Saliency Prediction with Active Semantic Segmentation

Ming Jiang[1]
mjiang@u.nus.edu
Xavier Boix[1,3]
elexbb@nus.edu.sg
Juan Xu[1]
jxu@nus.edu.sg
Gemma Roig[2,3]
gemmar@mit.edu
Luc Van Gool[3]
vangool@vision.ee.ethz.ch
Qi Zhao[1]
eleqiz@nus.edu.sg

[1] Department of Electrical and Computer Engineering
National University of Singapore
Singapore

[2] CBMM, LCSL
Massachusetts Institute of Technology
Istituto Italiano di Tecnologia
Cambridge, MA

[3] Computer Vision Laboratory
ETH Zurich
Switzerland

Semantic-level features have been shown to provide a strong cue for predicting eye fixations. They are usually implemented by evaluating classifiers everywhere in the image. As a result, extracting the semantic-level features may become a computational bottleneck that may limit the applicability of saliency prediction in real-time applications.

To tackle this problem, we show a saliency prediction algorithm that uses active semantic segmentation [3] to exploit semantic-level features in a computationally efficient way (see Figure 1). The active semantic segmentation framework assumes that evaluating classifiers everywhere in the image is computationally more expensive than inferring the semantic labeling from a probabilistic model. Given a budget of time, active semantic segmentation evaluates classifiers in a subset of regions, and propagates the semantic information to the regions that have not been observed. In this paper, we introduce new semantic-level features for saliency prediction based on the probabilistic output from active semantic segmentation. To carry the analysis of our model, we collected eye tracking data on two popular segmentation datasets, the PASCAL VOC07 [1] and the MSRC-21 [4]. Experiments show that the semantic-level features extracted from active semantic segmentation improve the saliency prediction from low- and regional-level features, and it allows controlling the computational overhead of adding semantics to the saliency predictor.

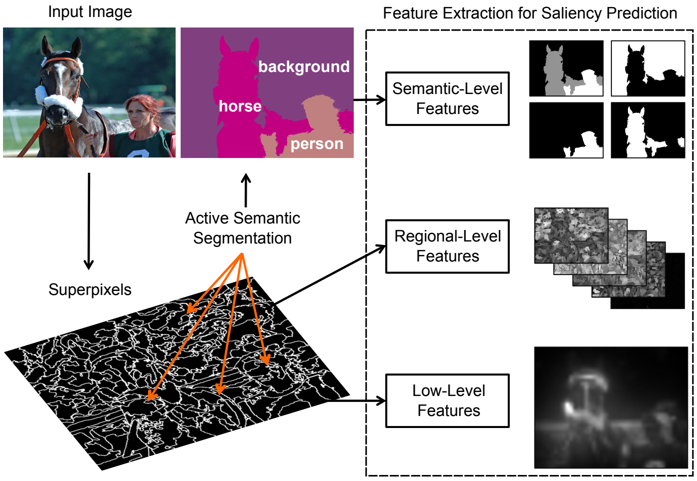

Figure 1: We propose a saliency prediction framework incorporating low-level, regional-level and new semantic-level features based on an active semantic segmentation of the input image. This allows to control the computational overhead of adding semantics to the saliency prediction.

Method

We use a Support Vector Regression (SVR) to learn a feature integration for saliency prediction. Our proposed model incorporates various saliency features at three levels:

Low-Level Features. We incorporate the state-of-the-art GBVS [2] method to calculate a low-level saliency feature, with combined color, intensity and orientation channels.

Regional-Level Features. Suggested by the Gestalt principles, locally coherent regions or proto-objects, are more likely to attract attention than others depending on the size and shape. We use five regional-level features (*i.e. size*, *solidity*, *convexity*, *complexity*, *eccentricity*) proposed by [5], which are independent of the semantics.

Semantic-Level Features. Based on the output of the active semantic segmentation (*i.e.* a set of semantic labelings of the full image that define the probability distribution of the semantics given few observations on the image), we compute a set of semantic-level features, including *label probability*, *semantic uncertainty*, *semantic rarity*, and *object center*.

Experiments and Dataset

We conducted eye-tracking experiments on the PASCAL VOC07 [1] and the MSRC-21 [4] datasets. Evaluation results demonstrated the effectiveness of the semantic features for saliency prediction under computational time constraints (see Figure 2). The eye-tracking data is publicly available at `http://www.ece.nus.edu.sg/stfpage/eleqiz/bmvc15.html`.

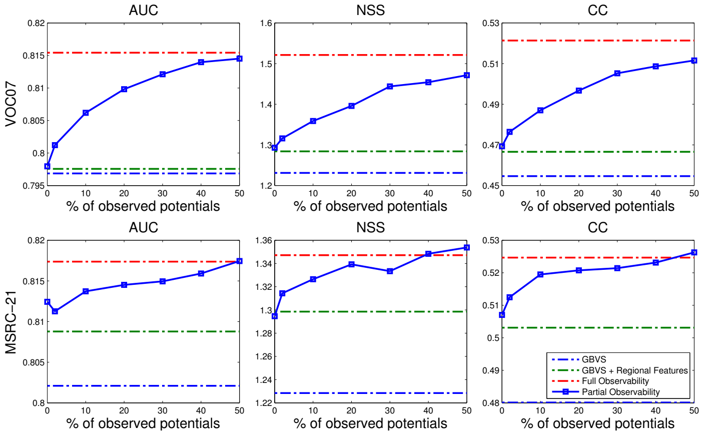

Figure 2: Performance evaluation of saliency prediction in the VOC07 and MSRC-21 datasets, with various percentages of observations. The semantic-level features are able to capture useful semantic information for saliency prediction, given a budget of time.

[1] M. Everingham, L. Van Gool, C. KI Williams, J. Winn, and A. Zisserman. The pascal visual object classes (VOC) challenge. *Int. Journal of Computer Vision*, 88(2):303–338, 2010.

[2] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. In *NIPS*, 2007.

[3] G. Roig, R. Boix, X.and de Nijs, S. Ramos, K. Kühnlenz, and L. Van Gool. Active map inference in crfs for efficient semantic segmentation. In *Proc. IEEE Int. Conf. on Computer Vision*, 2013.

[4] J. Shotton, J. Winn, C. Rother, and A. Criminisi. Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *Int. Journal of Computer Vision*, 2009.

[5] J. Xu, M. Jiang, S. Wang, M. S. Kankanhalli, and Q. Zhao. Predicting human gaze beyond pixels. *J. of Vision*, 14(1):1–20, January 2014.

Dr. Qi Zhao is the corresponding author.