NEURAL INFORMATION PROCESSING SYSTEMS

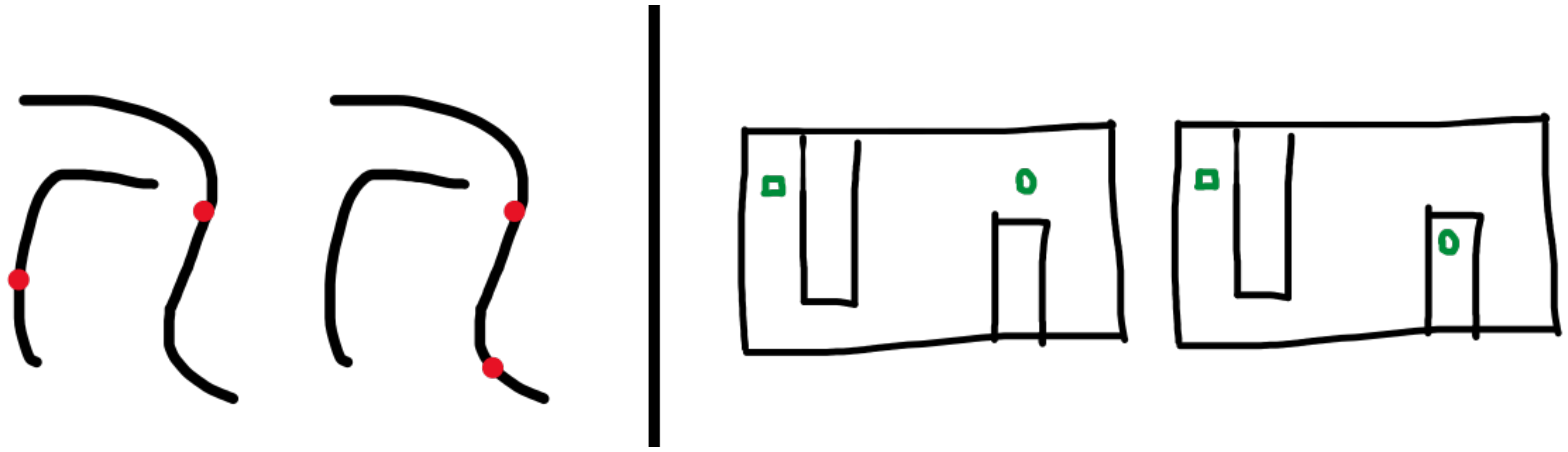
Adaptive recurrent vision performs zero-shot computation scaling to unseen difficulty levels

Vijay Veerabadran*, Srinivas Ravishankar*, Yuan Tang, Ritik Raina, Virginia R. de Sa

University of California San Diego

37th Conference on Neural Information Processing Systems (NeurIPS 2023)

Spatial visual reasoning



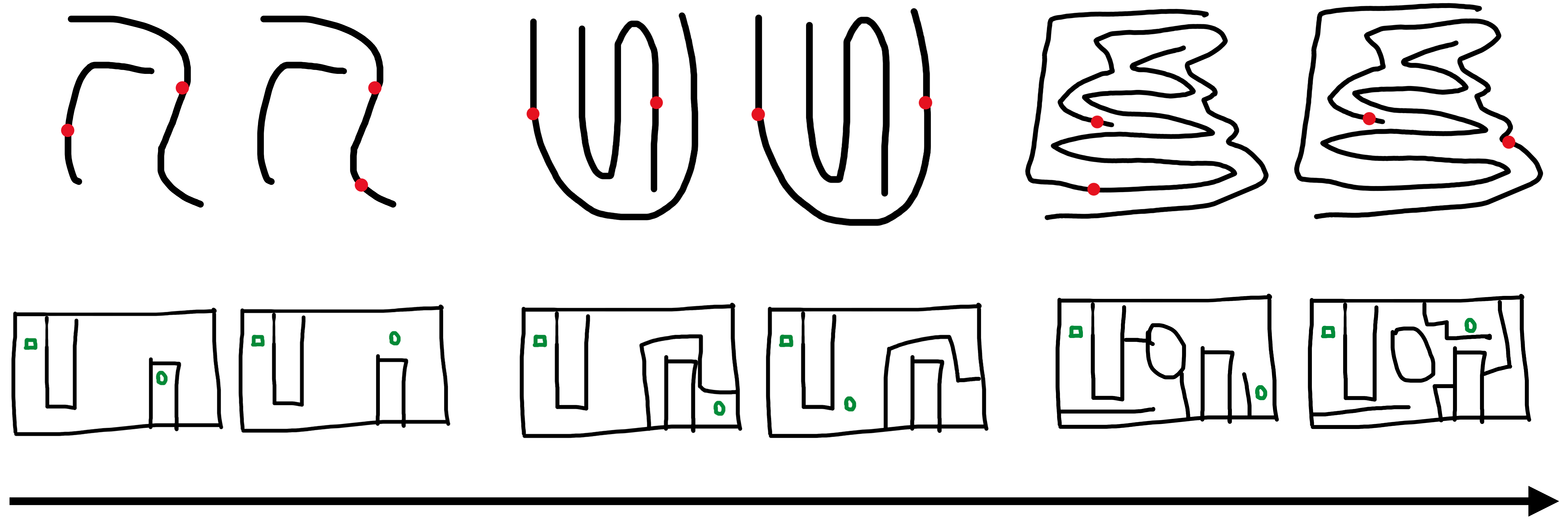
A) Curve tracing

B) Path integration

Spatial visual reasoning

Curve tracing

Path integration



Increasing task difficulty

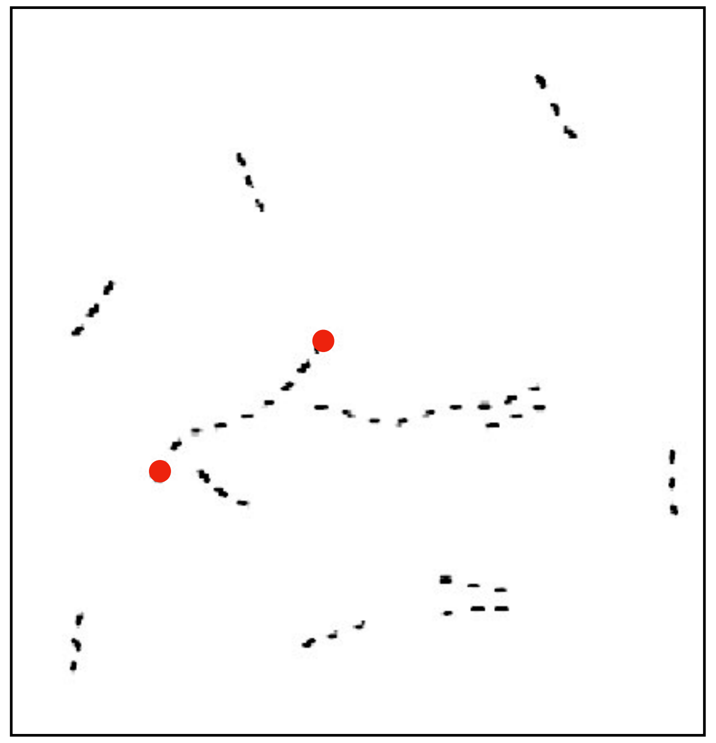
Human vision generalizes across difficulty levels in a zero-shot manner.

Ullman, S. (1987). Visual routines. In Readings in computer vision (pp. 298-328). Morgan Kaufmann.

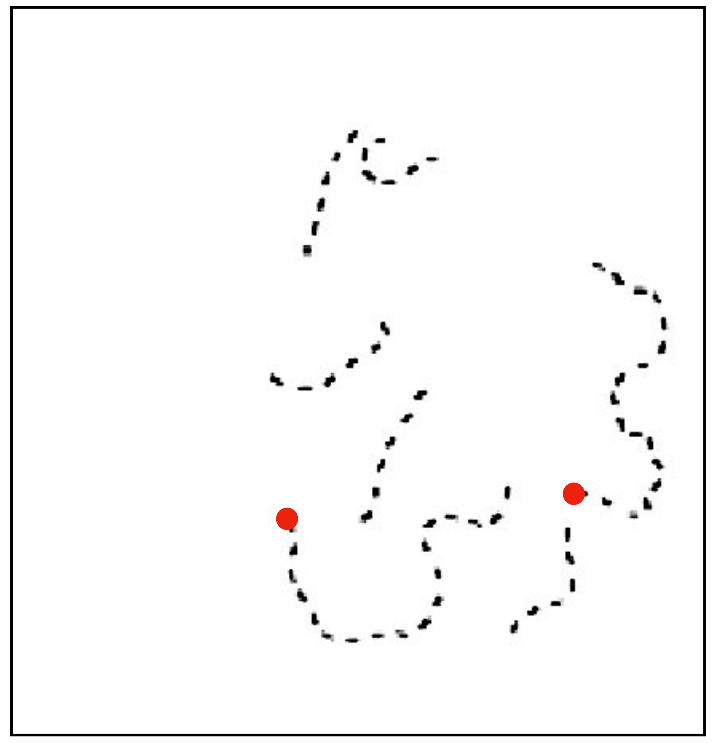
Can neural network models of visual processing show such generalization?

