J. Vis. Commun. Image R. 55 (2018) 789–794

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Deep network for visual saliency prediction by encoding image composition ☆

Bo Dai *, Weijing Ye, Jing Zheng, Qianyi Chai, Yiyang Yao

State Grid Zhejiang Electric Power Company, Information & Telecommunication Branch, China

ARTICLE INFO

Article history:
Received 25 June 2018
Revised 6 August 2018
Accepted 14 August 2018
Available online 20 August 2018

Keywords:
Deep network
Visual saliency regions
Image composition

ABSTRACT

This article will be visual significance into the graphical guidance (the chart is a medium-sized join subgraph). Deep structure, from the level of learning a significant map. The original image pixel to the object level graphic (oGL), and further Space level graphics (sGL). In particular, we first sample Super pixels from each image, and they are used as buildings Block of each object. In order to seamlessly describe different objects, the number of oGLs is generated by spatial adjacent links. The super pixel oGL object response mapping is obtained by obtaining, Transfer, the semantics of the image tag to oGL. As space, the layout of the object plays an important role in the prominence of the object based on the relevant learning distribution proposed sGL OGL position between. Finally, in order to imitate the "winner of all" Biological vision mechanism, the largest majority of voting programs, The sGL of the image, is probabilistically combined into a significant graph. Experimental results show that oGLs and sGLs capture the object level well and space-level visual cues, resulting in competitiveness significant detection accuracy.

1. Introduction

Effectively dealing with a large number of visual information in nature, The world, the human visual system has caused a significant mechanism Called a visual attention to choose a small part Follow-up information about the process, such as the object Classification and vehicle tracking [45–50]. To imitate the choice Mechanism, many computer vision attention models have been raised and can be used to improve many computer vision and Image processing tasks [30,15,38]. However, the computational visual significance is still a challenging issue for the following reasons Fig. 1:

- Psychophysics Researchers [12] show object-level hints and predominant attention; more observers May consider "interesting" objects, not height compared with nonobject areas. But the current calculation, The ability to model the model, is limited Object level prompt. A common way to integrate object levels, The prompt, is to add the object detector, but it does not extend well Unreliable.
- Objects between the spatial layouts affect their Significant values ??were significant, but the existing saliency models were not available and clearly reflected. As shown in Fig. 2 The observers experience, the relative position between the sky, the sailboat, while Siemens made sailing and Siemens significant (c). And if so Sky water level reversed, sky and water The sailing and water doors in the most conspicuous way Area (d).
- Object, The existing object development model, usually contains several manual visual features. This one and human visual neural mechanism does not match System, where only the original image pixels are used as input. The saliency map is calculated by the multilayer process in human vision Cortex.

To solve three questions, inspiration from the profound hierarchical learning of multi-level image architecture [8] Abstract visual task, we propose a model to learn the image of the different layers of the map. As shown in Fig. 2, the deep architecture starts from the beginning and the image associated with its semantic tag [37,32]. To seamlessly describe the different objects in the image, we generate oGLs probably random walk in space [7,27]. The image is super-pixel and further transmits the image tag OGL which is used to calculate its object-level response mapping. As the spatial layout between objects significantly affects them significantly, we present sGL, where each node represents the distribution of semantic objects is relative to each edge measure the displacement between objects. Given a new image, we extract all the sGLs and further probably integrate them into the final significant graph.

The contribution of this work can be summarized as follows: (1) profound visual attention to the learning frame is significant and

☆ This article is part of the Special Issue on TIUSM.

* Corresponding author.
E-mail address: 553641039@qq.com (B. Dai).

https://doi.org/10.1016/j.jvcir.2018.08.010
1047-3203/

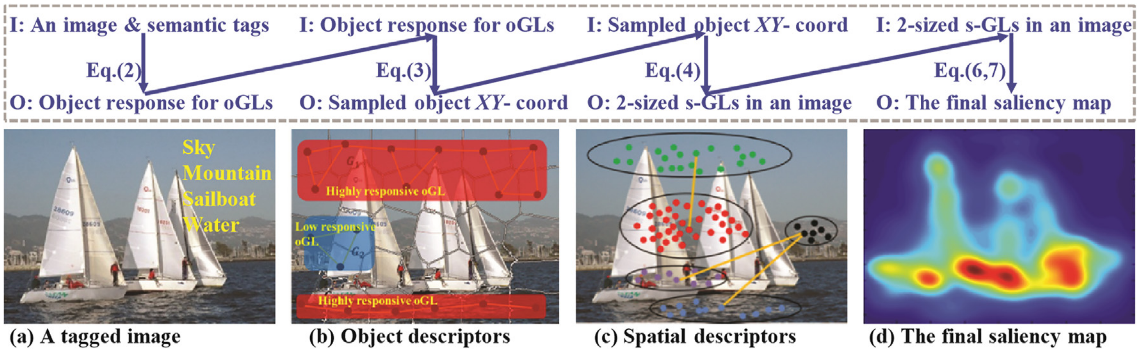

Fig. 1. Our deep architecture flow chart can be decomposed into four limited Boltzmann machines (RBM). The blue text represents the input (I) and output (O) of each RBM, and the blue arrow indicates the sampling process. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

