CNNの順･逆伝搬値とCRFを利用した弱教師領域分割

MIRU 2016 at Hamamatsu, Japan

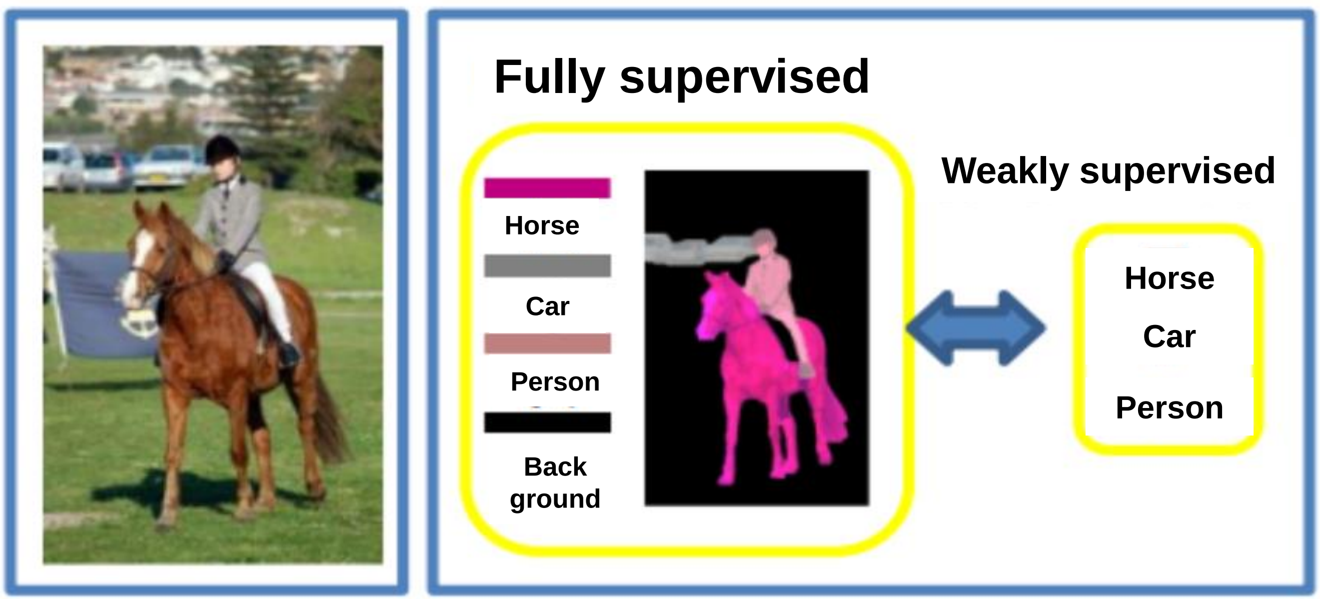
Wataru Shimoda and Keiji Yanai

The University of Electro-Communications, Tokyo, Japan

国立大学法人 電気通信大学

Introduction

- Pixel-wise annotation is costly
- Our goal is weakly supervised segmentation
  - Train with only image-level-label

