Integrated Deep and Shallow Networks for Salient Object Detection

Jing Zhang, Bo Li, Yuchao Dai, Fatih Porikli, Mingyi He

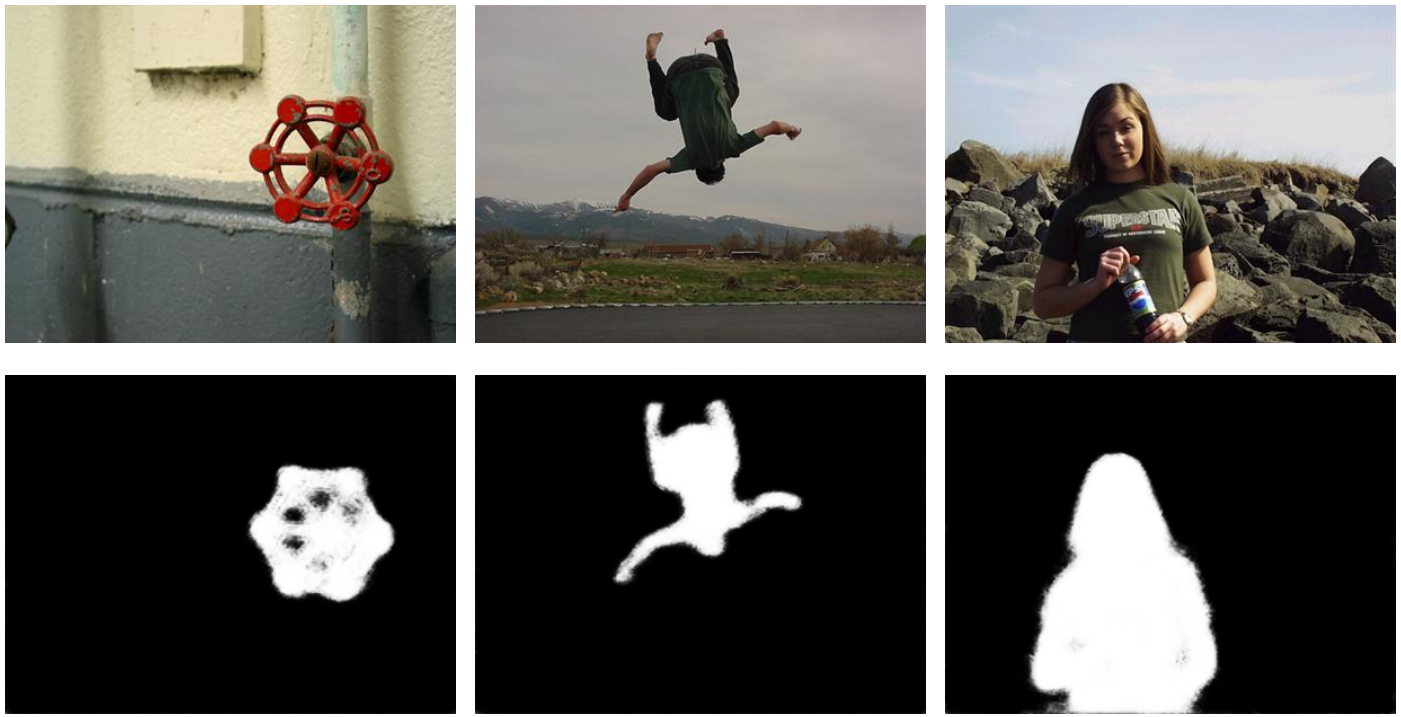
Jing Zhang[1,2], Bo Li[1], Yuchao Dai[2], Fatih Porikli[2], Mingyi He[1]
[1]Northwestern Polytechnical University [2]Australian National University

zjnwpu@gmail.com
robert_libo@qq.com
yuchao.dai@anu.edu.au
fatih.porikli@anu.edu.au
myhe@nwpu.edu.cn

2017-08-17

What is salient object detection?