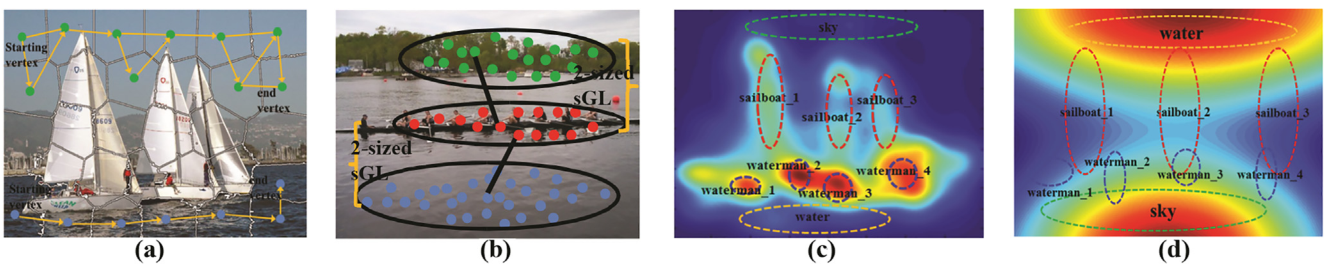

Fig. 2. An example of generating oGLs by probabilistically random walking (a), obtaining two 2-sized sGLs by sampling highly responsive oGLs (b), and the influence of spatial interaction among objects on their saliency (c and d).

the images are hierarchically calculated according to different types chart, (2) object level and space level charts describe the object and its spatial distribution separately using our significance model, and (3) the enhanced first saliency model using semantic significance to detect accuracy Image label.

2. Related work

A number of computationally significant models have been proposed recent.[1] Among them, Bayesian theorem is dominant and the method is closely related to the method based on the graphical model, the proposed method [34,35]. Probabilistic Bayesian Rule Model combines multiple visual cues, such as center base and statistics the natural scene into a significant pattern [51–54]. In citebay [10], Borji et al. simulates top-down visual effects under the influence of different tasks. Multiple factors such as scene background, physical behavior and bottom-up significance integrate through popular statistical tools such as kNN or a conventional Bayesian network (BN). Using more effective integration of visual cues, Borji et al. [6] further designed a new BN to describe task-driven visual attention, where the global scene context, previously participated in the location and car operations has been integrated together time to predict the next participating position. As described above Einhaser et al. [12], and low-level visual features, the object is more suitable for the foundation highlights the detection feature. Therefore, Borji et al. [5] present an object-level task-driven dynamic and the Bayesian Network (DBN) is used to predict human eye movement and they use the auxiliary oracle database to detect the object. In [22], Li et al. made a probability a multitasking framework for estimating visual significance. Low level visual features and task-related factors, textit i.e., are a scene of the "stimulating significant" function combined to form a final significant figure.

Indent, as a generalized Bayesian rule, the graphical model represents the conditional dependency structure between random variables. Thereby, based on the significance of the graphical model, we can provide a unified framework used to capture the complex dependencies between image areas. In [25], Liu et al. developed a prominent object detection as an image segmentation, where the conditional random field (CRF) [23] is configured to highlight the detection as one binary mark process. In Harel et al. [16] a well-known graphics-based prominence model (GBVS), a fully connected graphic model, constructs all grid locations for the image. Based on that, its surrounding height is not similar to the vertex and it is denoted as A; large value is assigned, and the activation map is integrated into the final significant figure. In [3] Avraham et al. proposed a probability measure regional significance, which is predicted according to preferences, the number of "interesting" objects in the scene. Especially, they propose a BN to represent the different label dependencies segmented area and then a small part of the most likely joint select the assignment to calculate the significance of each area.

In this paper, we introduce a method for visual saliency prediction. Visual saliency prediction now attracts many researchers since it is an indispensable feature in computer vision. It can reveal human visual mechanism. There is no doubt that leveraging human visual mechanism into computer vision is a great progress. What's more, visual saliency can be used in many intelligent systems, such as image/video retarget, image/video scene categorization.

3. The proposed approach

3.1. Object-level graphlets to seamlessly construct an object

As mentioned above, the significance of the object is below the low level visual features, so it is necessary in our coding significance model. We use super pixels as the basic element to construct a different object because: (1) the super pixel is the pixel, a cluster

[1] More comprehensive review of the significance the model can be found in Borji *et al.* Found in T-PAMI [4].