A Meta-Learned Neural Network Model of the Hippocampus with Brain-like Hierarchically-Structured Sparsity and Learning Rate Patterns

Joy Gong[1], Zhenglong Zhou[2], Anna Schapiro[2]

[1]University of Pennsylvania, College of Arts and Sciences Class of 2025, [2]Department of Psychology, University of Pennsylvania

Introduction

Multiple Learning Systems

- Computations in the brain are efficient, but energy consumption is high in deep neural networks
- The brain consists of hierarchically structured multiple learning systems, which manage episodic and statistical learning
  - Monosynaptic pathway (MSP): statistical, overlapping, slow learning
  - Trisynaptic pathway (TSP): episodic, sparse, fast learning
- It is unclear how the brain arrives at multiple learning systems. What gives rise to these?

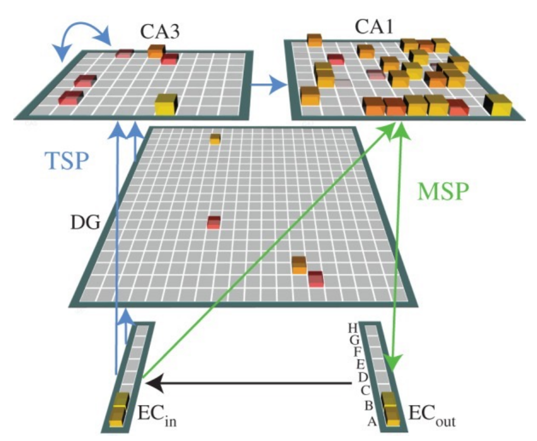


Figure 1. A deep neural network of the MSP and TSP. Adapted from "Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning," Schapiro *et al.,* 2017.

Meta-Learning

- Colloquially known as "learning to learn"
- Humans learn new concepts very efficiently. Is it possible for machines to do the same?
- Continual learning — where a model is fed a large stream of data needs to learn without forgetting — is difficult due to catastrophic forgetting or interference
  - A replay buffer and sparse representations can help

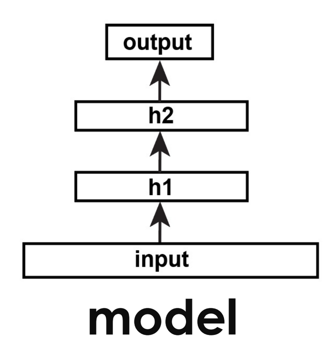


Figure 2. Our initial proposed meta-learned model.

Proposal

- Meta-learn parameters *m* (modulates within-layer k-winners-take-all) & *LR* (per-layer learning rate)

Methods

1. Initial baselines

- Base code off La-MAML, a continual online meta-learning model with a small replay buffer of examples from the past
- Bi-level optimization: outer (meta) loss which updates hyperparameters, and inner loss (traditional optimization)
- Train on 20 MNIST rotation tasks, each a 10-way classification
- Analyze sparse representations of weights and biases, unit variance of hidden layers

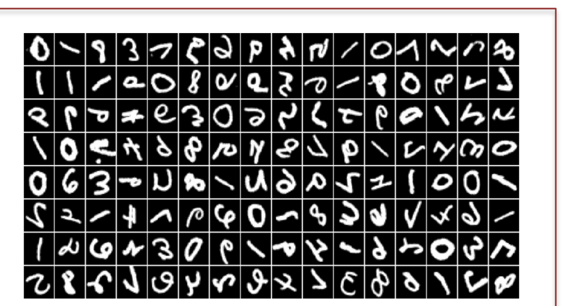


Figure 3. A subset of the MNIST rotation dataset. Each digit (0-9) has 20 different angles of rotation.

