Comparing Spatial Representations in Diverse Recurrent Neural Network Architectures Exploring the Gridworld

An Investigation into Internal Memory Mechanisms

Group: Space Invaders

Sofiya Zbaranska, *PhD Candidate (University of Toronto, CA)*
Ítalo Karmann Aventurato, *PhD Candidate (Universidade Estadual de Campinas, BR)*
Arthur Rodrigues, *MSc Student (Universidade Federal de Minas Gerais, BR)*
Setareh Metanat, *BSc in Cognitive Science, Research Assistant (UCalifornia San Diego, USA)*
Didi Ramsaran Chin, *BSc in Psychology (Universidad Católica Andrés Bello, VE)*
Kate Tabor, *PhD in Neuroscience (University of Washington, USA)*
Sara Asadi, *MSc Student (University of Lethbridge, CA)*
Souvik Bhattacharya, *Ph.D. Student in Computer Science (University of Illinois, USA)*
Leticia Cid, BASc, *Research Assistant (University of British Columbia, CA)*
Muhammad Mushhood Ur Rehman, *Masters in Public Health Student (University of Edinburgh, UK)*

Motivation: The "Black Box" of Spatial Memory

The Challenge:

- Learning previous locations in a certain environment is a memory task which additionally requires learning the spatial structure of the environment.

The Tool:

- Recurrent Neural Networks (RNNs) excel at processing sequential data, making them a natural choice for navigation tasks.

Our Question:

- While RNNs can learn structured tasks, it is unclear how they internally represent spatial information.
- The mechanism for creating and storing an internal "map" is not well understood.

Our Goal:

- To systematically investigate and compare how different RNN architectures encode the structure of the spatial environment and retrieve their past locations.