Datasets: PathFinder and Mazes

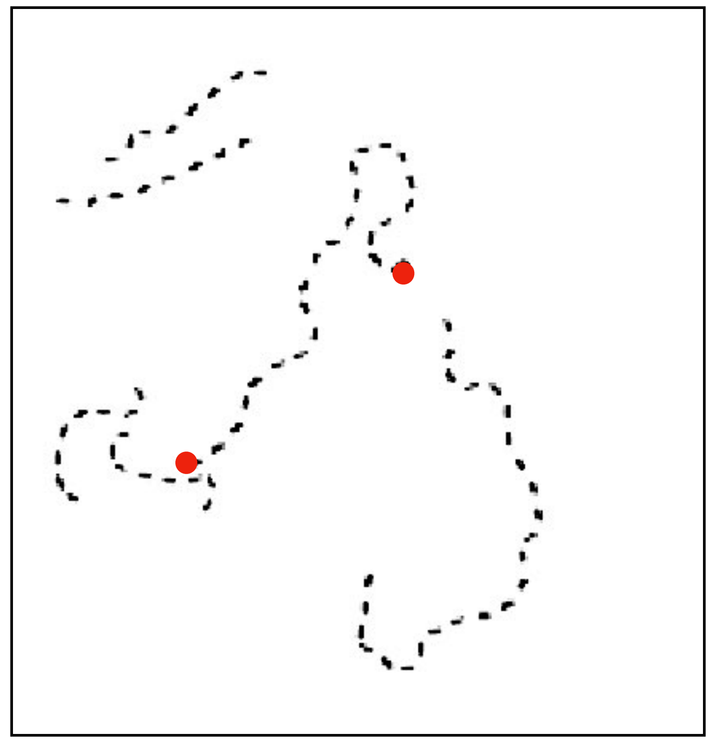
Example images from *PathFinder** (classification)



PathFinder-9



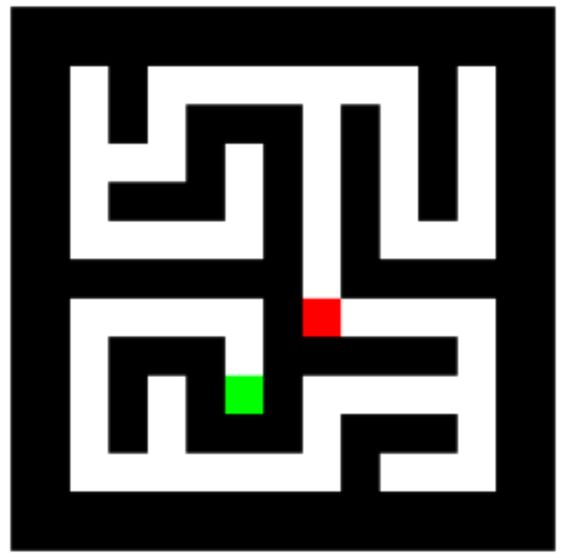
PathFinder-14



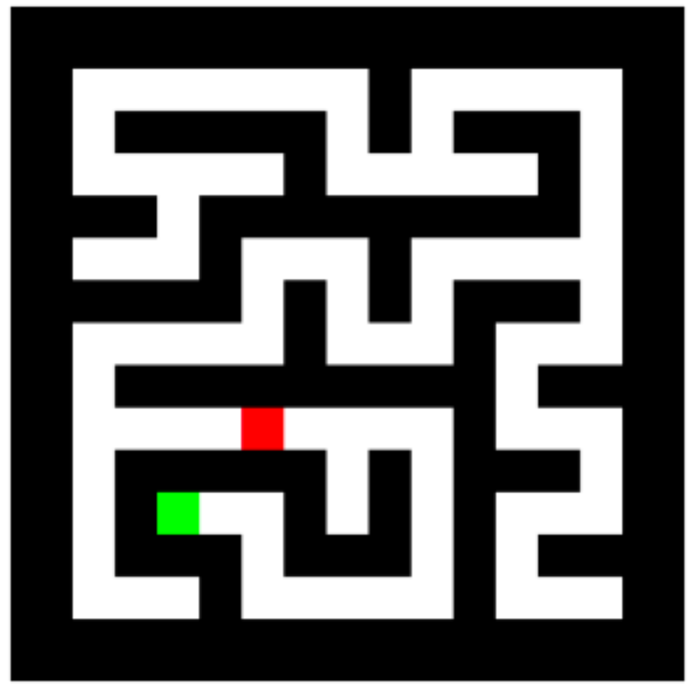
PathFinder-18

Example images from *Mazes*† (segmentation)

Mazes-S



Mazes-M



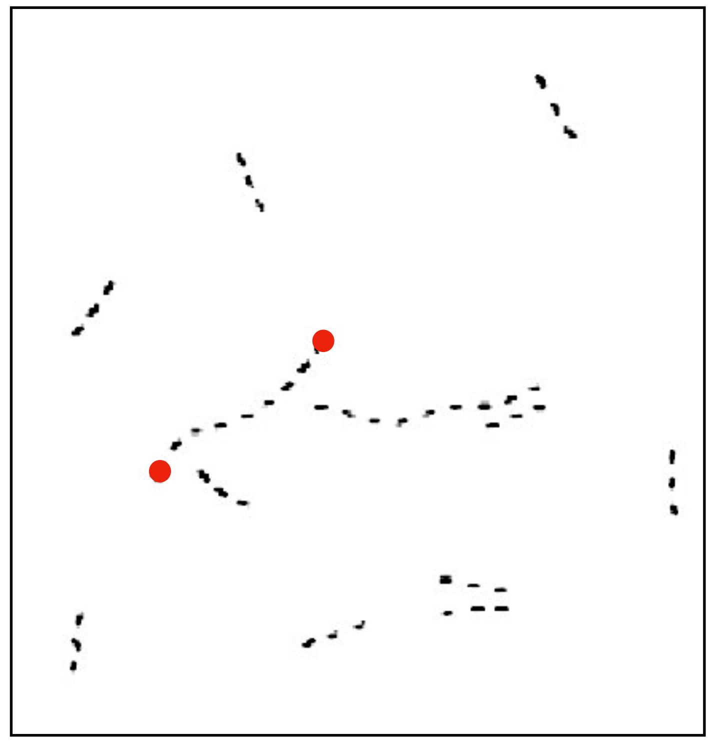
Mazes-L

* Linsley, D., Kim, J., Veerabadran, V., Windolf, C., & Serre, T. (2018). Learning long-range spatial dependencies with horizontal gated recurrent units. Advances in neural information processing systems, 31.

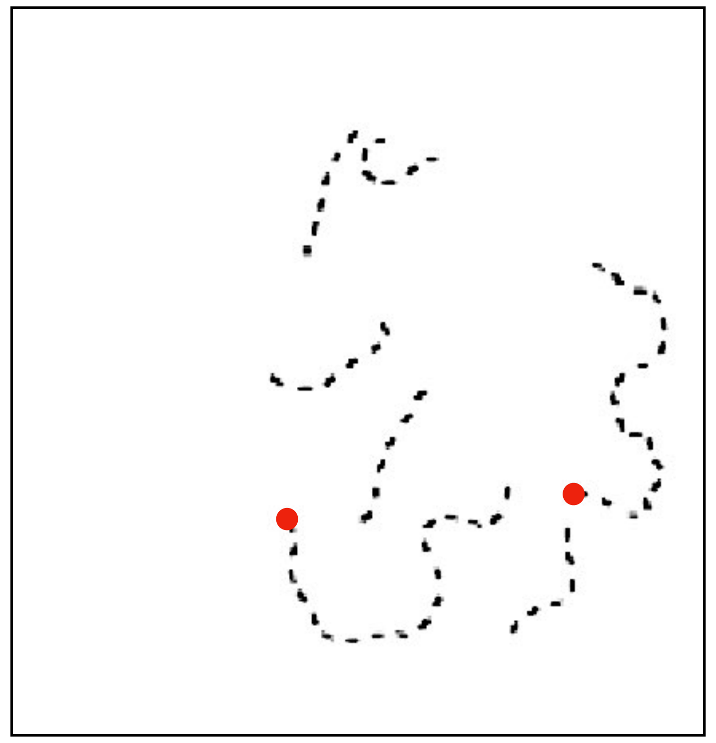
† Schwarzschild, A., Borgnia, E., Gupta, A., Bansal, A., Emam, Z., Huang, F., ... & Goldstein, T. (2021). Datasets for studying generalization from easy to hard examples. arXiv preprint arXiv:2108.06011.

