Biophysical Deep Learning: Laminar-resolved Cortical Columns as Neural Network Units

Dasja de Leeuw (dasja.deleeuw@maastrichtuniversity.nl)
Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, P.O. Box 616 Maastricht, 6200 MD, Limburg, the Netherlands

Rainer Goebel (r.goebel@maastrichtuniversity.nl)
Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, P.O. Box 616 Maastricht, 6200 MD, Limburg, the Netherlands

Mario Senden (mario.senden@maastrichtuniversity.nl)
Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, P.O. Box 616 Maastricht, 6200 MD, Limburg, the Netherlands

Abstract

Cognitive computational neuroscience strives to develop models that achieve both biological and cognitive fidelity. Propelled by the success of deep neural networks (DNNs) in emulating human functional capacities and neural representations, the field increasingly utilizes DNNs for generating, testing, and refining theories of the neurocomputational processes. However, these neuroconnectionist models often lack neuroanatomical detail and neuronal population dynamics, factors that provide important constraints on the neurocomputational solution space. To address this, we introduce a biophysics-informed neuroconnectionist modeling approach with powerful learning capabilities. Our approach constructs neural networks from laminar-resolved cortical columns with neuroanatomically-realistic internal connectivity. In a proof of concept, we show that these networks can be succesfully trained to reproduce firing rates of neuronal populations in a perceptual decision-making task and achieve high accuracy in classification tasks. This work demonstrates the feasibility of embedding function in biophysics-informed models and introduces a new class of neuroconnectionist models striking a meaningful balance between biological realism and cognitive function.

Keywords: biophysical modeling, deep learning, decision-making, neural ODEs

Introduction

Computational neuroscience knows two dominant approaches: bottom-up (biology-driven) and top-down (function-driven) modeling. DNNs have increasingly been used to integrate these approaches and achieve both biological and cognitive fidelity, a key challenge in the field (Kriegeskorte & Douglas, 2018). The value of DNNs lies in their incorporation of brain-inspired computational principles, combined with an unprecedented ability to solve perceptual tasks (Hassabis, Kumaran, Summerfield, & Botvinivk, 2017; Kriegeskorte, 2015; Kaligh-Razavi & Kriegeskorte, 2014; Yamins & DiCarlo, 2016). However, these networks often ignore neuroanatomical properties such as the connectivity structure of local neural circuits and neuronal cell densities. No distinction between excitatory and inhibitory populations are made, thereby amalgamating excitatory and inhibitory interactions and violating Dale's law.

In response, we introduce a biophysics-informed modeling approach. We use laminar-resolved cortical column networks with empirically-derived, brain-region-specific internal connectivity profiles and with realistic mean field population dynamics. The resulting networks essentially consist of coupled differential equations, and are trainable with the adjoint method originally developed for training neural ordinary differential equations (neural ODEs) (Chen, Rubanova, Bettencourt, & Duvenaud, 2019). The adjoint method executes backpropagation by solving a second neural ODE backwards in time, thus turning training into a continuous-time process and avoiding the drawbacks of backpropagation through time.

Methods

Mean field dynamics of cortical columns Our approach to modeling the cortical column builds upon the dynamic mean field (DMF) model of a cortical microcircuit for bistable perception, as described and developed by Evers, Peters, and Senden (2023). The DMF implementation simulates the firing rates of eight neuronal populations in a cortical column, structured in four cortical layers, L2/3, L4, L5 and L6. Each layer contains an excitatory and inhibitory population. The empirically derived intralaminar connectivity is adopted from Potjans and Diesmann (2014). This internal connectivity structure ensures targeted excitation and inhibition between cortical layers within a single column.

During training, external connections between the modeled columns are established and adjusted to minimize loss on the specified task. For both of our use cases, the internal, empirically-derived connectivity of the columns remain fixed during training. This approach ensures a high level of biological realism while training the network in a goal-driven manner.

Two-alternative forced choice dynamics As a first use case, we test whether a two-column network can reproduce the winner-take-all dynamics typical of perceptual decision making. The network is trained to minimize the loss between its L2/3 excitatory (L2/3e) firing rates and an effective inhibition decision-making model without layer separation as a training target (Wong & Wang, 2006). L2/3e is chosen as the readout layer, as activity in superficial cortical layers has been shown to determine perceptual decision outcomes (Changdrasekaran, Peixoto, Newsome, & Shenoy, 2017). While the model by Wong & Wang simulates lateral intraparietal (LIP) firing rates, the resulting winner-take-all behavior can be applicable to other brain areas than LIP. For our two-column network, middle temporal (MT) neuronal population counts are used to learn lateral connectivity and reproduce perceptual decision-making dynamics.