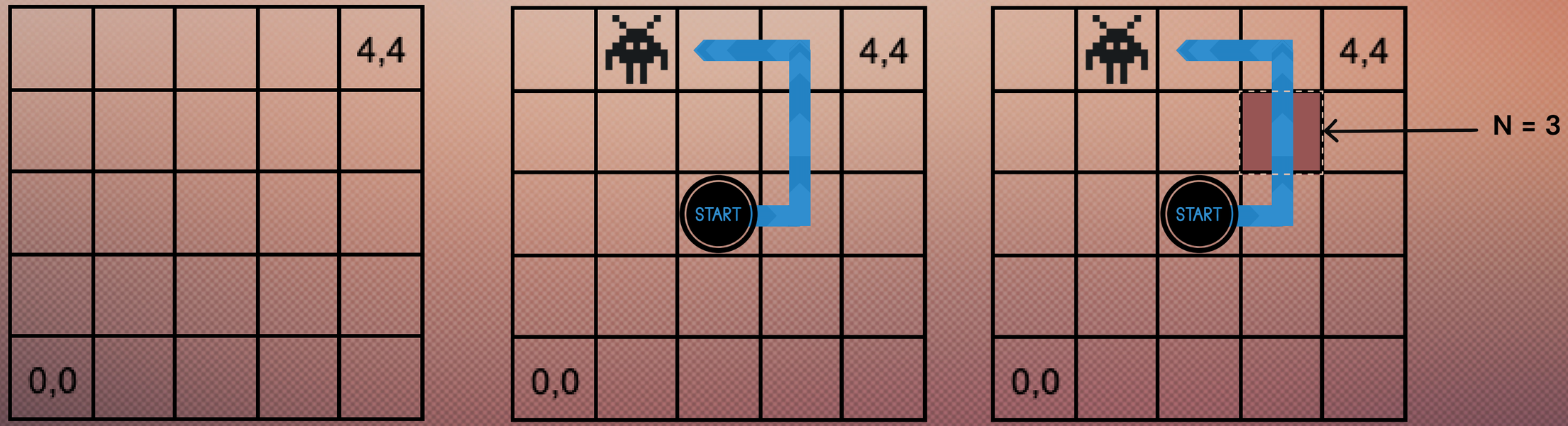
Our Questions

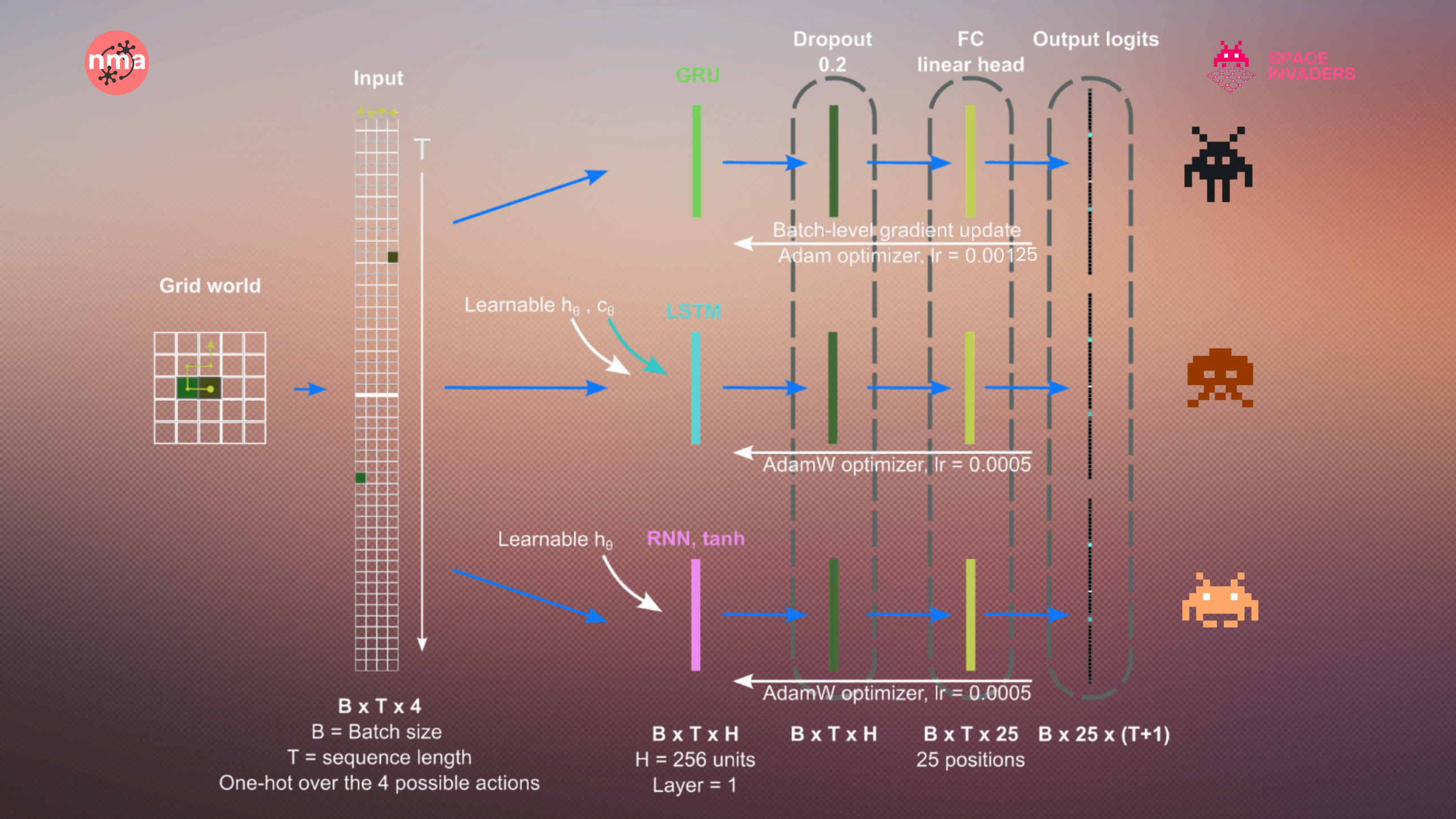
Memory Capacity:
Which architecture, vanilla RNN, GRU, or LSTM, most effectively retains spatial memory over longer time delays?

Internal Space Representation:
To what extent do the networks' internal representations align with the actual 2D layout of the environment?

Representational Differences:
How do the spatial maps encoded by a vanilla RNN, GRU, and LSTM compare with one another?