Datasets: PathFinder and Mazes

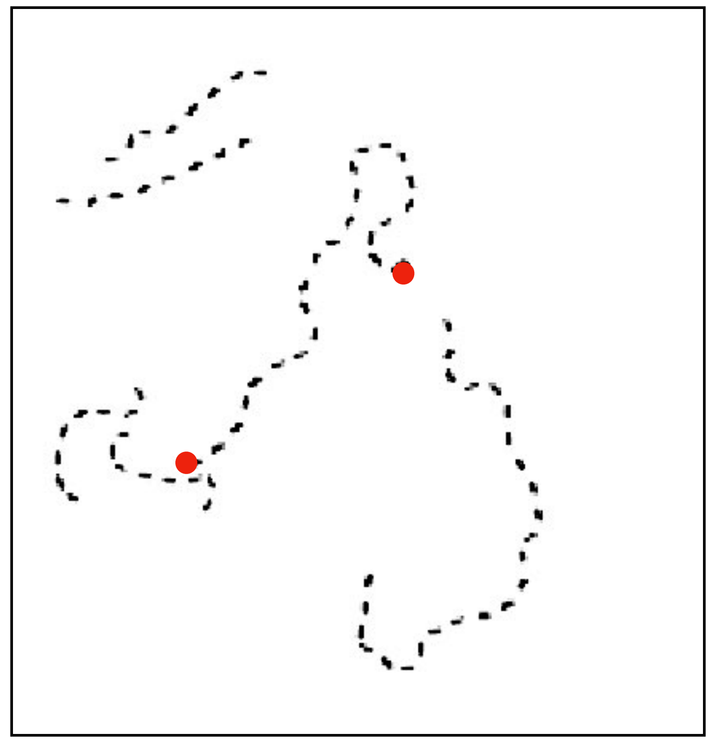
Example images from *PathFinder** (classification)



PathFinder-9



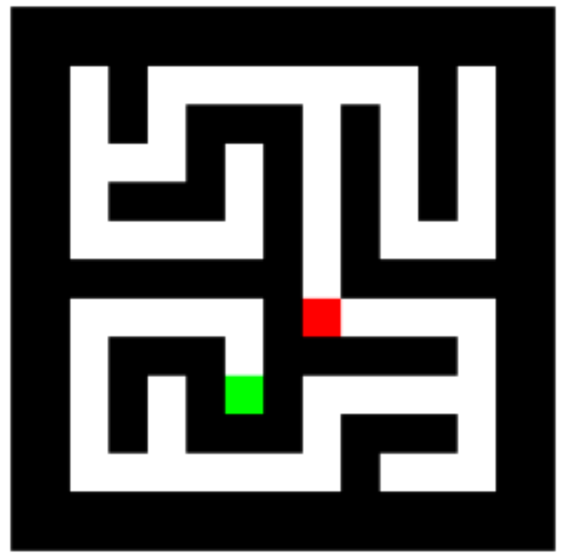
PathFinder-14



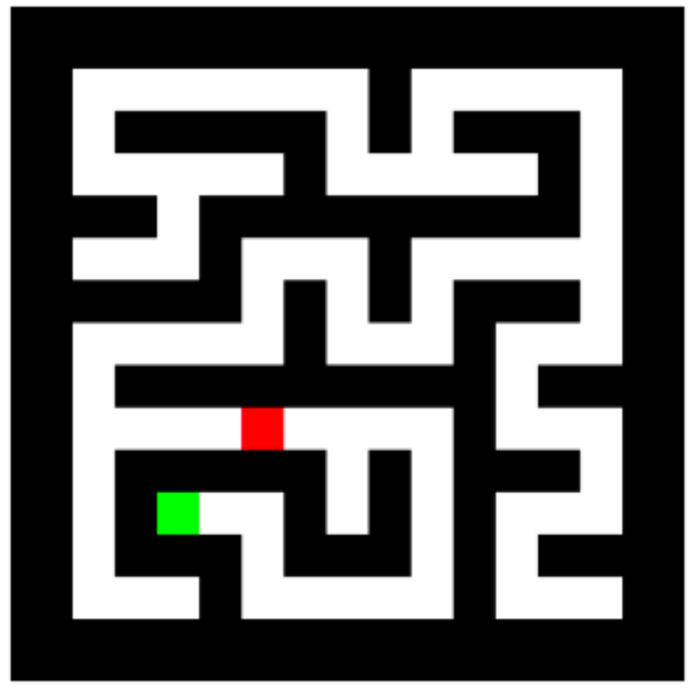
PathFinder-18

Example images from *Mazes*† (segmentation)

Mazes-S



Mazes-M



Mazes-L

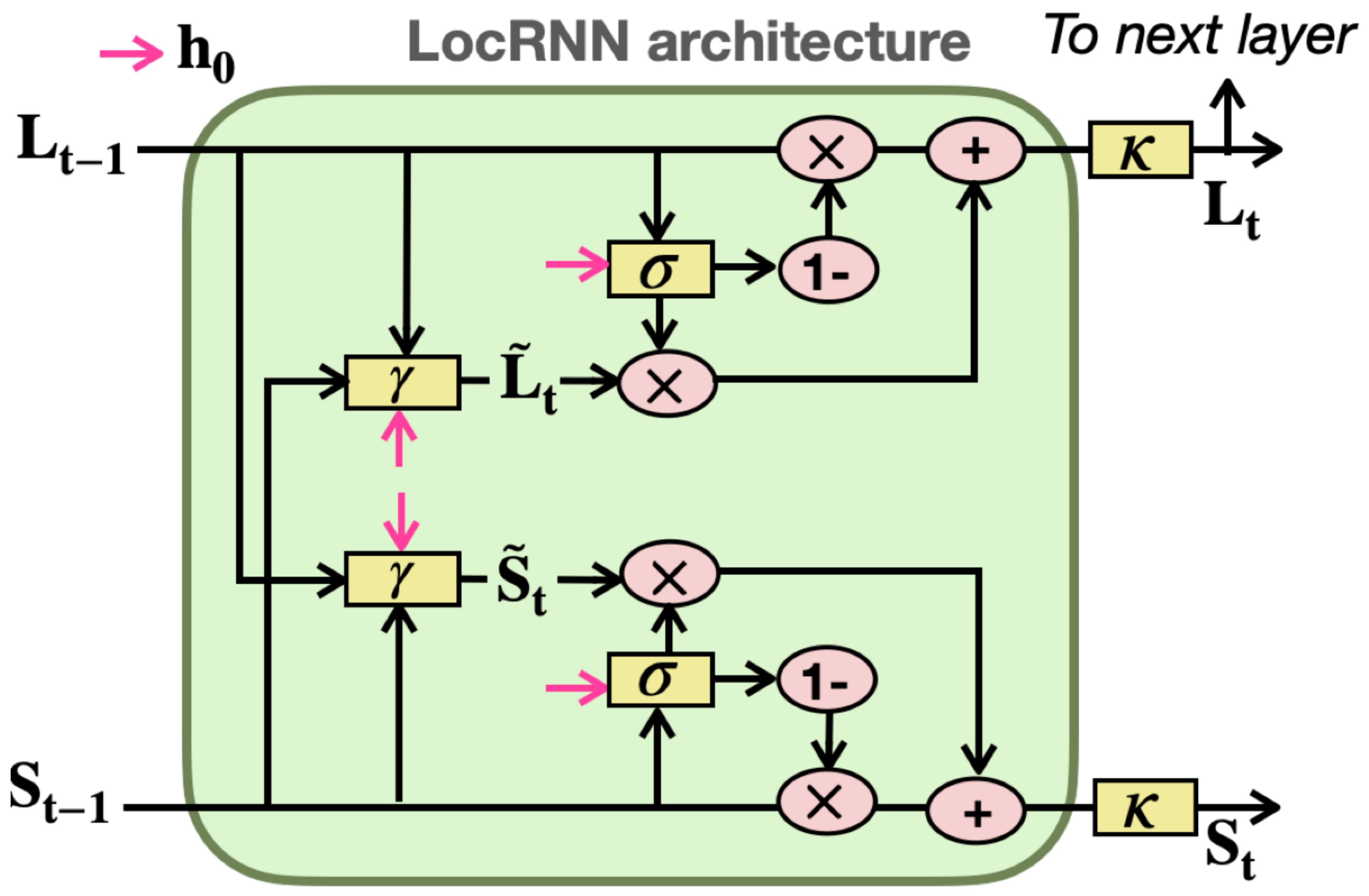
Segmentation labels

* Linsley, D., Kim, J., Veerabadran, V., Windolf, C., & Serre, T. (2018). Learning long-range spatial dependencies with horizontal gated recurrent units. Advances in neural information processing systems, 31.

