Attention-Dependent Perceptual Learning in a CNN Model of the Visual System

Thomas Maher (tm3566@nyu.edu), Grace Lindsay (Grace.Lindsay@nyu.edu)
Center for Data Science, New York University, 60 Fifth Avenue, 7th Floor, New York, NY 10011

Abstract

Perceptual learning is defined by increased performance on a perceptual task following prolonged practice. Many studies have observed that rates of perceptual learning can depend on the strength of visual attention deployed during the learning. Despite attempts to relate perceptual learning and attention, the exact mechanism behind this relationship remains unknown. Here, we propose a convolutional neural network (CNN) model of visual perceptual learning for the purpose of elucidating this relationship. Our model uses an attention system along with a local learning rule that, through weight updates, solidifies the impact of attention. We found that the model's performance on a precise visual task increased as a result of the local learning rule and that this effect was dependent on the magnitude of the attention modulation. This suggests that modulatory attention and plasticity in early visual areas are sufficient for inducing perceptual learning.

Keywords: Attention, Perceptual Learning, CNN

Introduction

Perceptual learning, resulting from prolonged exposure or practice on a task involving a particular stimulus, is characterized by increased performance on the same or other tasks involving the stimulus. While synaptic plasticity is thought to be a key factor in visual perceptual learning (Fahle, 2004; Gluck & Granger, 1993), it has been shown that attention to the perceptual stimulus during learning may be required for the effect to occur (Byers & Serences, 2012; Mukai et al., 2011).

Perceptual learning is also thought to induce highly spatially selective changes, indicating plasticity in early visual regions (Jehee et al., 2012; Mukai et al., 2011). Interestingly, recent work has shown that feature-based attention (which acts in a spatially global way), can cause performance benefits from perceptual learning to generalize to untrained spatial locations (Hung & Carrasco, 2021).

Recent computational models of perceptual learning have used Hebbian-like updating rules to re-weight model representations (Dosher et al., 2013; Petrov et al., 2005). Such models open a pathway to exploring perceptual learning mechanistically but tend be shallower and can't capture the full visual hierarchy. At least one previous study has used deep CNNs to replicate effects of perceptual learning, but did not incorporate attention and relied on backpropagation for weights updates, which is not biologically realistic.

Here we build on previous models that have incorporated feature attention (in the form of gain modulation) into CNNs (Lindsay & Miller, 2018; Martinez-Trujillo & Treue, 2004), and build on these to replicate perceptual learning by adding local Hebbian learning rules in early layers. Through this we hope to show how local learning can solidify attention-induced activity changes, leading to a better performing model.

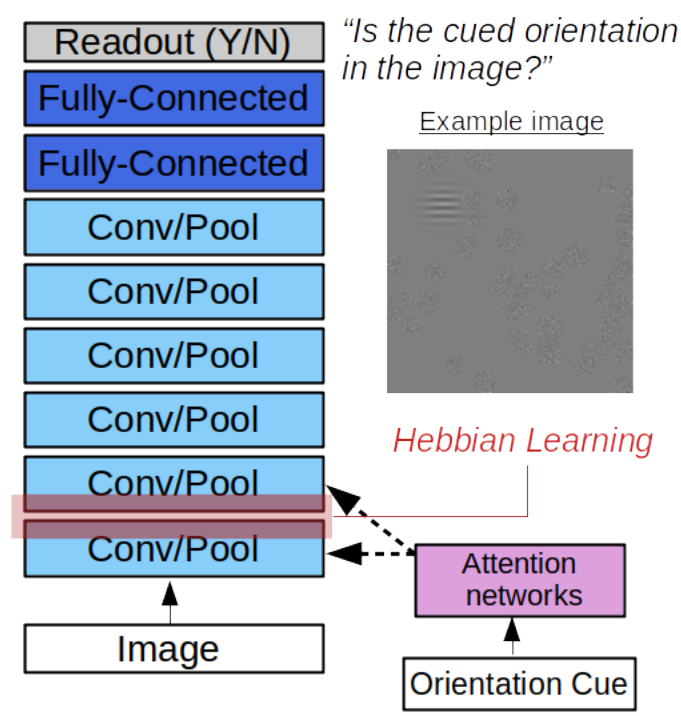

Figure 1: Architecture of the perceptual learning model. The left-hand side shows an AlexNet architecture which receives an input image containing a Gabor patch. Two attention networks take a Gabor orientation cue as input and modulate the first two convolutional layers, allowing the model to flexibly focus on the specialized task of discriminating whether the cued orientation is in the input image. Finally, a Hebbian update rule uses the correlation of activity between the first and second layers to update the model weights, thereby replicating perceptual learning.