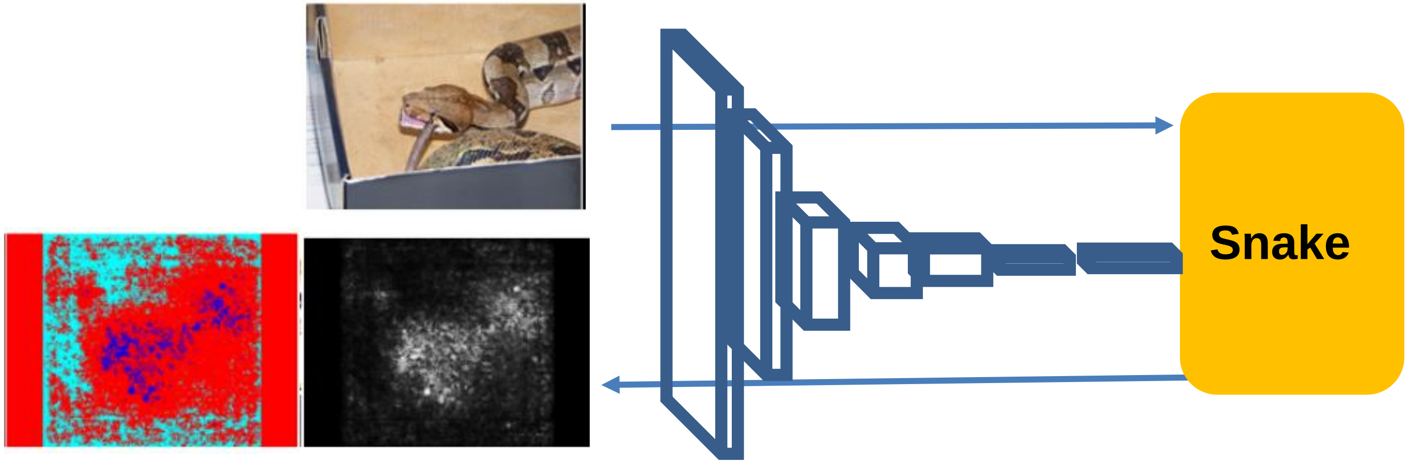


Our contribution

- We improved backpropagation(BP)-based saliency maps
  - By taking in some techniques used in forward-based semantic segmentation
- We showed BP-based saliency maps can help object localization
  - (1) We verified BP-based saliency maps can enhance forward-based coarse object heat maps
  - (2) We achieved semantic segmentation with only gradient by subtracting each class gradient

BP-based saliency maps

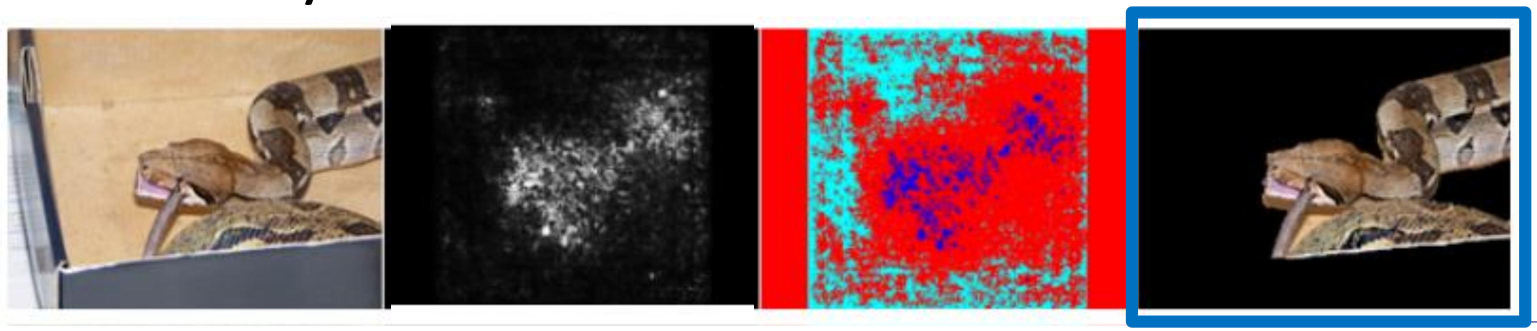
- Propagate class signal through backpropagation
- Visualize image-level-gradient as saliency maps
  - – saliency maps respond to object location



[Simonyan et al. ICLR 2014]

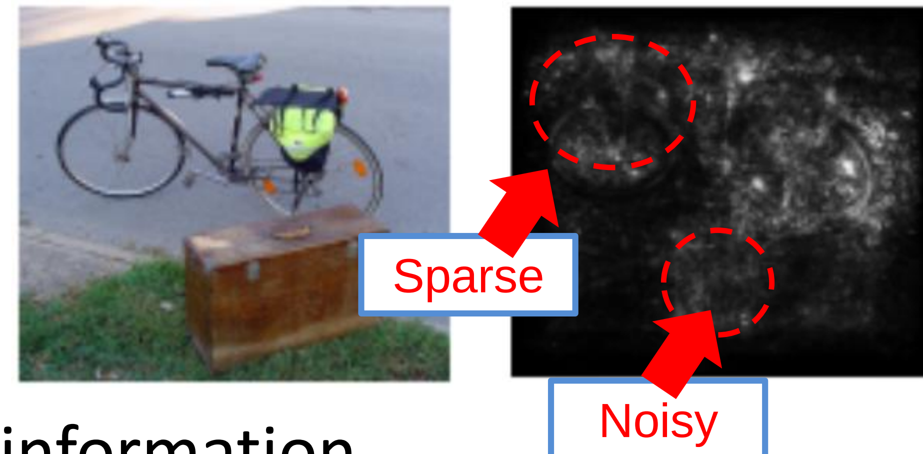
Visualization for Segmentation

- Visualization mean revealing object location
  - Computed using classification CNN, trained on image labels
  - Weakly supervised methods
- Simonyan et al. tried deal saliency maps as GrabCut seeds and achieved segmentation
  - But they didn’t show numerical results

