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

Pre-Print

Motivation: Recognition Given Visual Sequences

- Biological vision based on (and robust to) saccades [3]

Problem:

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

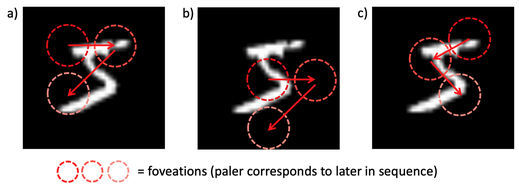


Fig 1: a) Sequence of hypothetical foveations during learning. b) During inference, the initial sensation may begin on a different part of the object (translation). Classifiers with a rigid input sequence requirement won't adequately sample the image. c) Regardless of where inference begins on the object (i.e. translation invariance), the trajectory should sample the object. This requires the classifier i) can integrate arbitrary spatial sequence paths, and ii) relies on spatial locations of features in an internal reference frame.

Key Ideas:

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

Cortical Networks with Grid Cells

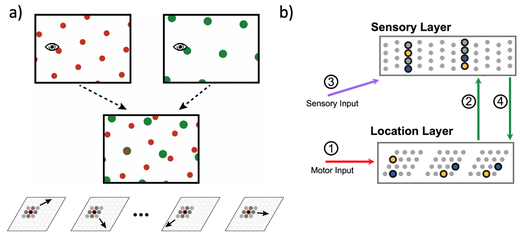


Fig. 2: a) Activity from multiple grid cell modules (different scale and orientation) uniquely encodes position. Locations updated with self-movement (**path integration**) b) Sensorimotor network w/ two layers (sensory: columns, spatial: grid cell modules) that reciprocally predict sparse representations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

Our approach ("GridCellNet")

- Core architecture as in [2], but pre-process images (MNIST) w/ CNN to create a map approximating foveal responses.
- Also extend [2] classification algorithm to integrate multiple learned examples (supports generalization)

Classification Algorithm

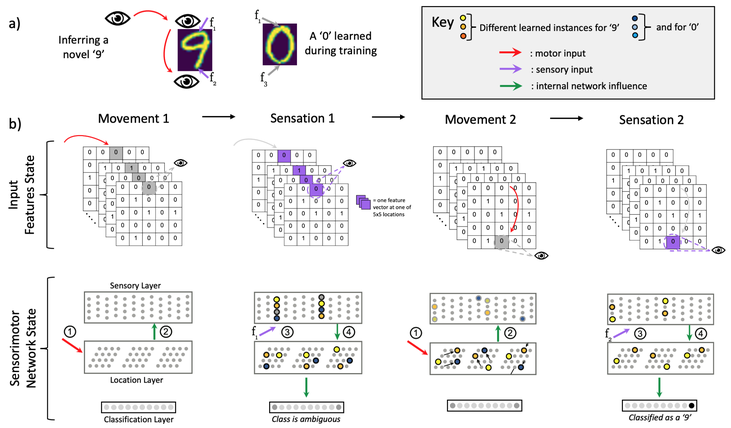
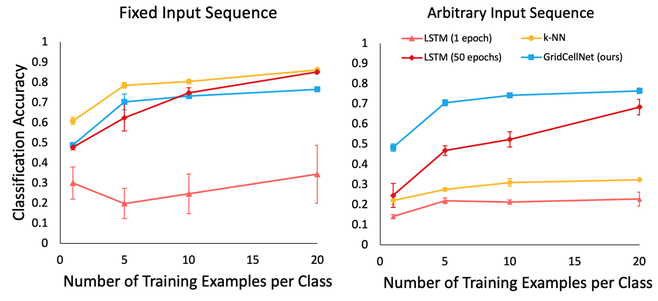


Fig.3 : a) Intuition: integrate features across space that, in isolation, are ambiguous. b) Features (sparse vectors) extracted from CNN feature map (upper row) are input to sensorimotor network (bottom row). Progresses through stages of (1) movement, (2) sensory prediction, (3) sensation, (4) location prediction. Classified if location representation drives a particular class node above threshold relative to other classes.

Results

- Evaluate classification w/ sequences of MNIST feature regions (same CNN input for all classifiers). Starting point and sequence either fixed during all training and evaluation, or arbitrary for any given object.



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. Unlike [1], uses internal reference frame to infer positions, supporting translation invariance.

Results

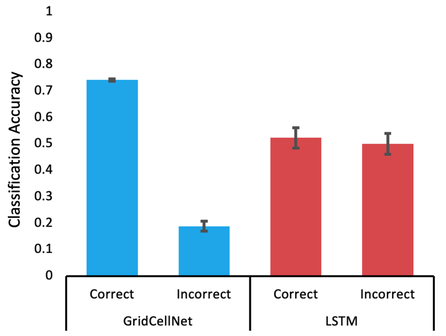


Fig. 5: Accuracy of an LSTM and our GridCellNet classifier given 10 training-examples per class and arbitrary sequence inputs. Shown are results either with correct (i.e. true) or incorrect information about the sensor's movement.

Key Results

- Performance of GridCellNet is crucially related to integration of features across space. Given false self-movement information, accuracy drops considerably
- LSTM is also provided self-movement to allow a fair comparison, but providing fabricated movement information during inference has virtually no effect on performance (i.e. bag-of-feature integration)

Takeaways

Summary

- Present a biologically motivated network that uses grid cell computations (**path integration**) and sensory inputs to recognise objects
- Path integration endows the system with **robustness to arbitrary input starting positions and sequences** as would be expected during naturalistic vision
- System uses self-movement to integrate features across 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.