† Schwarzschild, A., Borgnia, E., Gupta, A., Bansal, A., Emam, Z., Huang, F., ... & Goldstein, T. (2021). Datasets for studying generalization from easy to hard examples. arXiv preprint arXiv:2108.06011.

Models

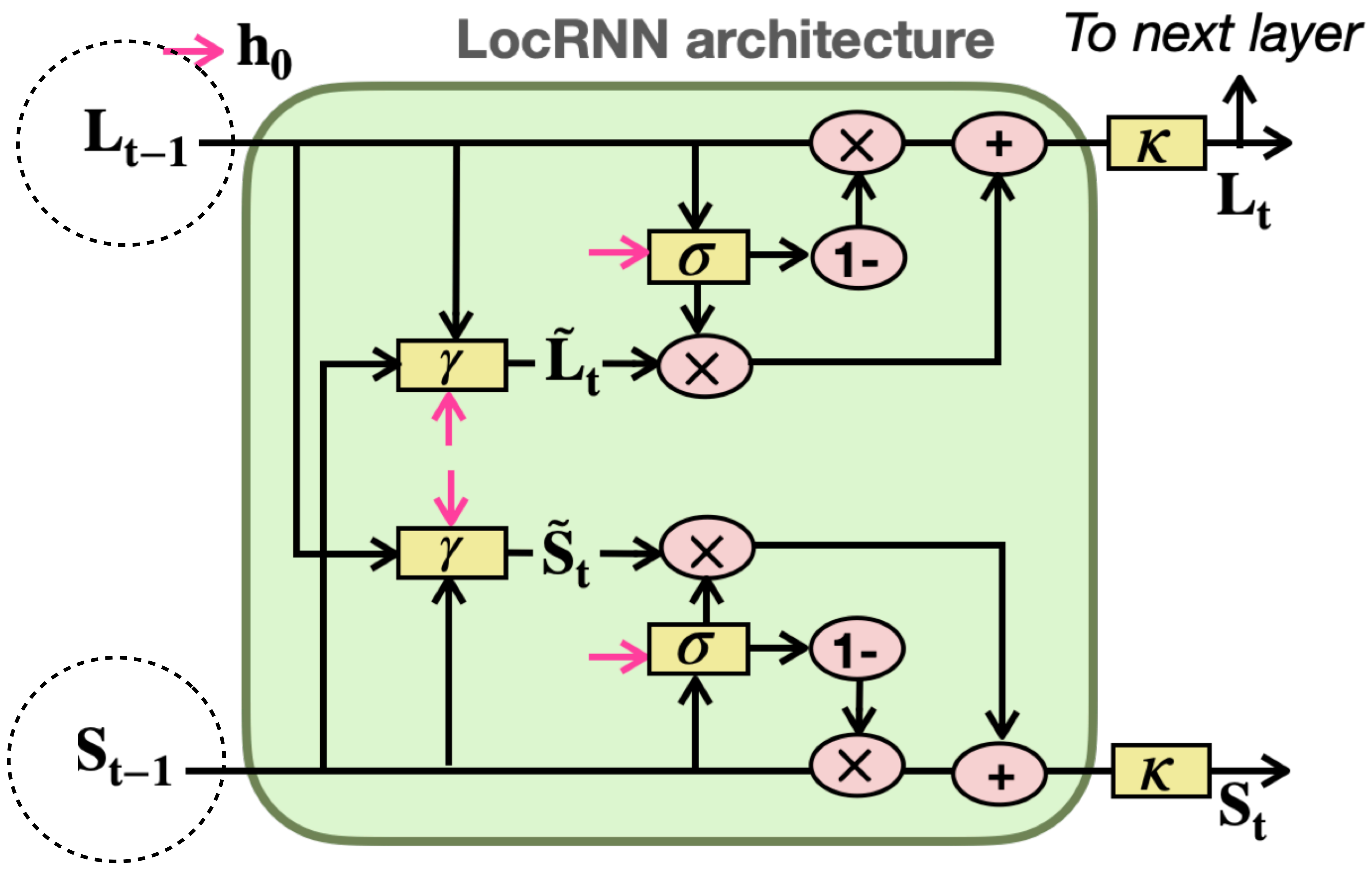
Introducing Locally connected RNN - LocRNN



- We introduce LocRNN, a bio-inspired RNN circuit implementing long-range lateral connections in CNNs

Li, Z. (1998). A neural model of contour integration in the primary visual cortex. Neural computation, 10(4), 903-940.

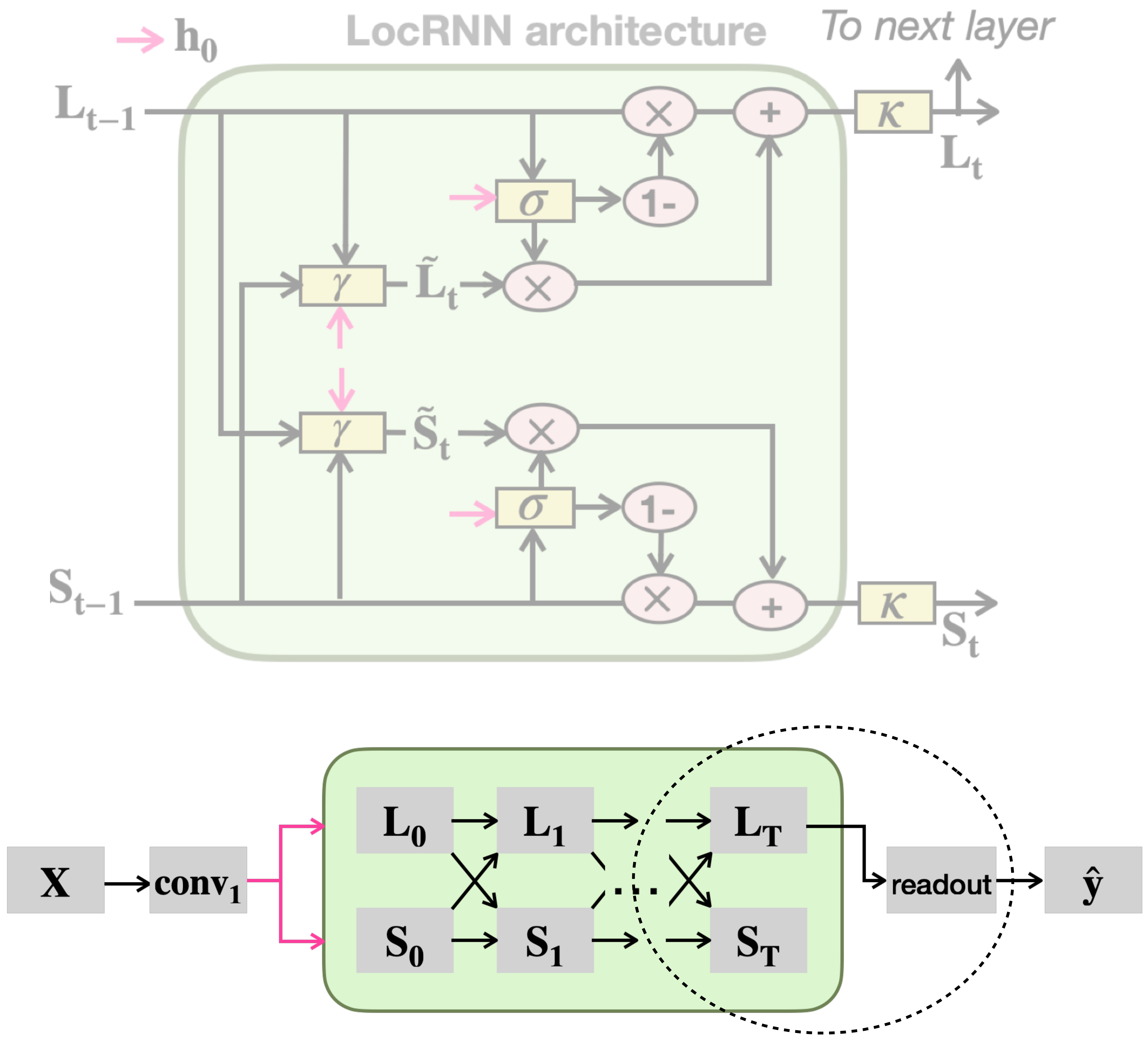
Introducing Locally connected RNN - LocRNN



- We introduce LocRNN, a bio-inspired RNN circuit implementing long-range lateral connections in CNNs
- Computation is performed by two populations of neurons $\mathbf{L}$ and $\mathbf{S}$ with gating