2. K-winners-take-all & per-layer learning rate

- Induce sparsity/inhibition in each layer through a meta-learned continuous multiplier; top k neurons are active
- Learn per-layer learning rates to mirror pathways in the brain

$$\mathbf{a}^* = [\mathbf{a} - I]_+$$
$$I = \text{descending-sort}([\mathbf{a}]_+)_{(k_t+1)}$$
$$k_t = \max\left(k_{\text{target}}, \lfloor k_{\max} - E_t(k_{\max} - k_{\text{target}})/s \rfloor\right),$$

$$\mathbf{a}^* = [\mathbf{a} - I]_+$$
$$I = \text{descending-sort}([\mathbf{a}]_+)_{(k_t+1)}$$
$$I = m * I$$

Figure 4. a) Traditional k-wta implementation. b) Our modification, with a meta-learned continuous multiplier *m*.

3. Various architectures

- 2-layer
- 4-layer
- 4-layer skip two-path; mimics two learning pathways
- 4-layer split two-path; mimics two learning pathways

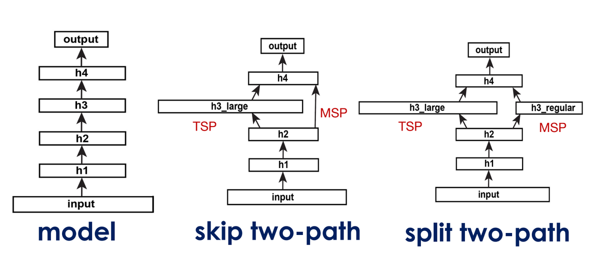
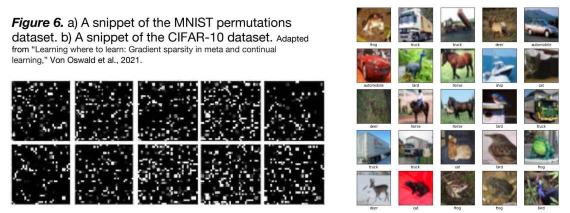


Figure 5. a) A 4-layer model. b) A skip two-path cortical-hippocampal architecture with a large h3 layer. c) A split two-path cortical hippocampal architecture with two options for h3 layer size.

4. New tasks and datasets

- Use task incremental learning instead of class incremental learning
- Add a set or individual label to each image; have model identify both image and label
- Test on MNIST permutations and CIFAR datasets (real world images)

Figure 6. a) A snippet of the MNIST permutations dataset. b) A snippet of the CIFAR-10 dataset. Adapted from "Learning where to learn: Gradient sparsity in meta and continual learning," Von Oswald et al., 2021.



Results

Figure 7. In a 2-layer network, meta-learning the inhibition multiplier *m* and learning rate LR leads to a, b) efficient task representations and c) benefits ANN performance.

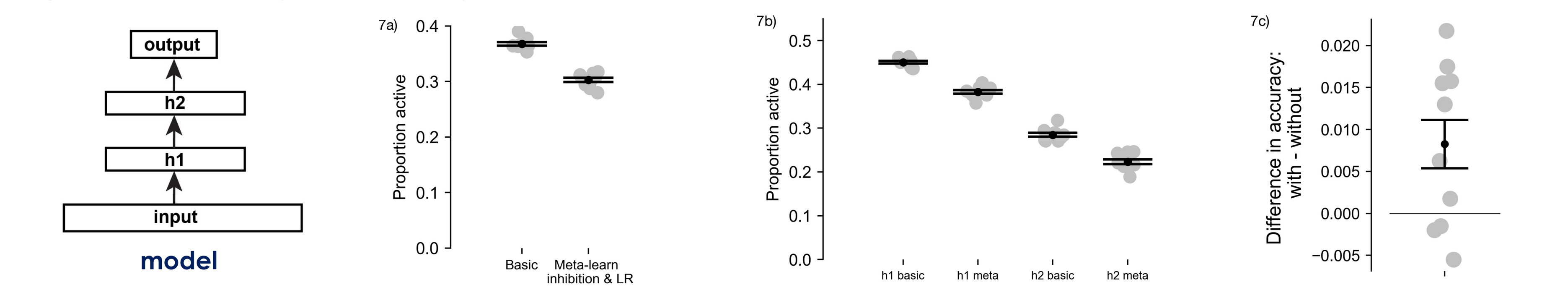


Figure 8. In a meta-learned 2-layer network, hierarchically structured a) sparsity and b) learning rates emerge.

