Wharton
University of Pennsylvania

Neuroscience-Inspired Deep Learning

15 Nov 2018

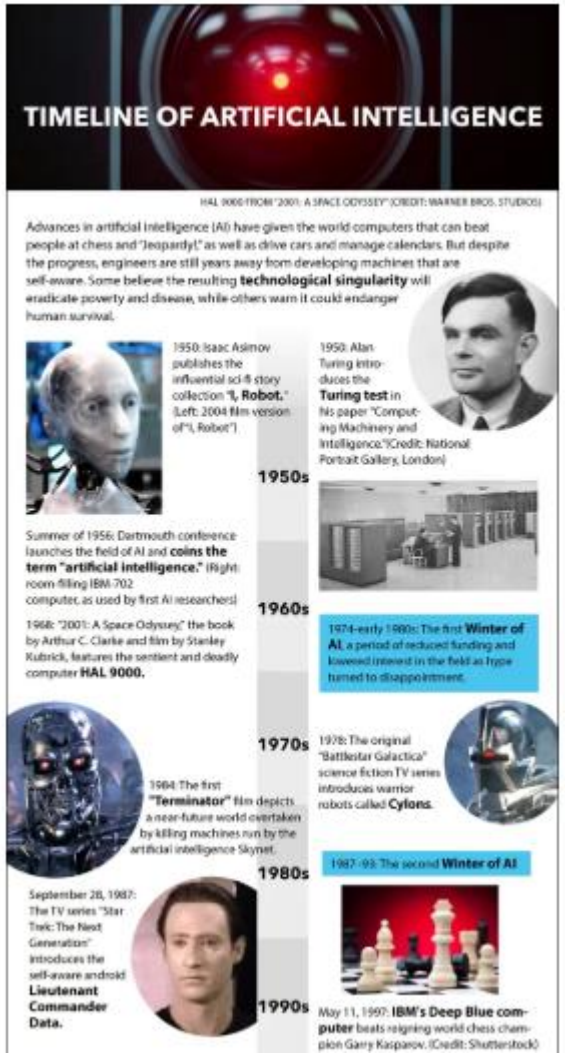
Executive Summary

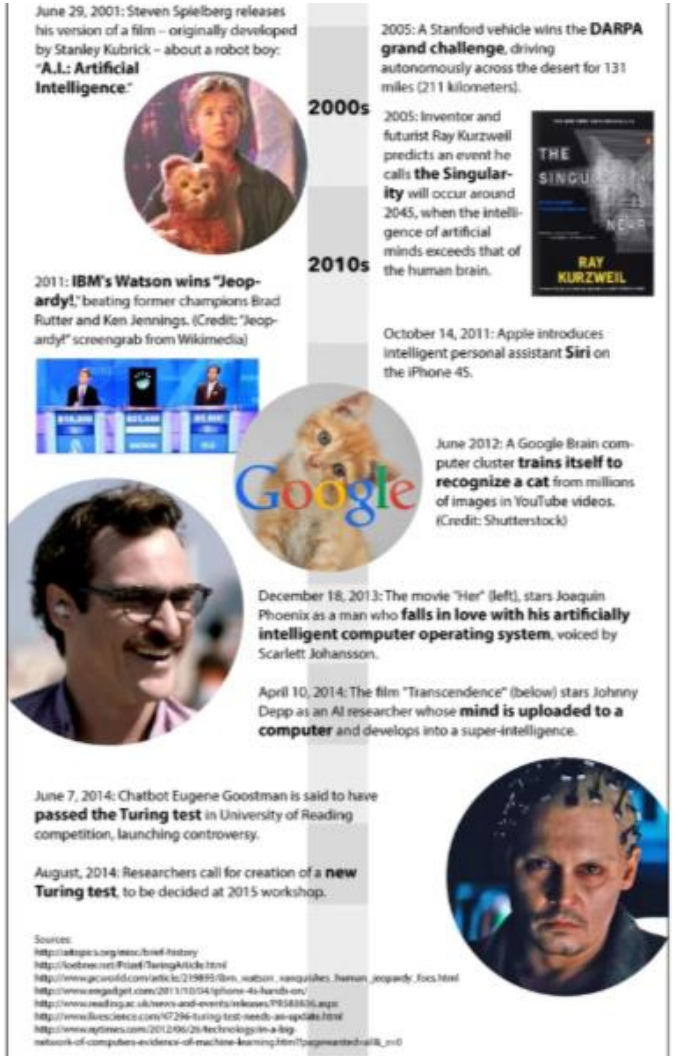
- Early work in AI was intertwined with neuroscience, with many pioneers straddled both fields and collaborations between these disciplines proving productive. However, the interaction between AI and neuroscience has become significantly less common, as both subjects have gained enormous complexity. As AI/ deep learning have laudable achievements, there seems to be **mixed attitude of contemporary AI scholars** towards the **potential impact of neuroscience findings on the development of AI**.
- This presentation aims to present the perspective in **Hassabis et al. (2017)** that **neuroscience provides a wealth of inspiration for new types of algorithms and architectures** that **complement the mathematical and logic-based approaches** that have largely dominated traditional approaches to AI through highlighting **the key arguments** and **selected neuroscience ideas/ findings** that **have had profound impact on or can potentially impact future AI/ deep learning research.**
- This presentation also briefly introduces the **vector-based navigation** model **using grid-like representations in artificial agents**, as featured in Banino et al. (2018), in order to showcase a recent example in which **promising results in navigation were achieved** through **inspiration from the neuroscience findings concerning grid cells.**

Agenda

- Background & Motivation
- Neuroscience-inspired AI – The past
  - Convolutional neural network
  - Reinforcement learning
- Neuroscience-inspired AI – The present
  - Attention
  - Episodic memory
- Neuroscience-inspired AI – The future
- Case study: vector-based navigation using grid-like representations in artificial agents

