New Perspectives in Computational Modeling of Human Attention

Organizers & Contributors
Ilker Yildirim (ilker.yildirim@yale.edu)
Mario Belledonne (mario.belledonne@yale.edu)
Yale University, New Haven, CT, USA

Contributors
Ruth Rosenholtz (rruth@mit.edu)
Massachusetts Institute of Technology and NVIDIA, Cambridge, MA, USA
Eivinas Butkus (eb3407@columbia.edu) & Nikolaus Kriegeskorte (nk2765@columbia.edu)
Columbia University, New York City, NY, USA
Seoyoung Ahn (ahnseoyoung@gmail.com) & Gregory Zelinsky (gregory.zelinsky@stonybrook.edu)
University of California, Berkeley, CA, USA & Stony Brook University, Stony Brook, NY, USA

Keywords: attention; computational modeling; goal-driven processing; tasks; visual perception; XXX

Introduction

This symposium will present a set of four talks and a panel discussion that will together take the audience inside a scientific revolution that has been (largely quietly) unfolding in the field of attention: A set of recent computational modeling approaches that allow us to think about human attention in fundamentally new ways.

In cognitive science, studies of attention stand out in at least two dimensions. First, and most bluntly, it is an outright confusing area to work in. "Attention" is a term ascribed to many sorts of mechanisms and phenomena. Case in point: there are at least three papers all published in 2024 presenting ongoing active (and, surprisingly, topically, largely non-overlapping) debates: Rosenholtz (2024), Theeuwes (2024), and Wu (2024).

Second, attention stands out in the extent of the gap between the rich empirical phenomena integrated into conceptual theories, versus formal computational models, with most influential models dating back at least a decade (e.g., Bruce & Tsotsos, 2005; Bundesen et al., 2015; Reynolds & Heeger, 2009; Wolfe, Cave, & Franzel, 1994; Dosher & Lu, 2000), rather than keeping up with the advances in experimental work.

This 90-minute-long gathering will show how the field of attention has been radically changing along both dimensions --- how models of attention have been carving new and productive ways of better drawing the contours of what attention is and enabling progress toward a more integrated research landscape of experiments and modeling.

Adaptive computation as a new mechanism of human dynamic attention

Mario Belledonne & Ilker Yildirim

A key role for attention is to continually focus visual processing to satisfy our goals. How does this work in computational terms? This talk will introduce adaptive computation — a new computational mechanism of human attention that bridges the momentary application of perceptual computations with their impact on planning outcomes. Adaptive computation is a dynamic algorithm that rations perceptual computations across objects on-the-fly, enabled by a novel and general formulation of task relevance. We evaluate adaptive computation in a case study of multiple object tracking (MOT) — a paradigmatic example of selection as a dynamic process, where observers track a set of target objects moving amidst visually identical distractors. Adaptive computation explains the attentional dynamics of object selection with unprecedented depth. It not only recapitulates several classic features of MOT (e.g., trial-level tracking accuracy, localization error of targets), but also captures properties that haven't previously been modeled — including both the sub-second patterns of attentional deployment between objects, and the resulting sense of subjective effort. Adaptive computation thus provides a new type of mechanistic model for the dynamic operation of visual attention.

Recurrent attention enables efficient and flexible use of energy in vision

Eivinas Butkus, Zhuofan Ying & Nikolaus Kriegeskorte

Vision is energetically costly. Visual attention may save energy by selecting, on the basis of a cursory initial analysis, the features and locations of a particular image that deserve scrutiny. In addition, attentional mechanisms might enable a neural network to invest energy flexibly, expending more energy when it is abundant and/or when accuracy matters especially. Here we investigate these ideas using convolutional neural network models with recurrent attentional mechanisms, implemented as multiplicative gain on time ("when gain"), feature map ("what gain"), and location ("where gain"). We compare the behavior of a variety of models to humans in the context of a visual search task. Humans and models were presented with cluttered

