Learning to attend in a brain-inspired deep neural network

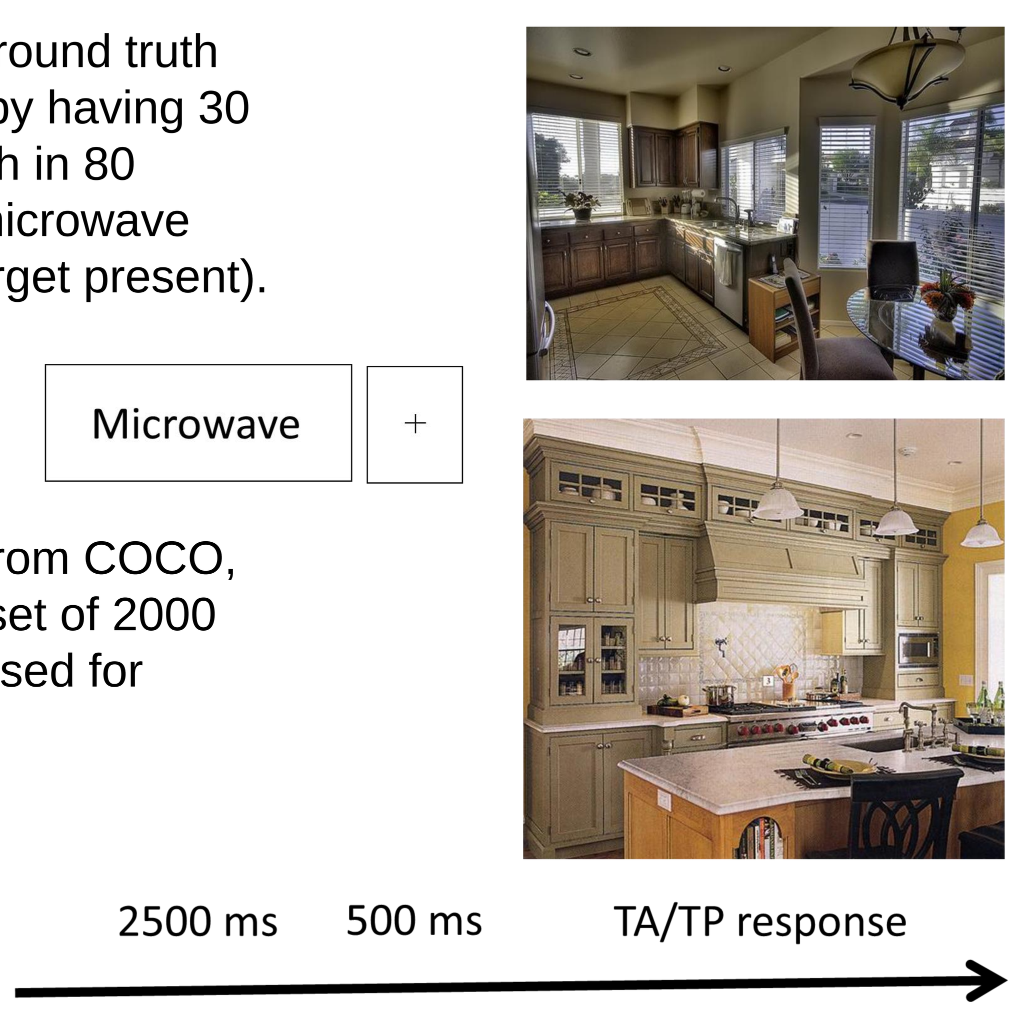
Hossein Adeli[1], & Gregory Zelinsky[1,2]
[1]Department of Psychology; [2]Department of Computer Science; Stony Brook University

