

Learning to Detect Salient Objects with Image-level Supervision

Chaonan Song

May 29, 2018

Abstract

Today, I read a thesis written by Professor Wang, *Learning to Detect Salient Objects with Image-level Supervision*. In this paper, Dr. Wang and his team leverage the observation that image-level tags provide important cues of foreground salient objects, and develop a weakly supervised learning method for saliency detection using image-level tags only. In the first stage of our training method, a global smoothing pool was proposed to enable the FCN to assign object class labels to the corresponding object areas, and the FCN could use the predicted saliency map to paste all potential foreground areas. In the second phase, an iterative conditional random field was developed to enhance the consistency of spatial tags and further improve performance. Their proposed approach eases the workload of annotations and allows the use of existing large training sets with image-level tags.

1. Introduction

Recently, more and more teams are training DNNs using samples with accurate pixel-level annotations for significance testing Li and Yu [4]. Compared with the unsupervised method Kong *et al.* [3], DNNs learned from comprehensive supervision can more effectively capture semantically prominent foreground areas and produce accurate results in complex scenarios. DNN's superior performance depends heavily on a large number of data sets and pixel-level annotations. However, the annotation work is very tedious, and the training set with accurate annotation is still scarce and expensive. To alleviate the need for large-scale pixel annotations, Dr. Wang's team explored the weak supervision of image-level tags to train saliency detectors. Image-level tags indicate the presence or absence of object categories in the image and are easier to collect than pixel-wise annotations. The task of predicting image-level labels focuses on the object class in the image and is not responsible for the target location (Figure 1 left), whereas the significance test aims to highlight the full expansion of foreground objects and ignore their categories (Figure 1 right). The two tasks

seem conceptually different but have inherent connections. The contributions of their team are shown in Table 1.

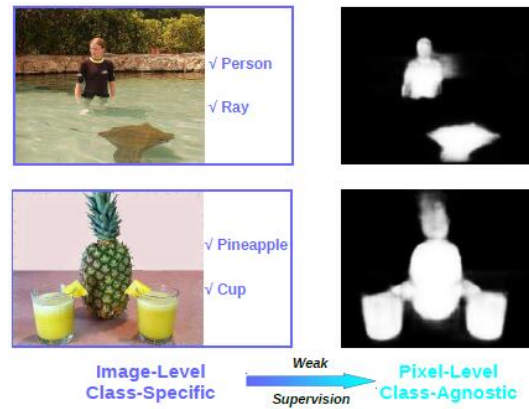


Figure 1. Image-level tags (left panel) provide informative cues of dominant objects, which tend to be the salient foreground. We propose to use image-level tags as weak supervision to learn to predict pixel-level saliency maps (right panel).

Contribution 1	An example of a saliency detector is provided
Contribution 2	Global smoothing pooling layer network is provided
Contribution 3	A new CRF algorithm is proposed

Table 1. Contributions of Dr. Wang and his team.

2. Related Work

Fully Supervised Saliency Detection. Many surveillance algorithms, such as CRF proposed by Liu *et al.* [5], Random Forest proposed by Jiang *et al.* [1], have been successfully applied to significant tests.

Weakly Supervised Learning. Weakly supervised learning has attracted more and more attention, such as object detection proposed by Song *et al.* [7], semantic segmentation proposed by Pathak *et al.* [6] and boundary detection proposed by Khoreva *et al.* [2].

The exploration of weakly supervised learning for significant tests is limited. Dr. Wang and his team were the first to learn significant object detectors using object class labels.

References

- [1] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li. Salient object detection: A discriminative regional feature integration approach. In *CVPR*, 2013.
- [2] A. Khoreva, R. Benenson, M. Omran, M. Hein, and B. Schiele. Weakly supervised object boundaries. *arXiv*, 2015.
- [3] Y. Kong, L. Wang, X. Liu, H. Lu, and X. Ruan. Pattern mining saliency. In *ECCV*, 2016.
- [4] G. Li and Y. Yu. Visual saliency based on multiscale deep features. In *CVPR*, 2015.
- [5] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H. Y. Shum. Learning to detect a salient object. *PAMI*, 33(2):353--367, 2011.
- [6] D. Pathak, P. Krahenbuhl, and T. Darrell. Constrained convolutional neural networks for weakly supervised segmentation. In *ICCV*, 2015.
- [7] H. O. Song, R. Girshick, S. Jegelka, J. Mairal, Z. Harchaoui, T. Darrell, et al. On learning to localize objects with minimal supervision. In *ICML*, 2014.