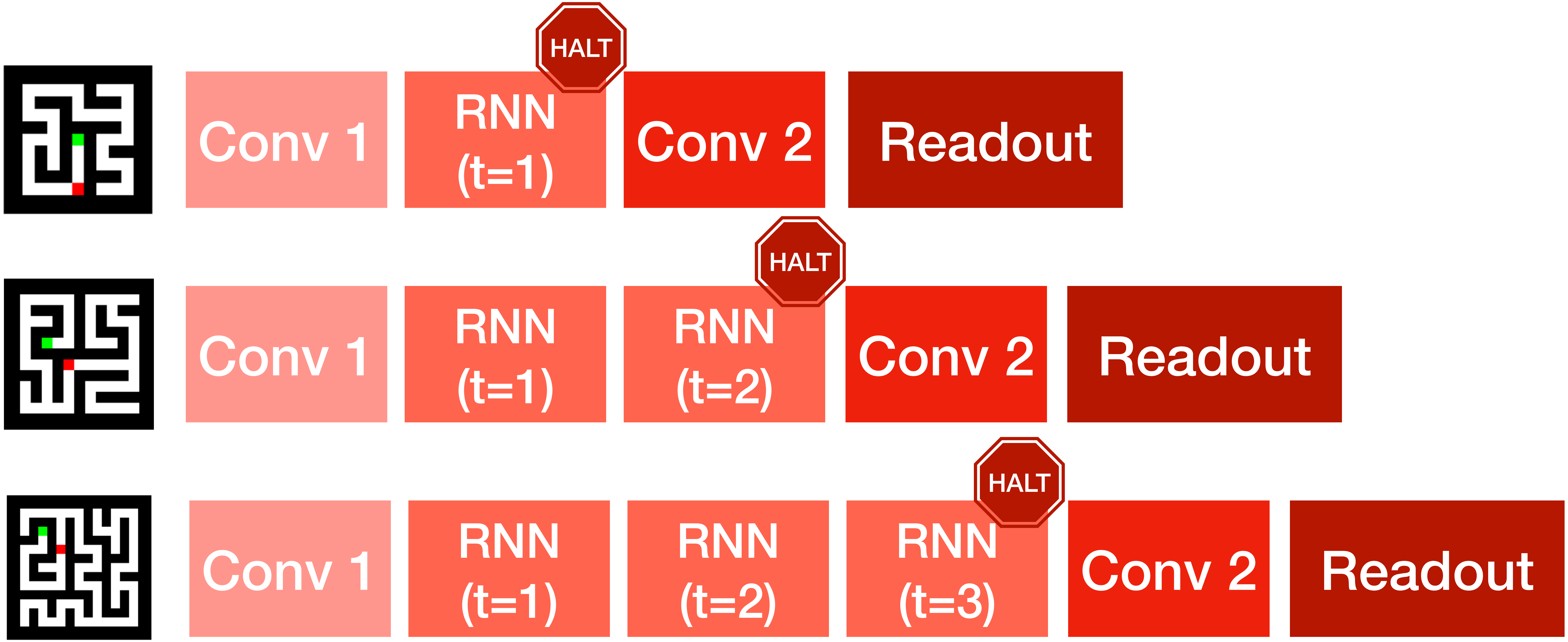
Li, Z. (1998). A neural model of contour integration in the primary visual cortex. Neural computation, 10(4), 903-940.

Introducing Locally connected RNN - LocRNN



- We introduce LocRNN, a bio-inspired RNN circuit implementing long-range lateral connections in CNNs
- Computation is performed by two populations of neurons $\mathbf{L}$ and $\mathbf{S}$ with gating
- $\mathbf{S}$ is an interneuron population similar to Li, Z. (Neural computation, 1998).

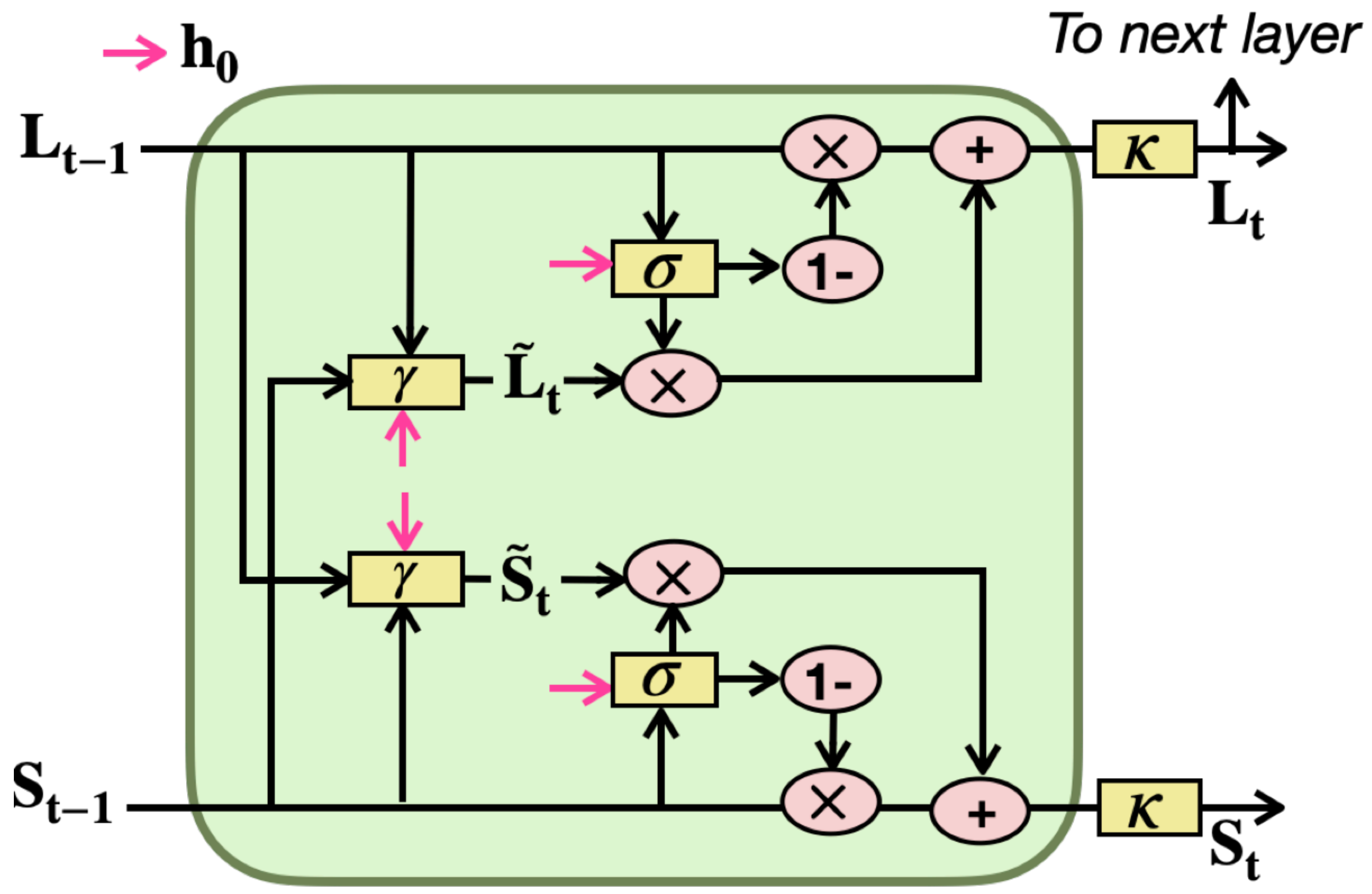
Combining ConvRNNs with Adaptive Computation Time



Graves, A. (2016). Adaptive computation time for recurrent neural networks. arXiv preprint arXiv:1603.08983.

Banino, A., Balaguer, J., & Blundell, C. (2021). Pondernet: Learning to ponder. arXiv preprint arXiv:2107.05407.