AI has evolved through decades…

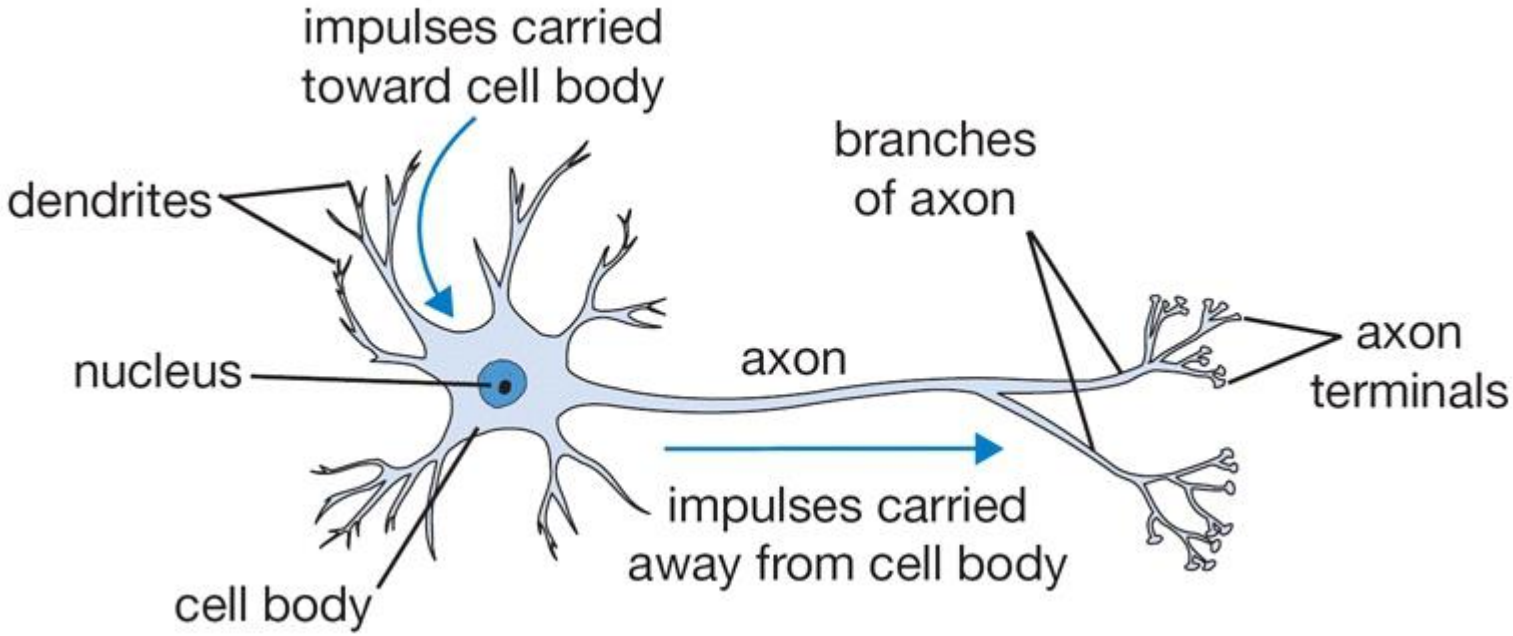




Source: https://www.livescience.com/47544-history-of-a-i-artificial-intelligence-infographic.html

A neuron – The basic computational unit of the brain

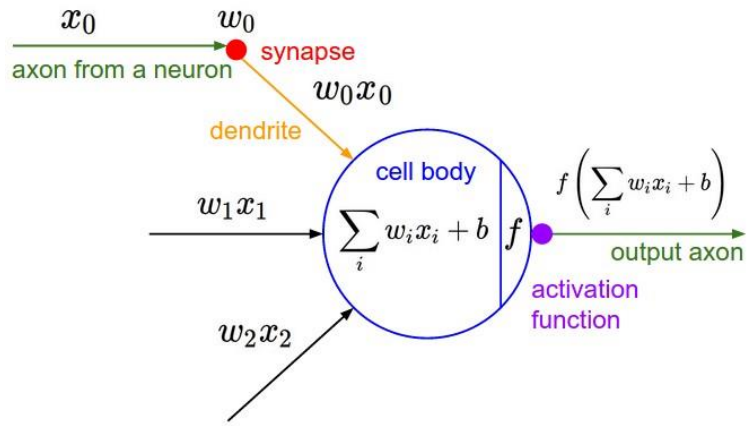
- Approximately 86 billion neurons can be found in the human nervous system and they are connected with approximately 10^14 - 10^15 synapses.
- Each neuron receives input signals from its dendrites and produces output signals along its (single) axon.
- The axon eventually branches out and connects via synapses to dendrites of other neurons.
- The dendrites carry the signal to the cell body where they all get summed. If the final sum is above a certain threshold, the neuron can fire, sending a spike along its axon.



Source: http://cs231n.github.io/neural-networks-1/#quick

A rather coarse model is used for modeling one neuron in AI

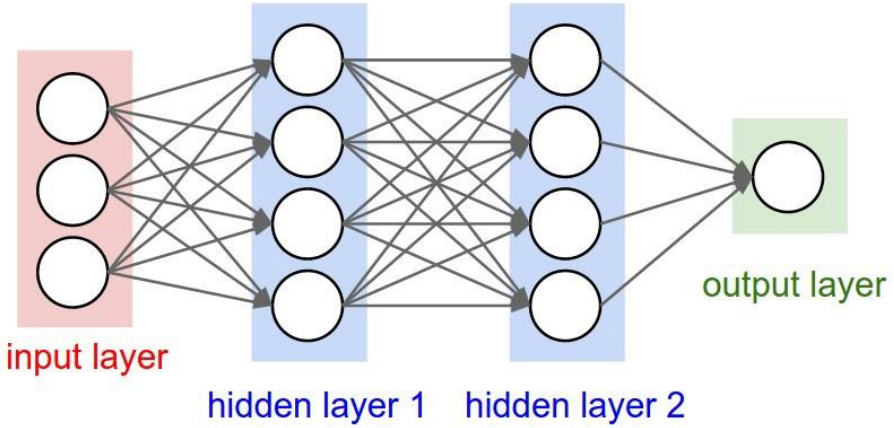
- In the computational model of a neuron, the signals that travel along the axons (e.g. x0) interact multiplicatively (e.g. w0x0) with the dendrites of the other neuron based on the synaptic strength at that synapse (e.g. w0).
- The idea is that the synaptic strengths (the weights w) are learnable and control the strength of influence (and its direction: excitory (positive weight) or inhibitory (negative weight)) of one neuron on another.
- In the computational model, we assume that the precise timings of the spikes do not matter, and that only the frequency of the firing communicates information. Based on this rate code interpretation, we model the firing rate of the neuron with an activation function f, which represents the frequency of the spikes along the axon.
- This model of a biological neuron is very coarse. For example, there are many different types of neurons, each with different properties. The dendrites in biological neurons perform complex nonlinear computations. The synapses are not just a single weight, but a complex non-linear dynamical system. The exact timing of the output spikes in many systems is important, suggesting that the rate code approximation may not hold.

