ARTICLE IN PRESS

Neurocomputing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Saliency detection via extreme learning machine

Lu Zhang, Jianhua Li, Huchuan Lu *

School of Information and Communication Engineering, Dalian University of Technology, Dalian 116023, China

ARTICLE INFO

Article history:
Received 15 December 2015
Received in revised form
31 March 2016
Accepted 23 August 2016
Communicated by Wang Qi

Keywords:
Saliency detection
Extreme learning machine
Multi-scale superpixels

ABSTRACT

In this paper, we propose an effective algorithm based on Extreme Learning Machine (ELM) for salient object detection. First, saliency maps generated by existing methods are taken as prior maps, from which training samples are collected for an ELM classifier. Second, the ELM classifier is learned to detect the salient regions, and the final results are generated by fusing multi-scale saliency maps. This ELM-based model can improve the performance of different state-of-the-art methods to a large degree. Furthermore, we present an integration mechanism to take advantages of superiorities of multiple saliency maps. Extensive experiments on five datasets demonstrate that our method performs well and the significant improvement can be achieved when applying our model to existing saliency approaches.

1. Introduction

Visual attention, a remarkable capability of early primate visual system, which helps human complete scene analysis in real time. Inspired by it, saliency detection which aims to identify the most salient regions from background has drawn much attention recently. As one of the most fundamental problems in computer vision, salient object detection has been applied to numerous visual tasks including image segmentation [1], image compression [2], object detection and recognition [3] and so on.

Saliency detection models have been developed in two research areas: eye fixation prediction (e.g., [4–7]) and salient object detection (e.g., [8–11]). Eye fixation prediction, which focuses on predicting human eye fixations on natural images, is helpful to understand human attention. Salient object detection aiming to simultaneously detect the location and the continuous shape of salient object is widely used for many high-level computer vision tasks. In general, saliency detection methods can be grouped into two categories: top-down methods and bottom-up methods. Bottom-up methods [4,8,9,11–19] are rapid, data-driven, task-independent, which construct saliency maps based on low-level visual information such as colors, features and space distance. While top-down approaches [3,20–22] are slower, volition-controlled, task-driven and require supervised learning based on training samples with manual labels. Bottom-up methods are more effective in detecting fine details rather than global shapes of objects based on low-level visual information. In contrast, top-down saliency models, which are usually based on representative features from training samples, are able to detect objects of certain sizes and categories. However, most of the existing top-down learning-based methods [3,23,24] are time-consuming in off-line training process or labeling positive samples manually.

In this paper, we propose an effective algorithm for salient object detection via Extreme Learning Machine (ELM) [25] to address the aforementioned problem. Firstly, an input image is segmented into superpixels, and saliency maps generated by existing saliency methods are regarded as prior maps, from which the training set containing both positive and negative samples are collected. Secondly, a strong classifier based on ELM is learned to measure saliency where three feature descriptors (RGB, CIELab, and the Local Binary Pattern) are extracted to exploit rich feature representations. Furthermore, multi-scale superpixels are used to detect the salient objects of variable sizes. Results from multi-scale saliency maps are integrated to further improve the detection performance. The proposed algorithm (Single-layer Extreme Learning Machine) can optimize the existing saliency methods and achieve favorable results.

For the purpose of taking advantages of different methods, we propose an integration method called Multi-layer Extreme Learning Machine (MELM). With different saliency maps as prior maps, the integrated results generated via Multi-layer Extreme Learning Machine are superior to the original methods.

In summary, the main contributions of our work are as follows:

(1) We propose an effective algorithm based on ELM to optimize existing saliency methods via learning a strong classifier model.
(2) Single-layer Extreme Learning Machine can greatly improve the performance of existing saliency detection methods. Multi-layer Extreme Learning Machine can integrate multiple saliency maps and get a favorable promotion.

* Corresponding author.
E-mail address: lhchuan@dlut.edu.cn (H. Lu).

http://dx.doi.org/10.1016/j.neucom.2016.08.066
0925-2312/© 2016 Published by Elsevier B.V.

Please cite this article as: L. Zhang, et al., Saliency detection via extreme learning machine, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2016.08.066

2. Related work

Numerous salient object detection methods have been proposed in recent years. Most existing methods can be grouped into two categories: contrast-based methods and learning-based methods.

Contrast-based methods: Contrast-based models focus on various appearance between salient objects and background. Itti et al. [4] propose a saliency model based on visual attention, which integrates multiscale image features into a single saliency map for rapid scene analysis. In [26], saliency measure is formulated using a statistical framework and local feature contrast in illumination, color, and motion information. In [11], Jiang et al. use an iterative energy minimization measure to integrate low-level saliency stimuli and shape prior. Compared with the local contrast based saliency methods aforementioned, global contrast based methods evaluate saliency of an image region using its contrast with respect to the entire image. Achanta et al. [8] put forward a fast frequency-tuned method which extracts color and luminance as features to estimate center-surround contrast. In [9], Cheng et al. adopt global contrast strategy as well as spatial relationship to identify region saliency instead of local contrast. A contrast-based saliency filter is presented for measuring saliency by the uniqueness and spatial distribution of regions over an image [16]. Fang et al. [27] come up with a novel saliency detection model in the compressed domain and measure saliency by Hausdorff distance calculation and feature map fusion. In [28], a soft abstraction approach is used to remove unnecessary image details and generate perceptually accurate salient regions. In [29], Wang et al. propose a multi-spectrum based saliency detection algorithm and novelly incorporate near-infrared clues into the detection framework.

