Understanding the functional and structural differences across excitatory and inhibitory neurons

Sun Minni (sunminni1031@gmail.com)[*,1], Li Ji-An (jian.li.acad@gmail.com)[*,2], Theodore Moskovitz (thm2118@columbia.edu)[3], Grace Lindsay (gracewlindsay@gmail.com)[3], Kenneth Miller (kendmiller@gmail.com)[3], Mario Dipoppa (mario.dipoppa@gmail.com)[3], Guangyu Robert Yang (gyyang.neuro@gmail.com)[3]

1 Department of Machine Intelligence, Peking University, Beijing 100871 China

2 Department of Statistics and Finance, University of Science and Technology of China, Hefei, Anhui 230027 China

3 Zuckerman Institute, Columbia University, New York, NY 10027 USA

Abstract

One of the most fundamental organizational principles of the brain is the separation of excitatory (E) and inhibitory (I) neurons. In addition to their opposing effects on post-synaptic neurons, E and I cells tend to differ in their selectivity and connectivity. Although many such differences have been characterized experimentally, it is not clear why they exist in the first place. We studied this question in an artificial neural network equipped with multiple E and I cell types. We found that a deep convolutional recurrent network trained to perform an object classification task was able to capture salient distinctions between E and I neurons. We explored the necessary conditions for the network to develop distinct selectivity and connectivity across cell types. We found that neurons that project to higher-order areas will have greater stimulus selectivity, regardless of whether they are excitatory or not. Sparser connectivity is required for higher selectivity, but only when the recurrent connections are excitatory. These findings demonstrate that the differences observed across E and I neurons are not independent, and can be explained using a smaller number of factors.

Keywords: visual cortex; Dale's principle; artificial neural networks

Introduction

Deep neural networks have become powerful tools to model the brain (Yamins & DiCarlo, 2016). However, standard deep networks lack fundamental architectures of the brain, in particular Dale's law, the separation of excitatory (E) and inhibitory (I) neurons. In the brain, excitatory and inhibitory neurons differ in several important ways. There are several times (4-10x) more excitatory neurons than inhibitory neurons. Neurons that project to other areas, the so-called "principal neurons", are all excitatory in the cortex. In the mammalian sensory cortex, excitatory neurons are overall more selective to stimuli than inhibitory neurons in the same area (Znamenskiy et al., 2018). Finally, excitatory neurons are more sparsely connected with each other (Harris & Shepherd, 2015), compared to inhibitory neurons.

It is not clear whether these differences across excitatory and inhibitory neurons serve computational purposes. It is also unknown whether these differences are mutually independent or not. Here we address these questions using artificial neural networks.

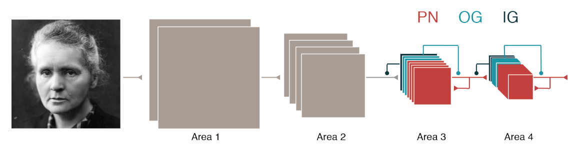

Figure 1: **Convolutional recurrent network with multiple cell types.** Principal neurons (PN) are excitatory, while output-gating (OG) and input-gating (IG) neurons are inhibitory.

Multi-cell Convolutional Recurrent Network

The networks we use to model the visual cortex start with 2 layers of purely feedforward, convolutional processing, followed by 2 layers of recurrent processing (Figure 1). Each recurrent layer consists of excitatory and inhibitory neurons (channels). In the brain, excitatory and inhibitory neurons can be further divided into many subtypes that differ in their inputs, output targets, and gene expression. Two major types of inhibitory neurons target input-receiving and output-generating areas of excitatory neurons respectively.

Here we modified the structure of a convolutional LSTM unit (Xingjian et al., 2015) to introduce two distinct types of inhibitory neurons into the network (Figure 1). We use multi-layer perceptrons (MLPs) with a single hidden layer to recurrently generate the input and output gates. The two additional sets of neurons in the hidden layers become the input-gating (IG) and output-gating (OG) neurons. In the standard version of this network, we impose Dale's law in the recurrent layers by ensuring that principal neurons (PN) only make excitatory connections. In comparison, the gate neurons (IG, OG) all make inhibitory connections.

Reproducing functional and structural differences across cell types

We trained the multi-cell network on the image classification dataset CIFAR10 (Krizhevsky & Hinton, 2009). We ask whether training develops qualitative differences between excitatory and inhibitory neurons in the network, as observed in the brain. We measured the selectivity to natural images for model excitatory and inhibitory neurons. The selectivity is quantified as the skewness of the distribution of responses to various stimuli (Samonds, Potetz, & Lee, 2014). The skewness is higher if a neuron is strongly selective to a small number of stimuli. After training, the selectivity is substantially higher for excitatory neurons (PN) compared to inhibitory neurons (IG, OG), in both recurrent layers (areas 3 and 4). Then, we quantified the proportion of connection weights that exceed a threshold chosen to separate the strongest weights from the weaker weights. We found that sparser excitatory connectivity emerged in the network after training, in agreement with biological observations.

Why do excitatory and inhibitory neurons have distinct selectivity and connectivity?

We have shown that neural networks with excitatory and inhibitory neurons develop different selectivity and connectivity, qualitatively reproducing long-standing findings in the brain. The emergent differences across our model E and I neurons can only be explained by their built-in structural asymmetry. We have incorporated three major forms of asymmetry that exist in the brain. In this section, we will remove individual asymmetry, and test which one led to the observed differences in selectivity and connectivity.

