Cortically-Embedded RNNs for integration of cortex-wide neuroscience data into recurrent neural network models

Eva Sevenster* (eva.sevenster@bristol.ac.uk)
School of Engineering Mathematics and Technology, University of Bristol
Bristol, UK

Aswathi Thrivikraman* (ash.thrivikraman@bristol.ac.uk)
School of Engineering Mathematics and Technology, University of Bristol
Bristol, UK

Guy Davies (guy.davies@bristol.ac.uk)
School of Engineering Mathematics and Technology, University of Bristol
Bristol, UK

Ulysse Klatzmann (ulysseklatzmann@gmail.com)
Université de Paris Cité, Paris, France

Dabal Pedamonti (dabal.pedamonti@bristol.ac.uk)
School of Engineering Mathematics and Technology, University of Bristol
Bristol, UK

Sean Froudist-Walsh (sean.froudist-walsh@bristol.ac.uk)
School of Engineering Mathematics and Technology, University of Bristol
Bristol, UK

Abstract

Current state-of-the-art recurrent neural network models can capture complex neural dynamics during the performance of higher cognitive tasks (Yang, Joglekar, Song, Newsome, & Wang, 2019; Driscoll, Shenoy, & Sussillo, 2024). However, they largely overlook anatomy, limiting their ability to make species-specific and anatomically-precise predictions for experimentalists. Cortex-wide dynamical models increasingly integrate anatomical features including connectivity, dendritic spines and receptors (Froudist-Walsh et al., 2021; Mejias & Wang, 2022; Cabral, Hugues, Sporns, & Deco, 2011), but are incapable of solving most cognitive tasks. Here, we introduce Cortically-Embedded Recurrent Neural Networks (CERNNs), which embed artificial neural networks into a species-specific cortical space, facilitating direct comparisons to empirical neuroscience data across the entire cortex and allowing the incorporation of biologically-inspired constraints. We trained CERNNs, with macaque or human anatomy, to perform multiple cognitive tasks (e.g. working memory, response inhibition). CERNNs were trained with different architectural constraints and biologically-inspired loss functions. We evaluated CERNNs on (1) task performance, (2) alignment of connectivity with the macaque mesoscopic connectome, and (3) task-evoked activity patterns. The best performing models penalized both long-distance connections and deviations from empirical spine density. These results suggest that distributed cognitive networks may arise naturally as the brain attempts to solve complex tasks under wiring constraints with systematic gradients of single neuron properties. More broadly, CERNNs constitute a framework by which artificial neural networks can be integrated with cortex-wide neuroanatomy, physiology and imaging data to produce anatomically-specific testable hypotheses across species.

Keywords: neural network; cortex; large-scale; cognitive; RNN; multi-task

Model architecture

CERNN integrates recurrent neural networks with cortex-wide neuroscience data through four key brain-inspired constraints: 1. Artificial units (neural populations) are embedded at locations defined by the cortical geometry 2. Stimuli are input only to units in primary sensory areas 3. Output is read only from units in frontal eye fields (FEF, for tasks requiring a saccadic response, as in this work) or primary motor cortex (otherwise). 4. Penalties are imposed during the learning process for proposing biologically-unrealistic solutions.

To ensure that only units in V1 or S1 receive input, and only units in FEF produce output, we define three binary masks M_{vis}, M_{som}, and M_{out}. Each mask is 1 for units in its corresponding cortical subset (V1, S1, or FEF/M1) and 0 otherwise. We then form the masked weight matrices by elementwise (Hadamard) multiplication:

$$\widetilde{W}_{\text{in}} = W_{\text{in}} \circ \left(M_{\text{vis}} + M_{\text{som}}\right), \quad \widetilde{W}_{\text{out}} = W_{\text{out}} \circ M_{\text{out}}.$$