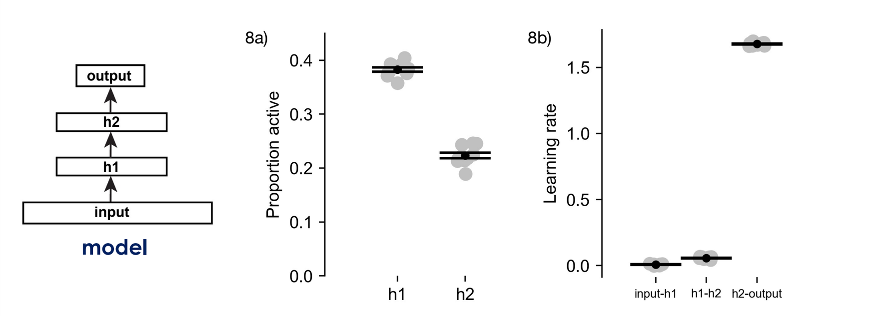


Figure 9. In a deeper 4-layer network, similar patterns of a) sparsity and b) learning rate are achieved through meta learning.

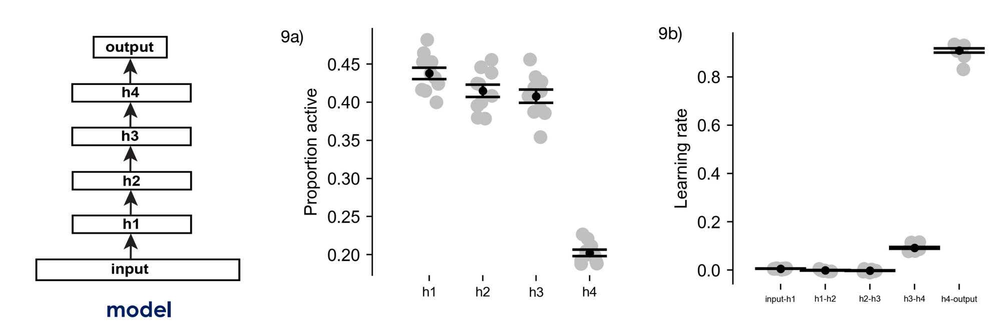


Figure 10. In the skip two-path architecture, the large intermediate layer h3_large becomes sparser over tasks through meta-learning and remains sparsest of all layers. a) Average sparsity for all layers. b) Sparsity over task index.

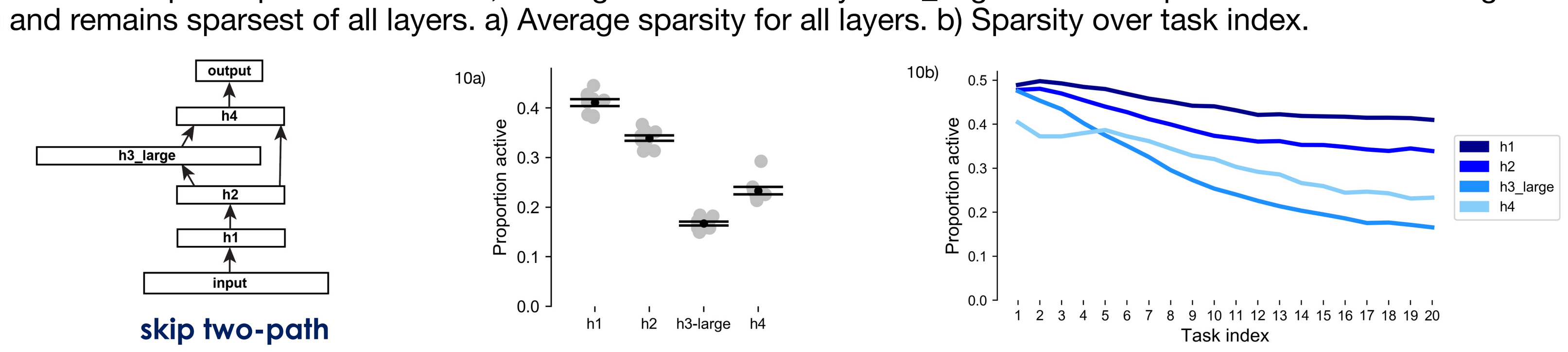


Figure 11. In the skip two-path architecture, higher learning rates for the large intermediate layer h3_large emerge through meta-learning. a) Average *LR* for all layers. b) Average *LR* for all layers except for the last. c) *LR* over task index.