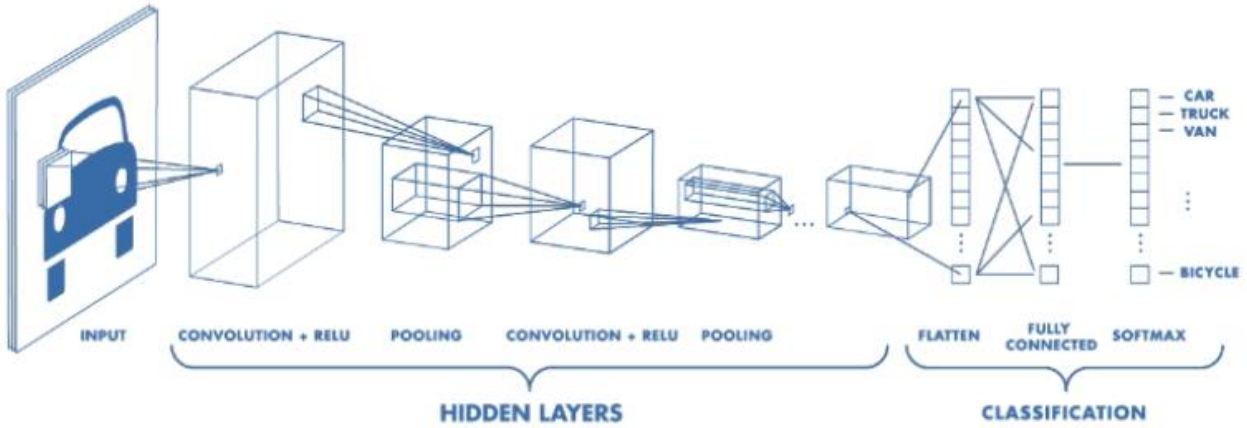


Source: http://cs231n.github.io/neural-networks-1/#quick

AI has grown tremendously in complexity…

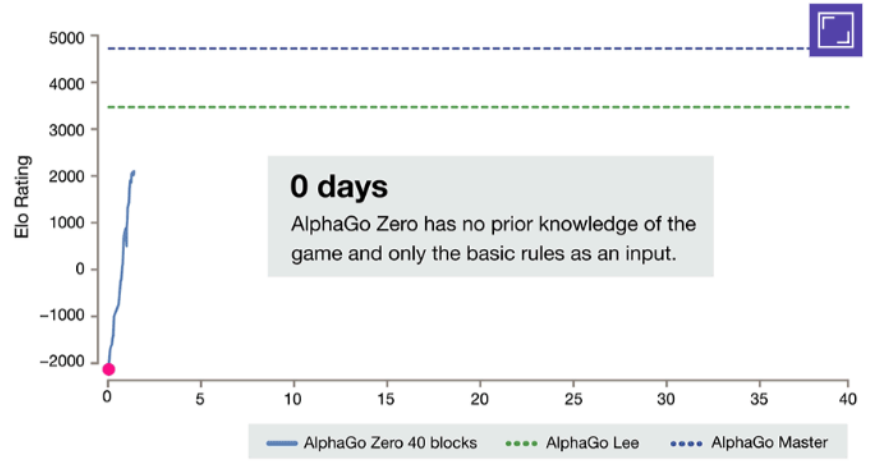
Pictorial illustration only





$$
\begin{aligned}
\log p_\theta(x^{(i)}) &= \mathbf{E}_{z\sim q_\phi(z|x^{(i)})}\left[\log p_\theta(x^{(i)})\right] && (p_\theta(x^{(i)}) \text{ Does not depend on } z)\\
&= \mathbf{E}_z\left[\log \frac{p_\theta(x^{(i)} \mid z)p_\theta(z)}{p_\theta(z \mid x^{(i)})}\right] && \text{(Bayes' Rule)}\\
&= \mathbf{E}_z\left[\log \frac{p_\theta(x^{(i)} \mid z)p_\theta(z)}{p_\theta(z \mid x^{(i)})}\frac{q_\phi(z \mid x^{(i)})}{q_\phi(z \mid x^{(i)})}\right] && \text{(Multiply by constant)}\\
&= \mathbf{E}_z\left[\log p_\theta(x^{(i)} \mid z)\right] - \mathbf{E}_z\left[\log \frac{q_\phi(z \mid x^{(i)})}{p_\theta(z)}\right] + \mathbf{E}_z\left[\log \frac{q_\phi(z \mid x^{(i)})}{p_\theta(z \mid x^{(i)})}\right] && \text{(Logarithms)}\\
&= \mathbf{E}_z\left[\log p_\theta(x^{(i)} \mid z)\right] - D_{KL}(q_\phi(z \mid x^{(i)}) \,||\, p_\theta(z)) + D_{KL}(q_\phi(z \mid x^{(i)}) \,||\, p_\theta(z \mid x^{(i)}))
\end{aligned}
$$

… and has impressive achievements



0 days

AlphaGo Zero has no prior knowledge of the game and only the basic rules as an input.

21 days

AlphaGo Zero reaches the level of AlphaGo Master, the version that defeated 60 top professionals online and world champion Ke Jie in 3 out of 3 games in 2017.

Captured Stones

3 hours

AlphaGo Zero plays like a human beginner, forgoing long term strategy to focus on greedily capturing as many stones as possible.

Captured Stones

70 hours

AlphaGo Zero plays at super-human level. The game is disciplined and involves multiple challenges across the board.

Source: https://deepmind.com/blog/alphago-zero-learning-scratch/

Mixed views of contemporary AI scholars on the potential impact of neuroscience on AI

"So **what does deep learning have to do with the brain**? At the risk of giving away the punchline, I would say **not a whole lot**."[1]

- Andrew Ng -

"Let's be **inspired by nature, but not too much**"[2]

- Yann LeCun -

"One is to **use neuroscience as a source of inspiration for algorithmic and architectural ideas**. The human brain is the only existing proof we have that the sort of general intelligence we're trying to build is even possible, so we think it's worth putting the effort in to try and understand how it achieves these capabilities. Then **we can see if there are ideas we can transfer over into machine learning and AI**."

- Demis Hassabis -

Source: 1. "Neural Networks and Deep Learning - What does this have to do with the brain?". Retrieved at https://www.coursera.org/lecture/neural-networks-deep-learning/what-does-this-have-to-do-with-the-brain-obJnR

2. "Deep Learning Tutorial", ICML, Atlanta, 2013-06-16. Retrieved at https://cs.nyu.edu/~yann/talks/lecun-tutorial-icml-2013.pdf

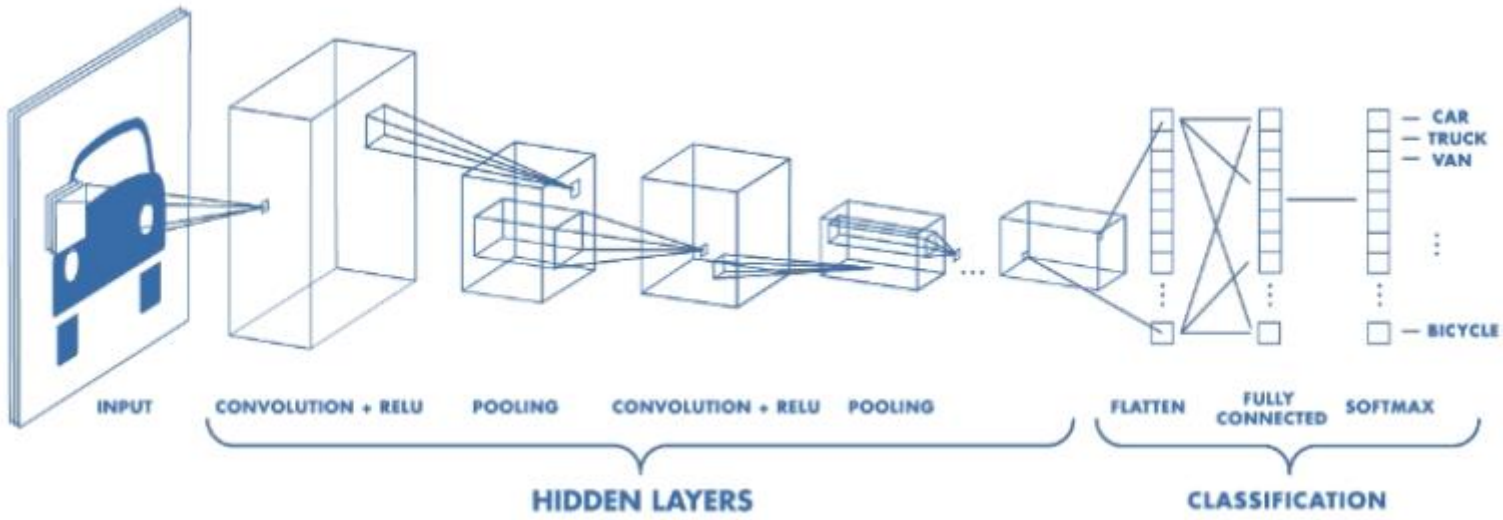
3. https://www.theverge.com/2017/7/19/15998610/ai-neuroscience-machine-learning-deepmind-demis-hassabis-interview