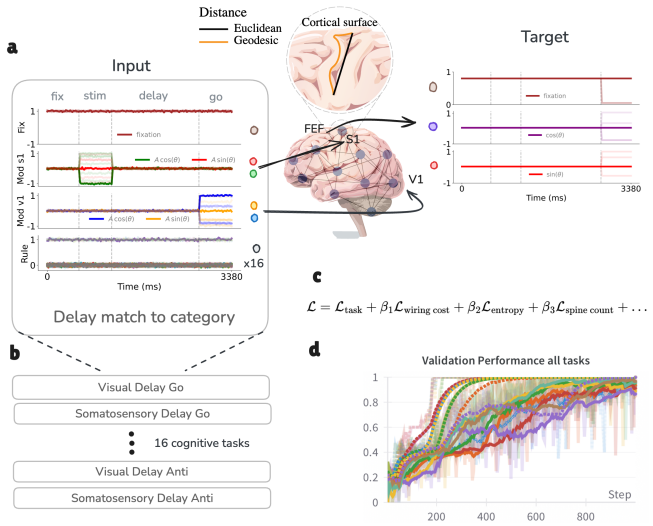

Figure 1: CERNN architecture and training approach. a) Recurrently connected units each have a spatial location on the cortical surface. Inputs target only primary sensory areas, and saccadic responses are read only from Frontal Eye Fields. b) Single CERNNs were trained to perform multiple (16-26) neuroscience tasks. c) CERNN models were trained to balance task performance with biologically-inspired constraints. d) CERNN models achieved good performance on all tasks (shown in different colours).

Let $\mathbf{h}(t) \in \mathbb{R}^N$ be the hidden state vector (including all cortical locations). Neural dynamics unfold following:

$$\tau \frac{d\mathbf{h}}{dt} = -\mathbf{h}(t) + \text{ReLU}\Big(W_{\text{rec}}\,\mathbf{h}(t) + \tilde{W}_{\text{in}}\,\mathbf{x}(t), + b_{\text{in}} + \xi(t)\Big),$$

where $\mathbf{x}(t) \in \mathbb{R}^D$ is the external input.

The output $\hat{\mathbf{y}}(t) \in \mathbb{R}^K$ is obtained by:

$$\hat{\mathbf{y}}(t) = \tilde{W}_{\text{out}}\,\mathbf{h}(t).$$

This ensures that the readout is exclusively from FEF units.

This set-up forces the network to deal with brain-like problems that are alien to typical RNNs, such as the propagation of sensory activity across the cortex to the frontal eye fields.

Anatomical locations

The spatial location of recurrently connected units was taken from standard group average structural MRI scans from the macaque (Yerkes19) (Donahue et al., 2016) and human (Human Connectome Project FS-LR) (Elam et al., 2021). Within each standard surface, we placed units at locations at the centre of cortical areas defined by popular parcellations for each species (Markov et al., 2014; Glasser et al., 2016). We calculated the geodesic distances between pairs of units along the cortical surface, to penalize long-distance connections (Achterberg, Akarca, Strouse, Duncan, & Astle, 2023).

Integration of dendritic spine count gradients in the macaque and human

Dendritic spine counts are the locations of excitatory synaptic connections, and in primates there is a systematic increasing gradient of spines along the cortical hierarchy (Elston, 2007). These spine gradients are critical for cortex-wide dynamical models to reproduce realistic cognitive activity patterns (Froudist-Walsh et al., 2021; Mejias & Wang, 2022). We aimed to capture this spine count gradient in the CERNNs.

In the macaque, we used spine-count data from 27 regions (Elston, 2007), and inferred the spine count for the remaining regions based on the cortical hierarchy (Froudist-Walsh et al., 2021). For the human, where spine data is sparsely available, we capitalized on the inverse correlation between the T1w/T2w ratio from structural imaging data and the spine count (Pereira-Obilinovic, Froudist-Walsh, & Wang, 2024) to infer the estimated spine count. In model variants with the spine loss L_{spine}, the models were penalized during training if the total absolute incoming connections to an area deviated from the spine count.

Training CERNNs to solve cognitive tasks

We compared CERNNs trained to perform 12-26 cognitive tasks while balancing task performance with distinct biologically-inspired loss functions (e.g. with wiring cost minimisation, entropy maximisation, Figure 1), and tested which trained networks most closely match the empirical mesoscopic connectivity of the macaque cortex.

We define the task loss L_{task} as the time-averaged mean squared error between the network's output $\hat{y}(t)$ and the target output $y^*(t)$. Let i index the output units and t index time. We trained CERNNs with backpropagation-through-time to minimise the total loss L, which contained L_{task} and the other brain-inspired losses (Figure 1). After training, many single CERNN models with anatomical constraints learned the tasks well (above 95% performance).