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

images containing a single handwritten digit among multiple letters. The task was to determine the class (what) and location (where) of the handwritten digit. Humans viewed brief image presentations (100-400 ms stimulus duration) and also rated the difficulty of each search. Models were pre-trained for object classification, and only their attentional and readout mechanisms were trained on our task. The loss encouraged high accuracy and low energy use (defined as the total convolutional activation summed over space and time). We found that recurrent attention mechanisms enabled networks to achieve the same accuracy at lower energetic cost. In addition, models receiving an additional input indicating the relative cost of energy and errors could flexibly trade energy for accuracy. Finally, models with recurrent attention explained human errors and trial difficulty judgements better than a no-attention baseline. Our work demonstrates the importance of resource costs for understanding the computational mechanisms of biological vision.

Active Object Generation as Attentional Guidance

Seoyoung Ahn & Gregory Zelinsky

Many search theories explain attentional guidance through template matching, where an internal representation of the target—stored in high-level brain areas like working memory—guides attention by matching sensory input to the template. While this approach works well for simple features like color or shape (e.g., searching for a red circle), it does not generalize well to real-world object search, where the target template is inherently ambiguous (e.g., searching for an insect could mean looking for a red, round ladybug or a green, spiky grasshopper). In this talk, I propose that categorical search operates through the active generation of target-like features from the current visual input, with search guidance determined by the ease of generating the target. This "generation cost" is quantified using deep neural networks for image synthesis, providing a computational measure of how much an image must be modified to resemble a given target. Despite not being explicitly trained for target detection, the model achieved ~90% accuracy in localizing targets, demonstrating robustness even under blurry noise. Furthermore, it closely aligned with human eye-movement data, particularly excelling in explaining target-absent search, where conventional feature-matching models fail due to the lack of clear target-defining features. This approach fundamentally reconceptualizes template-based theories by suggesting that attention is drawn to an object not merely because they match a stored template, but because they facilitate the generative process of constructing the target's features.

Visual Attention in Crisis

Ruth Rosenholtz

Early in the study of visual attention, it appeared promising that understanding of preattentive and attentional processes could provide a unifying explanation of a wide range of visual phenomena, by elucidating a critical capacity limit faced by visual processing. Since then, researchers have uncovered significant anomalies, frustrating hopes of a single predictive mechanism. Vision science needs to rethink visual attention from the ground up. What behavioral phenomena demonstrate capacity limits, and require additional mechanisms one might want to call "attention"? Which supposedly attentional phenomena can be better explained by perceptual processes such as peripheral vision? What does the general success of real-world vision suggest about capacity limits? Considering these questions helps enumerate the critical phenomena that an attentional model needs to explain and provides insight into the nature of mechanisms and capacity limits. With this house cleaning there may once again be hope for a unifying theory, in which all perception results from performing a task, where tasks face a limit on complexity.

References

Bundesen, C., Vangkilde, S., & Petersen, A. (2015). Recent developments in a computational theory of visual attention (TVA). Vision research, 116, 210-218.

Bruce, N., & Tsotsos, J. (2005). Saliency based on information maximization. Advances in Neural Information Processing Systems, 18.

Dosher, B. A., & Lu, Z. L. (2000). Mechanisms of perceptual attention in precuing of location. Vision research, 40(10-12), 1269-1292.

Reynolds, J. H., & Heeger, D. J. (2009). The normalization model of attention. Neuron, 61(2), 168-185.

Rosenholtz R. Visual Attention in Crisis. Behavioral and Brain Sciences. Published online 2024:1-32. doi:10.1017/S0140525X24000323

Theeuwes, J. (2024). Attentional capture and control. Annual Review of Psychology, 76.

Wu, W. (2024). We know what attention is! Trends in Cognitive Sciences, 28(4), 304-318.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: an alternative to the feature integration model for visual search. Journal of Experimental Psychology: Human perception and performance, 15(3), 419.