training only single label image
- Our method is high speed than previous proposal base detection method