Salient object detection aims at identifying the **visually interesting objects** regions that stand out relative to their neighbors and are consistent with **human perception**.

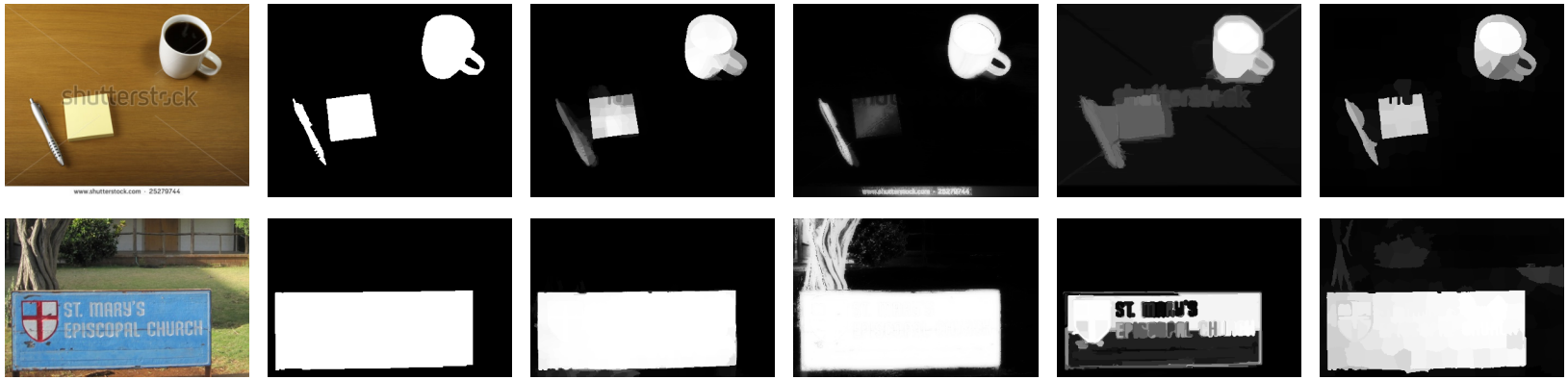


Sample images and their corresponding saliency maps.

Deep features vs handcrafted features

- Deep features can efficiently capture semantic information.
- Handcrafted features, which is summarized and described with human knowledge, are pivotal for simple scenarios.
- Deep features based salient object detection achieves the state-of-the-art performance;
- There exist situations where handcrafted saliency methods would outperform deep saliency methods.

