The Effect of Attention on Contrast Response Functions in Convolutional Neural Networks

Sudhanshu Srivastava (sudhanshu@ucsb.edu)
Dynamical Neuroscience,
University of California, Santa Barbara

Miguel P. Eckstein (migueleckstein@ucsb.edu)
Psychological and Brain Sciences
University of California, Santa Barbara

Abstract

Covert attention results in contrast-dependent influences on neuronal activity with three distinct signatures (Carrasco, 2011): response gain, contrast gain, and baseline shift. Convolutional Neural Networks (CNNs) have recently been used to model covert attention tasks, capturing some known results from psychology and neurophysiology (Srivastava et al., 2024b). This work uses CNNs to understand emergent attention-related neuronal Contrast Response Functions (CRFs). We trained 10 CNNs on the Posner cueing task to optimize target detection with a central cue pointing to the likely target location, with varying contrasts of the display elements. With no explicit attention mechanisms built in, the networks show a behavioral cueing effect and all three gain types emerged in the deeper layer neurons of the networks: Response Gain (17.4%), Contrast Gain (2.3%), and Baseline Shift (22.8%). Using ROC analysis for each neuron, we assessed whether different gain types are associated with different target and cue sensitivities. Response gain neurons had the highest target sensitivity and the lowest cue sensitivity. Baseline shift neurons had the highest cue sensitivity and the lowest target sensitivity. Contrast gain neurons had target/cue intermediate range sensitivities. Together, we show that the diversity of neuronal gain types reported in the literature might arise as an emergent property of task optimization and neurons with different CRFs neurons might be associated with representations of the predictive cue and target.

Keywords: Convolutional Neural Networks; covert attention; Contrast Response Functions

Introduction

Covert Attention refers to the ability to select a part of an image or a scene without moving one's eyes. The Posner cueing task (Posner, 1980) and its variants are the most prominent tasks in the behavioral and neurophysiological studies of covert attention. In this task, observers find a target that appears at one of two possible locations, with a cue, typically an arrow or a box, pointing to the likely target location. Human observers are more accurate when the cue points to the target location (valid trials) than when it points opposite to the target location (invalid trials). In terms of neuronal signatures, a common manipulation is varying the contrast of the items at the two locations and characterizing the response as a function of contrast at the cued and uncued locations. Three signatures have been reported in the literature: an increasing effect of attention with contrast (response gain); a peak effect at middle contrasts and low effect at extreme contrasts (contrast gain); and a fixed effect across contrasts (baseline shift). The normalization model of attention (Reynolds & Heeger, 2009) aimed to model a common framework that explained both response and contrast gain. Here, we investigate the type of contrast response functions that emerge in a CNN trained to maximize target detection on a cueing task (Srivastava et al., 2024b). We show that with no divisive normalization or explicit attention mechanism, the three gain types emerge in the neurons in the deeper layers of the CNNs and each gain type neuron is uniquely related to the cue and target tuning properties of the neurons.

Methods

Stimuli: Each stimulus contained two tilted lines with a central cue: a horizontal line, pointing to the left or right half of the image. The target, a line tilted 20 degrees was present with a 50% probability, and when present, appeared at the location pointed to by the cue 80% of the time and opposite to the cue 20% of the time. The Weber contrast levels of the tilted lines were uniformly sampled from 0.0625, 0.125, 0.25, 0.5, and 1.0 and both the tilted lines always had the same contrast within a stimulus. The cue contrast was fixed at 0.8. Additive white Gaussian noise was added to each stimulus.

Architecture and Training details: The CNNs had 3 convolution layers with 10, 18, and 24 kernels (size 3 by 3) each followed by max pooling (2 by 2). The last two layers were dense layers of sizes 2000 and 2 respectively. Tanh activation function was used except for SoftMax in the output layer. The CNNs were trained via gradient descent to predict if the target was present in an image. All contrasts were trained jointly.

Neuronal Gain Curves: 10 sets of 1000 test images with target and cue at left, target, and cue at right, target left with the cue at right, target right with the cue at left, per contrast, were shown to the networks and the responses of each neuron were obtained. For each location, we compared the mean across the 1000 images in each set for the valid vs. invalid conditions. Significance testing was done across the 10 sets with FDR correction. We used the Naka-Rushton equations to classify the gain curves as Response Gain, Contrast Gain, and Baseline Shift:

$$R(C) = R_{max}\frac{C^n}{C^n + C_{50}^n} + R_{offset} \tag{1}$$

Where $R(C)$ denotes the neuron's response at contrast C, and R_{max}, C_{50}, n, and R_{offset} are fitting parameters. The fitting parameters are obtained first for the neuron's response to invalid stimuli. For fitting the CRF for valid condition, the n parameter is kept the same as the invalid condition. Then, one parameter at a time is varied, keeping the other two the same as the invalid CRF, to fit valid CRF. The R^2 value for all three fits is calculated, and a neuron is classified as response gain, contrast gain, or baseline shift if the best fit is obtained on varying only R_{max}, C_{50}, or R_{offset} respectively. Only neurons with $R^2 > 0.9$ on both valid and invalid conditions were considered.