Two types of connections can be learned: (1) lateral inhibition connections from the excitatory population in L2/3 (L2/3e) to the inhibitory population (L2/3i) in the neighboring column and (2) self-excitation connections in L2/3e in the same column. Input currents are randomly initialized in range [15, 40], with relatively small differences between them (small relative evidence) to ensure that the learned connectivity is essential to exhibit winner-take-all behavior.

XOR classification As a second use case, we test whether a small network of our columnar units can be trained on classification tasks rather than continuous activity. A three-column network is trained for XOR classification, where targeting inhibitory populations in neighboring columns enables learning nonlinear functions like XOR.

The input currents the model receives are chosen as either 0Hz or approximately 20Hz to represent input types 0 and 1, respectively. The two input currents are passed to the excitatory (L4e) and inhibitory (L4i) populations in columns A and B. Input currents are weighted by learnable feedforward weights,

which represent the number of synapses x synapse strength between the presynaptic and postsynaptic columns. The firing rates of the excitatory population in L2/3 (L2/3e) of columns A, B is passed to layers L4e and L4i of the final column C. The feedforward weights from A and B to C are again learnable parameters. The final firing rates of L2/3e in column C serve as the model's output and are used for optimization with respect to the XOR targets (0, 1). Figure 1A shows the architecture of the network.

Results

Two-alternative forced choice decision-making Figure 2 shows firing rates across cortical layers for varying levels of relative evidence. Winner-take-all dynamics are successfully reproduced in L2/3, with clear firing rate separation for all nonzero relative evidence levels. L2/3 also shows faster and more pronounced divergence with increasing relative evidence, consistent with findings in superficial layers by Changdrasekaran et al. (2017). This relationship persists in L4 and L5, while L6 does not show elevated firing for the dominant column. Instead, L6 exhibits reduced decision-related activity, aligning with Changdrasekaran et al. (2017). Overall, these results demonstrate that meaningful behavior emerges across all layers, even when training only targets L2/3.

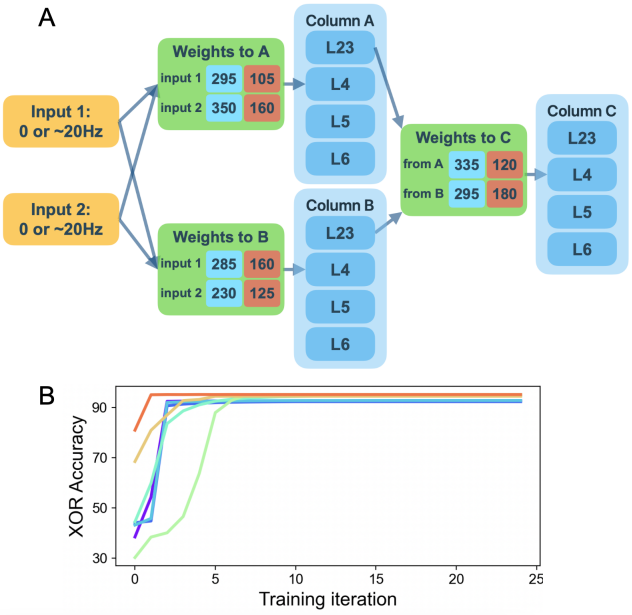

Figure 1: (A) Three-column network architecture able to perform XOR classification after learning the feedforward weights in green. Weights targeting the excitatory L4 population are colored in blue and those targeting the inhibitory population are colored in red. (B) XOR test accuracy for different random initializations, shown in distinct colors.