Gridworld Spatial Memory Task



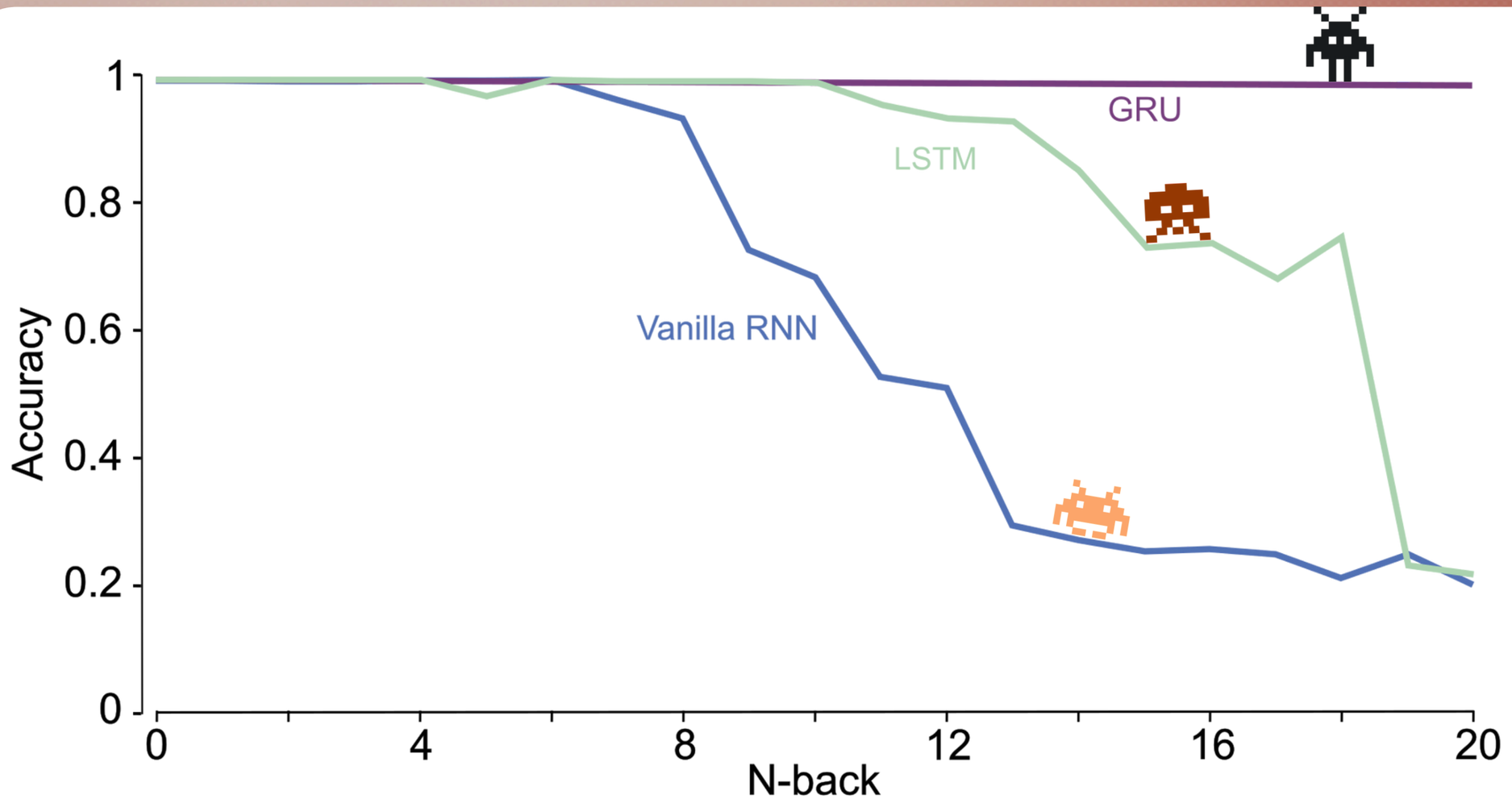

Input
GRU
Dropout
0.2
FC
linear head
Output logits
SPACE
INVADERS
T
Batch-level gradient update
Adam optimizer, lr = 0.00125
Grid world
Learnable h_θ , c_θ
LSTM
AdamW optimizer, lr = 0.0005
Learnable h_θ
RNN, tanh
AdamW optimizer, lr = 0.0005
B x T x 4
B = Batch size
T = sequence length
One-hot over the 4 possible actions
B x T x H
H = 256 units
Layer = 1
B x T x H
B x T x 25
25 positions
B x 25 x (T+1)

Results: Comparing Memory Performance

Objective: To determine which architecture has the most robust memory.

