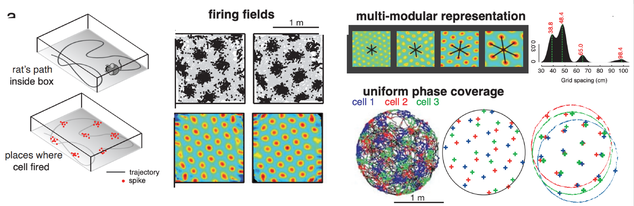
Self-Supervised Learning of Representations for Space Generates Multi-Modular Grid Cells

Rylan Schaeffer*, Mikail Khona*, Tzuhsuan Ma, Cristobal Eyzaguirre, Sanmi Koyejo, Ila R Fiete
rschaef@cs.stanford.edu, mikail@mit.edu, *denotes equal contribution

Introduction: Grid cells

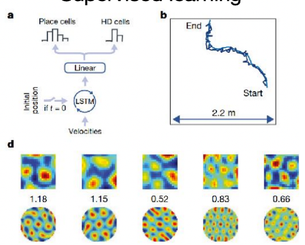
*Q: What are good neural representations for **spatial navigation?***



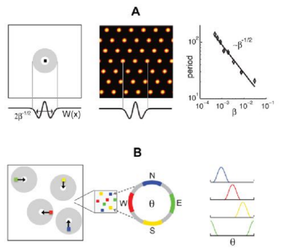
*Q:Why has the **mammalian lineage** evolved this multi-periodic, non-local representation of space, which is an aperiodic and local variable?*

Previous Approaches

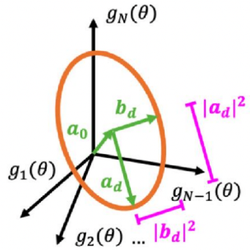
Supervised learning



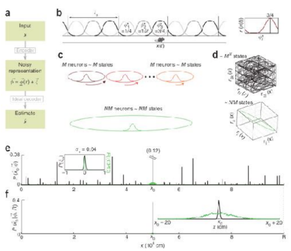
Continuous attractors



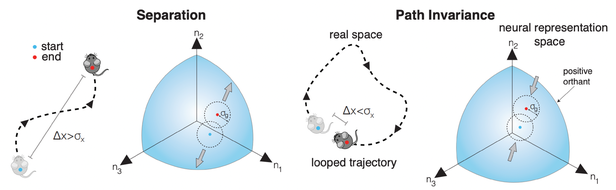
Basis function optimization



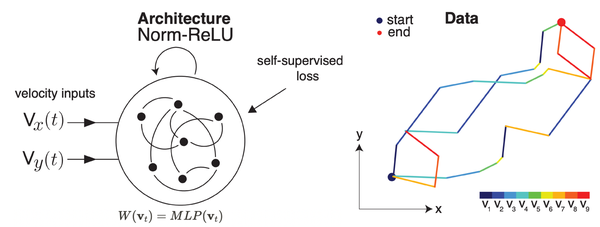
Coding theory



Self-Supervised Learning Loss for Space



Architecture and Training Data

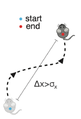


$\sigma(\cdot) = Norm(ReLU(\cdot)) = ReLU(\cdot) / ||ReLU(\cdot)||$

Normalization of neural population activity: prevent trivial solutions often found by contrastive SSL

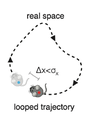
$$\mathcal{D}_{\text{gradient step}} = \left\{ (\mathbf{v}_{\pi_b(1)}, \mathbf{v}_{\pi_b(2)}, \ldots, \mathbf{v}_{\pi_b(T)}), (\mathbf{g}_{\pi_b(1)}, \mathbf{g}_{\pi_b(2)}, \ldots, \mathbf{g}_{\pi_b(T)}) \right\}_{b=1}^{B}$$

with shared initial state $\mathbf{g}_0$



$$\mathcal{L}_{Sep} = \sum_{\substack{\forall \pi_b, \pi_{b'}, t, t': \\ ||\mathbf{x}_{\pi_{b'}(t)} - \mathbf{x}_{\pi_b(t')}||_2 > \sigma_x}} \exp\left(- \frac{||\mathbf{g}_{\pi_{b'}(t)} - \mathbf{g}_{\pi_b(t')}||_2^2}{2\sigma_g^2} \right)$$

1 coarse grained bit of information about *relative*, and not absolute, spatial location



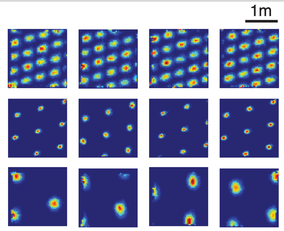
$$\mathcal{L}_{Inv} = \sum_{\substack{\forall \pi_b, \pi_{b'}, t, t': \\ ||\mathbf{x}_{\pi_{b'}(t)} - \mathbf{x}_{\pi_b(t')}||_2 < \sigma_x}} \left|\left| \mathbf{g}_{\pi_b(t)} - \mathbf{g}_{\pi_{b'}(t')} \right|\right|_2^2$$

