Pre-Print

Grid cell path integration for movement-based visual object recognition

Niels Leadholm[1,2], Marcus Lewis[1] and Subutai Ahmad[1]

1: Numenta, Inc. 2: The University of Oxford

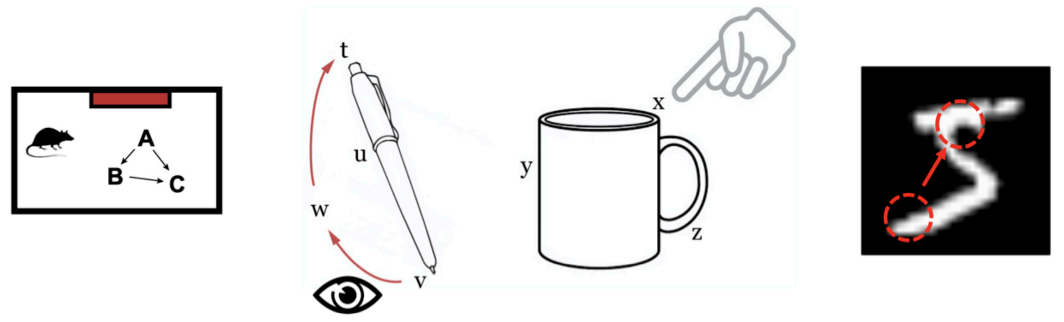
niels.leadholm@seh.ox.ac.uk, {mlewis, sahmad}@numenta.com

Motivation: Recognition Given Visual Sequences

- Biological vision based on saccades [3]
- Vision models focus on feed-forward parallel processing

Problem:

- Given a sequence of visual inputs, its currently unclear how the brain integrates these for object recognition
- Integration is challenging as the sequence across space is not fixed



Key Ideas:

- Grid cells are thought to encode space and enable path integration
- Could grid cells enable visual object recognition with arbitrary sequences of sensations?

Cortical Networks with Grid Cells

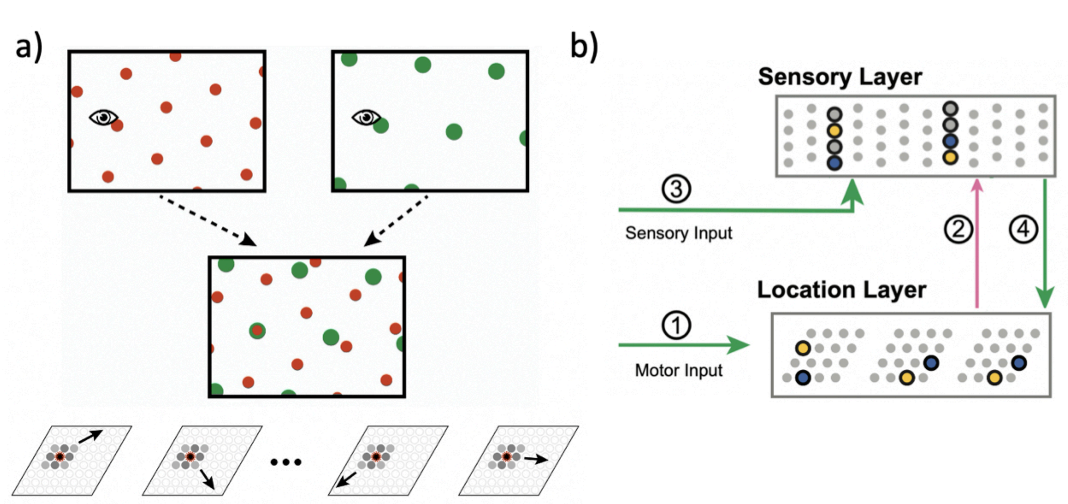


Fig. 2: a) Activity from multiple grid cell modules (different scale and orientation) uniquely encodes position. Locations updated with self-movement (**path integration**) b) Sensorimotor network w/ two layers (sensory: columns, spatial: grid cell modules) that reciprocally predict sparse representations in one another (steps 2 and 4).