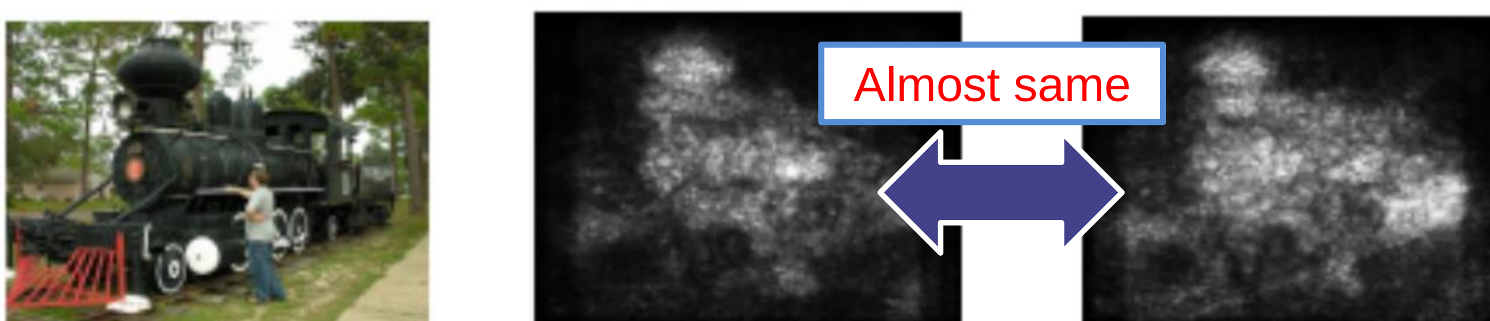


Problems of gradient obtained by backpropagation

- Previous BP-based segmentation accuracy is poor due to following factors
  - Gradient often become sparse and noisy



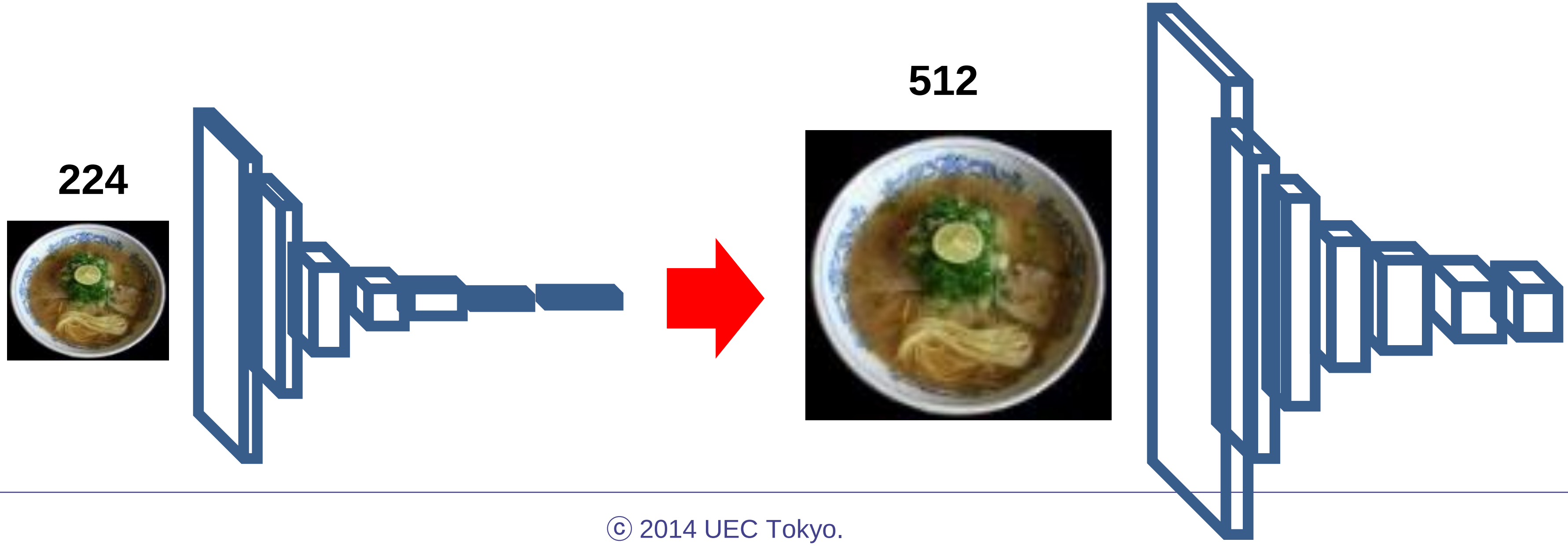
  - Gradient lose semantic information



Gradient for person　　Gradient for train

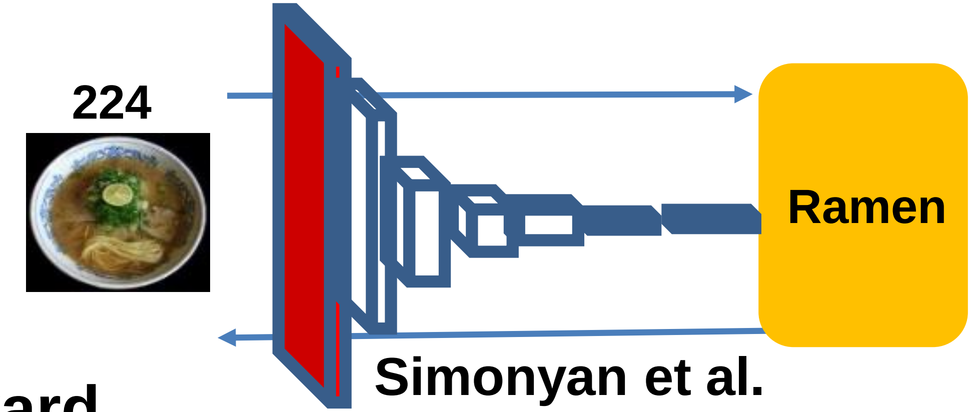
Fully Convolutional Network(FCN)

