A Predictive Model for Eye Fixations

Lauren Arnett and Chengzhi Mao
Columbia University in the City of New York
{lba2138,cm3797}@columia.edu

Abstract

Applications of tracking eye fixation location span from neuroscience and the study of human vision to advertising and human computer interaction. By creating a model to generate a prediction of where the eye may be most attracted to, industries can circumvent the expense of running studies on eye-tracking with actual human subjects. We look to improve upon existing models of saliency by using Bayesian methods with deep learning techniques.

1. Introduction

It is estimated that 80% of all external sensory input processed by the human brain is processed by the visual pathway [1]. As such, optimizing image layout for processing by the human brain allows for better information retrieval and retention across the image. Studying how humans' eyes move across images is thus relevant for fields from neuroscience to advertising and art. Being able to predict where humans are most likely to look provides a guideline as to whether the image has an effective layout, what humans are attracted to in viewing art, or how an image should be cropped to feature the subject. Using an existing dataset of eye movements, we build a predictive model to generate the most likely fixation locations on a new image.

2. Related Work

Traditionally, studies on eye movements have been carried out such that viewers look at images on a monitor while an eye tracker records the eye-fixations that stay within a threshold angle of movement. This procedure is very costly, and necessitated formulating a method to predict where users will look. Thus, models of saliency—the likelihood of a location to attract the visual attention of a human—developed that are modeled mathematically using biologically plausible linear filters. For example, linear combinations of filters for low-level features such as color, intensity, and orientation filters can be used to compute a total saliency map for an image, providing a bottom-up understanding of the image [2].

These models do not account for particular tasks that the viewer may have in looking at the image, and often do not align with the ground truth fixation locations. Judd *et al*. [3] propose using deep learning for this task rather than deriving mathematical models and show that training from a large database of eye-tracking data outperforms existing models. Kümmerer *et al*. [4] also employ deep learning for predicting fixation locations, using the AlexNet architecture in *DeepGaze I* and *DeepGaze II*, built upon the VGG-19 network.

Developing a dataset for use by saliency models is also a field of exploration. Judd *et al*. [3] creates a dataset of 1003 images with the fixation locations from 15 viewers each and makes it publicly available. Jiang *et al*. [5] relies on an assumption of eye-mouse coordination—they simulate recording eye-tracking data with instead recording mouse-tracking data using Amazon Mechanical Turk. This provides a less expensive and training-intensive method of developing a simulated saliency map. Finally, the MIT300 dataset [6] looks to provide a performance benchmark for new predictive models for saliency, with performance statistics for over 80 models at the time of writing.

3. Dataset of Eye-Tracking Data

4. Learning a Model

4.1. Training

4.2. Performance

5. Conclusion

References

[1] R. Jerath, M. W. Crawford, V. A. Barnes. Functional representation of vision within the mind: A visual consciousness model based in 3D default space. *Journal of Medical Hypotheses and Ideas*, 9(1):45-56, 2015.

[2] L. Itti and C. Koch. A Saliency-Based Search Mechanism for Overt and Covert Shifts of Visual Attention. *Vision Research*, 40(10-12):1489–1506, 2000.

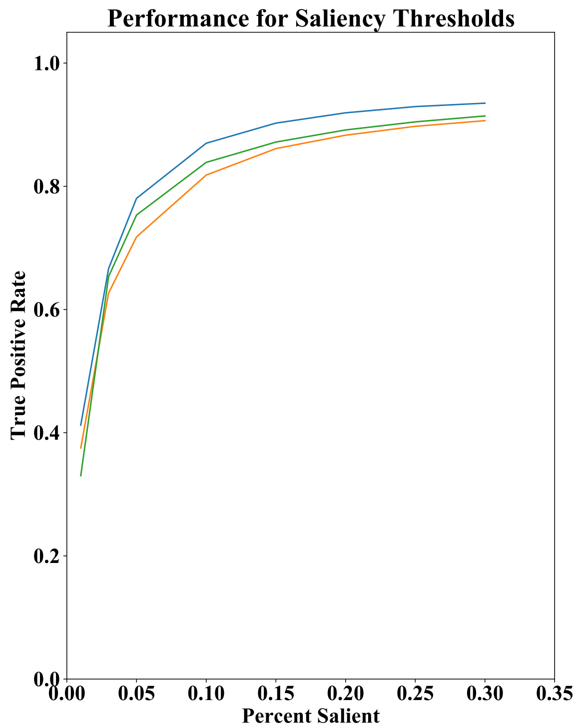

Figure 1. Example of a short caption, which should be centered.

[3] T. Judd, K. Ehinger, F. Durand, and A. Torralba. Learning to predict where humans look. *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 2106–2113, 2009.

[4] M. Kümmerer, T. S. A. Wallis, L. A. Gatys, M. Bethge. Understanding Low- and High-Level Contributions to Fixation Prediction. *The IEEE Int. Conf. Comput. Vis.*, pp. 4789-4798, 2017.

[5] M. Jiang, S. Huang, J. Duan, Q. Zhao SALICON: Saliency in Context. *CVPR*, pp. 1072-1080, 2015.

[6] Z. Bylinskii, T. Judd, A. Borji, L. Itti, F. Durand, A. Oliva, and A. Torralba. MIT Saliency Benchmark.

[7] N. Wilming, S. Onat, J. Ossandón, A. Acik, T. C. Kietzmann, K. Kaspar, R. R. Gamiero, A. Vormberg, P. König. Data from: An extensive dataset of eye movements during viewing of complex images. `https://doi.org/10.5061/dryad.9pf75`. *Dryad Digital Repository*, 2017.

[8] N. Wilming, S. Onat, J. Ossandón, A. Acik, T. C. Kietzmann, K. Kaspar, R. R. Gamiero, A. Vormberg, P. König. An extensive dataset of eye movements during viewing of complex images. `https://doi.org/10.1038/sdata.2016.126`. *Nature Scientific Data*, 4(1):160126, 2017.