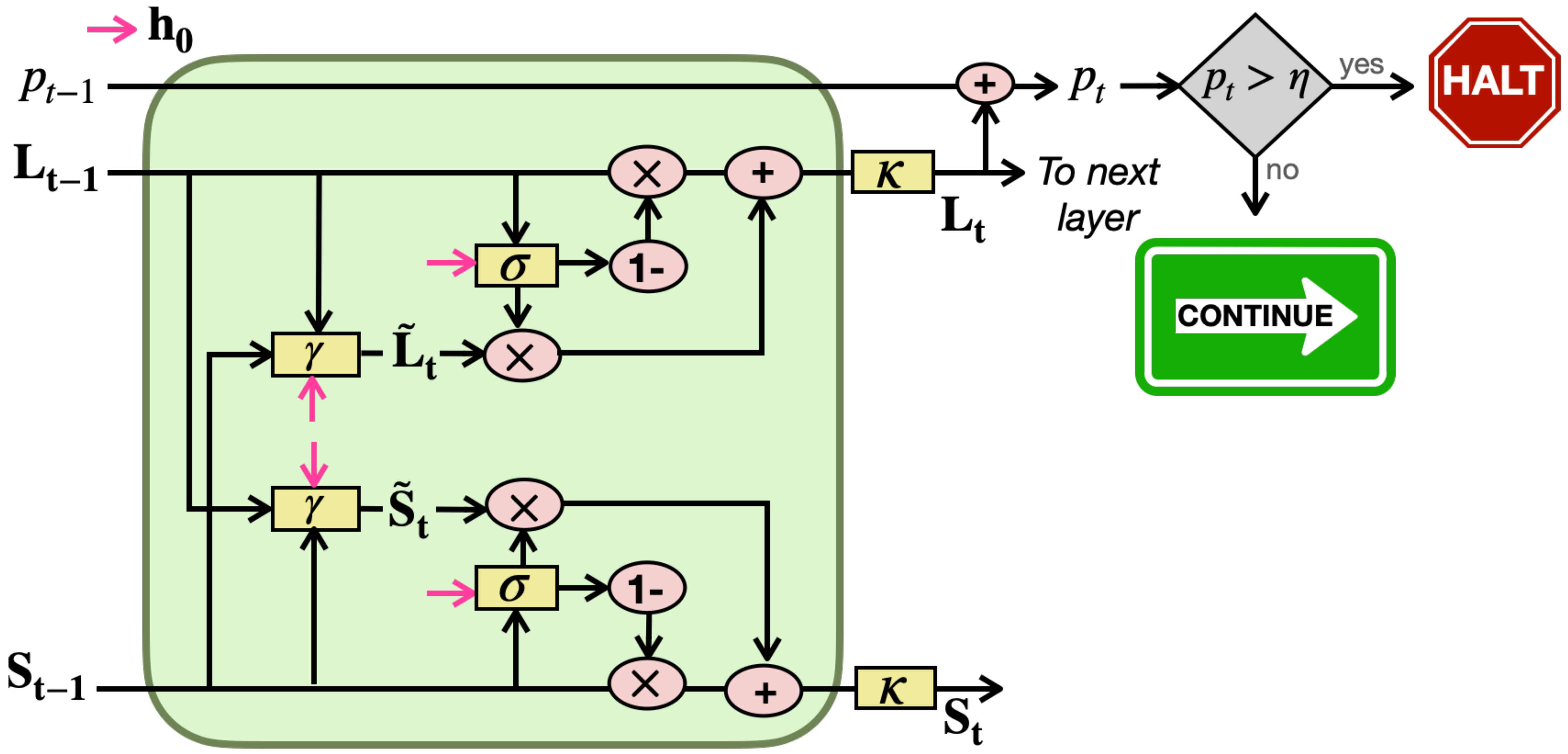
Combining ConvRNNs with Adaptive Computation Time



Prediction error

$$\mathcal{L} = \boxed{\sum_{i=0}^{i=||\mathcal{D}||} \frac{1}{||\mathcal{D}||} ||\mathbf{y}^i - \hat{\mathbf{y}}^i \quad ||_2}$$

Combining ConvRNNs with Adaptive Computation Time



- Learnable halting convolution projection computes a cumulative halting quantity p_t as a function of $\mathbf{L}_{<t}$
- If cumulative halting quantity p_t reaches/exceeds threshold η, ConvRNN halts processing

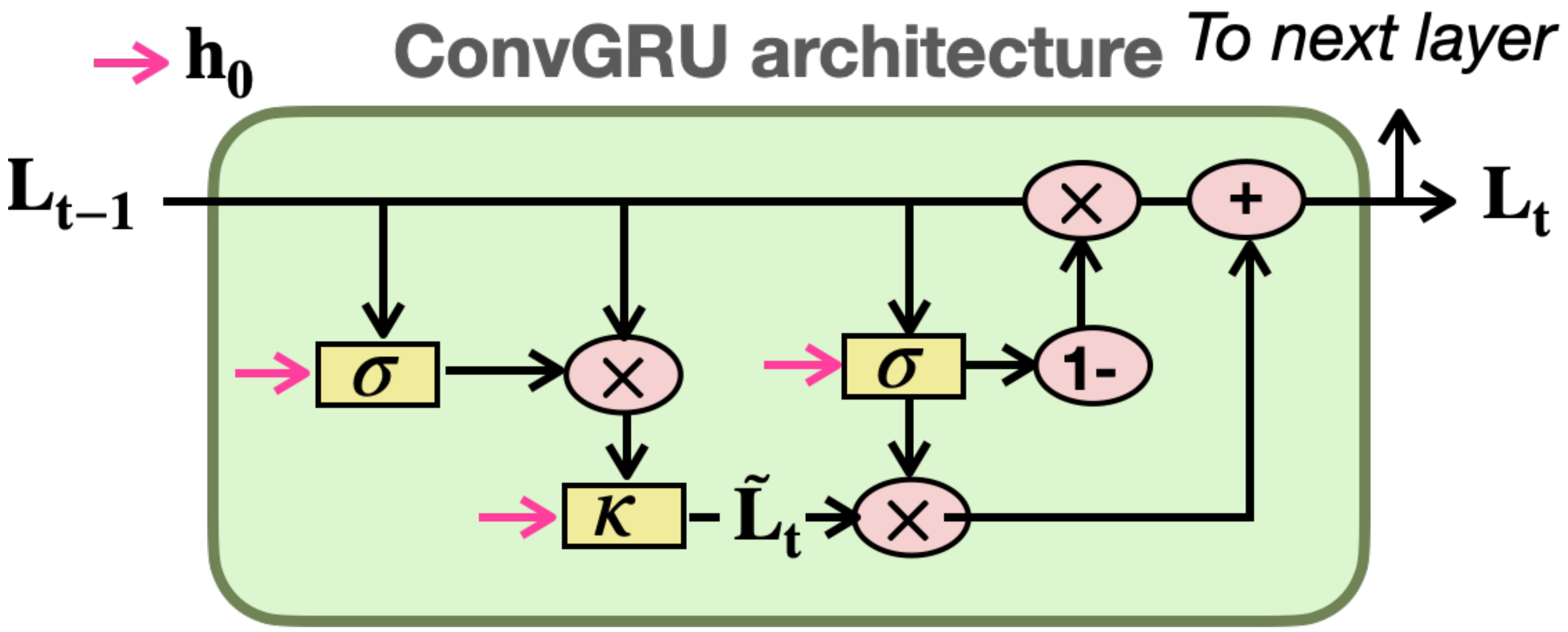
Prediction error | Ponder cost

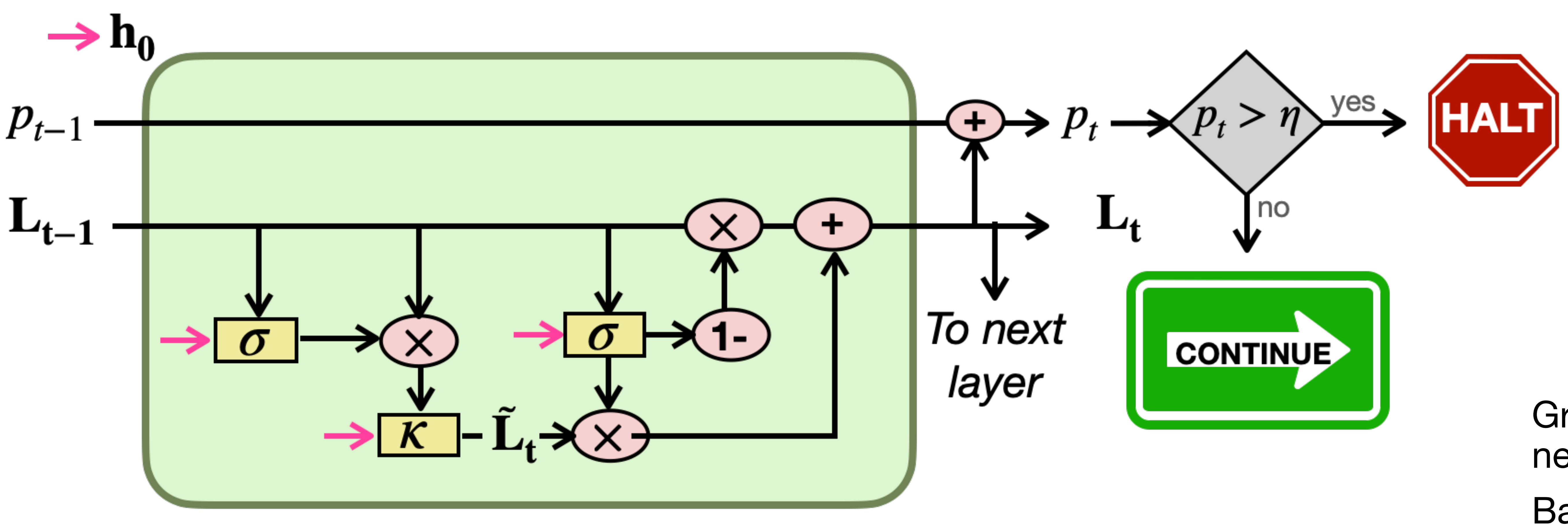
$$\mathcal{L} = \boxed{\sum_{i=0}^{i=||\mathcal{D}||} \frac{1}{||\mathcal{D}||} ||\mathbf{y}^i - \hat{\mathbf{y}}^i_{act}||_2} - \boxed{\tau * p^i_{t_{halt}-1}}$$

Graves, A. (2016). Adaptive computation time for recurrent neural networks. arXiv preprint arXiv:1603.08983.