Deep features and handcrafted features together

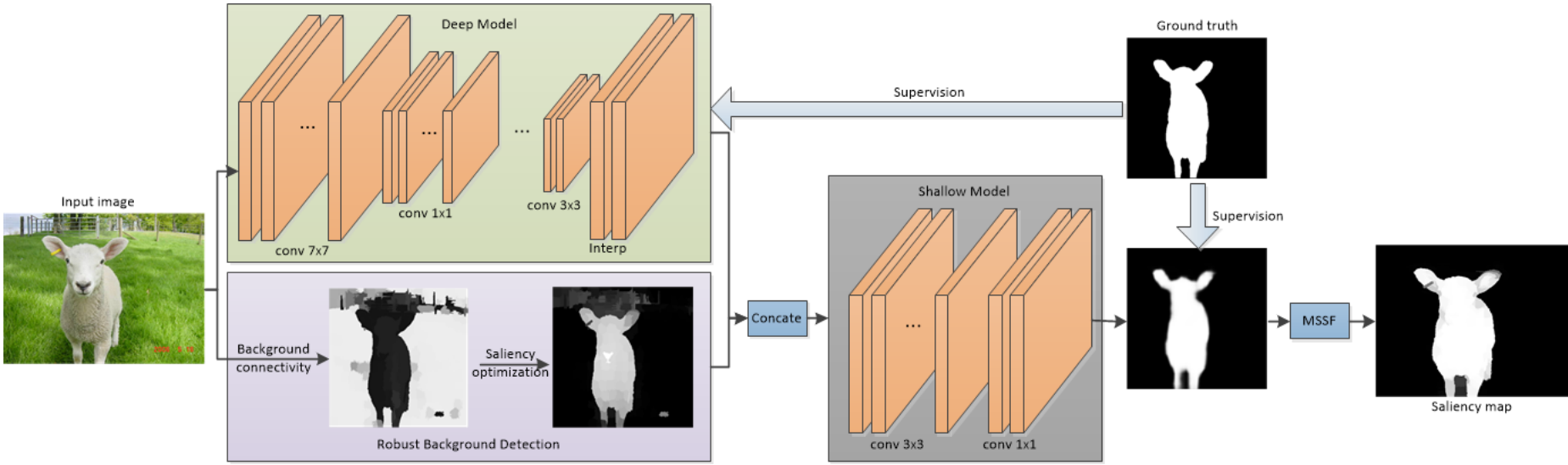


- Whether data-driven (e.g. deep learning) based saliency detection methods sufficiently exploit statistical information?
- Whether unsupervised saliency and data-driven saliency can be combined to achieve even better performance?

Motivation

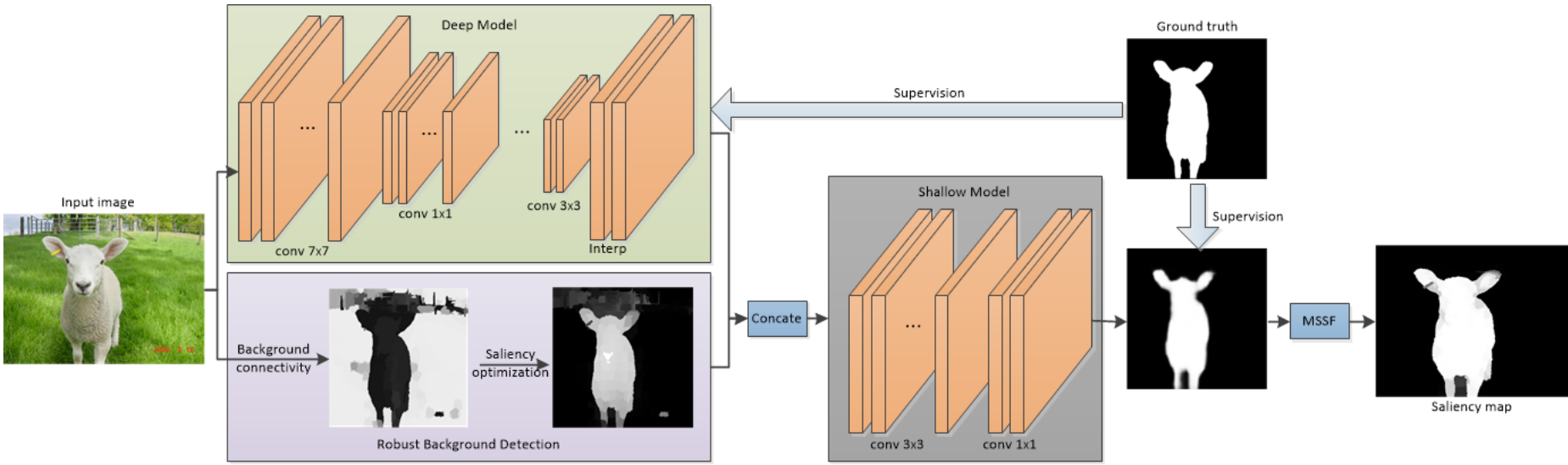
- Deep features can be a double-edged sword:
  - Deep features provide high-level semantic cues critical for saliency detection, however
  - Structure information may be neglected in high-level deep features,
  - Existing FCNN based deep saliency methods cannot incorporate handcrafted prior knowledge,
  - Feature maps from FCNN are usually blurred around edges.