Methods

Our goal is to test if a local Hebbian learning rule combined with top-down feature attention can replicate observed effects of perceptual learning and its dependence on attention. Our modeling approach has three training stages. First, we start with an Alexnet model pre-trained on ImageNet. Second, we train an attention system to modulate the activity of the first two layers of the CNN based on an orientation cue. At the output, the model must learn to respond positively if the cued orientation is present and negatively if not. In the third phase, we mimic perceptual learning by letting a Hebbian learning rule change the weights between the first and second layers of the network while attention modulates activity at these layers (see Figure 1). Further details of this process are provided below.

Task and Stimuli For attention training, the model is given a binary classification task in which it indicates whether a cued orientation is present in an input image. Input images consist of one Gabor patch in one out of four quadrants and with one of four orientations; pixel noise patches are also randomly placed on each image (see Figure 1) for an example. The cue is represented as a one-hot encoded vector indicating one of the four possible orientations, and this is passed into the attention networks. For the perceptual learning phase, the orientation cue is held constant, as is the stimulus location. On each trial, the stimulus orientation is either the cued one or 90 degrees opposite it (leading to chance performance of 50%).

Model We used an Alexnet model pre-trained on a 1000-way classification task with Imagenet implemented in Tensorflow and swapped out the final layer for a binary classifier. Two attention networks (each a 2-layer feedforward MLP) were used to implement feature-based attention in the model, one for each of the convolutional layers modulated by attention. When given an orientation cue, the attention networks each output a set of gain modulation values, f_i one for each feature channel in their respective layers. At a given layer, activity in feature channel i is multiplied by $(1+\beta f_i)$, with β being a strength parameter (held constant at 1 during attention training). Thus, as in previous work (Lindsay & Miller, 2018), feature attention works globally across space, but modulates different feature channels differently. The attention network and binary output layer of the model are trained jointly via backpropagation to perform the binary classification task.

Perceptual Learning

To implement perceptual learning, backpropagation is turned off and the second convolutional layer is coded as a 'locally connected' layer, i.e. the weight sharing normally used in convolutional layers is turned off to allow spatially specific learning. While the network performs the constant orientation detection task described above, a Hebbian-inspired learning rule (wherein correlated pre-post activity leads to weight increases and anti-correlated leads to decreases) is applied to the weights that connect the first pooling layer to the second convolutional layer. This procedure is done either with $\beta = 1$ (strong attention) or $\beta = .8$ (weak attention). Initial and post-learning performance on this task is evaluated with strong attention and stimuli either at the trained quadrant or the diagonally opposite one.

Results and Conclusion

In the 3 networks (different colors) shown in Figure 2, we can see that performance consistently increased after the perceptual learning phase. However, this increase was dependent on the strength of attention used during learning, with performance increases weaker for weaker attention (dotted lines). To test if the feature-based attention we deployed here causes learning to generalize to other spatial locations as observed experimentally (Hung & Carrasco, 2021), we also evaluated post-learning performance in an untrained quadrant. We can see that perceptual learning does still increase performance here, but the increases are less than those for the trained quadrant. Furthermore, weaker attention leads to weaker performance enhancement on average here as well, indicating that this learning is dependent on top-down feature-based attention.

In total, we are able to replicate the impact of perceptual learning on behavior using a local learning rule in an attention-modulated CNN. We demonstrated the dependence of perceptual learning on attention by modulating attention strength. Traditionally, local learning rules have struggled to enhance performance in deep neural networks due to the credit assignment problem. However, we show here how attentional modulation can assist with credit assignment, making the application of local learning rules (even very early in the network) suitable for enhancing performance.

