Supplementary

A Training details

On Mazes, we trained models using per-pixel binary cross entropy with a training minibatch size of 64 images (and inference batch size of 50 images) and a learning rate schedule starting with warmup followed by step learning rate decay as indicated in Schwarzschild et al. (2021) for 50 total epochs of training. On PathFinder, we used binary cross-entropy to train models with a minibatch size of 256 images and a constant learning rate of 1e-4 for all models for a total of 20 epochs of training. All models were trained on NVIDIA RTX A6000 GPUs and implemented using PyTorch.(Paszke et al., 2017).

B Instability of other baseline ConvRNNs

ConvRNN training is often faced with instability issues that lead to sensitivity with respect to random seeds or lack of convergence of models on downstream tasks. We tested a suite of ConvRNNs previously introduced that are similar to LocRNN on three difficulty levels of PathFinder (in-difficulty evaluation, i.e., training and testing on each difficulty level independently). This evaluation highlighted the above issue especially on the difficult levels of PathFinder where LocRNN was the only model which could converge to stable solutions across different random seeds unlike the other networks which performed at chance as shown below in Fig. 5.

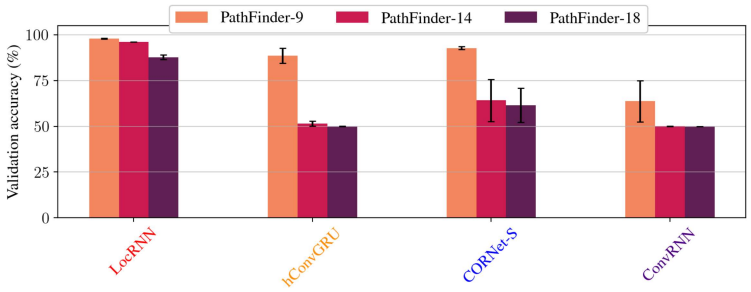

Figure 5: Performance of various ConvRNN models on PathFinder-9, PathFinder-14, and PathFinder-18.

C Input and output format for PathFinder and Mazes

Each example maze is an $n \times n$ RGB matrix, with colored squares indicating the start (green) and end (red) positions in the maze. The output is a binary matrix of size $n \times n$ with the segmented path indicating the maze solution. An example is shown in Figure 6 (top).

Each PathFinder example is an $n \times n$ binary matrix as shown in Figure 6 (bottom). The output for one sample is a pair of probabilities denoting which class the sample belongs to (negative, meaning the disks are at the end of disconnected paths, or positive, meaning the disks are connected through the contour).

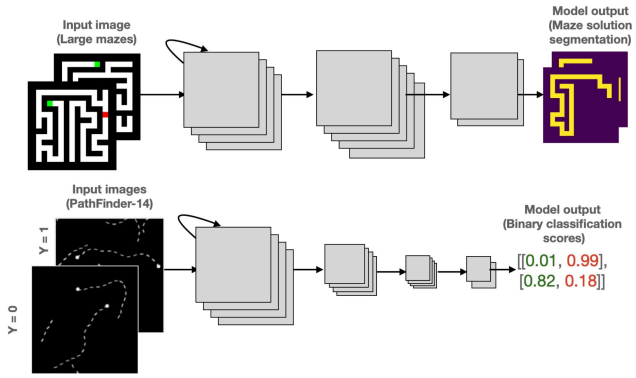

Figure 6: (Top) Example images from 11×11 Mazes processed by a model to produce the solution as a segmentation prediction. (Bottom) Example input images from PathFinder-14 processed by a classifier to produce binary classification output.