Integrating deep features and handcrafted features



Given an input image, our deep model produces a coarse saliency map. Then a shallow model integrates deep saliency and handcrafted saliency. Finally, a multi-scale superpixel level fusion (MSSF) obtains a spatially coherent saliency map.

Fully convolutional neural networks for saliency detection



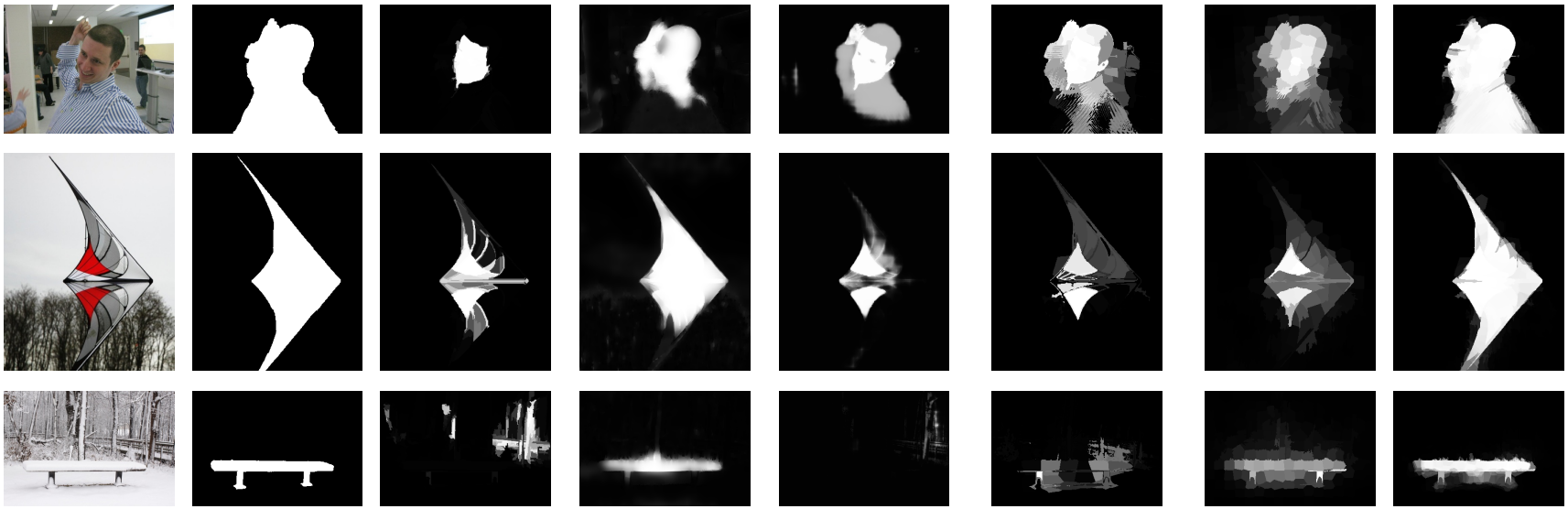
- Finetune an FCNN [Chen, 2016] [He, 2016] with dilated convolutional layers for semantic segmentation to adapt it to salient object detection.
- 3,000 images from the MSRA10K for training.

Multi-scale superpixel level fusion

Steps for multi-scale superpixellevel fusion:

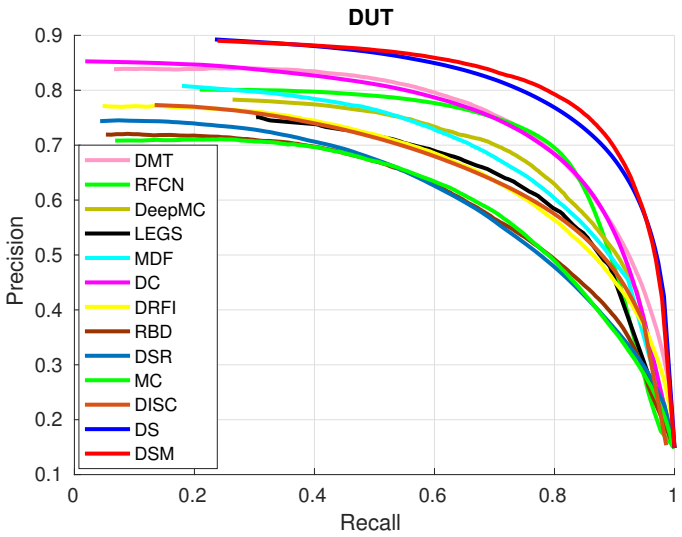
- SLIC for image over-segmentation $X = \{X_1, X_2, \cdots, X_N\}$, where $N = 100, 200, 300, 400$ to achieve multi-scale image over-segmentation;
- Per-superpixel saliency map S_k, $k = 1, 2, 3, 4$ where saliency value of each superpixel is defined as median saliency prediction score of saliency map from our deep-shallow model S_{DS};
- Saliency fusion: $S_{DSM} = \sum S_k$

