GPU Computing for Cognitive Robotics

Martin Peniak, Davide Marocco, Angelo Cangelosi

GPU Technology Conference, San Jose, California, 25 March, 2014

Acknowledgements

- This study was financed by:
  - EU Integrating Projects - ITALK and Poeticon++ within the FP7 ICT programme Cognitive Systems and Robotics
  - ARIADNA scheme of The European Space Agency
- Thanks to my supervisors Prof Angelo Cangelosi, Dr. Davide Marocco and Prof Tony Belpaeme for their support
- Thanks to Calisa Cole and Chandra Cheij from NVIDIA for their help

New position at Cortexica

Imperial College London

CORTEXICA
vision systems
The Fashion Image Recognition Company

- Leading provider of visual search and image recognition technology for mobile device
- Creators of a bio-inspired vision system enabling intelligent image recognition using principles derived from the human sight

www.cortexica.com

Overview

- Action and language acquisition in humanoid robots
- Biologically-inspired Active Vision system
- Software development


esa
space for europe


AQUILA
Software Architecture for Cognitive Robotics

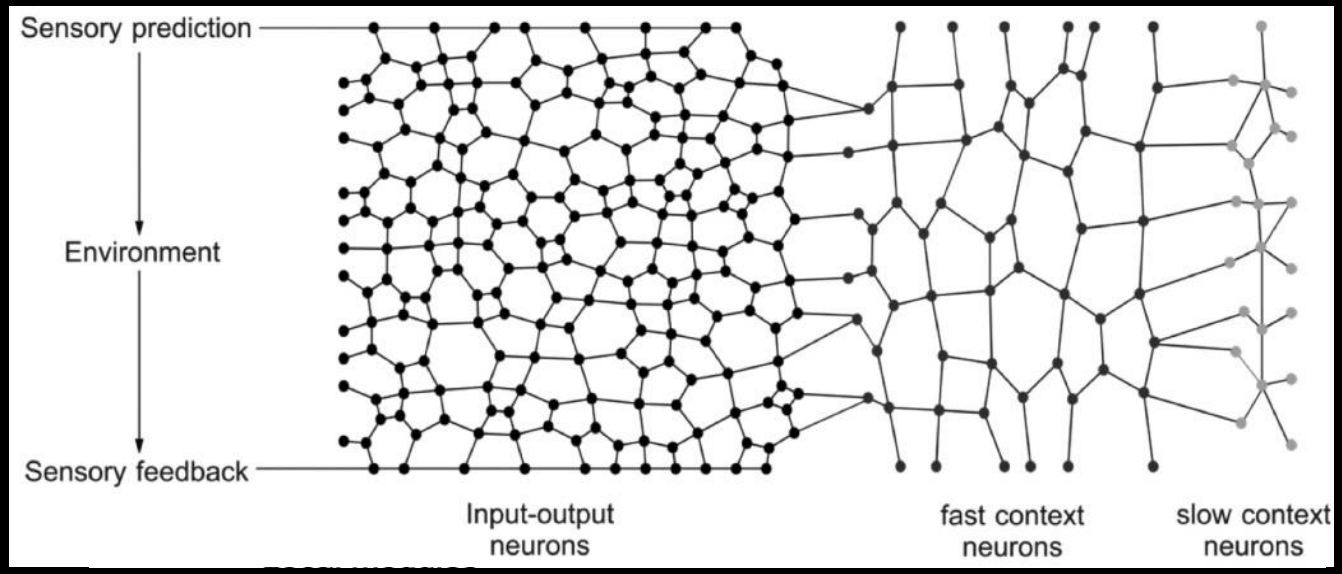

NVIDIA

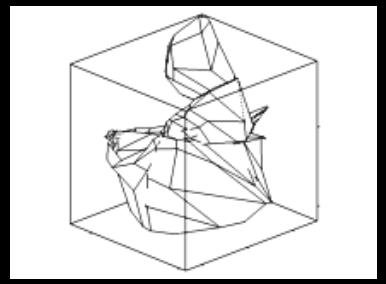
Action and Language Acquisition in Humanoid Robots

