



# Modeling Human Visual Search in Natural Scenes: A Combined Bayesian Searcher and Saliency Map Approach

Gaston Bujia<sup>1,2†</sup>, Melanie Sclar<sup>1†‡</sup>, Sebastian Vita<sup>1</sup>, Guillermo Solovey<sup>2</sup> and Juan Esteban Kamienkowski<sup>1,3\*</sup>

<sup>1</sup> Laboratorio de Inteligencia Artificial Aplicada, Instituto de Ciencias de la Computación, Universidad de Buenos Aires – CONICET, Ciudad Autónoma de Buenos Aires, Argentina, <sup>2</sup> Instituto de Cálculo, Universidad de Buenos Aires – CONICET, Ciudad Autónoma de Buenos Aires, Argentina, <sup>3</sup> Maestría de Explotación de Datos y Descubrimiento del Conocimiento, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Autónoma de Buenos Aires, Argentina

## OPEN ACCESS

### Edited by:

Adrian G. Palacios,  
Universidad de Valparaíso, Chile

### Reviewed by:

Thierry Viéville,  
Institut National de Recherche en  
Informatique et en Automatique  
(INRIA), France  
Tom Foulsham,  
University of Essex, United Kingdom  
Emmanuel Dauce,  
Centrale Marseille, France

### \*Correspondence:

Juan Esteban Kamienkowski  
juank@dc.uba.ar

### †Present address:

Melanie Sclar,  
Paul G. Allen School of Computer  
Science & Engineering, University of  
Washington, Seattle, WA,  
United States

‡These authors have contributed  
equally to this work

Received: 23 February 2022

Accepted: 26 April 2022

Published: 27 May 2022

### Citation:

Bujia G, Sclar M, Vita S, Solovey G  
and Kamienkowski JE (2022)  
Modeling Human Visual Search in  
Natural Scenes: A Combined  
Bayesian Searcher and Saliency Map  
Approach.  
Front. Syst. Neurosci. 16:882315.  
doi: 10.3389/fnsys.2022.882315

Finding objects is essential for almost any daily-life visual task. Saliency models have been useful to predict fixation locations in natural images during a free-exploring task. However, it is still challenging to predict the sequence of fixations during visual search. Bayesian observer models are particularly suited for this task because they represent visual search as an active sampling process. Nevertheless, how they adapt to natural images remains largely unexplored. Here, we propose a unified Bayesian model for visual search guided by saliency maps as prior information. We validated our model with a visual search experiment in natural scenes. We showed that, although state-of-the-art saliency models performed well in predicting the first two fixations in a visual search task (90% of the performance achieved by humans), their performance degraded to chance afterward. Therefore, saliency maps alone could model bottom-up first impressions but they were not enough to explain scanpaths when top-down task information was critical. In contrast, our model led to human-like performance and scanpaths as revealed by: first, the agreement between targets found by the model and the humans on a trial-by-trial basis; and second, the scanpath similarity between the model and the humans, that makes the behavior of the model indistinguishable from that of humans. Altogether, the combination of deep neural networks based saliency models for image processing and a Bayesian framework for scanpath integration proves to be a powerful and flexible approach to model human behavior in natural scenarios.

**Keywords:** visual search, eye movements, ideal observer, saliency maps, human behavior

## 1. INTRODUCTION

Visual search is a natural task that humans perform in everyday life, from looking for someone in a photograph to searching your favorite mug in the kitchen. The ability to find a target relies on our capability to gather visual information through a sequence of eye movements; that perform a discrete sampling of the scene. This sampling of information is not carried out on random points.