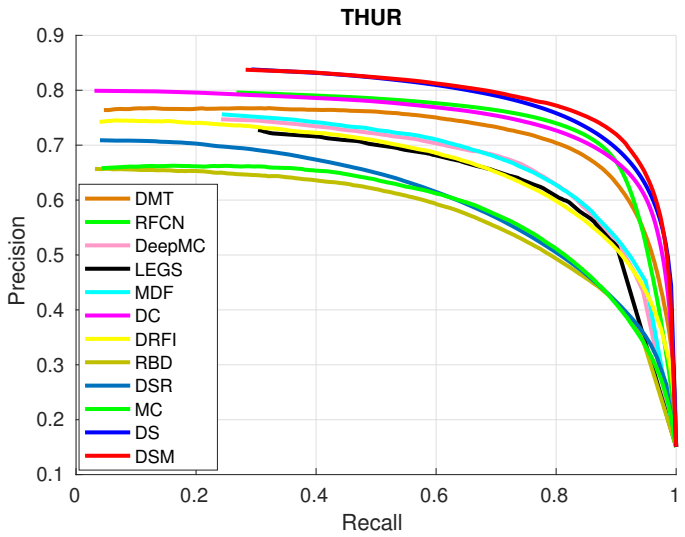
Experimental results

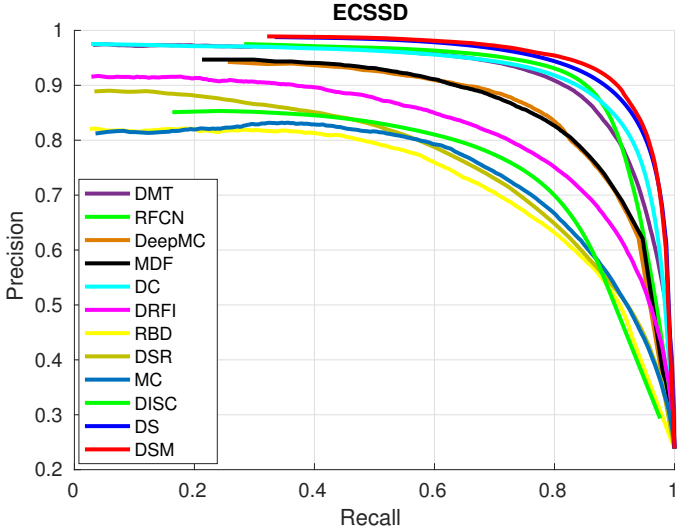


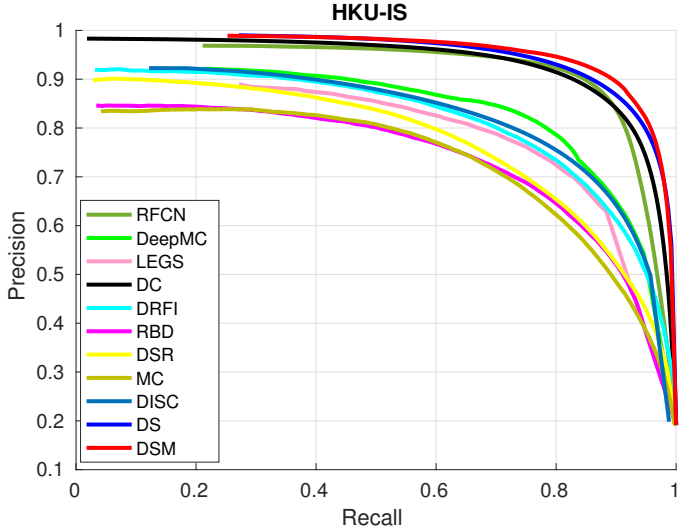
Salient object detection results on challenging images by different methods

Experimental results

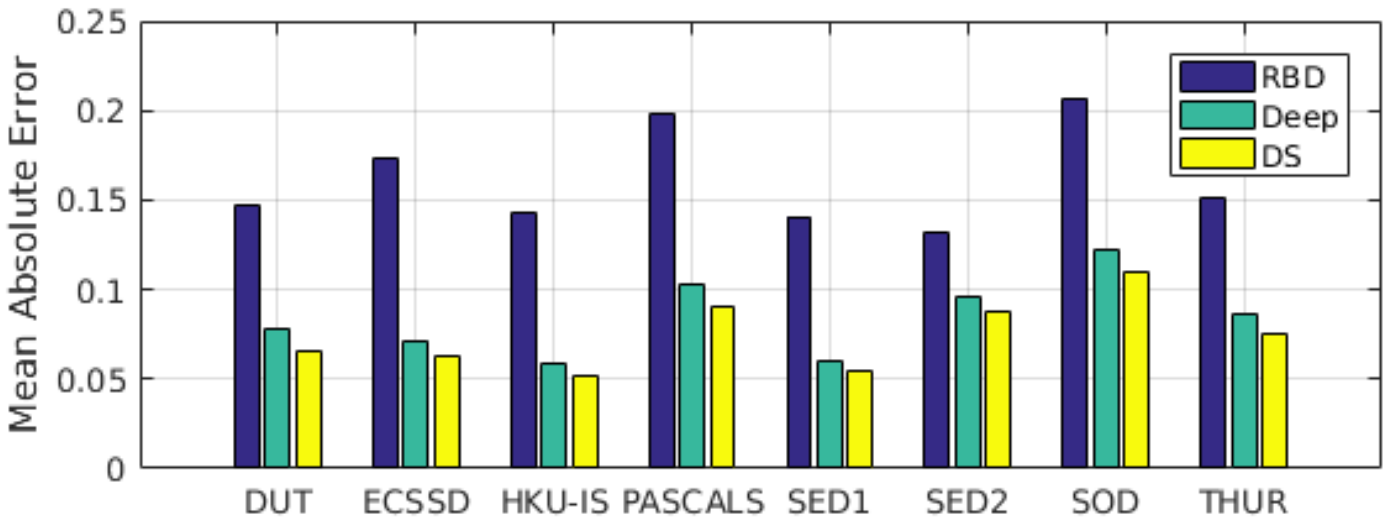








Model Analysis



MAE on eight benchmark datasets.

Conclusion

- An end-to-end FCNN based approach for saliency detection
- Multi-level superpixel level saliency fusion to enhance saliency maps
- Small and relatively simple training dataset with state-of-the-art performance
- Efficient for saliency prediction in testing stage, 0.4 sec per image with 0.2 sec for image over-segmentation.

Key references


L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," arXiv, 2016


K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition," CVPR 2016

Thanks!