Regularization

$$\mathcal{L}_{Cap} = -\left|\left| \frac{1}{BT} \sum_{\pi_b, t} \mathbf{g}_{\pi_b(t)} \right|\right|_2^2$$

$$\mathcal{L}_{ConIso} \stackrel{\text{def}}{=} \mathbb{V}\left[\left\{ \frac{||\boldsymbol{g}_t - \boldsymbol{g}_{t-1}||}{||\boldsymbol{v}_t||} \right\}_{t\,:\,0 < ||\boldsymbol{v}_t|| < \sigma_x} \right]$$

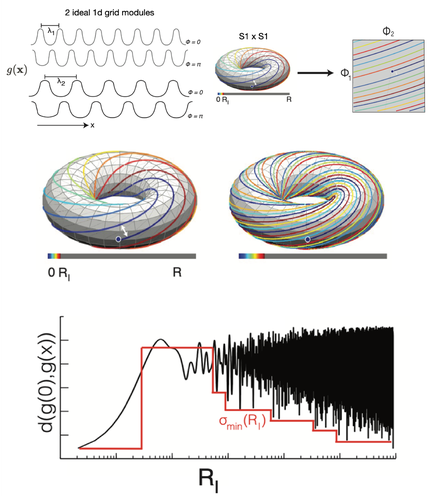
Result: Multi-Modular Grid Cells

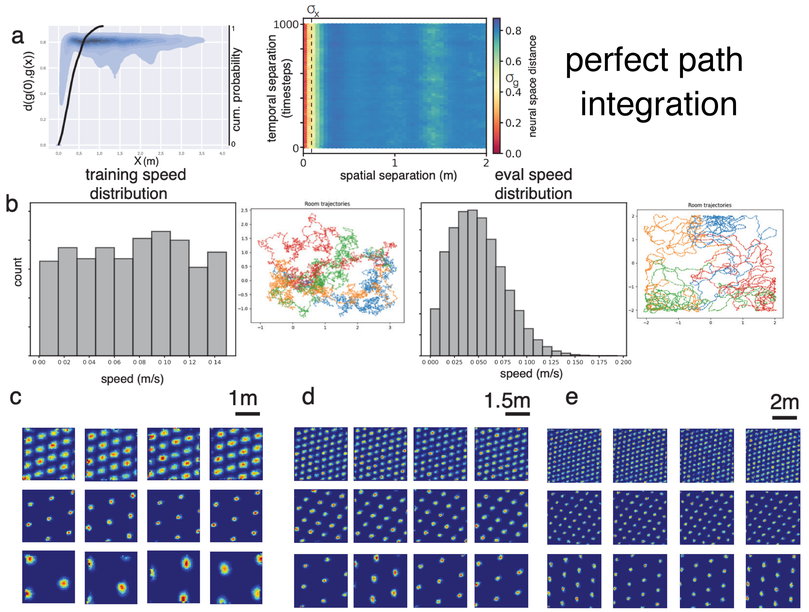


12 cells from 128 unit RNN

Out of Distribution Generalization

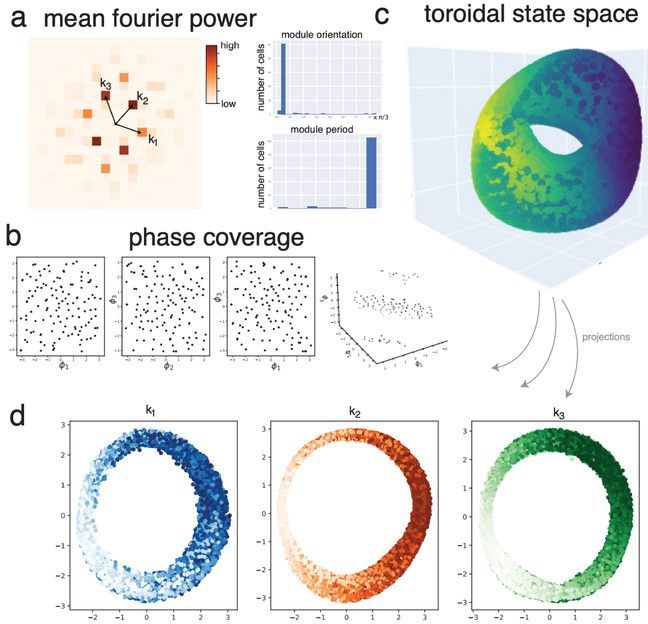
High-capacity state space





perfect path integration

Mechanistic Analyses Reveal Toroidal Topology



Conclusions

- Grid cells emerge from learning a self-consistent, high capacity representation of space in a recurrent neural network.
- SSL is useful for explaining neural responses beyond sensory brain regions.

References

1 Nobel Prize for Physiology/Medicine 2014.
2. Burak and Fiete, PLOS Comp. Bio (2009)
3. Sreenivasan et al, Nat. neurosci. (2011)
4. Sorscher, Mel et al, Neuron (2023)
5. Dorrell et al (ICLR 2023)
6. Schaeffer et al (NeurIPS 2022)
7. Banino et al, Nature (2018)
8. Xu, Gao et al (2022)