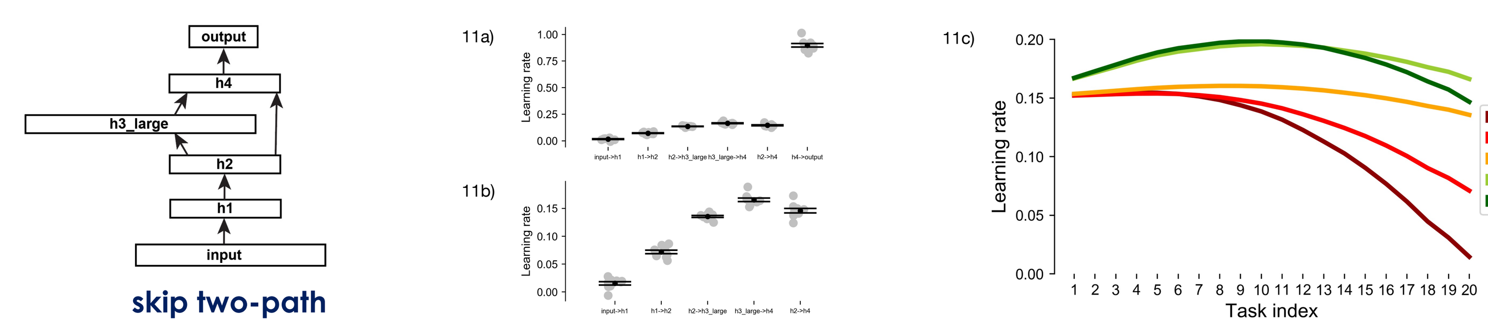


Figure 12. In the split two-path architecture, h3_large exhibits a higher level of sparsity than h3_regular and becomes increasingly sparse through meta-learning. a) Average sparsity for all layers. b) Sparsity of task index.

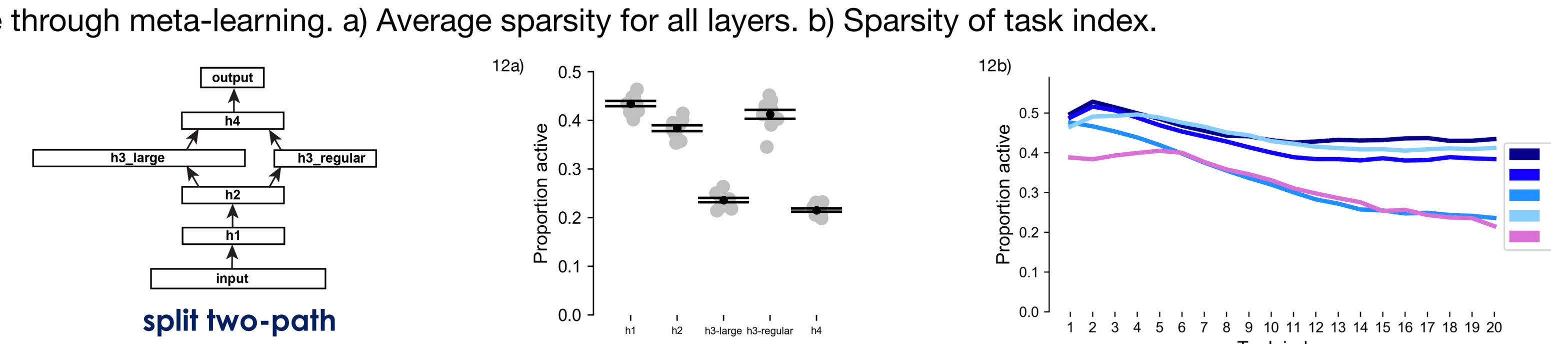


Figure 13. In the split two-path architecture, h3_large decorrelates task representations more than h3_regular, similar to the TSP vs. MSP.