Comparison with macaque connectivity data

Macaque CERNN model connectivity weights were compared to retrograde tract-tracing data (Markov et al., 2014), specifically the complete 40-area subgraph (Froudist-Walsh et al., 2021). Human CERNN model connectivity weights were compared to diffusion MRI tractography data from the Human Connectome Project (Demirtaş et al., 2019).

In networks that penalized long-distance connections with an L_2-style regularizer, and penalized deviations from the spine count, we observed several salient features from real brain connectivity. 1) Both macaque and human data demonstrated an exponential decay of connectivity strength with distance. Notably, this was only observed for L_2-style distance penalties, and not L_1-style, as previously proposed (Achterberg et al., 2023). 2) The density of the human matrix was considerably lower than the macaque matrix, as predicted by comparative neuroanatomy studies (Magrou et al., 2024). 3) Higher cortical areas formed strong recurrently connected networks (Markov et al., 2014).

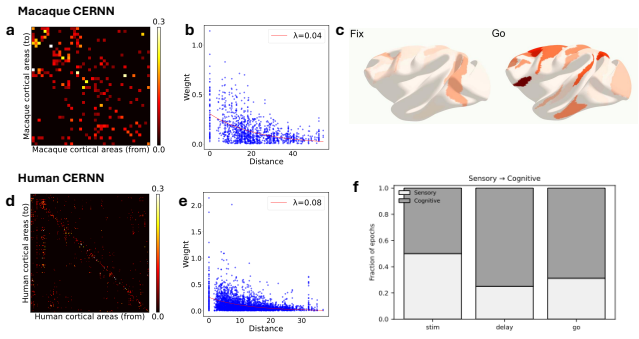

Figure 2: CERNN connectivity, task performance and example cortical activity patterns. Trained CERNN models show realistic sparsity in the connectivity for both Macaques (a) and Humans (d), and exponential decay with distance (d), (e). (c) CERNNs solved the cognitive tasks through distributed cortical activity. (f) Delay-period activity is characterised by distributed network activity distant from primary sensory areas. Here activity patterns across are classified as sensory or cognitive in different task periods by comparison with canonical resting-state networks (Yeo et al., J. Neurophysiol. 2011).

Conclusions

CERNN is a framework that can be extended to integrate many other types of neuroscience data (e.g. receptor densities and anatomical connectivity). By adapting this framework to the comparison of multiple species (and comparison with cross-species data), CERNNs can be a valuable tool to identify species-specific and general cognitive mechanisms.

Through continued integration with neuroscience data, CERNNs will enable prediction of cortex-wide mechanisms of cognition that are species-specific and anatomically-precise. This will accelerate the 'virtuous cycle' between model-guided cortex-wide experiments and experimentally-driven model improvements.

References

Achterberg, J., Akarca, D., Strouse, D. J., Duncan, J., & Astle, D. E. (2023, December). Spatially embedded recurrent neural networks reveal widespread links between structural and functional neuroscience findings. *Nature Machine Intelligence*, *5*(12), 1369–1381. Retrieved 2024-02-06, from https://www.nature.com/articles/s42256-023-00748-9 (Number: 12 Publisher: Nature Publishing Group) doi: 10.1038/s42256-023-00748-9

Cabral, J., Hugues, E., Sporns, O., & Deco, G. (2011, July). Role of local network oscillations in resting-state functional connectivity. *NeuroImage*, *57*(1), 130–139. Retrieved 2023-04-19, from https://www.sciencedirect.com/science/article/pii/S105381 doi: 10.1016/j.neuroimage.2011.04.010

Demirtaş, M., Burt, J. B., Helmer, M., Ji, J. L., Adkinson, B. D., Glasser, M. F., ... Murray, J. D. (2019, March). Hierarchical heterogeneity across human cortex shapes large-scale neural dynamics. *Neuron*, *101*(6), 1181–1194.e13. Retrieved 2021-04-28, from https://www.sciencedirect.com/science/article/pii/S089662 doi: 10.1016/j.neuron.2019.01.017

Donahue, C. J., Sotiropoulos, S. N., Jbabdi, S., Hernandez-Fernandez, M., Behrens, T. E., Dyrby, T. B., ... Glasser, M. F. (2016). Using diffusion tractography to predict cortical connection strength and distance: a quantitative comparison with tracers in the monkey. *J. Neurosci.*, *36*, 6758–6770.