While the above-mentioned contrast-based methods are effective, difficulty still remains in distinguishing salient objects from background accurately without discriminative prior knowledge. In [17], a unified model for saliency detection is constructed which combines low-level features and high-level priors based on low rank matrix recovery theory. A hierarchical model for saliency detection is put forward to deal with the scale problem [30]. In [31], Tian et al. propose a novel salient model which adopts color contrast and orientation contrast to calculate the bottom-up feature maps and top-down cue of depth-from-focus from the same single image to guide the generation of final salient regions. Recently, Wang et al. [32] present a contrast-based saliency model which takes both background and foreground into consideration.

Learning-based methods: Methods in this category are constructed based on machine learning, where high-level information and supervised methods are incorporated to improve the accuracy of saliency maps. Judd et al. [23] propose an eye fixation prediction model, where a linear Support Vector Machine (SVM) classifier are trained based on a dataset containing fixation locations of 15 experimenters. In [3], a binary saliency estimation model is proposed where a conditional random field is trained to combine a set of novel features. Yang et al. [20] present a model that learns a Conditional Random with latent variables as well as a discriminative dictionary for saliency detection. In [33], a multiple-instance learning model is presented where low-, mid- and high-level features are incorporated in the detection procedure. Jiang et al. [22] formulate saliency detection as a regression problem and learn a regression forest which maps the descriptors to saliency scores. In [34], Lu et al. put forward a method to learn optimal seeds which are then propagated using a diffusion process. Tong et al. [35] propose a salient object detection method via bootstrap learning, instead of training classifier in a large dataset, they train a set of weak SVM classifiers based on initial maps and then a strong classifier is obtained by integrating the weak classifier via multi-kernel boosting algorithm.

Extreme learning machine: Extreme Learning Machine (ELM), which is first put forward by Huang et al. [36,37], is a simple learning algorithm whose learning speed is fast and generalization performance is good. It has been applied in many image processing tasks [38–40]. ELM algorithm can be conducted in two learning processes: first, input data are mapped into the hidden layer by either random feature mapping or kernel learning. Second, final output are obtained by multiplying the middle results with corresponding output weights. In the learning process of traditional three-layer neural networks, all the parameters need to be tuned iteratively and thus there exists the dependency between different layers of parameters. Different from that, the hidden layer in ELM is not parametric and need not be tuned. Besides, ELM not only tends to reach the smallest training error but also the smallest norm of weights, which are both important for achieving better generation performance. In a word, ELM could provide satisfying performance at extremely fast learning speed [37].

Compared with SVM, ELM requires fewer optimization constraints and results in simpler implementation, faster learning and better generalization performance. Furthermore, ELM is able to provide a unified solution to different practical applications and has milder optimization constraints compared to SVM, though they have the same cost function [25]. In summary, ELM performs better and faster than SVM generally in classification.

3. Structure of proposed algorithms

The main steps of our algorithm are shown in Fig. 1. We first segment the input image into multi-scale superpixels, and the saliency maps of existing saliency detection methods are regarded

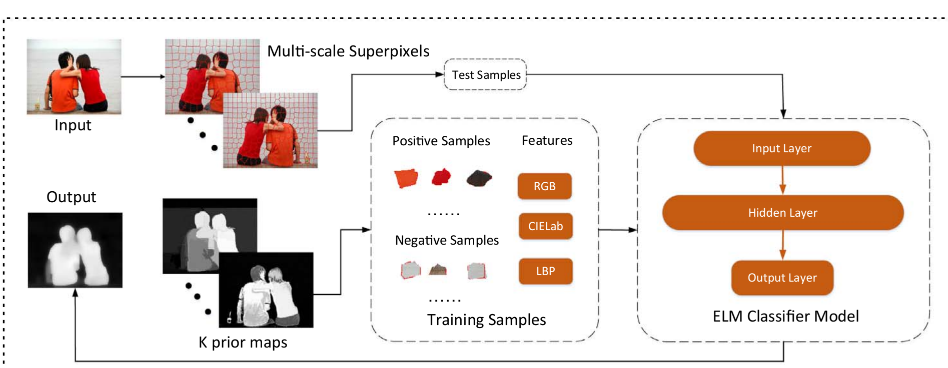

Fig. 1. Structure of salient object detection via ELM. After segmenting an input image into multi-scale superpixels, training samples for a strong classifier are collected from prior maps at each scale. A strong classifier based on ELM is learned to detect salient objects where three feature descriptors are extracted to exploit rich feature representations. The final saliency map is generated by linear combining saliency results at multiple scales.

Please cite this article as: L. Zhang, et al., Saliency detection via extreme learning machine, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2016.08.066