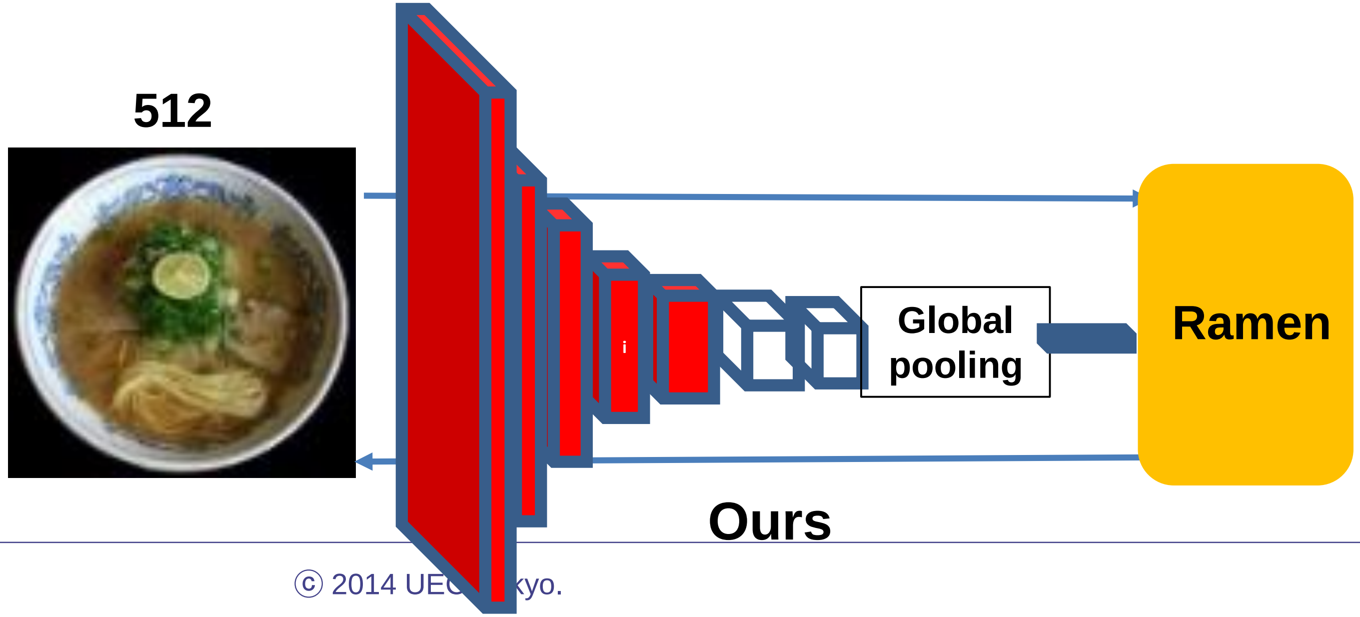
- Replace Fc layer to Convolution layer
- FCN accept arbitrary input image size
- Output and intermediate feature maps become more dense



Change points

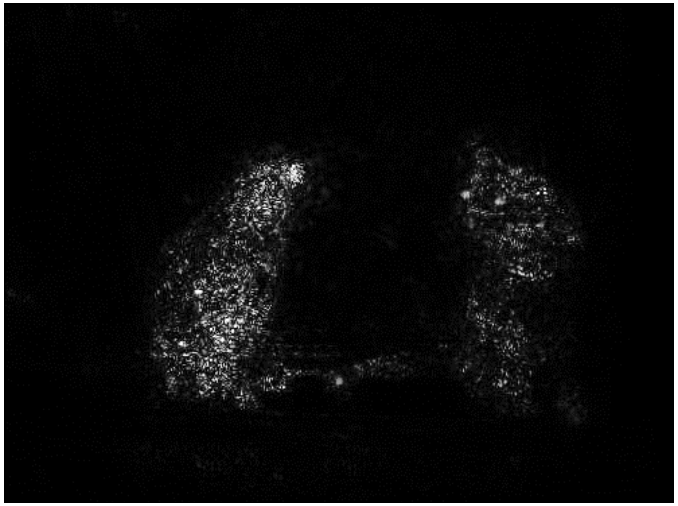
- FCN + Global Pooling
- Input image size
- Intermediate layer
- ReLU function in Backward