Banino, A., Balaguer, J., & Blundell, C. (2021). Pondernet: Learning to ponder. arXiv preprint arXiv:2107.05407.

Combining ConvRNNs with Adaptive Computation Time





Prediction error — Ponder cost

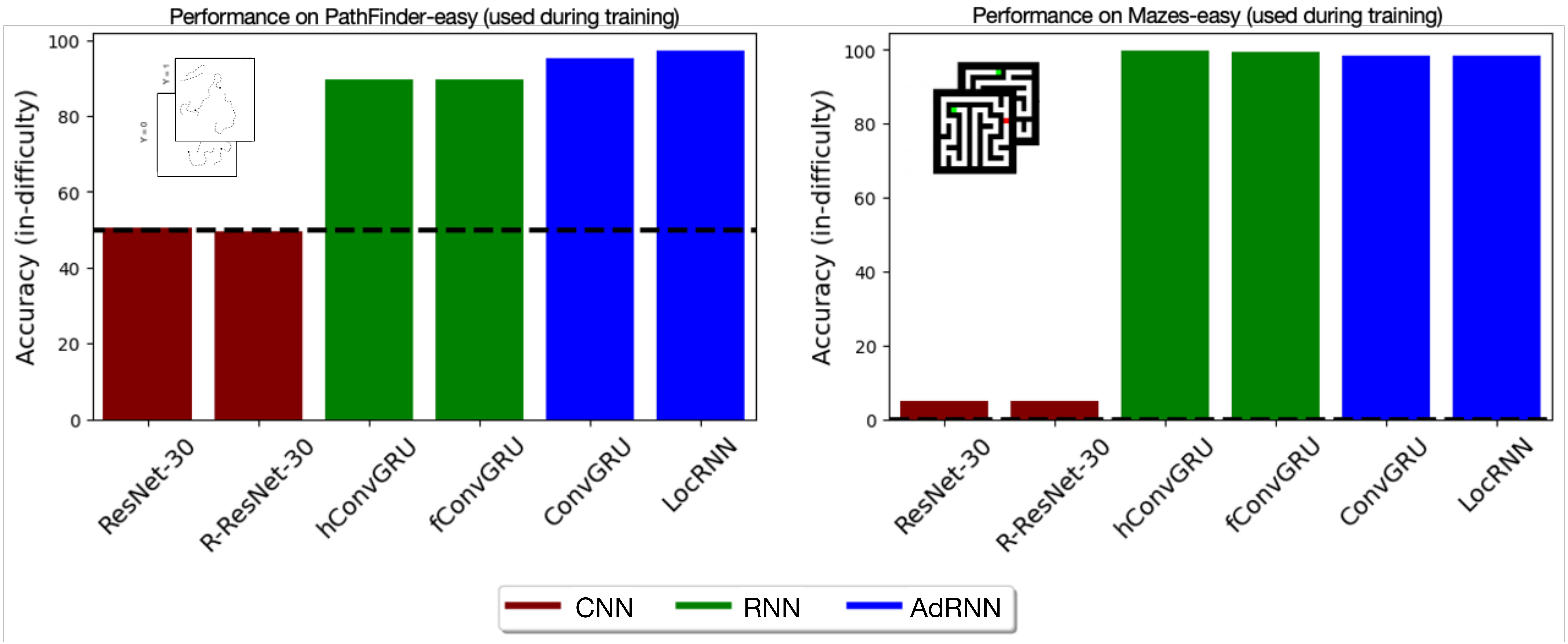
$$\mathcal{L} = \boxed{\sum_{i=0}^{i=||\mathcal{D}||} \frac{1}{||\mathcal{D}||} ||\mathbf{y}^i - \hat{\mathbf{y}}^i_{act}||_2} - \boxed{\tau * p^i_{t_{halt}-1}}$$

Graves, A. (2016). Adaptive computation time for recurrent neural networks. arXiv preprint arXiv:1603.08983.

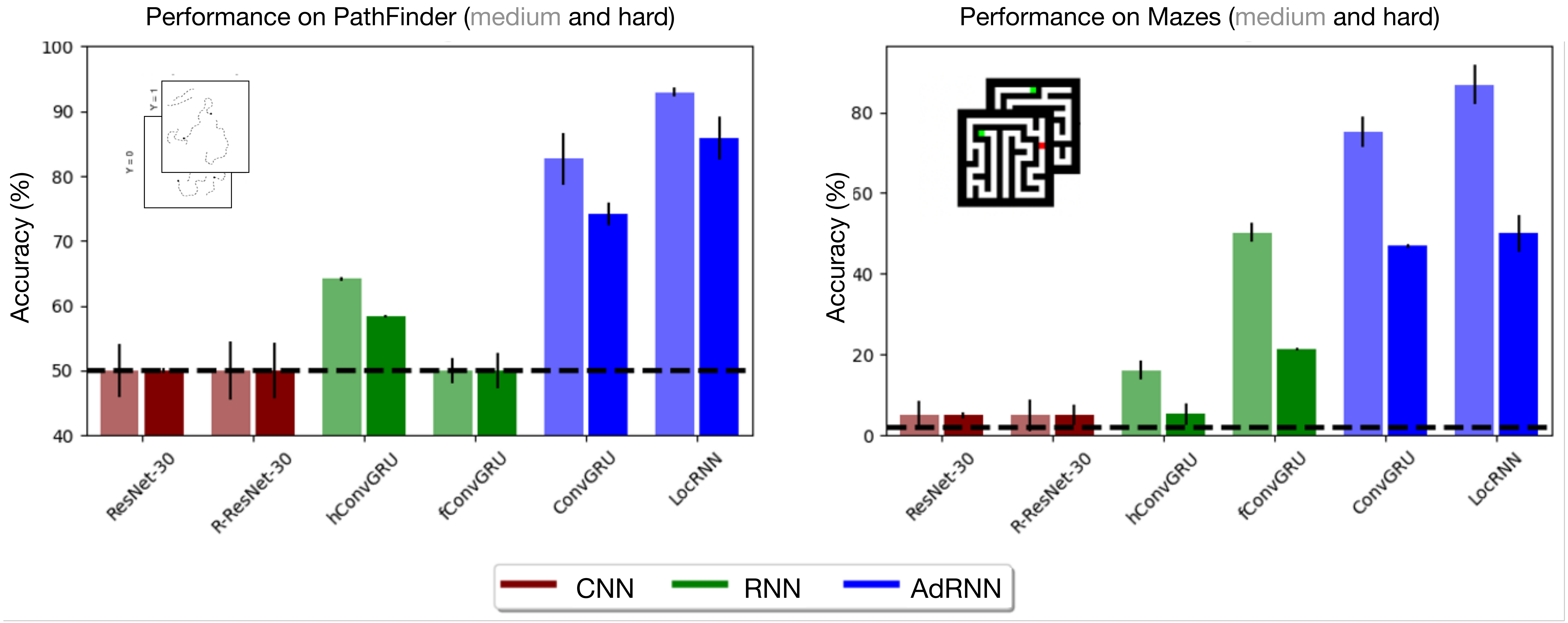
Banino, A., Balaguer, J., & Blundell, C. (2021). Pondernet: Learning to ponder. arXiv preprint arXiv:2107.05407.