Neuroscience-inspired AI – The past

Convolutional Neural Network – Architecture

CNNs have two components

- Feature extraction component: perform a series of convolutions and pooling operations during which the features are detected
- Classification component: assign a probability for the object on the image being what the algorithm predicts it is

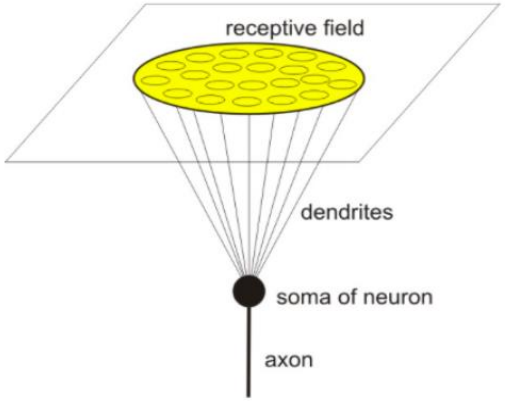


Source: https://www.mathworks.com/videos/introduction-to-deep-learning-what-are-convolutional-neural-networks--1489512765771.html

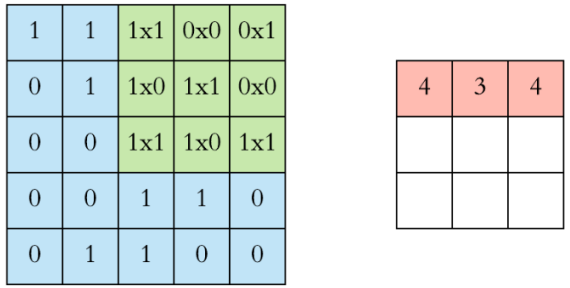
Neuroscience-inspired AI – The past

Convolutional Neural Network – Neuroscience inspiration

- Research done by D.H Hubel and T.N Wiesel on mammal brains in the 1950s and 1960s suggested a new model for how mammals perceive the world visually.
- The single-cell recordings from the visual cortex of mammals that revealed how visual input is filtered and pooled in simple and complex cells (with larger receptive fields than simple cells, receiving projections from a number of cells with simple fields; stimulus being effective wherever it was placed in the field provided that the orientation was appropriate) inspired CNN and associated operations.



The receptive field of an individual sensory neuron is the region of the sensory space in which a stimulus will modify the firing of that neuron (e.g. a piece of retina in the case of vision)



The filter (the green square, called receptive field) is sliding over the input (the blue square); the sum of the convolution goes into the feature map (the red square)

Source: 1. http://neuroclusterbrain.com/neuron_model.html

2. https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2

Neuroscience-inspired AI – The past

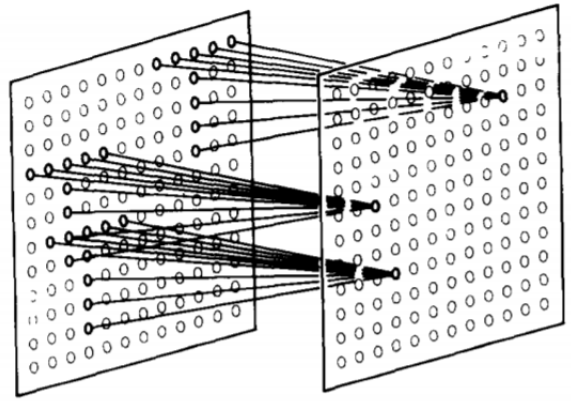
Convolutional Neural Network – Neuroscience inspiration

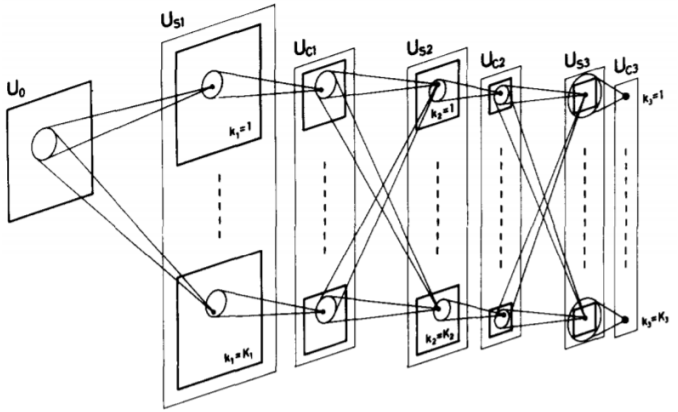
- Current network architectures mimic the hierarchical organization of mammalian cortical systems, following the ideas first proposed in early neural network models of visual processing by Fukushima in early 1980s.
- The neocognitron, proposed by Fukushima, has a hierarchical structure

➢ The information of the stimulus pattern given to the input layer of the neocognitron is processed stepwise in each stage of the network

➢ A cell belonging to a deeper stage tends to respond selectively to a more complicated feature of the stimulus patterns, has a larger receptive field and is less sensitive to shifts in position of the stimulus patterns

➢ Each cell in the deepest stage responds only to a specific stimulus pattern without being affected by the position or the size of the stimulus patterns



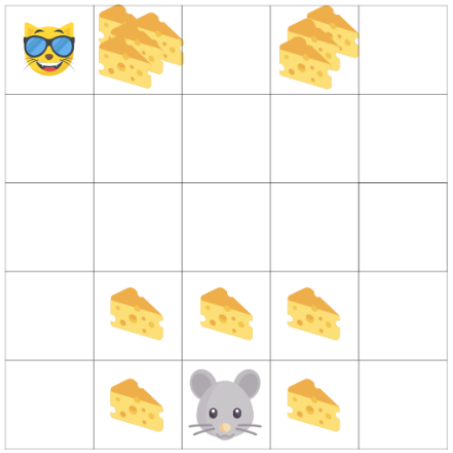


Source: Fukushima and Miyake. "Neocognitron: a new algorithm for pattern recognition tolerant of deformations and shifts in position ". *Pattern Recognition*, vol. 15, no 6, 1982, pp. 455-469.

Neuroscience-inspired AI – The past

Reinforcement Learning – Neuroscience inspiration

- Reinforcement learning (RL) addresses the question of how to maximize future reward by mapping states in the environment to actions. Although not widely appreciated among AI researchers, RL methods were originally inspired by research in animal learning. Particularly, the development of temporal-difference (TD) methods was intertwined with animal behavior in conditioning experiments.



$$V(S_t) \leftarrow \underbrace{V(S_t)}_{\text{Previous estimate}} + \alpha[\underbrace{R_{t+1}}_{\text{Reward t+1}} + \underbrace{\gamma V(S_{t+1})}_{\text{Discounted value on the next step}} - V(S_t)]$$

TD Target

- TD learning, instead of waiting until the end of the episode to update the expected future reward estimation, immediately forms a TD target based on observed reward R_{t+1} and estimate $V(S_{t+1})$.
- The origins of TD learning are partially in animal learning psychology. According to Sutton and Barto (1998), learning to predict via a TD algorithm corresponds to classical conditioning (a.k.a. Palovian conditioning, in which the reinforcing stimulus is contingent upon the animal's behavior, as opposed to instrumental conditioning).
- This above model can be further generalized by including the temporal dimension where events within individual trials influence learning, and it provides an account of second-order conditioning, where predictors of reinforcing stimuli become reinforcing themselves.

Source: Sutton and Barto. "Reinforcement Learning", 1998. MIT Press.