Next steps: We hope to incorporate spatial attention into this model. We will also work to refine our attention training procedure as it sometimes produces models strong with orientation or location biases that are not suitable for the perceptual learning phase.

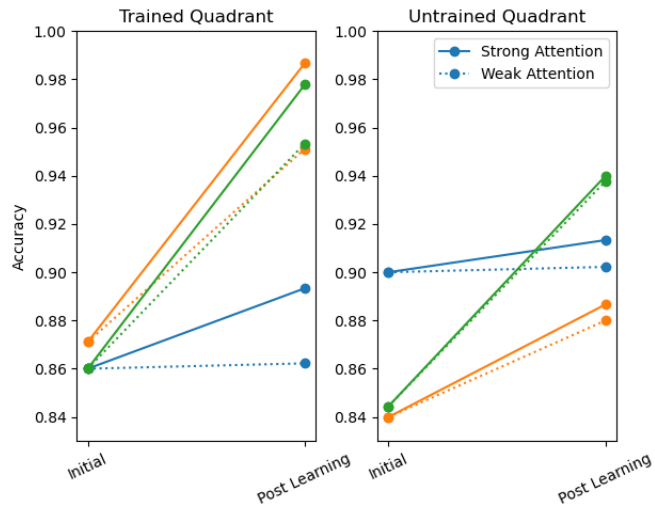

Figure 2: Results for several models with different initializations before and after perceptual learning. Dotted lines indicate weaker attention strength. Left: the models ability to perform the task with the cued orientation increased after perceptual learning. Right: The same effect is seen to a lesser degree when the task is performed on a region other than the trained region. In both cases, stronger attention led to a larger effect of perceptual learning.

References

Byers, A., & Serences, J. T. (2012). Exploring the relationship between perceptual learning and top-down attentional control [Visual Attention 2012 Volume I]. *Vision Research*, *74*, 30–39. https://doi.org/https://doi.org/10.1016/j.visres.2012.07.008

Dosher, B. A., Jeter, P., Liu, J., & Lu, Z.-L. (2013). An integrated reweighting theory of perceptual learning. *Proceedings of the National Academy of Sciences*, *110*(33), 13678–13683. https://doi.org/10.1073/pnas.1312552110

Fahle, M. (2004). Perceptual learning: A case for early selection. *Journal of Vision*, *4*(10), 4–4. https://doi.org/10.1167/4.10.4

Gluck, M. A., & Granger, R. (1993). Computational models of the neural bases of learning and memory. *Annual Review of Neuroscience*, *16*(Volume 16, 1993), 667–706. https://doi.org/https://doi.org/10.1146/annurev.ne.16.030193.003315

Hung, S.-C., & Carrasco, M. (2021). Feature-based attention enables robust, long-lasting location transfer in human perceptual learning. *Scientific reports*, *11*(1), 13914.

Jehee, J. F., Ling, S., Swisher, J. D., van bergen, R. S., & Tong, F. (2012). Perceptual learning selectively refines orientation representations in early visual cortex. *The Journal of Neuroscience*. 10.1523/JNEUROSCI.6112-11.2012

Lindsay, G. W., & Miller, K. D. (2018). How biological attention mechanisms improve task performance in a large-scale visual system model. *eLife*, *7*. https://doi.org/10.7554/eLife.38105

Martinez-Trujillo, J. C., & Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Current biology*, *14*(9). https://doi.org/https://doi.org/10.1016/j.cub.2004.04.028

Mukai, I., Bahadur, K., Kesavabhotla, K., & Ungerleider, L. G. (2011). Exogenous and endogenous attention during perceptual learning differentially affect post-training target thresholds. *Journal of vision*, *11*(1), 25–25.

Petrov, A. A., Dosher, B. A., & Lu, Z.-L. (2005). The dynamics of perceptual learning: An incremental reweighting model. *The Psychological Review*, *112*(4). https://doi.org/10.1037/0033-295X.112.4.715