Forgetting index (0 is perfect recall)*

- Vanilla RNN: 0.397
- LSTM: 0.153
- GRU: 0.00002

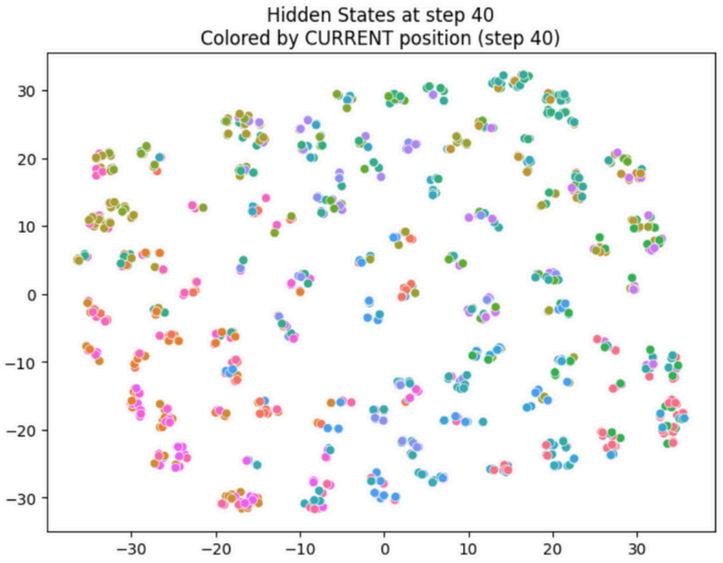


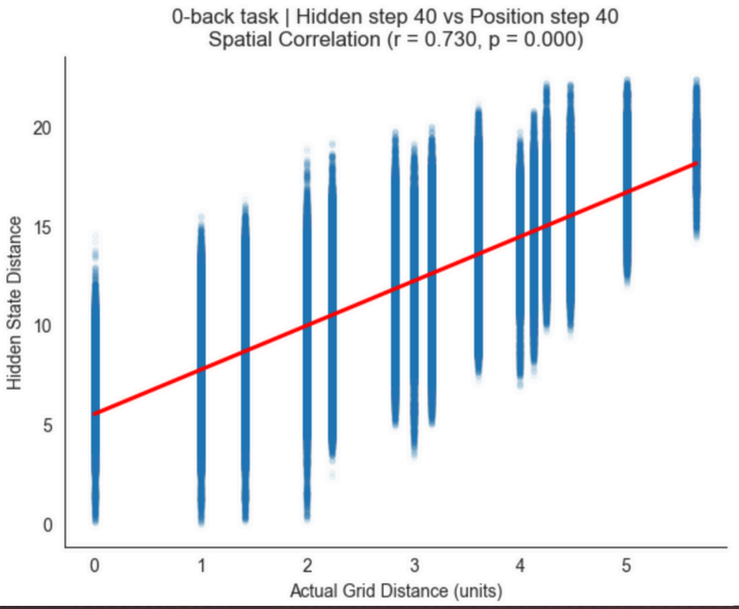
$$\text{*Index} = \frac{\text{Area above the curve}}{\text{Total area of plot (1x20)}}$$

Results: Spatial Representations in Models

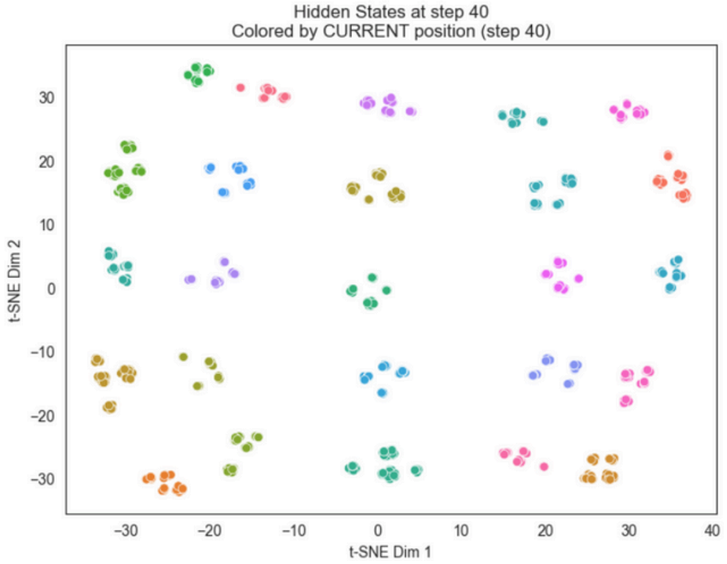
Objective: To "see" if the networks learn and represent the spatial map internally