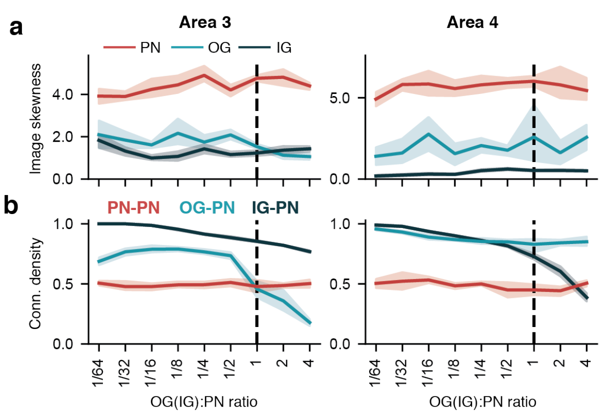

Figure 2: **Selectivity and connectivity in networks with various ratios of excitatory and inhibitory neurons.** image skewness (a), and connection density (b) for networks where the number of inhibitory neurons (both IG and OG) varied from 1/64 to 4x of the number of excitatory neurons.

Asymmetry in numbers: In our standard network, the ratio between the number of OGs(IGs) and PNs is 1:4. We varied the number of inhibitory channels in the recurrent layers, while maintaining the number of excitatory channels. We found little evidence that the difference in E and I selectivity depends on the ratio of their numbers (Figure 2a). In comparison, the density of inhibitory connections decreases as the number of inhibitory channels increase, while the connection density of excitatory neurons remain steady (Figure 2b).

Asymmetry in projection: In both the cortex and our standard network, principal neurons are exclusively excitatory. Here, we trained *InhPN* networks with inhibitory principal neurons and excitatory interneurons. In InhPN networks, the excitatory neurons (interneurons) are less selective to natural images compared to the inhibitory neurons (principal neurons) (Figure 3). This result argues that whichever type of neuron serves as the principal neurons would demand higher selectivity, presumably because the principal neurons need to carry detailed stimulus information to the next layer. Interestingly, in the InhPN networks, the connectivity among inhibitory (principal) neurons remains dense despite a heightened selectivity (Figure 3). This result is in stark contrast with the sparser connectivity needed for highly selective excitatory neurons in the standard network.

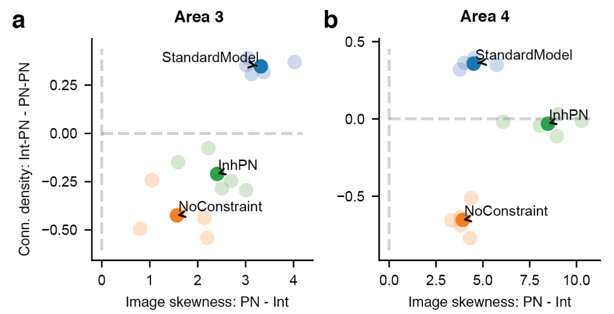

Figure 3: **Removing asymmetry between excitatory and inhibitory neurons.** Difference in connection density against difference in image selectivity for areas 3 (a) and 4 (b) of three types of networks. Light circles: individual networks; dark circles: average.

Asymmetry in action: We removed Dale's law in *NoConstraint networks*. The formerly-excitatory principal neurons developed higher selectivity compared to the formerly-inhibitory interneurons (Figure 3), consistent with our previous finding that principal neurons have higher selectivity regardless of the sign of their outputs. Similar to previous results, the principal neurons no longer have sparser connectivity (Figure 3).

Conclusion

We have shown that recurrent neural networks equipped with multiple cell types are capable of capturing several important features of the brain, including higher selectivity and sparser connectivity among excitatory neurons. These qualitative features emerged solely from the pressure to perform the task, suggesting that these qualities are indeed beneficial to task performance. This allows us to study what aspects of the network give rise to this distinction between excitatory and inhibitory neurons. We found that the higher selectivity of excitatory neurons is mainly driven by their role as the principal neurons that transmit information to upper layers. When Dale's law is obeyed, a higher selectivity necessitates sparser connectivity among excitatory neurons.

Acknowledgments

We thank Dan Yamins, Aran Nayebi, and Daniel Bear for helpful conversations. We thank Jochen Weber, Sander Antoniades, and Raj Bose for the GPU cluster. Funding sources: Simons Foundation, NSF DBI-1707398, and the Gatsby Charitable Foundation.

References

Harris, K. D., & Shepherd, G. M. (2015). The neocortical circuit: themes and variations. *Nature neuroscience*, *18*(2), 170.

Krizhevsky, A., & Hinton, G. (2009). *Learning multiple layers of features from tiny images* (Tech. Rep.). Citeseer.

Samonds, J. M., Potetz, B. R., & Lee, T. S. (2014). Sample skewness as a statistical measurement of neuronal tuning sharpness. *Neural computation*, *26*(5), 860–906.

Xingjian, S., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., & Woo, W.-c. (2015). Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems* (pp. 802–810).

Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature neuroscience*, *19*(3), 356.

Znamenskiy, P., Kim, M.-H., Muir, D. R., Iacaruso, M. F., Hofer, S. B., & Mrsic-Flogel, T. D. (2018). Functional selectivity and specific connectivity of inhibitory neurons in primary visual cortex. *bioRxiv*, 294835.