Deep learning for saliency prediction

Where: Rennes, Sirocco, IRISA – University of Rennes 1

Contacts :

Olivier Le Meur - Associate Professor
ESIR - University of Rennes 1
olemeur@irisa.fr
http://people.irisa.fr/Olivier.Le_Meur/

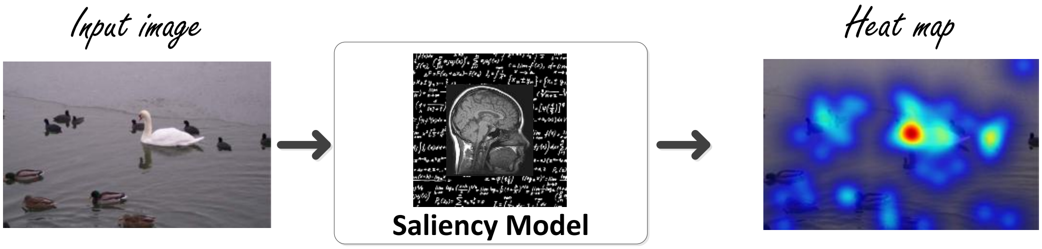
Keywords: saliency, eye fixation, machine learning and deep learning

Context:

Our visual environment contains far more information than we are able to process at one time. To deal with this large amount of data, we must focus our visual processing resources on relevant visual information. This strategy is called the selective visual attention. There are two kinds of visual attention: one involves eye movements (overt orienting) whereas the other occurs without eye movements (covert orienting). Most research activities relevant to visual attention have dealt with the understanding and modelling of overt attention. Two well recognized attentional mechanisms control overt visual attention. Bottom-up mechanism refers to the ability of an area to attract our attention unconsciously and effortlessly. It relies on the low-level characteristics of visual stimuli, such as color, luminance, texture, motion, to name a few. The bottom-up guidance source is classically represented by a saliency map which indicates the most visually interesting parts of our visual field. At the opposite, there are top-down contributions which encompass a number of factors. Top-down contributions are obviously related to the observers' goals but also to the prior knowledge, motivations, mood and experience of observers.

Computational modelling of visual attention:

The computational modelling of visual attention consists in predicting where an observer look within a scene. This prediction is mainly based on low-level visual features. There exist now a number of models. A recent taxonomy has been proposed by [Borji & Itti, 2013]. The figure below illustrates the main objective of a saliency model. From an input image, a heat (or saliency) map is computed. It indicates where the most salient areas are located within the scene. Blue color corresponds to non-salient areas whereas red color represents the most salient areas.



A new breakthrough: saliency prediction based on convolutional networks: **A VERY HOT TOPIC IN COMPUTER VISION**

The rise of deep learning methods has shown that many difficult computer vision problems can be solved by machine learning algorithms and more specifically by Convolution Neural Networks (CNNs).

When applied on images, CNNs consist of multiple layers of small neuron collections which process portions of the input image. The convolutional layer is the core building block of a CNN. The layer's parameters consist of a set of learnable kernels; those weights are learned during the back-propagation step, which aims to reduce the predictor error, i.e. the difference between the prediction and the actual value.

A number of deep learning-based saliency model has been very recently proposed. Some references are given below:

> Kruthiventi, S. S., Ayush, K., & Babu, R. V. (2015). ***Deepfix: A fully convolutional neural network for predicting human eye fixations***. *arXiv preprint arXiv:1510.02927*.

Jetley, S., Murray, N., & Vig, E. (2016). ***End-to-End Saliency Mapping via Probability Distribution Prediction***. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 5753-5761).

Pan, J., McGuinness, K., Sayrol, E., O'Connor, N., & Giro-i-Nieto, X. (2016). ***Shallow and Deep Convolutional Networks for Saliency Prediction***. *arXiv preprint arXiv:1603.00845*.

Vig, E., Dorr, M., & Cox, D. (2014). ***Large-scale optimization of hierarchical features for saliency prediction in natural images***. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2798-2805).

Jiang, M., Huang, S., Duan, J., & Zhao, Q. (2015, June). ***SALICON: Saliency in context***. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1072-1080). IEEE.

Objectives:

The objective of the proposed internship is to propose **a new deep-learning-based model for predicting eye fixations**. This objective can be split into the following subtasks:

1. **Bibliography**. The candidate will have to understand how deep learning models work, what they can be used for, how they are trained, and their strengths and weaknesses. Note that there are many lectures and tutorials on the web (for instance: http://cs231n.github.io/, https://en.wikipedia.org/wiki/Convolutional_neural_network ,...).
2. **Caffe framework**. The candidate will use the Caffe framework (http://caffe.berkeleyvision.org/) to design its own model and to perform the training. Before designing his own model, the candidate will test several models proposed on Caffe webpage.
3. **Application to eye-fixation prediction**.
4. **Evaluation** of the model and comparison with existing methods (see http://saliency.mit.edu/index.html benchmark web site)

References:

- Saliency models:

  [Borji & Itti, 2013] Borji, A., & Itti, L. (2013). ***State-of-the-art in visual attention modeling***. *IEEE transactions on pattern analysis and machine intelligence*, *35*(1), 185-207.

  [Le Meur & Liu, 2015] Le Meur, O., & Liu, Z. (2015). ***Saccadic model of eye movements for free-viewing condition***. *Vision research*, *116*, 152-164.

- Deep learning :

  [LeCun et al., 2015] LeCun, Y., Bengio, Y., & Hinton, G. (2015). ***Deep learning***. *Nature*, *521*(7553), 436-444.

  [Krizhevsky et al., 2012] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ***Imagenet classification with deep convolutional neural networks***. In *Advances in neural information processing systems* (pp. 1097-1105).