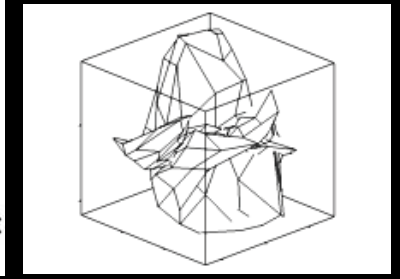
Learning Actions

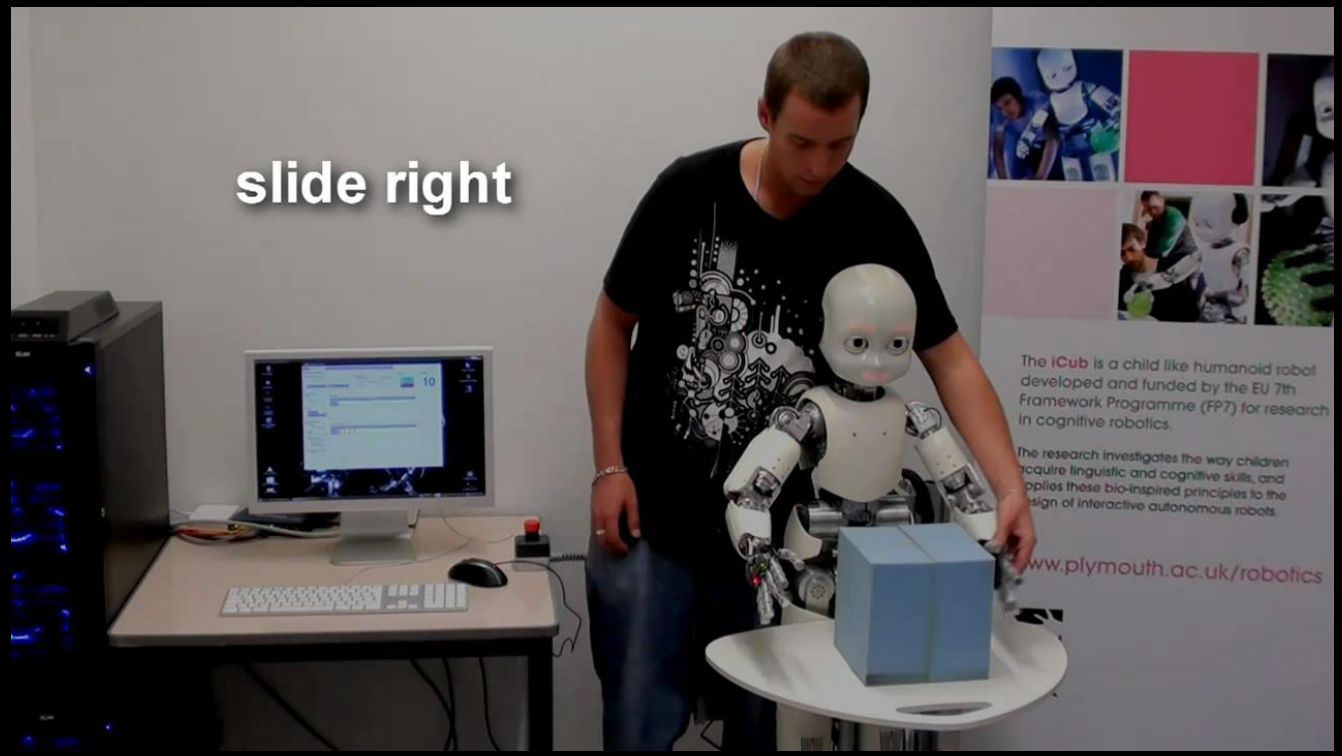
- Humans are good at learning complex actions
- Constant repetition of movements with certain components segmented as reusable elements
- Motor primitives are flexibly combined into novel sequences of actions
- Human motor control system known to have motor primitives implemented as low as at the spinal cord and hi-level planning and execution takes place in primary motor cortex