Neuroscience-inspired AI – The present

Attention – Neuroscience inspiration

- Up until recently, most convolutional neural networks worked with entire images or video frames, with equal priority given to all image pixels at the earliest stage of processing.
- However, primate visual systems are believed to work more strategically. Rather than processing all input in parallel, visual attention shifts among locations and objects, centering processing resources and representational coordinates on a series of regions in turn.
- There have been neuro-computational models that show how this piecemeal approach benefits behavior by prioritizing and isolating the information that is relevant at any given moment.
- Inspired by the idea above has been a source of inspiration for AI architectures that take glimpses of the input image at each step, update internal state representations and then select the next location to sample.

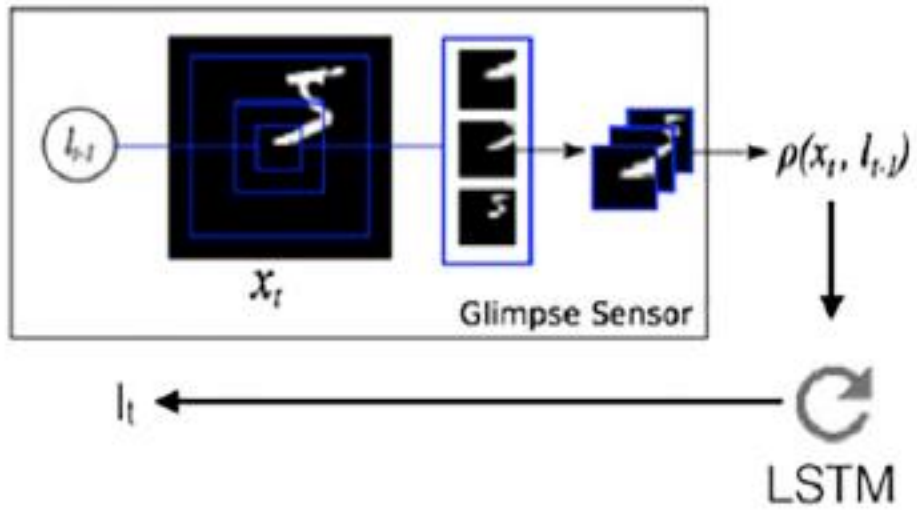
➢ Mnih et al. (2014) proposes a network that uses this selective attentional mechanism to ignore irrelevant objects in a scene, allowing the network to perform well in object classification tasks with clutter.

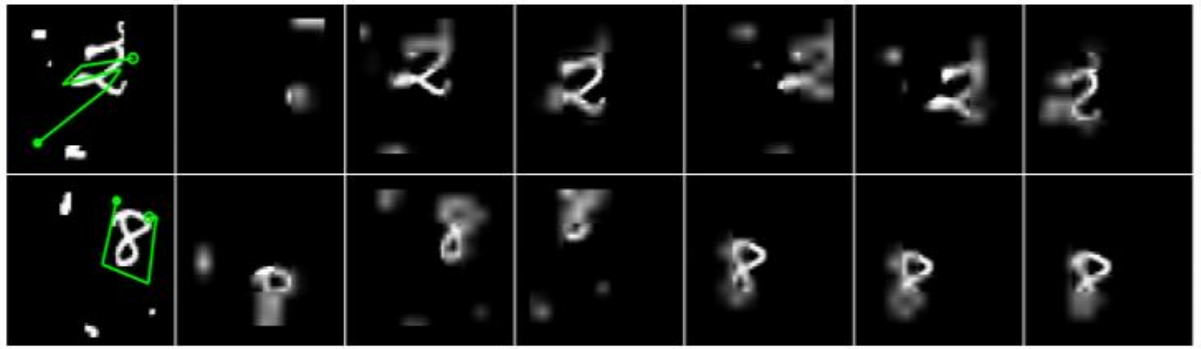
Source: Hassabis et al. “Neuroscience-inspired artificial intelligence”. Neuron, vol 95, issue 2, 2017, pp. 245-258.

Neuroscience-inspired AI – The present

Attention – Recurrent Attention Model (RAM) - Mnih et al. (2014)

- **Glimpse sensor:** Given the coordinates of the glimpse and an input image, the sensor extracts a retina-like representation ρ(xt, lt−1) centered at lt−1 that contains multiple resolution patches.
- **Rest of model:** This is the input to a glimpse network, which produces a representation that is passed to the LSTM core, which defines the next location to attend to (lt) (and classification decision).
- **Result:** The glimpse paths clearly show that the learned policy avoids computation in empty or noisy parts of the input space and directly explores the area around the object of interest.



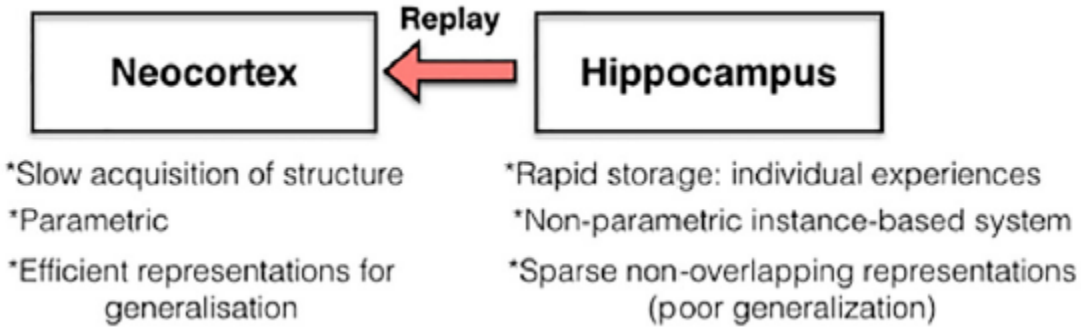


Source: Mnih et al. (2014). "Recurrent Models of Visual Attention". Retrieved at http://papers.nips.cc/paper/5542-recurrent-models-of-visual-attention

Neuroscience-inspired AI – The present

Episodic Memory – Neuroscience inspiration

- A prominent view in neuroscience is animal learning is supported by parallel or complementary learning systems in the hippocampus and neocortex.
- The hippocampus acts to encode novel information about a single exposure, but this information is gradually consolidated to the neocortex in sleep or resting periods that are interleaved with periods of activity.
- This consolidation is accompanied by replay in the hippocampus and neocortex, which is observed as a reinstatement of the structured patterns of neural activity that accompanied the learning event.
- The above idea inspired experience replay, a key ingredient in the deep Q-network (DQN), which exhibits expert play on Atari 2600 video games and illustrates the successful integration of reinforcement learning with deep learning.



Source: Hassabis et al. "Neuroscience-inspired artificial intelligence". Neuron, vol 95, issue 2, 2017, pp. 245-258.

Neuroscience-inspired AI – The present

Episodic Memory – Experience replay

- Learning from batches of consecutive samples is problematic because samples are correlated, which results in inefficient learning.
- In order to address the above problem, experience replay, whereby the network stores a subset of the training data in an instance-based way and then replays it offline, learning anew from successes or failures that occurred in the past.

➢ Continually update a replay memory table of transitions containing data on state, action and reward (st, at, rt, st+1) as game (experience) episodes are played

➢ Train Q-network on random minibatches of transitions from the replay memory, instead of consecutive samples

Source: 1. Hassabis et al. “Neuroscience-inspired artificial intelligence”. Neuron, vol 95, issue 2, 2017, pp. 245-258.