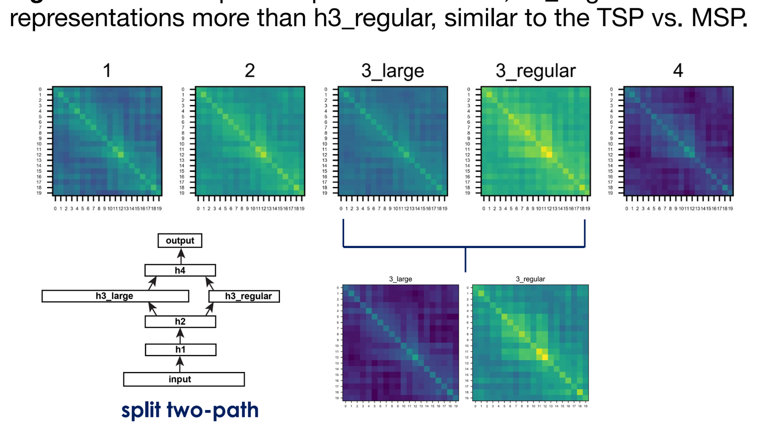
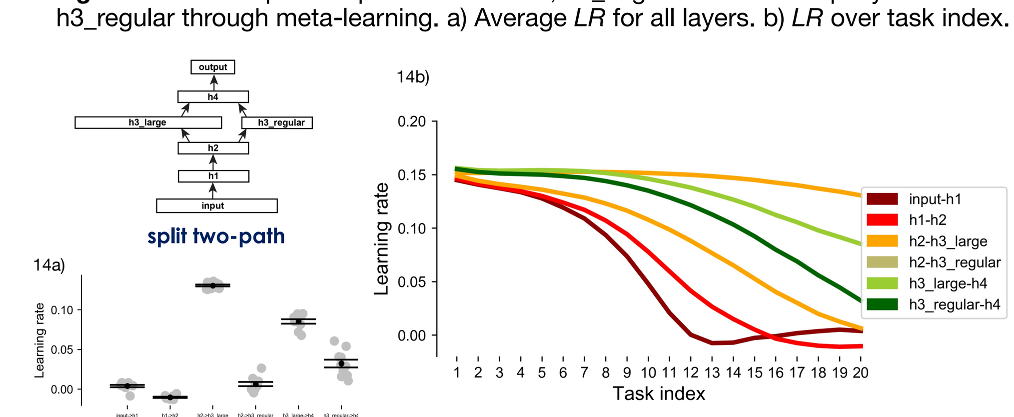


Figure 14. In the split two-path architecture, h3_large learns more rapidly than h3_regular through meta-learning. a) Average *LR* for all layers. b) *LR* over task index.



Conclusions

A Brain-like Model

- Meta-learning parameters that modulate interference and learning results in efficient, sparse task representations
- Resulting architecture shows a hierarchal differentiation of sparsity and learning rates that resembles the brain!
  - MSP vs. TSP
  - Hippocampus vs. neocortex