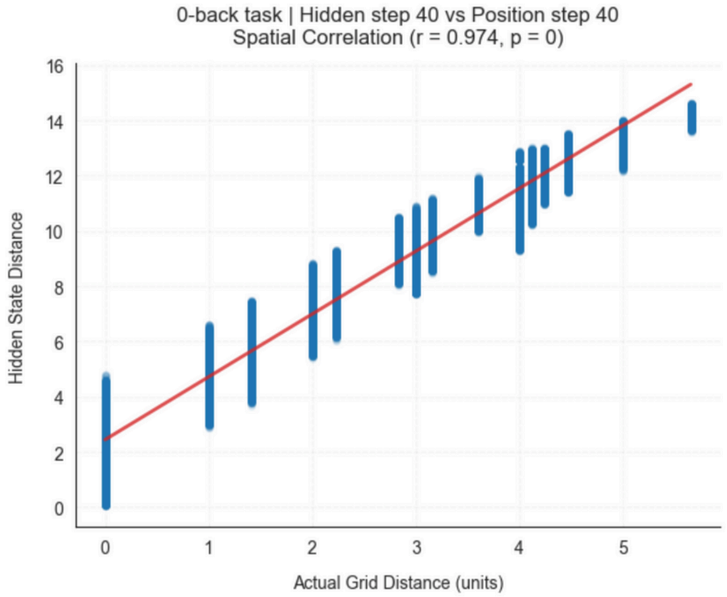
Vanilla RNN





LSTM





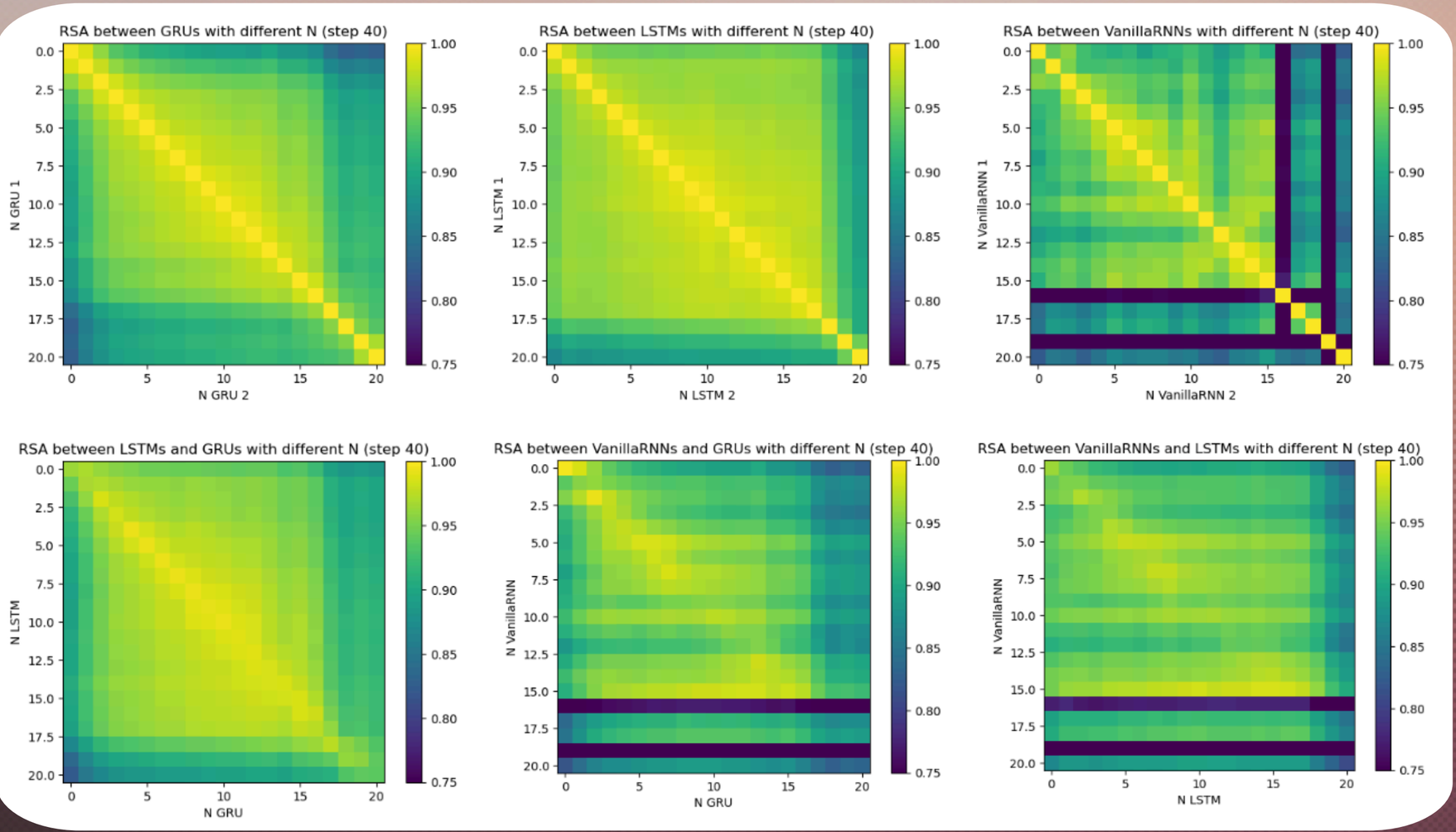
GRU

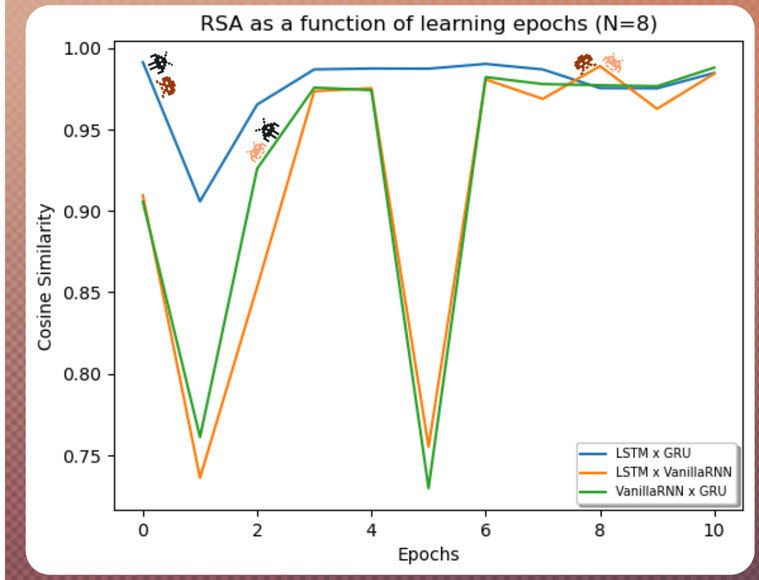
Content is no longer available

Content is no longer available

Results: Spatial Representations Across Models

Objective: To quantitatively measure how similarly the architectures represent space.





SPACE INVADERS

Main Conclusions

- Under similar conditions and parameters, the GRU model exhibits the best performance in the N-back task, achieving perfect recall.

- All RNNs implicitly learn and store a detectable spatial map in their hidden states, with the LSTM and GRU models exhibiting a more grid-like structure (as shown by t-SNE) and reliable grid space vs. hidden states scalability.

- LSTM and GRU share the highest representational similarity of the learned task.

Questions?

SPACE INVADERS

[Access Our Repo Here:]

What we did: “Import torch”
How we felt:

neur1s/**GridVaders**

Comparing neural networks performing spatial memory task in a 2D gridworld

11 Contributors · 0 Issues · 22 Stars · 2 Forks

neur1s/GridVaders: Comparing neural networks performing spatial memory task in a 2D gridworld

Comparing neural networks performing spatial memory task in a 2D gridworld - neur1s/GridVaders

GitHub

Thank You!

SPACE INVADERS