2. http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture14.pdf

Neuroscience-inspired AI – The future

Toward human-level intelligence – Neuroscience inspiration

- In neuroscience, the advent of new tools for brain imaging and genetic bioengineering have begun to offer a detailed characterization of the computations occurring in neural circuits, promising better understanding of mammal brain function.
- The relevance of neuroscience, both as a roadmap for AI research and a source of computational tools, is salient in the following areas:

➢ **Intuitive understanding of the physical world:** knowledge of core concepts relating to the physical world, such as space, number and objectness, which allow people to construct compositional mental models that can guide inference and prediction

➢ **Efficient learning:** ability to rapidly learn about new concepts from only a handful of examples, leveraging prior knowledge to enable flexible inductive inferences

➢ **Transfer learning:** ability to generalize or transfer generalized knowledge gained in one context to novel, previously unseen domains

➢ **Imagination and planning:** ability to flexibly select actions based on forecasts of long-term future outcomes through simulation-based planning, which uses predictions generated from an internal model of the environment learned through experience

➢ **Virtual brain analytics:** applying tools from neuroscience to AI systems, synthetic equivalents of single-cell recording, neuroimaging, and lesion techniques, we can gain insights into the key drivers of successful learning in AI research and increase the interpretability of these systems

Source: Hassabis et al. “Neuroscience-inspired artificial intelligence”. Neuron, vol 95, issue 2, 2017, pp. 245-258

Case study - Vector-based navigation using grid-like representations in artificial agents

Motivation & Overall approach

- Despite impressive successes in such fields as object recognition and complex games, navigation remains a substantial challenge for deep neural networks.
- ➢ Deep neural networks trained by reinforcement learning fail to rival the proficiency of mammal spatial behavior.
- Banino et al. (2018) seeks to leverage the computational functions of grid cells to develop a deep reinforcement learning agent with mammal-like navigational abilities.
- ➢ Grid cells are thought to provide a representation that functions as a metric for coding space and is critical for integrating self-motion (path integration) and planning direct trajectories to goals (vector-based navigation).
- ➢ The study first trained a recurrent network to perform path integration, leading to the emergence of representations resembling grid cells, as well as other entorhinal cell types.
- ➢ The study then aimed to show that this representation provided an effective basis for an agent to locate goals in challenging, unfamiliar, and changeable environments—optimizing the primary objective of navigation through deep reinforcement learning.

Source: Banino et al. "Vector-based navigation using grid-like representations in artificial agents". Nature, vol.557, issue 7705, 2018, pp. 429-433.

Case study - Vector-based navigation using grid-like representations in artificial agents

Neuroscience foundation

- An internal map of the environment and a sense of place are needed for recognizing and remembering our environment and for navigation. This navigational ability, which requires integration of sensory information, movement execution and memory capacities, is one of the most complex of brain functions.
- John O'Keefe discovered place cells in the hippocampus that signal position and provide the brain with spatial memory capacity. May-Britt Moser and Edvard I. Moser discovered in the medial entorhinal cortex, a region of the brain next to hippocampus, grid cells that provide the brain with an internal coordinate system essential for navigation. Together, the hippocampal place cells and the entorhinal grid cells form interconnected nerve cell networks that are critical for the computation of spatial maps and navigational tasks.
- The grid cells showed an astonishing firing pattern. They were active in multiple places in the open box that together formed nodes of an extended hexagonal grid (Figure 2), similar to the hexagonal arrangements of holes in a beehive. Using recordings from multiple grid cells in different parts of the entorhinal cortex, the Mosers also showed that the grid cells are organized in functional modules with different grid spacing ranging in distance from a few centimeters to meters, thereby covering small to large environments.

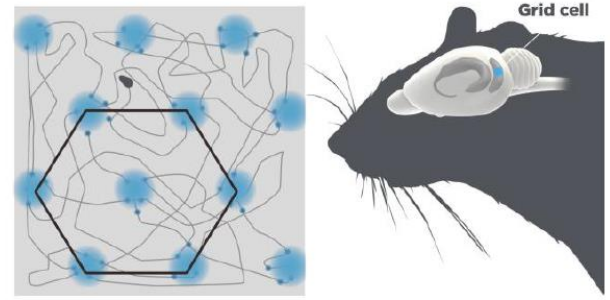


Figure 2. Grid cells. *The grid cells are located in the entorhinal cortex depicted in blue. A single grid cell fires when the animal reaches particular locations in the arena. These locations are arranged in a hexagonal pattern.*

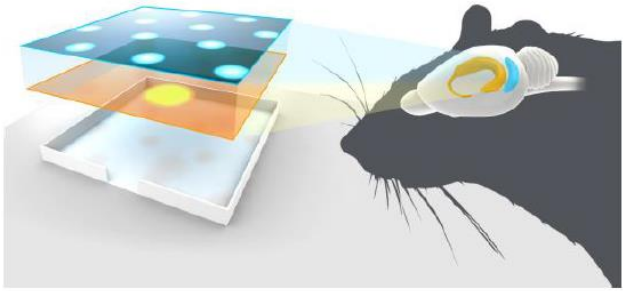
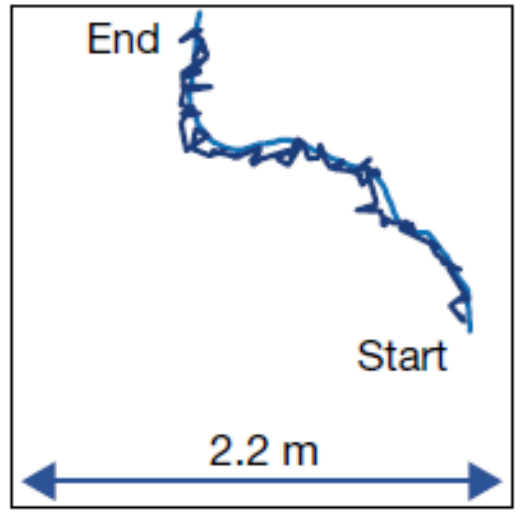


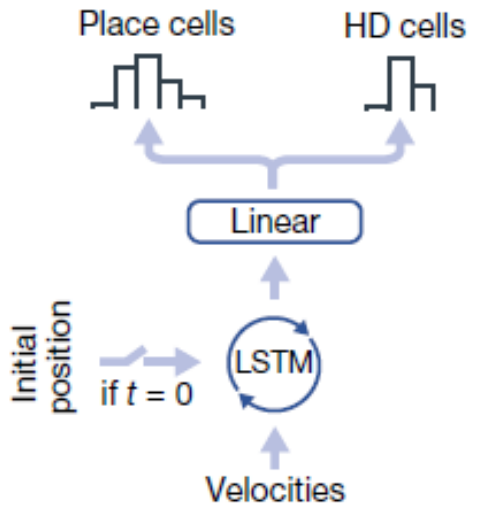
Figure 3. A schematic showing grid cells (blue) and place cells (yellow) in the entorhinal cortex and hippocampus, respectively.