Explicit hierarchical structure vs multiple timescales

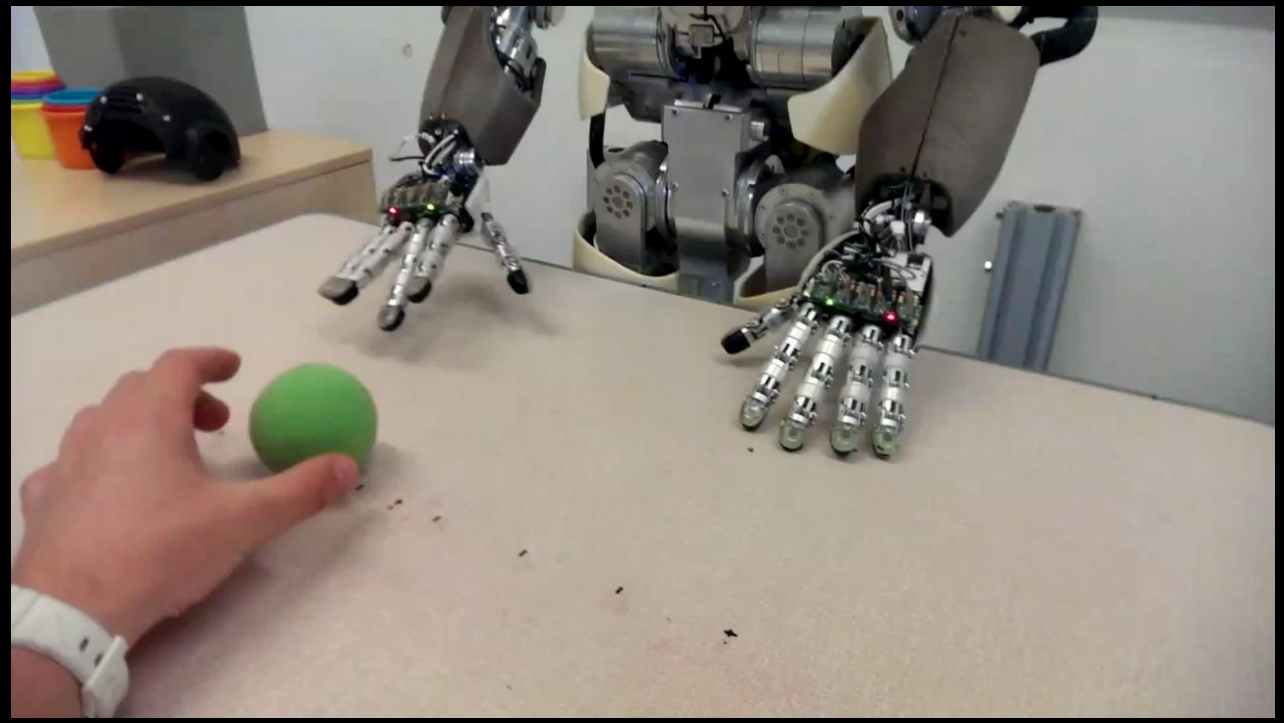







slide right
The iCub is a child like humanoid robot developed and funded by the EU 7th Framework Programme (FP7) for research in cognitive robotics.
The research investigates the way children acquire linguistic and cognitive skills, and applies these bio-inspired principles to the design of interactive autonomous robots.
www.plymouth.ac.uk/robotics