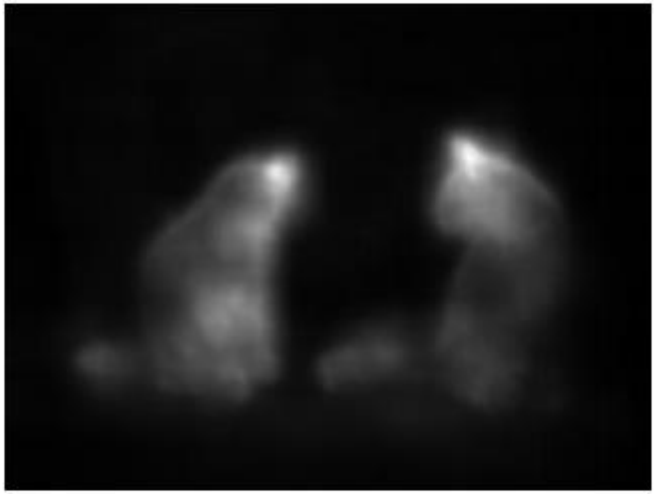




Change result

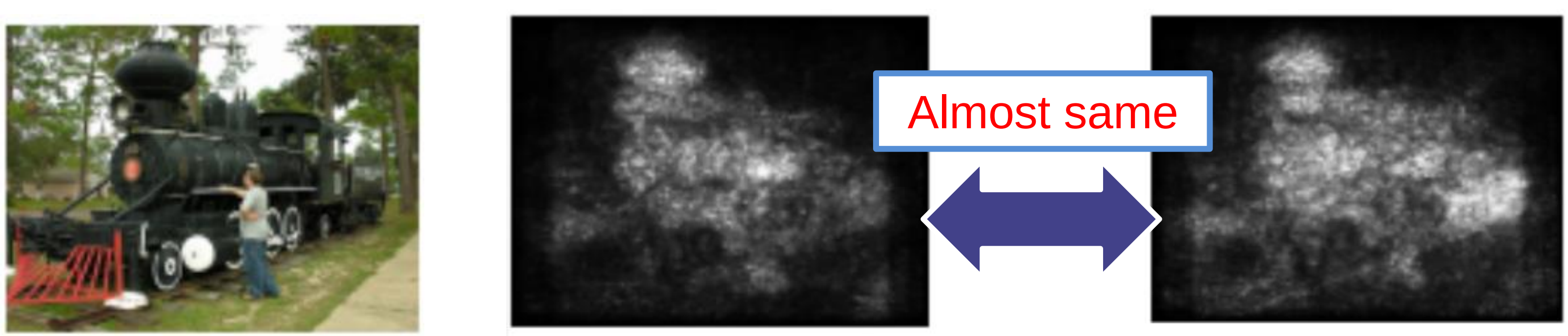
- Saliency maps become more dense and clear





To obtain semantic information

- Gradient loses semantic information
- To solve this problem
  - (1) We combine forward-based feature maps
  - (2) We subtract each class gradient



Gradient for person

Gradient for train

(1) Combining forward-based coarse object heat maps

- We use BP-based saliency maps to enhance forward-based coarse object heat maps
- Forward-based feature maps
  - Zoom out feature(ZOF)
    - CNN + Super Pixels
    - Train SVM with MIL
  - Fully convolutional networks(FCN)
    - Replace Fc layer to Conv layer
    - Output matrix has semantic inofrmation

(1) Experiment

- Dataset
  - Pascal VOC 2012
  - 21 general object class (including background)
  - 10532 training images
- Training
  - We fine-tune VGG16 FCN model with image-level-label by global pooling
  - We adopt Sigmoid cross entropy loss for multi class label
  - We randomly resize input image to avoid overfitting

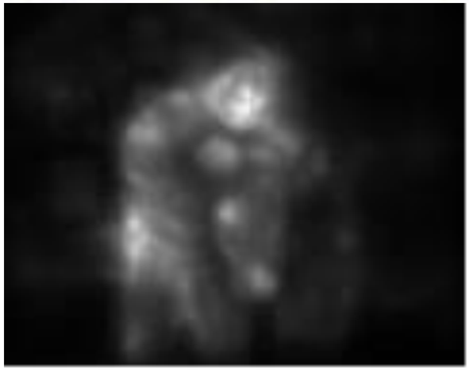
(1) Experimental results

- BP-based saliency maps enhance forward-based feature maps clearly.