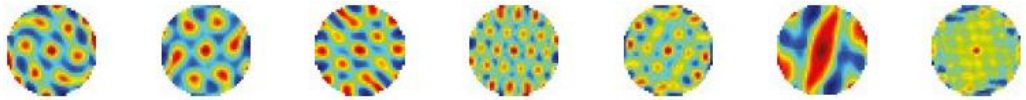
Case study - Vector-based navigation using grid-like representations in artificial agents

Phase 1. Training deep neural network to path integrate

- The network is trained in a square arena (2.2m x 2.2m), using simulated trajectories modelled on those of foraging rodents.
- The network was required to update its estimate of location and head direction using translational and angular velocity signals, mirroring those available to the mammalian brain.
- Velocity was provided as input to a recurrent network with a long short-term memory (LSTM) architecture, which was trained using backpropagation through time allowing the network to dynamically combine current input signals with activity patterns reflecting past events







- The LSTM projected to place and head direction units via a linear layer—units with activity defined as a simple linear function of their input
- The vector of activities in the place and head direction units, corresponding to the current position, was provided as a supervised training signal at each time step
- The network was able to path integrate accurately

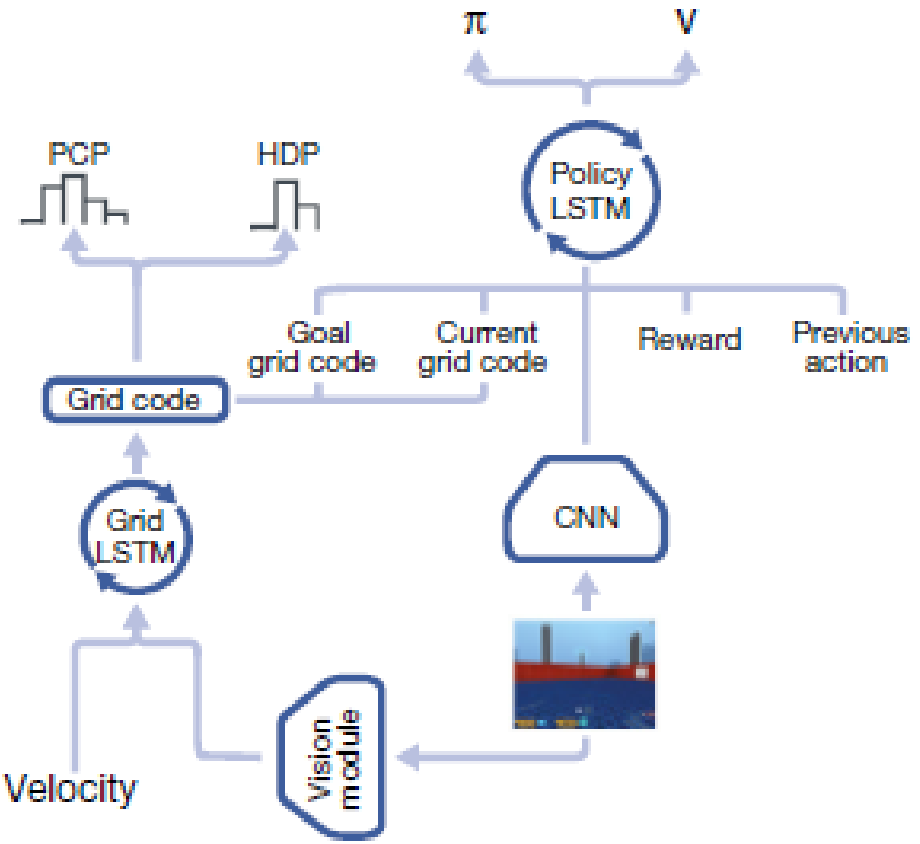
Source: Banino et al. "Vector-based navigation using grid-like representations in artificial agents". Nature, vol.557, issue 7705, 2018, pp. 429-433.

Case study - Vector-based navigation using grid-like representations in artificial agents

Phase 2. Test the hypothesis that the emergent representations provide an effective basis function for goal-directed navigation in novel challenging and changeable environments, when trained through deep reinforcement learning.

- The grid network was trained using supervised learning but, to better approximate the information available to navigating mammals, it now received velocity signals perturbed with random noise as well as visual input.
- Place cell input to grid cells is suggested, based on experimental evidence, to correct for drift and anchors grids to environmental cues.

  To parallel this, visual input was processed by a 'vision module' consisting of a convolutional network that produced place and head direction cell activity patterns that were provided as input to the grid network 5% of the time—akin to a moving animal making occasional, imperfect observations of salient environmental cues

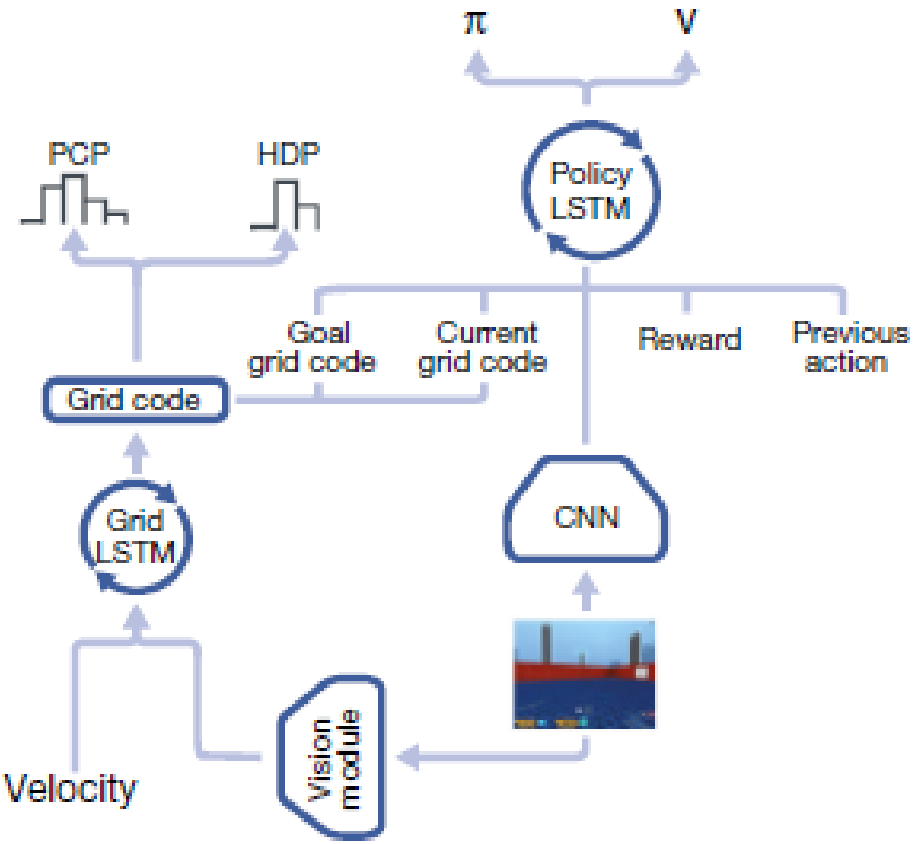


Source: Banino et al. "Vector-based navigation using grid-like representations in artificial agents". Nature, vol.557, issue 7705, 2018, pp. 429-433.

Case study - Vector-based navigation using grid-like representations in artificial agents

Phase 2. Test the hypothesis that the emergent representations provide an effective basis function for goal-directed navigation in novel challenging and changeable environments, when trained through deep reinforcement learning.

- The output of the linear layer of the grid network, corresponding to the agent's current location, was provided as input to the 'policy LSTM', a second recurrent network that both controls the agent's actions and outputs a value function.
- Additionally, whenever the agent reached the goal, the 'goal grid code'– activity in the linear layer—was subsequently provided to the policy LSTM during navigation as an additional input.