XOR classification Figure 1B shows rapid XOR accuracy convergence and no indication of local minima, showing feedforward weights are effectively learned to solve XOR. Figure

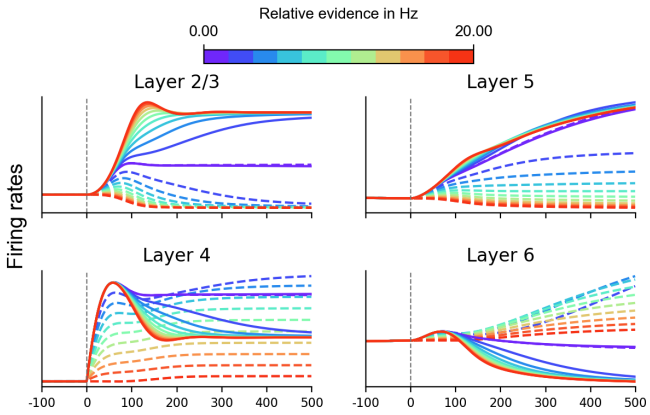

Figure 2: Laminar differences in decision-related firing rates in excitatory populations (averaged and normalized). Colors show the range of relative evidence from high (red) to low (purple). Solid lines represent the high-evidence column, dashed lines the low-evidence column.

1A shows a model solution for the feedforward weights. In this solution, column A has a higher excitatory-to-inhibitory (E-I) ratio than column B for both inputs, allowing it to activate with just one input at 20Hz. Column B, with a lower E-I ratio, requires both inputs to be active for significant output. Column A then excites column C (also via a high E-I ratio), while input from column B tends to inhibit it. Functionally, this results in column A computing the OR case, column B the AND case, and column C the OR-and-not-AND case, successfully implementing the XOR operation. These results demonstrate that the learned connectivity is both functional and interpretable, offering a window into how neural circuits might implement logical operations like XOR.

Conclusion

Here we provide proof of concept that our modeling approach can achieve functional competence by learning a connectivity structure between cortical columns in two examples. The first example shows that the model can learn to reproduce typical dynamics observed in two-alternative forced choice tasks. The second example shows that the approach remains compatible with supervised learning procedures frequently employed in neuroconnectionism, and allows interpretation of learned weights. Together, these use cases also show that functional feedforward and lateral connections between columns can be learned while intra-columnar connections remain fixed. These results suggest potential for scaling to larger networks requiring layer-specific connectivity, such as those based on predictive coding or models of mental imagery and visual perception.

Acknowledgments

This work has received funding from the European Research Council for the project "MINDSEYE" under grant agreement no. ERC-AdG-2023 (101141800).

References

Changdrasekaran, C., Peixoto, D., Newsome, W., & Shenoy, K. V. (2017). Laminar differences in decision-related neural activity in dorsal premotor cortex. *Nature Communications*, *8*, 614. doi: 10.1038/s41467-017-00715-0

Chen, R., Rubanova, Y., Bettencourt, J., & Duvenaud, D. (2019). Neural ordinary differential equations. *arXiv*. doi: 10.48550/arXiv.1806.07366

Evers, K., Peters, J., & Senden, M. (2023). Layered structure of cortex explains reversal dynamics in bistable perception. *bioRxiv*. doi: 10.1101/2023.09.19.558418

Hassabis, D., Kumaran, D., Summerfield, C., & Botvinivk, M. (2017). Neuroscience-inspired artificial intelligence. , *95*(2), 245–258. doi: 10.1016/j.neuron.2017.06.011

Kaligh-Razavi, S. M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain it cortical representation. *PLoS Computational Biology*, *10*(11). doi: 10.1371/journal.pcbi.1003915

Kriegeskorte, N. (2015). Deep neural networks: A new framework for modeling biological vision and brain information processing. *Annual review of vision science*, *1*, 417–446. doi: 10.1146/annurev-vision-082114-035447

Kriegeskorte, N., & Douglas, P. (2018). Cognitive computational neuroscience. *Nature Neuroscience*, *21*, 1148–1160. doi: 10.1038/s41593-018-0210-5

Potjans, T. C., & Diesmann, M. (2014). The cell-type specific cortical microcircuit: relating structure and activity in a full-scale spiking network model. *Cerebal Cortex*, *24*(3), 785–806. doi: 10.1093/cercor/bhs358

Wong, K., & Wang, X. (2006). A recurrent network mechanism of time integration in perceptual decisions. *Journal of Neuroscience*, *26*(4), 1314–1328. doi: 10.1523/JNEUROSCI.3733-05.2006

Yamins, D., & DiCarlo, J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, *19*, 356-365. doi: 10.1038/nn.4244