Method	Mean IU
FCN-MIL [ICLR 2015] (FCN only)	24.9
ZOF with GBP (Ours)	37.7
FCN with GBP (Ours)	40.7

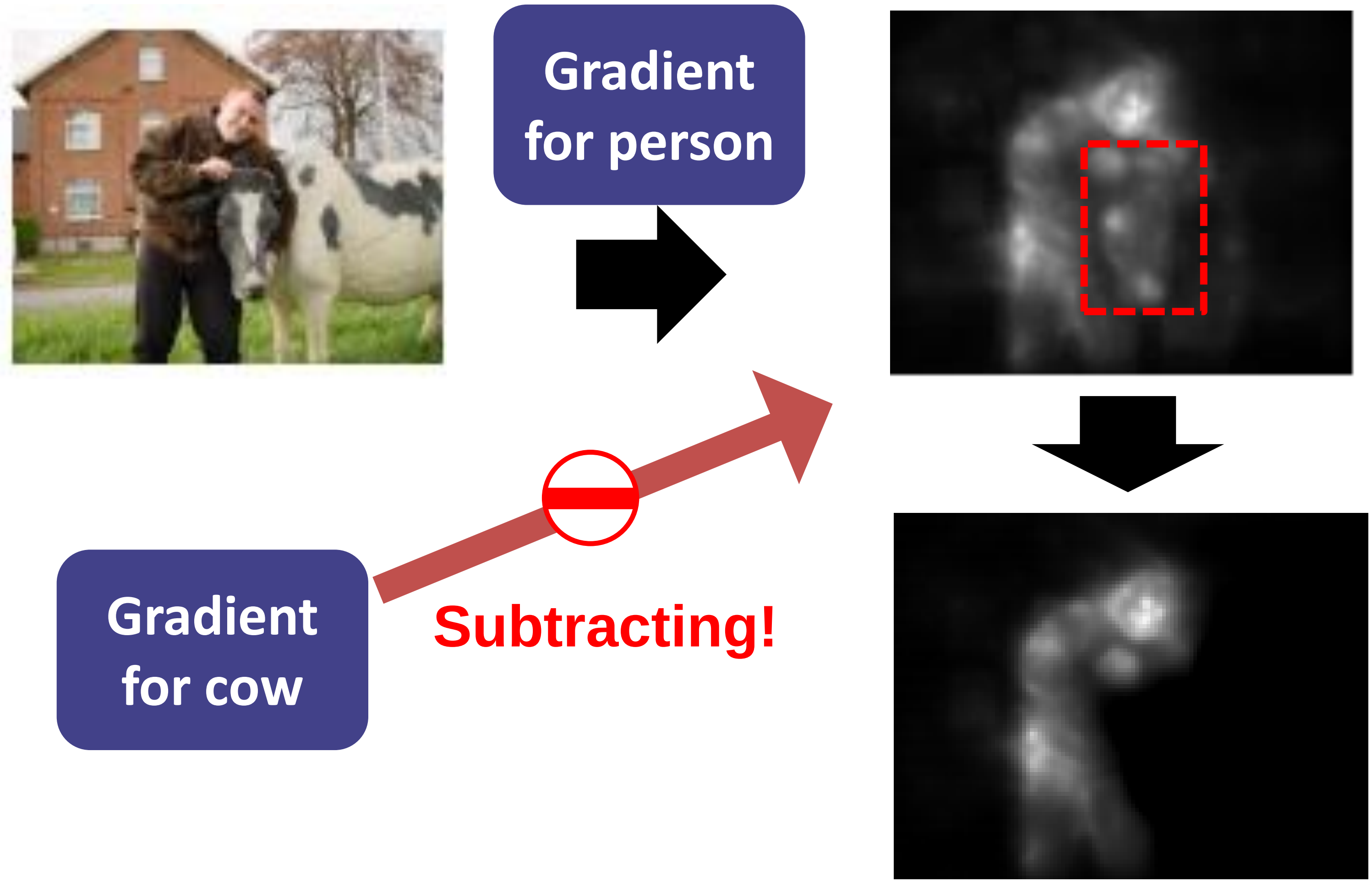
Why do gradient maps lose semantic information?

- Large gradient regions mean contributed to recognition of CNN
- Concern
  - Not-target class regions also respond
  - Background regions don't respond
- Does object-ness contribute CNN recognition even though not-target class regions due to training with general object datasets?



