Deep Networks for Saliency Detection via Local Estimation and Global Search

Lijun Wang[1], Huchuan Lu[1], Xiang Ruan[2] and Ming-Hsuan Yang[3]
[1]Dalian University of Technology. [2]OMRON Corporation. [3]University of California at Merced.

Existing saliency detection methods are mainly formulated as computational models in a bottom-up fashion with either a local or a global view. Local methods [2, 3], which compute center-surround differences in a local context to capture the regions locally standing out from their surroundings, often lack global information and tend to highlight the boundaries of salient objects rather than the interiors. Global methods [1, 8] take the entire image into consideration and can uniformly assign saliency values to the contained regions. However, they are less effective when the textured regions of salient objects are similar to the background.

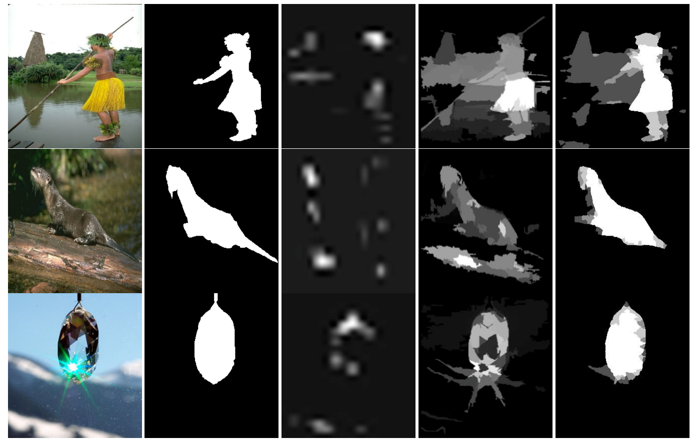
In order to address the above drawbacks, this paper presents a novel saliency detection algorithm by applying deep neural networks (DNNs) in both local estimation and global search (LEGS). Our methods provides two important insights. First, local features learned by a supervised scheme can effectively capture local contrast, texture and shape information for saliency detection. Second, the complex relationship between different global saliency cues can be captured by deep networks and exploited principally rather than heuristically

In the local estimation stage, we formulate a binary classification problem to determine whether each pixel is salient (1) or non-salient (0) based on its surrounding. We use a deep network, named as **DNN-L**, to conduct classification, since DNNs do not rely on hand-crafted features. The proposed **DNN-L** consists of six layers, with three convolutional layers and three fully connected layers. Each layer is followed by a non-linear mapping which is implemented by ReLUs. Max pooling is applied to all the three convolutional layers for translational invariance. The dropout procedure is used after the first and the second fully connected layers to avoid overfitting. The network takes a RGB image patch of 51×51 pixels as an input, and exploits a softmax regression model as the output layer to generate the probabilities of the central pixel being salient and non-salient. At test time, **DNN-L** is applied to the entire input image in a sliding window fashion to predict the foreground probability of each pixel and achieves the local saliency map. In order to maintain spatial consistency and enforce object level concepts, the local saliency maps are further refined by geodesic object proposal (GOP) [4] method. A subset of object candidates are selected from all the generated object segments based on the local saliency maps. The selected object candidates have high probabilities to be the potential objects. The refined local saliency map is generated by averaging the selected object masks.

In the global search stage, a 72-dimensional global feature vector is proposed to globally characterize each candidate region produced by the GOP method. The global feature vector covers global contrast, geometric information, and local saliency measurements evaluated by the local estimation. Another deep network, namely **DNN-G**, is developed to map the global features to the saliency value of each candidate region through a supervised learning scheme. The proposed **DNN-G** consists of six fully connected layers, each of which is followed by a ReLU and a dropout layer. Taken the global feature vector of the i-th candidate region, the output of **DNN-G** is a two dimensional vector $[\phi_i^1, \phi_i^2]$, where ϕ_i^1 and ϕ_i^2 predict the precision and overlap rate of the candidate region, respectively, according to the ground truth saliency map. The global confidence score of the region is defined as $conf_i^G = \phi_i^1 \times \phi_i^2$. Denote $\{\hat{\mathbf{O}}_1, \ldots, \hat{\mathbf{O}}_N\}$ as the mask set of all the candidate regions in the input image sorted by the global confidence scores in a descending order. The corresponding global confidence scores are represented by $\{conf_1^G, \ldots, conf_N^G\}$. The final saliency map is computed by a weighted sum of the top K candidate masks,

$$\mathbf{S}^G = \frac{\sum_{k=1}^{K} conf_k^G \times \hat{\mathbf{O}}_k}{\sum_{k=1}^{K} conf_k^G}. \tag{1}$$

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.



(a) (b) (c) (d) (e)

Figure 1: Saliency detection by different methods. (a) Original images. (b) Ground truth saliency maps. (c) Saliency maps by a local method [2]. (d) Saliency maps by a global method [1]. (e) Saliency maps by the proposed method.

Both the **DNN-L** and **DNN-G** network are trained on a subset of the MSRA-5000 [6] and the PASCAL-S [5] data set using stochastic gradient descent with momentum. Horizontal reflection and rescaling (5%) are applied to all the training images to augment the training data set. We evaluate the proposed algorithm on four benchmark data sets: MSRA-5000, SOD [7], ECCSD [9] and PASCAL-S, against ten state-of-the-art method. Our method performs favorably against the state-of-the-arts in terms of PR curves, F-measure as well as MAE scores.

[1] Ming-Ming Cheng, Guo-Xin Zhang, Niloy J Mitra, Xiaolei Huang, and Shi-Min Hu. Global contrast based salient region detection. In *CVPR*, pages 409–416, 2011.

[2] Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *PAMI*, 20(11):1254–1259, 1998.

[3] Dominik Alexander Klein and Simone Frintrop. Center-surround divergence of feature statistics for salient object detection. In *ICCV*, pages 2214–2219, 2011.

[4] Philipp Krähenbühl and Vladlen Koltun. Geodesic object proposals. In *ECCV*, pages 725–739. 2014.

[5] Yin Li, Xiaodi Hou, Christof Koch, James M. Rehg, and Alan L. Yuille. The secrets of salient object segmentation. In *CVPR*, pages 280–287, 2014.

[6] Tie Liu, Zejian Yuan, Jian Sun, Jingdong Wang, Nanning Zheng, Xiaoou Tang, and Heung-Yeung Shum. Learning to detect a salient object. *PAMI*, 33(2):353–367, 2011.

[7] Vida Movahedi and James H Elder. Design and perceptual validation of performance measures for salient object segmentation. In *POCV*, pages 49–56, 2010.

[8] Federico Perazzi, Philipp Krahenbuhl, Yael Pritch, and Alexander Hornung. Saliency filters: Contrast based filtering for salient region detection. In *CVPR*, pages 733–740, 2012.

[9] Qiong Yan, Li Xu, Jianping Shi, and Jiaya Jia. Hierarchical saliency detection. In *CVPR*, pages 1155–1162, 2013.