Results

Evaluation on held-out images *in-difficulty* (easy)



Extrapolation performance on *unseen-difficulty levels* (medium, hard)



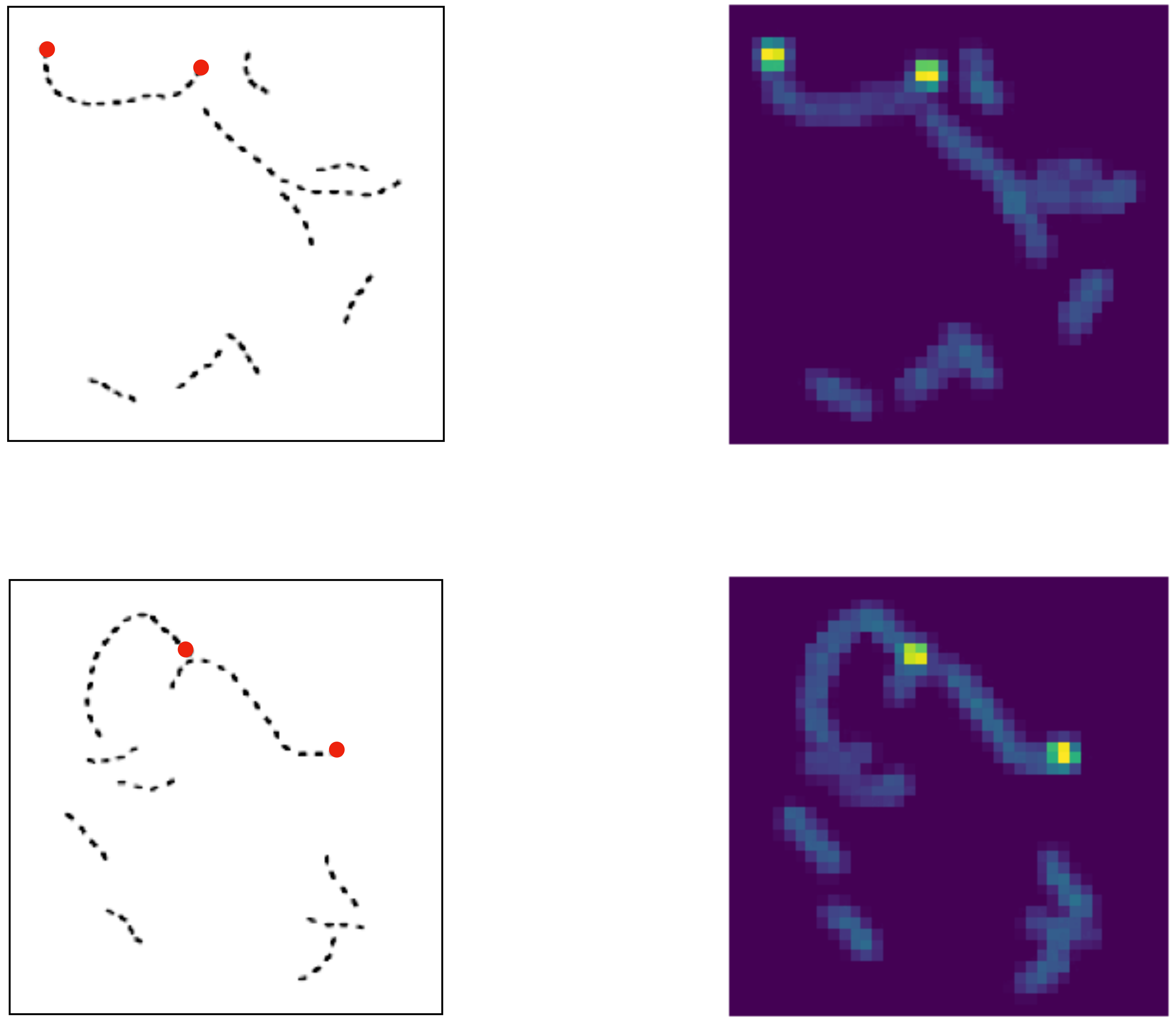
AdRNNs zero-shot generalize to novel test-time difficulties

Curve tracing and path integration in LocRNN

PathFinder

Image inputs

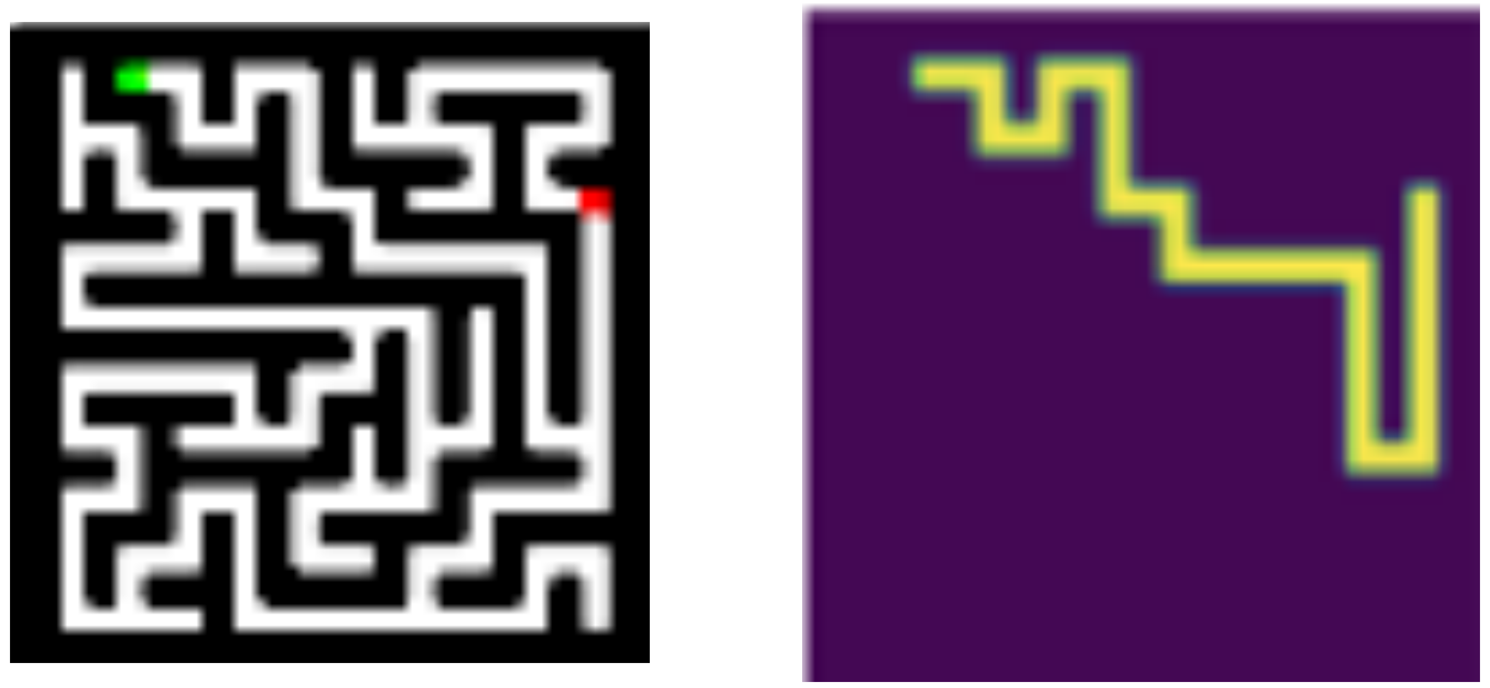
Activations, $\mathbf{L_t} \in [0,1]$



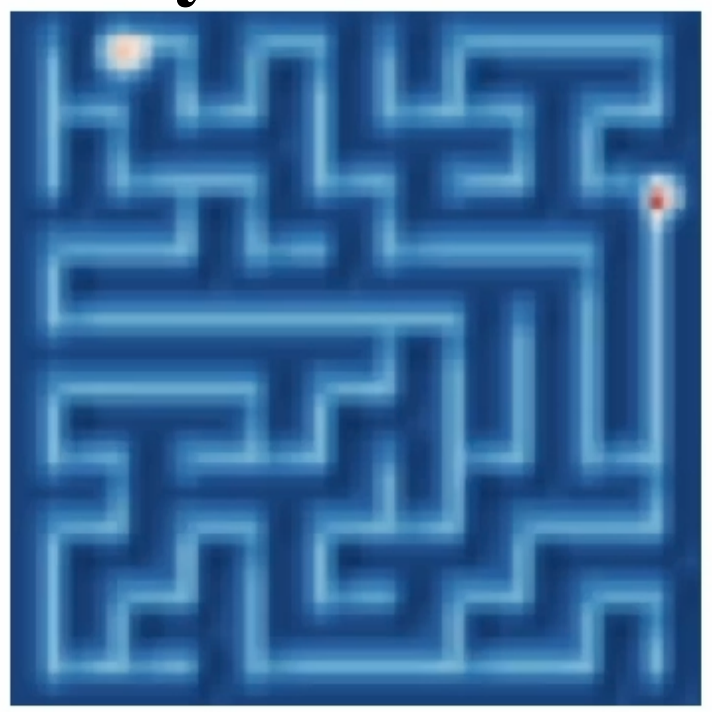
Mazes

Image inputs

Ground truth



Activations, $\mathbf{L_t} \in [-1,1]$



Adaptive recurrent vision performs zero-shot computation scaling to unseen difficulty levels

Vijay Veerabadran · Srinivas Ravishankar · Yuan Tang · Ritik Raina · Virginia de Sa

Great Hall & Hall B1+B2 #412

vveeraba@ucsd.edu

Thu 14 Dec 10:45 a.m. CST — 12:45 p.m. CST (NOLA time)

UCSD

NSF

UC San Diego
SOCIAL SCIENCES
Social Sciences Computing Facility

SONY

THE KAVLI FOUNDATION

NEURAL INFORMATION PROCESSING SYSTEMS