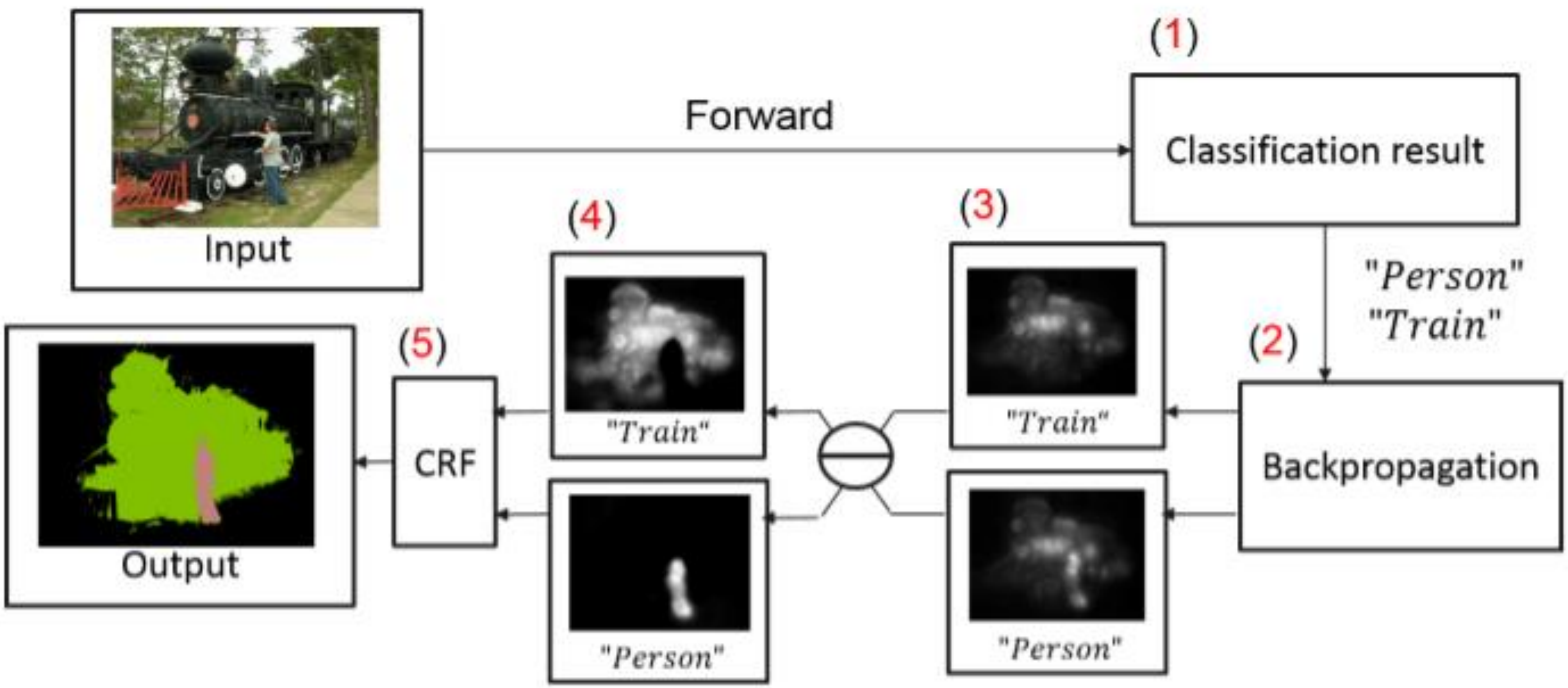
Gradient for person

(2) Subtracting each class gradient



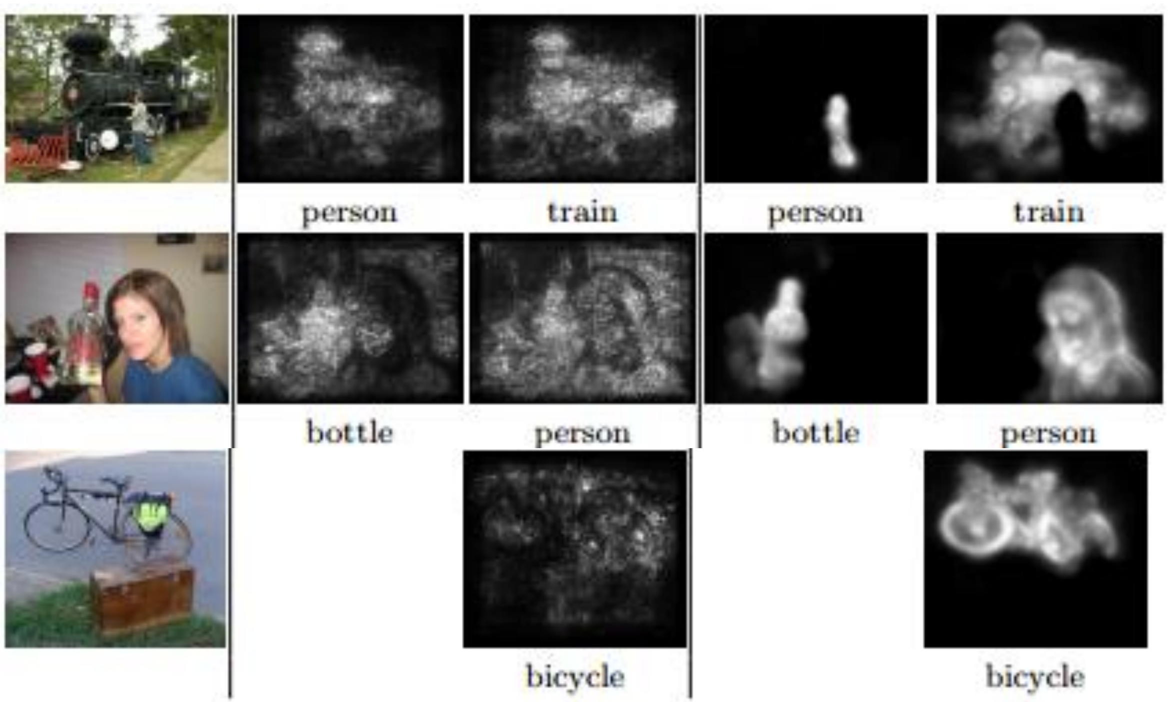
(2) Proposed method

- We achieved semantic segmentation with only gradient maps
- We obtain final regions by Dense CRF



(2) Compare with base-line method

- Saliency maps and numerical results



Simonyan et al. **Ours**

Method	Mean IOU
Sim et al. + CRF	33.8
Ours	44.1

(2) Effect of subtraction

- Test for subtraction class numbers
  - Note that we need N times backward computation
- Class N = 0 means no subtraction

Class N	0	1	2	3	4	5	10
Mean IU	38.2	42.2	43.5	44.1	**44.2**	44.0	43.7

(2) Comparison with previous works

Method	Mean IOU
MIL-FCN (iclr 2015)	**25.7**
EM-Adapt(iccv 2015)	**38.2**
CCNN (iccv 2015)	**34.5**
MIL-sppxl (cvpr2015)*	**36.6**
MIL-bb (cvpr2015)*	**37.8**
