GRID CELL PATH INTEGRATION FOR MOVEMENT-BASED VISUAL OBJECT RECOGNITION

Anonymous authors
Paper under double-blind review

ABSTRACT

Grid cells enable the brain to model the physical space of the world and navigate effectively via path integration, updating self-position using information from self-movement. Recent proposals suggest that the brain might use similar mechanisms to understand the structure of objects in diverse sensory modalities, including vision. In machine vision, object recognition given a sequence of sensory samples of an image, such as saccades, is a challenging problem when the transition order taken over the object must generalize to a novel, never-experienced sequence order - yet this is something humans do naturally and effortlessly. We explore how grid cell-based path integration in a cortical network can support reliable recognition of objects given an arbitrary sequence of sensory inputs. Our network (GridCellNet) uses grid cell computations to integrate visual information and make predictions based on movements. We use local Hebbian plasticity rules to learn rapidly from a handful of examples (few-shot learning), and consider the task of recognizing MNIST digits given a sequence of image feature patches. We compare GridCellNet to k-Nearest Neighbour (k-NN) classifiers as well as recurrent neural networks (RNNs), both of which lack explicit mechanisms for generalizing to arbitrary sequences of input samples. We show that GridCellNet can reliably perform classification, generalizing to both unseen MNIST digits and novel sequence trajectories over the digit. This work builds on previous efforts leveraging grid cells in visual object recognition (Bicanski & Burgess, 2019), but demonstrates generalization to novel objects rather than the recall of memorized examples. Additionally, our system goes further to utilize grid cells for an internal reference frame derived from sensory inputs and internal motor information alone, endowing the classification process with translation invariance. In particular, classification generalizes to sequences starting at arbitrary regions of the object without any recourse to an external reference frame.

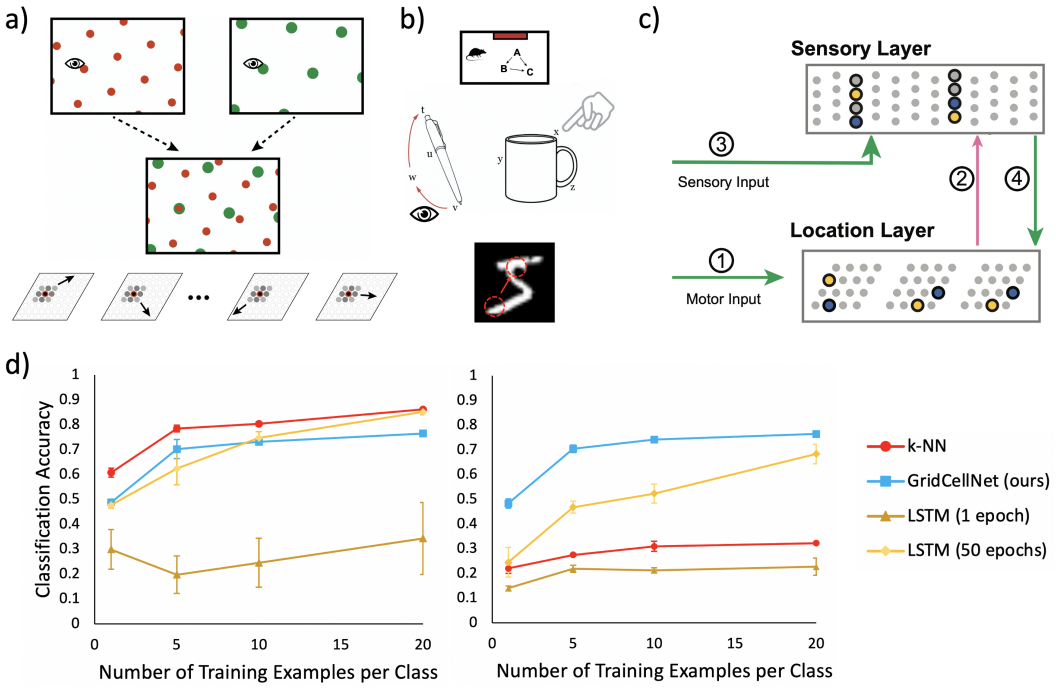

Figure 1: a) The combination of multiple grid-cells of different scale and phase (red and green) can uniquely encode the position of a sensor (e.g. retinal patch). Here we use grid cell modules with sparse activity (bottom) to encode and update the sensor's position. b) We hypothesize that this process can be used for object recognition with active sensors, and use sequences through a 5x5 grid of feature patches from MNIST images. c) GridCellNet takes in motor input when the sensor moves (1) and updates its location representation. The current location representation is used to predict incoming sensory information (2), before this is received (3). Correctly predicted sensory information is then used to update the location representation (4). Locations are initially ambiguous, represented using a union of locations, and disambiguated over time via sensory input. (Yellow and blue dots in the location layer indicate two different objects which are compatible with the current sequence.) Classification occurs when active grid cells drive a particular class neuron above threshold (not shown). The network is based on cells with sparse binary activity, dendritic segments, and Hebbian-like learning, following Marcus Lewis et al, 2019 (Frontiers in Neural Circuits 13, 22), and with figures reproduced with permission from the authors. d) Classification accuracy on a subset of the MNIST test set as a function of the number of training examples per-class. (Left) Accuracy when an identical sequence of passes over the input space is used for both training and inference. All three classifiers do well for few-shot MNIST, although note our inputs are pre-processed features rather than pixels. (Right) Performance when the sequence can be arbitrary and different between training and inference. GridCellNet maintains its accuracy, unlike the other classifiers. We include an LSTM with only 1-epoch of training as the k-NN and GridCellNet observe each feature only once. Error bars show the 95% confidence interval of the mean across three random seeds.