Driscoll, L. N., Shenoy, K., & Sussillo, D. (2024, July). Flexible multitask computation in recurrent networks utilizes shared dynamical motifs. *Nature Neuroscience*, *27*(7), 1349–1363. Retrieved 2024-08-10, from https://www.nature.com/articles/s41593-024-01668-6 (Publisher: Nature Publishing Group) doi: 10.1038/s41593-024-01668-6

Elam, J. S., Glasser, M. F., Harms, M. P., Sotiropoulos, S. N., Andersson, J. L. R., Burgess, G. C., ... Van Essen, D. C. (2021, December). The Human Connectome Project: A retrospective. *NeuroImage*, *244*, 118543. Retrieved 2022-04-19, from https://www.sciencedirect.com/science/article/pii/S105381 doi: 10.1016/j.neuroimage.2021.118543

Elston, G. N. (2007). Specialization of the neocortical pyramidal cell during primate evolution. In *Evolution of Nervous Systems* (pp. 191–242). Elsevier. Retrieved 2019-11-08, from https://linkinghub.elsevier.com/retrieve/pii/B012370878 doi: 10.1016/B0-12-370878-8/00164-6

Froudist-Walsh, S., Bliss, D. P., Ding, X., Rapan, L., Niu, M., Knoblauch, K., ... Wang, X.-J. (2021, November). A dopamine gradient controls access to distributed working memory in the large-scale monkey cortex. *Neuron*, *109*(21), 3500–3520.e13. Retrieved 2024-11-07, from https://www.cell.com/neuron/abstract/S0896-6273(21)00621- (Publisher: Elsevier) doi: 10.1016/j.neuron.2021.08.024

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., ... Van Essen, D. C. (2016, August). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171–178. Retrieved 2021-01-21, from https://www.nature.com/articles/nature18933 (Number: 7615 Publisher: Nature Publishing Group) doi: 10.1038/nature18933

Magrou, L., Joyce, M. K. P., Froudist-Walsh, S., Datta, D., Wang, X.-J., Martinez-Trujillo, J., & Arnsten, A. F. T. (2024, May). The meso-connectomes of mouse, marmoset, and macaque: network organization and the emergence of higher cognition. *Cerebral Cortex*, *34*(5), bhae174. Retrieved 2024-05-27, from https://doi.org/10.1093/cercor/bhae174 doi: 10.1093/cercor/bhae174

Markov, N. T., Ercsey-Ravasz, M. M., Ribeiro Gomes, A. R., Lamy, C., Magrou, L., Vezoli, J., ... Kennedy, H. (2014, January). A weighted and directed interareal connectivity matrix for macaque cerebral cortex. *Cerebral Cortex*, *24*(1), 17–36. Retrieved 2020-04-14, from https://academic.oup.com/cercor/article/24/1/17/272931 doi: 10.1093/cercor/bhs270

Mejias, J. F., & Wang, X.-J. (2022, February). Mechanisms of distributed working memory in a large-scale network of macaque neocortex. *eLife*, *11*, e72136. Retrieved 2022-03-02, from https://doi.org/10.7554/eLife.72136 doi: 10.7554/eLife.72136

Pereira-Obilinovic, U., Froudist-Walsh, S., & Wang, X.-J. (2024, December). *Cognitive network interactions through communication subspaces in large-scale models of the neocortex.* bioRxiv. Retrieved 2024-12-13, from https://www.biorxiv.org/content/10.1101/2024.11.01.621513v3 (Pages: 2024.11.01.621513 Section: New Results) doi: 10.1101/2024.11.01.621513

Yang, G. R., Joglekar, M. R., Song, H. F., Newsome, W. T., & Wang, X.-J. (2019, February). Task representations in neural networks trained to perform many cognitive tasks. *Nature Neuroscience*, *22*(2), 297–306. Retrieved 2022-04-28, from https://www.nature.com/articles/s41593-018-0310-2 (Number: 2 Publisher: Nature Publishing Group) doi: 10.1038/s41593-018-0310-2