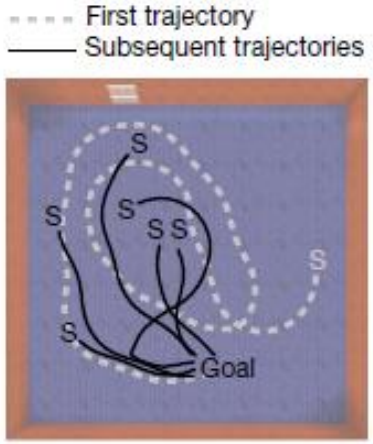


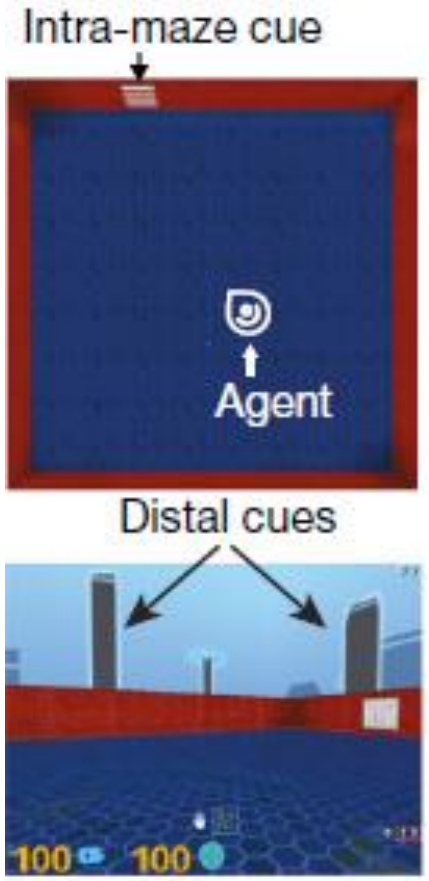
Source: Banino et al. "Vector-based navigation using grid-like representations in artificial agents". Nature, vol.557, issue 7705, 2018, pp. 429-433.

Case study - Vector-based navigation using grid-like representations in artificial agents

Phase 3. Examine the navigational capacities of the agent in a simple setting

- Examine in a 2.5 m × 2.5 m square arena
- The agent was still able to self-localize accurately in this more challenging setting, where ground truth information about location was not provided and velocity inputs were noisy
- Furthermore, the agent exhibited proficient goal-finding abilities, typically taking direct routes to the goal from arbitrary starting locations





Source: Banino et al. "Vector-based navigation using grid-like representations in artificial agents". Nature, vol.557, issue 7705, 2018, pp. 429-433.

Case study - Vector-based navigation using grid-like representations in artificial agents

Phase 3. Examine the agent's ability to perform vector-based navigation, enabling downstream regions to calculate goal-directed vectors by comparing current activity with that of a remembered goal

- In the agent, the authors expect these calculations to be performed by the policy LSTM, which receives the current activity pattern over the linear layer (termed 'current grid code') as well as that present the last time the agent reached the goal (termed 'goal grid code') and uses them to control movement.
- The authors demonstrated that withholding the goal grid code from the policy LSTM of the grid cell agent had a strikingly deleterious effect on performance.
- The authors also demonstrated that a targeted lesion (i.e. silencing) to the most grid-like units within the goal grid code should have a greater adverse effect on performance than a sham lesion (i.e. silencing of non-grid units).

Source: Banino et al. "Vector-based navigation using grid-like representations in artificial agents". Nature, vol.557, issue 7705, 2018, pp. 429-433.

Case study - Vector-based navigation using grid-like representations in artificial agents

Phase 3. Striking additional results

- A core feature of mammalian spatial behavior is the ability to exploit novel shortcuts and traverse unvisited portions of space, a capacity thought to depend on vector-based navigation.
- The grid cell agent robustly demonstrated these abilities in specifically designed neuroscience-inspired mazes, taking direct routes to the goal as soon as they became available.

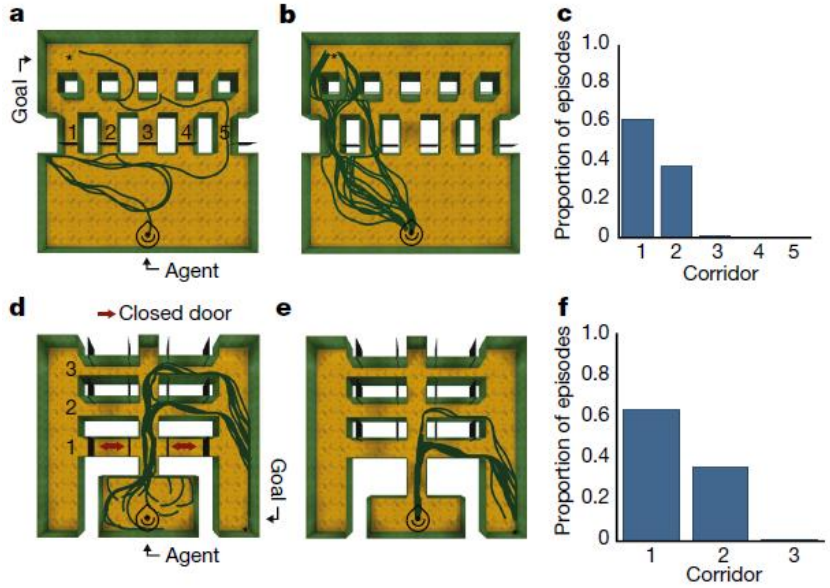


Fig. 4 | Flexible use of shortcuts. **a**, Example trajectory from grid cell agent in the linear sunburst maze (only door 5 open; icon indicates start location). **b**, Testing configuration with all doors open; grid cell agent uses the newly available shortcuts (100 episodes shown). **c**, Histogram showing agent's strong preference for most direct routes. **d**, Example grid cell agent trajectories (100) during training in the double E-maze (corridor 1 doors closed). **e**, Testing configuration with corridor 1 open, and 100 grid agent trajectories. **f**, Histogram analogous to **c** showing that agent prefers newly available shortest route. See Extended Data Fig. 10 for performance of place cell agent.

Source: Banino et al. "Vector-based navigation using grid-like representations in artificial agents". Nature, vol.557, issue 7705, 2018, pp. 429-433.

Reference

1. Hassabis et al. “Neuroscience-inspired artificial intelligence”. *Neuron,* vol 95, issue 2, 2017, pp. 245-258.
2. Hubel and Wiesel, “Receptive fields, binocular interaction and functional architecture in the cat's visual cortex”. *Physiol*, vol. 160, 1962, pp. 106-154.
3. Fukushima and Miyake. “Neocognitron: a new algorithm for pattern recognition tolerant of deformations and shifts in position ”. *Pattern Recognition*, vol. 15, no 6, 1982, pp. 455-469.
4. Sutton and Barto. “Reinforcement Learning”, 1998. MIT Press.
5. Mnih et al. “Recurrent Models of Visual Attention”, 2014. Retrieved at http://papers.nips.cc/paper/5542-recurrent-models-of-visual-attention
6. May-Britt Moser. “Grid Cells, Place Cells and Memory”, 2014. Retrieved at
7. Banino et al. “Vector-based navigation using grid-like representations in artificial agents”. *Nature,* vol.557, issue 7705, 2018, pp. 429-433.