Our approach ("GridCellNet")

- Core architecture as in [2], but pre-process images (MNIST) w/ CNN for feature extraction
- Also extend [2] classification algorithm to compare representation to multiple learned examples (supports generalization)
- Learning uses rapid Hebbian weight updates

Classification Algorithm

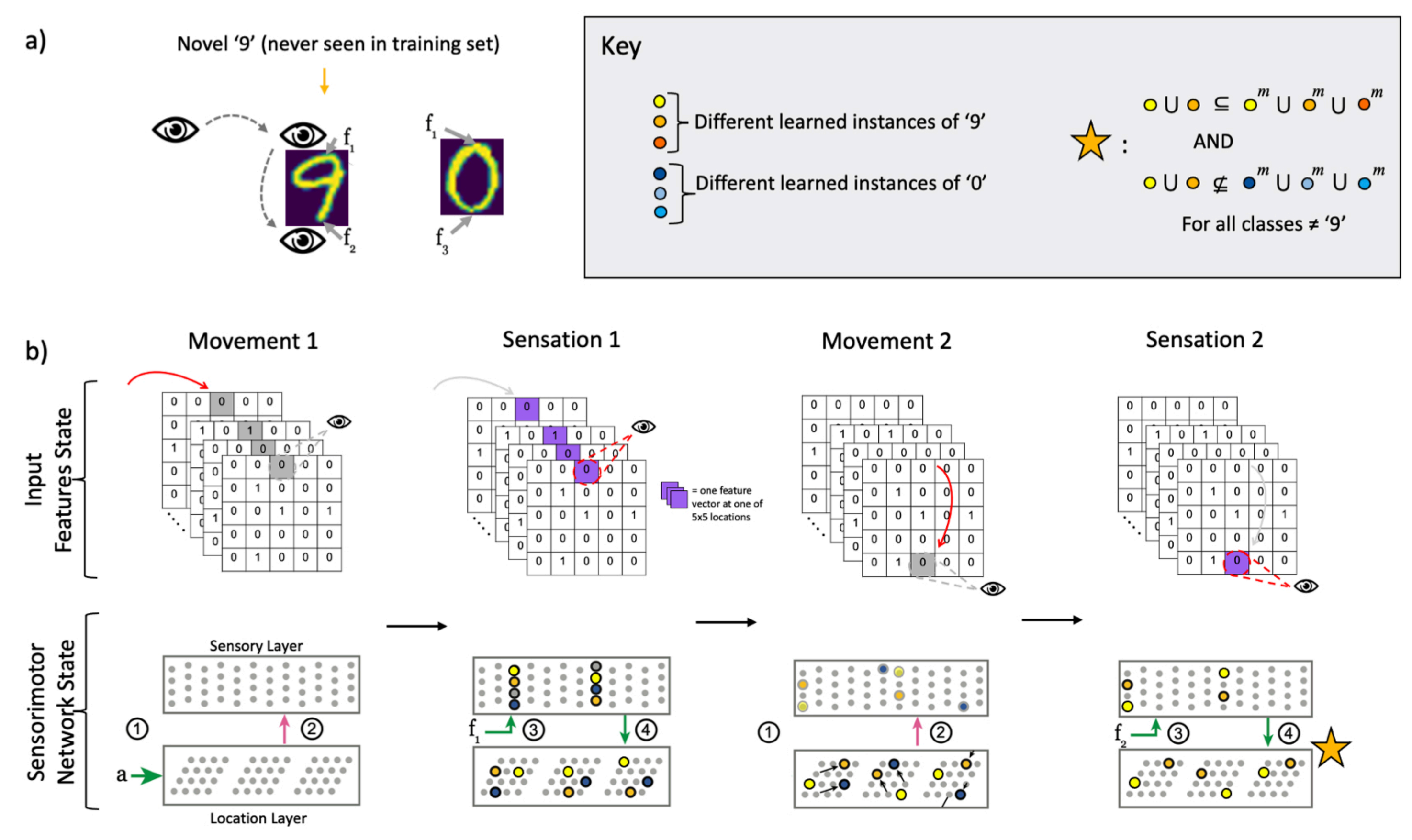
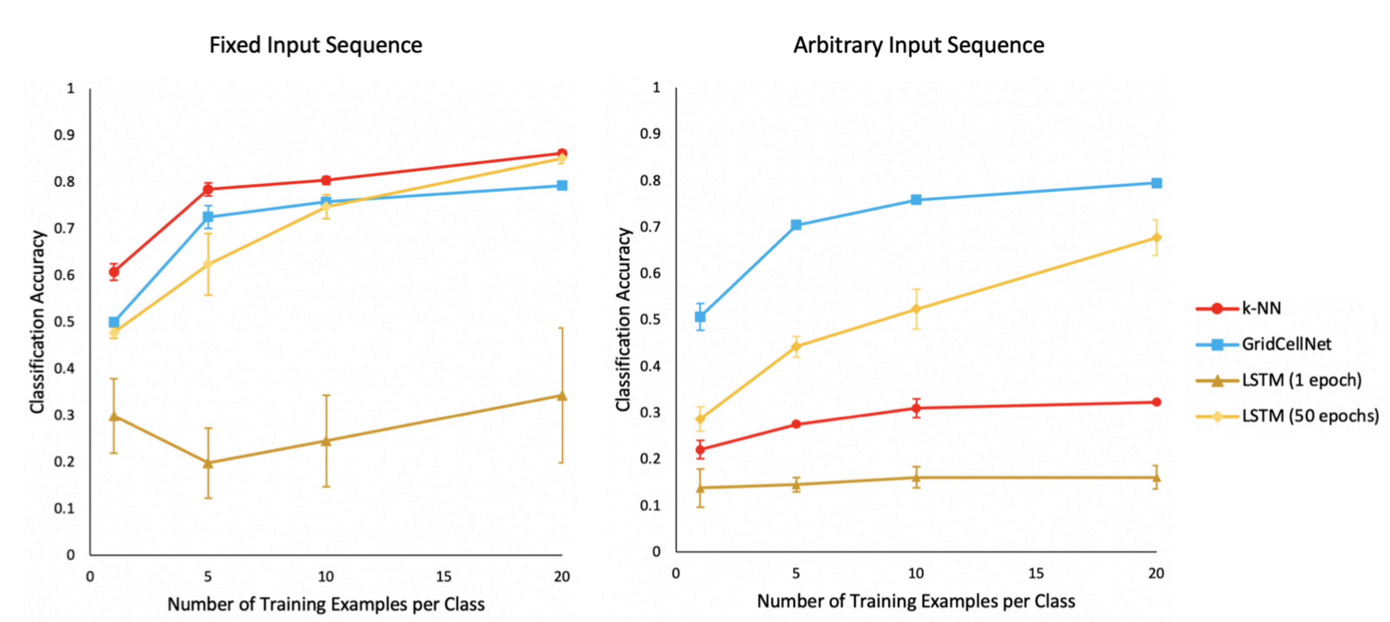