ROC Analysis: For target sensitivity, the CNN was shown 10 sets of 1000 noisy images, each with no cue and containing the following: only the target at right, only the distractor at right, only the target at left, and only the distractor at left. The response of each neuron to these stimuli was recorded. For each neuron's pair of response distributions (target present and absent), the area under the ROC (AUROC) for detecting the target vs. distractor presence at each location was calculated. For the cue sensitivity, the same procedure was repeated with only the cue present/absent at each location, with no target or distractors in the scene. Sensitivity was defined as |AUROC – 0.5|.

Results

Behavior Results: All networks showed performance between 64-75% on an unseen test set across contrasts, and a difference between valid and invalid trials for most contrasts.

ROC Results: The early convolution layers were retinotopic and showed separate subpopulations of neurons tuned to the cue and the target. In the third convolution layer, we find neurons tuned jointly to the target and the cue but only at one spatial location at a time, while the dense layer has neurons integrating the target and cue across locations. These results agree with (Srivastava et al., 2024a) but they used a peripheral cue and only one contrast level.

Neuronal Gain Curves: While the early layers have neurons whose responses fit the Naka-Rushton equations, they don't show a significant difference in response when a valid vs. an invalid cue is present, and neurons with the three gain types are found in the dense layer. We found that across models, baseline shift is the most prominent gain type (22.8±2.9 %), followed by response gain (17.4±1.9 %), followed by contrast gain (2.3±0.4 %).

New gain types: Besides the types of gain reported in the attention literature, we also find neurons with a negative slope with respect to contrast, and neurons with lower response to a valid target than to an invalid target at the same location. Among these, we find 20.4±2.2% fit the baseline shift equations, 16.1±1.3% fit the response gain equations, and 1.4±0.3% fit the contrast gain equations.

Relating gain types with cue and target sensitivities: We find that neurons classified as response gain have the highest target sensitivity (0.26±0.02) and the lowest cue sensitivity (0.05±0.001) while baseline shift neurons have the lowest target sensitivity (0.18±0.02) and the highest cue sensitivity (0.08±0.001). Contrast gain neurons had both sensitivities between the response gain and baseline shift neurons (0.21±0.05 target; 0.06 ±0.002 cue).

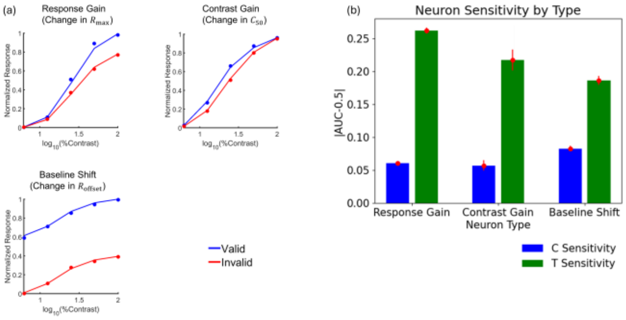

Figure 1: (a) Example of the three gain types observed in the CNN neurons. (b) Mean sensitivity (|AUROC-0.5|) to the cue (blue) and target (green) by neuronal gain type. Error bars indicate standard error over ten networks.

Conclusions

The diversity in the neuronal gain types is typically unified using the divisive normalization model of attention(Itthipuripat et al., 2014; Reynolds & Heeger, 2009). Here, we use CNNs with no explicit divisive normalization built-in, train them to maximize target detection across contrasts, and find that all three neuronal gain types reported in the literature emerge in the network. Further, the sensitivity of these neurons to the cue and the target are related to the neuronal gain type. Together, we provide an alternative plausible explanation for the diversity of neuronal gain types reported in the literature and a testing bed for neural network models of attention.

Acknowledgments

This work was supported by The Institute for Collaborative Biotechnologies through US Army Research Office Contract W911NF-19-D-001 and US Army Research Office Cooperative Agreement W911NF-19-2-0026.

References

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*(13), 1484–1525. https://doi.org/10.1016/j.visres.2011.04.012

Itthipuripat, S., Garcia, J. O., Rungratsameetaweemana, N., Sprague, T. C., & Serences, J. T. (2014). Changing the Spatial Scope of Attention Alters Patterns of Neural Gain in Human Cortex. *Journal of Neuroscience*, *34*(1), 112–123. https://doi.org/10.1523/JNEUROSCI.3943-13.2014

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Reynolds, J. H., & Heeger, D. J. (2009). The Normalization Model of Attention. *Neuron*, *61*(2), 168–185. https://doi.org/10.1016/j.neuron.2009.01.002

Srivastava, S., Wang, W. Y., & Eckstein, M. P. (2024a). *Bridging Neurons and Behavior in a Convolutional Neural Network with Emergent Human-like Covert Attention* (p. 2023.09.17.558171). bioRxiv. https://doi.org/10.1101/2023.09.17.558171

Srivastava, S., Wang, W. Y., & Eckstein, M. P. (2024b). Emergent human-like covert attention in feedforward convolutional neural networks. *Current Biology*, *34*(3), 579-593.e12. https://doi.org/10.1016/j.cub.2023.12.058