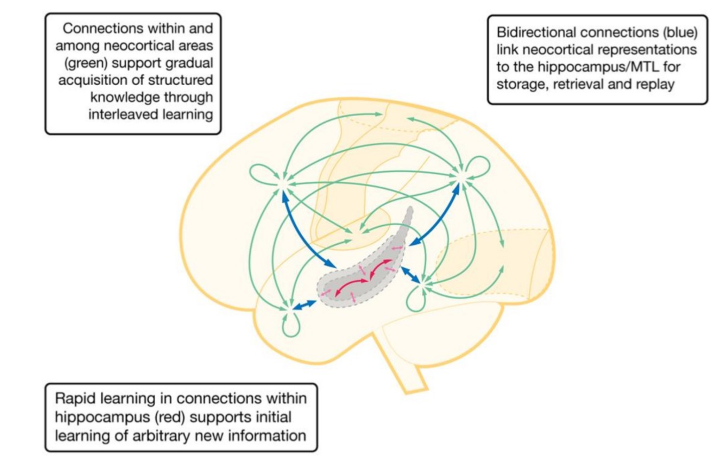
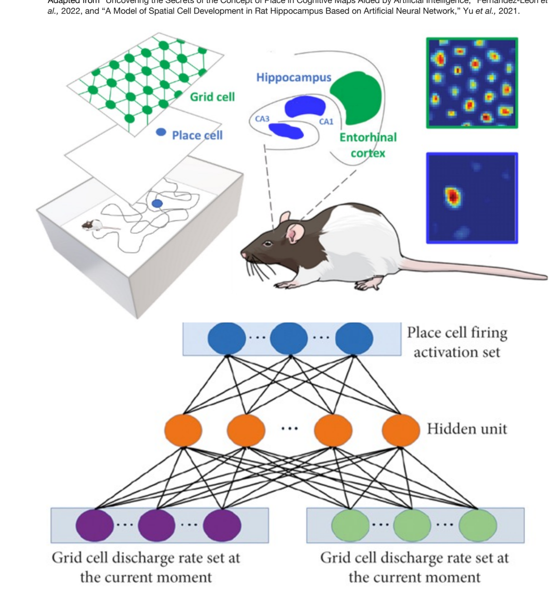


Figure 15. Hippocampal-cortical interactions with both rapid and gradual learning, of which the complementary learning systems within the hippocampus are a microcosm. Adapted from "What Learning Systems do Intelligent Agents Need? Complementary Learning Systems Theory Updated," Kumaran *et al.,* 2016.

Future Questions

- Is sparsity (% neurons active) dependent on the size of the layer?
- Why is learning rate generally highest in projections to the last layer?
- Is it possible to learn the optimal architecture of a model, including number of layers, size of layers, and pathway projections?
- Will meta-learning these parameters result in similar patterns and performance on real world images?
- Are there sparse representations in the widely used CNN and transformer (e.g., Chat GPT) models?

Figure 16. a) Grid cell and place cell systems, locations, and firing fields in the brain. b) A plausible neural network model of the place-grid cell system. Adapted from "Uncovering the Secrets of the Concept of Place in Cognitive Maps Aided by Artificial Intelligence," Fernandez-Leon *et al.,* 2022, and "A Model of Spatial Cell Development in Rat Hippocampus Based on Artificial Neural Network," Yu *et al.,* 2021.



Broader Implications

- This work provides a general framework for neural architectures in the brain, such as:
  - Visual and auditory processing systems
  - Grid and place cell (spatial navigation) systems
  - Hippocampal-neocortical learning during sleep
- Few-shot and reinforcement learning, as well as natural language processing, can be optimized with meta-learning methods

References

1. Gupta, G., Yadav, K., & Paull, L. (2020). Look-ahead meta learning for continual learning. *Advances in Neural Information Processing Systems, 33*, 11588-11598.
2. Javed, K., & White, M. (2019). Meta-learning representations for continual learning. *Advances in neural information processing systems*, *32*.
3. Kumaran, D., Hassabis, D., & McClelland, J.L. (2016). What Learning Systems do Intelligent Agents Need? Complementary Learning Systems Theory Updated. *Trends in Cognitive Sciences*, *20*, 512-534.
4. Riemer, M., Cases, I., Ajemian, R., Liu, M., Rish, I., Tu, Y., & Tesauro, G. (2018). Learning to learn without forgetting by maximizing transfer and minimizing interference. *arXiv preprint arXiv:1810.11910.*
5. Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *372*(1711), 20160049. https://doi.org/10.1098/rstb.2016.0049
6. Von Oswald, J., Zhao, D., Kobayashi, S., Schug, S., Caccia, M., Zucchet, N., & Sacramento, J. (2021). Learning where to learn: Gradient sparsity in meta and continual learning. *Advances in Neural Information Processing Systems*, *34*, 5250-5263.

Acknowledgements

- This project was funded by PI Anna Schapiro and the Penn Computational Cognitive Neuroscience Lab.
- Special thanks to graduate student Zhenglong Zhou for his leadership and mentorship on this project, as well as all members of the Schapiro Lab for their invaluable support.