MIL-seg (cvpr2015)*	**42.0**
Ours w/o CRF	**40.5**
Ours w/ CRF	**<u>44.1</u>**

* means that they use additional data

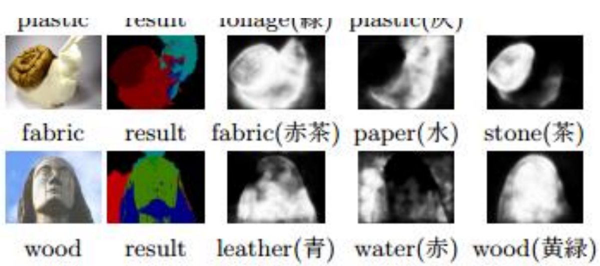
(2) Example of Results

W/o CRF   W/ CRF   Ground truth

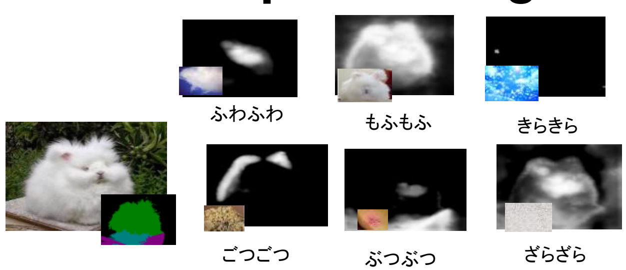
(2) Applications

- We can adapt this method for any CNN models
- Easy implementation!
- GitHub https://github.com/shimoda-uec/dcsm

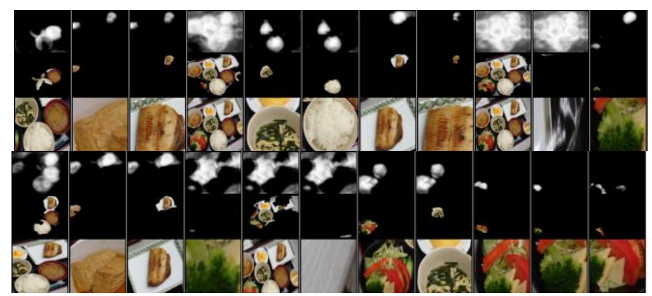
Material images



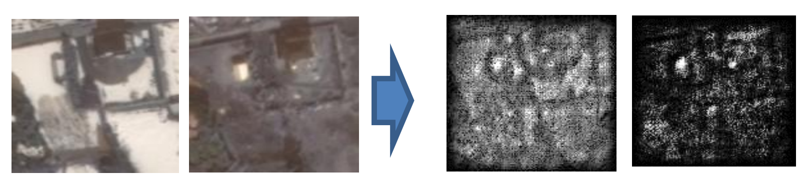
Onomatopoeia images



Food images



Satelite images（in AIST）



Conclusion

- We adapted visualization method to semantic segmentation method
- We improved a BP-based saliency maps
- We achieved semantic segmentation using only gradient maps by subtracting
- We achieved the state of the art in the weakly supervised semantic segmentation with Pascal VOC 2012.