Introduction

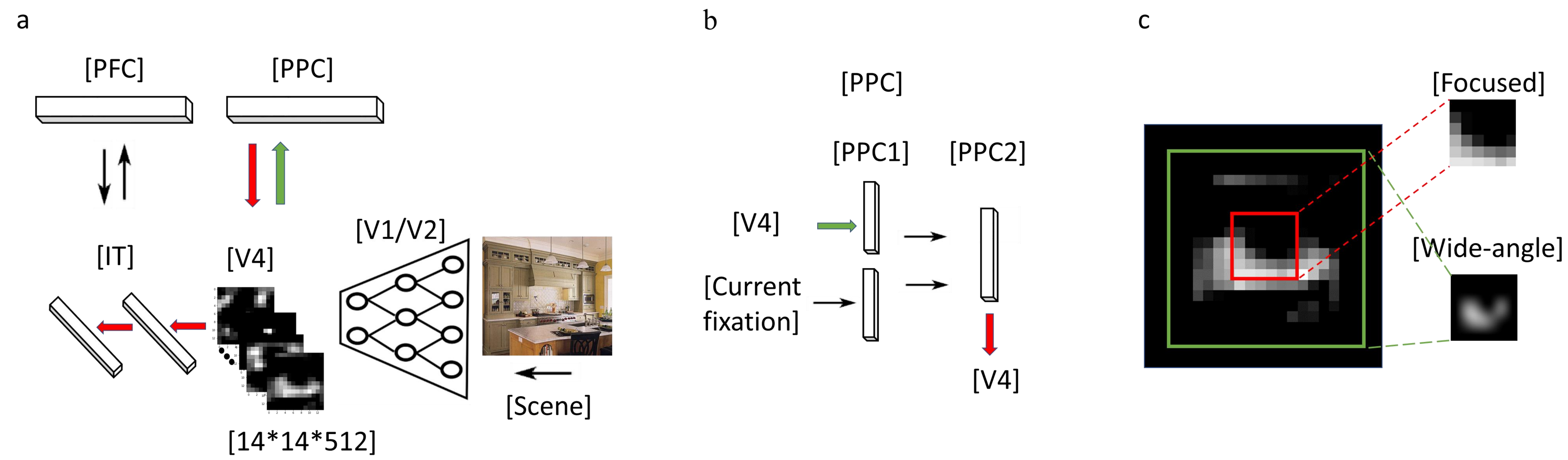
- Visual attention enables primates to prioritize the selection and further processing of visual inputs for the purpose of achieving behavioral goals, but how is this attention control learned?
- Most neural and cognitive models avoid asking this question, focusing instead on the effects that prioritization and selection have on neural and behavioral responses.
- Neural models of attention control are largely based on Biased Competition Theory (Desimone & Duncan, 1995), which focuses on the effects that attending to an object's location or features has on neural recordings and brain network dynamics.
- Similarly, cognitive computational models of attention (Tsotsos et al., 1995) aim to predict the selection and guidance of attention shifts to behavioral goals using image-computable methods and visually-complex inputs.
- These neural and cognitive models of attention control are therefore engineered to fit (or predict) behavior and/or neural data without addressing the more fundamental questions of how attention control signals emerge and function in the context of performing a task or why the brain might even find prioritizing visual inputs and shifting attention to be a useful thing to do.
- Here we leverage the potential of these three perspectives by introducing ATTNet, an image-computable DNN model of the ATTention Network. ATTNet is inspired by biased-competition theory and trained using deep reinforcement learning. Through the application of reward in the context of a search task, ATTNet learns to shift its attention to the locations of features of the rewarded object category.

Behavioral Methods

- A behavioral ground truth was obtained by having 30 subjects search in 80 images for a microwave target (50% target present).
- Images were from COCO, and a disjoint set of 2000 images were used for training.