Fig.3 : a) Intuition: integrate features across space that, in isolation, are ambiguous. b) Features (sparse vectors) extracted from CNN feature map (upper row) are input to sensorimotor network (bottom row). Classified (yellow star) if location representation a subset of learned examples for a class at a given position m.

Results

- Evaluate classification w/ sequences of MNIST feature regions (same CNN input for all classifiers). Sequence either fixed for all training and evaluation, or arbitrary



Key Results

- GridCellNet (ours) robust to different paths of sampled features across space, outperforming LSTM and k-NN
- Strong performance despite limited training examples
- Unlike [2], works with images; unlike [1] and [2], generalizes to unseen examples, rather than training set

Results

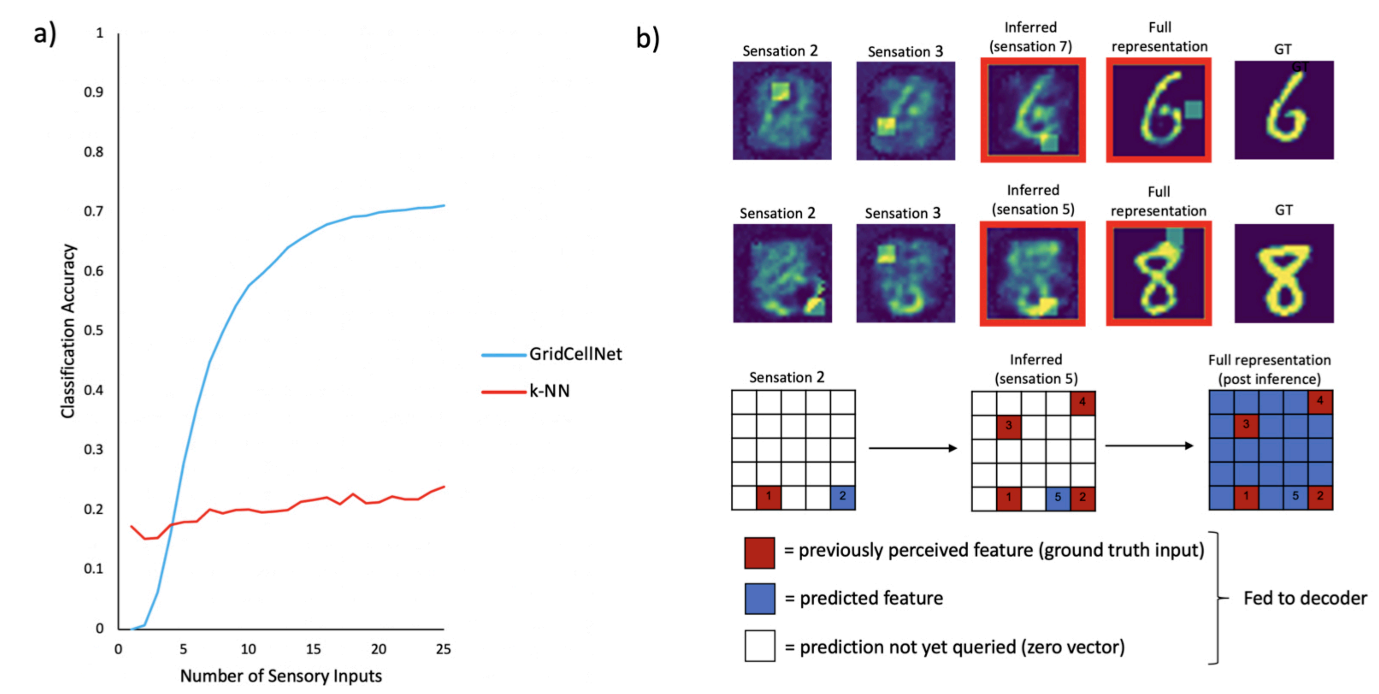


Fig. 5: a) Accuracy as a function of elements in input sequence observed. b) Prediction of unseen features in sequence. Conceptual depiction of representations at bottom.

Key Results

- Recognition often rapid (before all possible sensations experienced)
- GridCellNet also predicts unseen regions, pulling similar features from learned examples

Takeaways

Summary

- Present a biologically motivated network that uses grid cell computations (**path integration**) and sensory input to infer location representation
- Path integration endows the system's object recognition with **robustness to arbitrary sequence inputs** as would be expected during naturalistic vision
- Inference often occurs after only a few 'fixations', and network can predict features in unobserved input space

Key Takeaway

Grid-cell computations could underpin strong human performance in object recognition settings that challenge current machine learning systems. Employing these methods in artificial systems could bring benefits in robustness and flexibility.

References

[1] Andrej Bicanski and Neil Burgess. "A Computational Model of Visual Recognition Memory via Grid Cells". In: *Current Biology* (2019). ISSN: 09609822. DOI: 10.1016/j.cub.2019.01.077.

[2] Marcus Lewis et al. "Locations in the neocortex: A theory of sensorimotor object recognition using cortical grid cells". In: *Frontiers in Neural Circuits* (2019). ISSN: 16625110. DOI: 10.3389/fncir.2019.00022.

[3] Alfred L Yarbus. *Eye Movements During Perception of Complex Objects*. 1967. DOI: 10.1007/978-1-4899-5379-7{_}8.