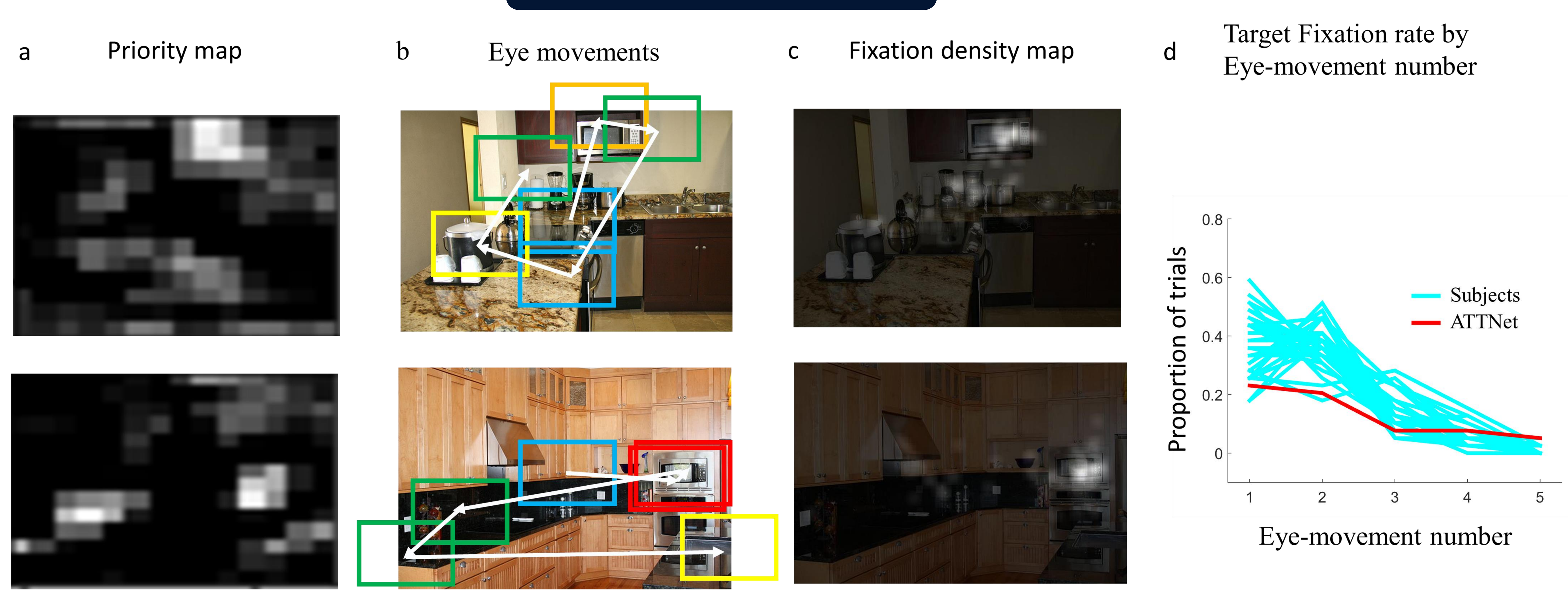


ATTNet Model



- **Anatomy of ATTNet**. (a) ATTNet consists of three interacting components: (1) early parallel visual processing [V1/V2, V4] modeled by the convolutional layers of a Convolutional Neural Network (CNN) trained for object classification (VGG16), (2) ventral processing, and (3) dorsal processing [PPC]. (b) Dorsal processing integrates V4 activity with current eye position to (c) spatially prioritize and select visual inputs for routing through the ventral pathway to IT and PFC.
- ATTNet was trained using policy gradient reinforcement learning (Mnih et al., 2014). The model makes the target present/absent decision after 5 eye-movements and only gets rewarded for correct responses.

Results



- (b) Like subjects, ATTNet started each trial fixating at the center of the scene. The information from this area is routed along the ventral and dorsal pathways (the boxes show the visual areas that are routed ventrally). The routing windows are colored based on the ventral response, with warmer colors showing more confidence in the routed pattern being the target. (a) A priority map generated in the dorsal pathway guides attention to a new location and the process repeats. Shown is the priority map that was generated in the dorsal pathway based on the initial fixations for two sample search displays; both indicate that the model learned to bias visual inputs reflecting evidence for the target-category. Over the course of training ATTNet learns that certain patterns are rewarding and attends to these patterns in order to make more informed decisions and collect more reward. (c) Corresponding fixation data from subjects.
- (d) A quantification of ATTNet's attention guidance to targets. Plotted is the proportion of trials where the target was first fixated, grouped by eye-movement serial number (first, second, … fifth) for ATTNet (red) and individual subjects (cyan). While not as strong as subjects, ATTNet's attention was also guided to the target, as evidenced by target fixations by early eye-movements.

Discussion

- Understanding computational principles of selective attention is key to understanding brain function and building brain-inspired AI systems. This work studies the computational benefit of attention as a dynamic selective routing of information for performing a visual search task.