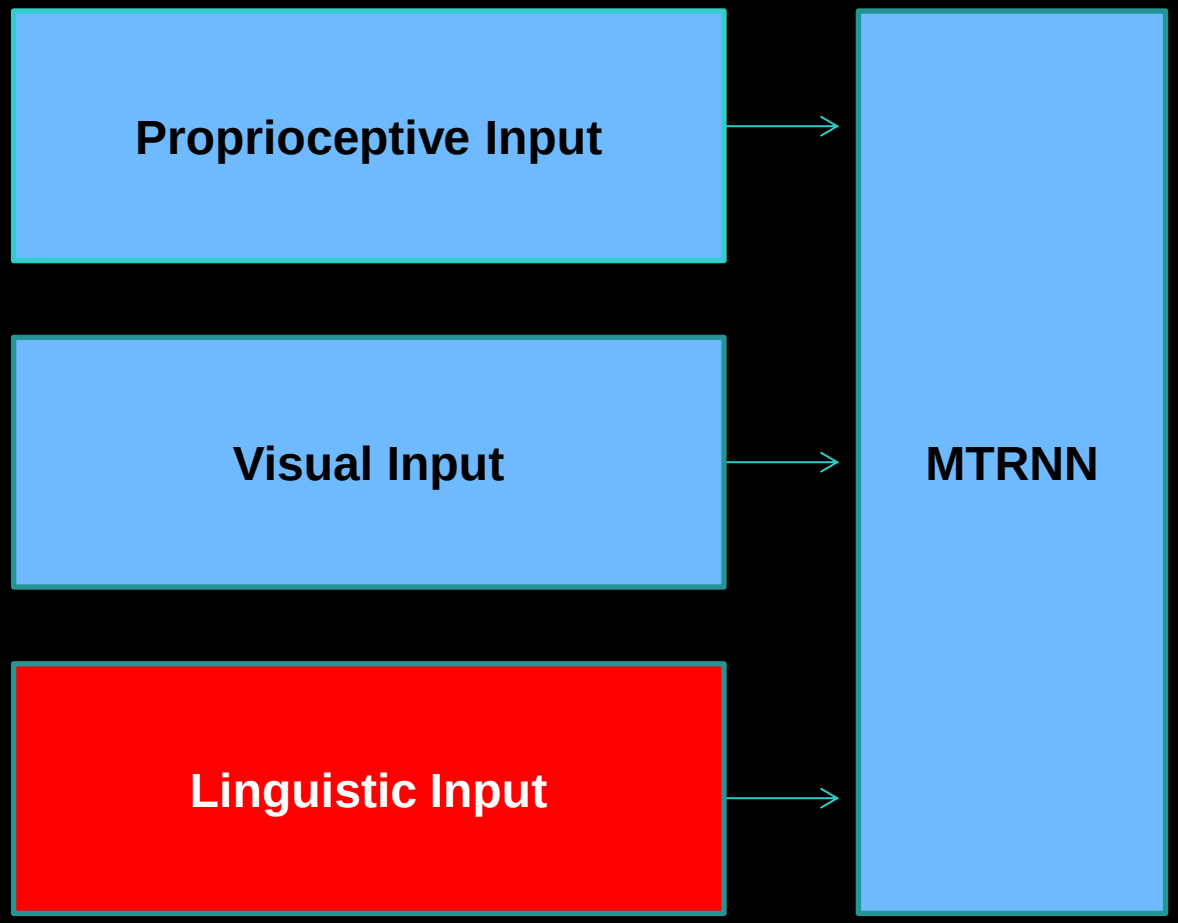
Initial testing of two actions



- Experimental Setup
  - SOM and MTRNN trained on 2 sequences each repeated 5x with different positions
  - Extended version of up to 9 action sequences
  - Left and Right hand used individually
  - MTRNN input: head, torso and arms (41 DOF)
  - Update rate: 50ms

Multiple Time-scales Recurrent Neural Network

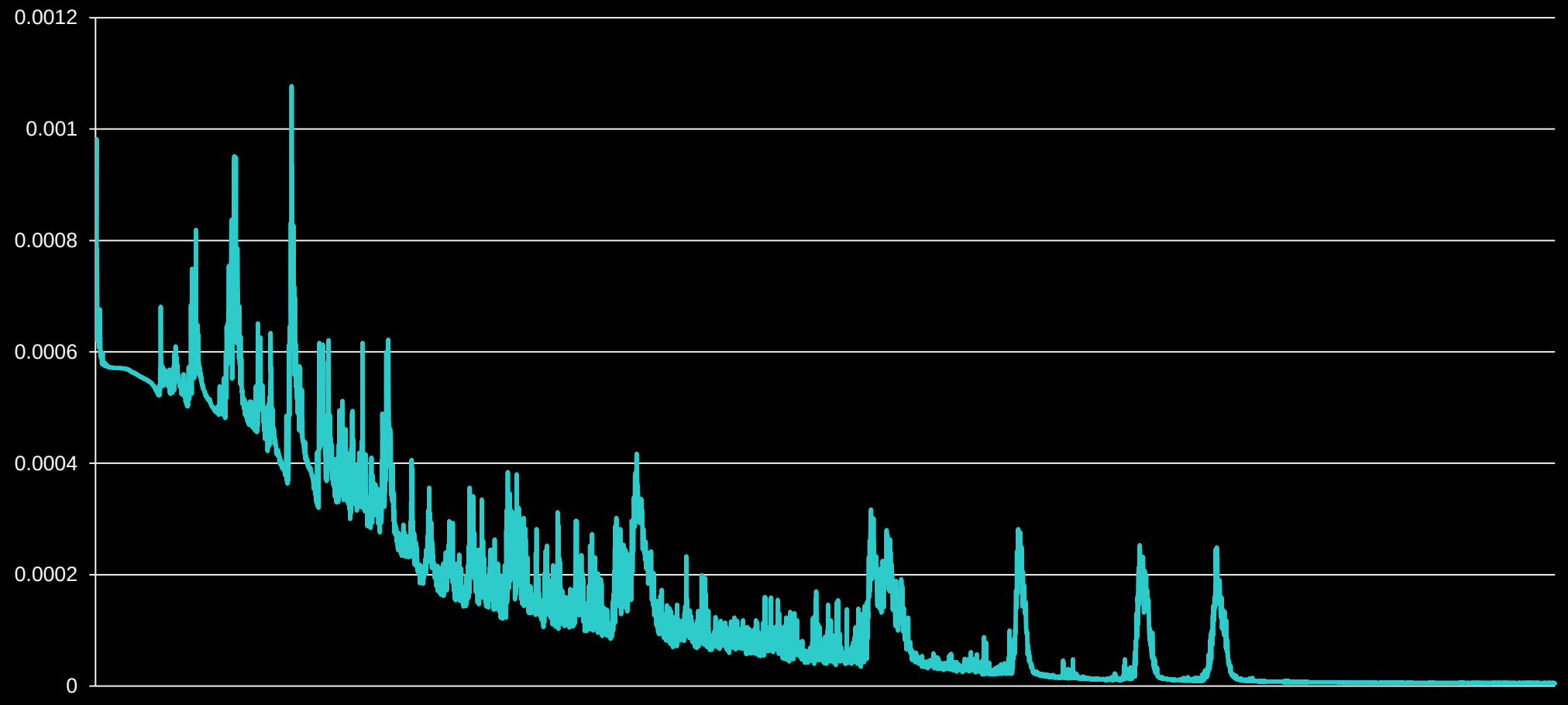
Experiment on action-language grounding – step 1



	Action 1	Action 2	Action 3
Object 1	trained		
Object 2		trained	
Object 3			trained

Results

20 trials conducted and each reached the threshold error of 0.000005



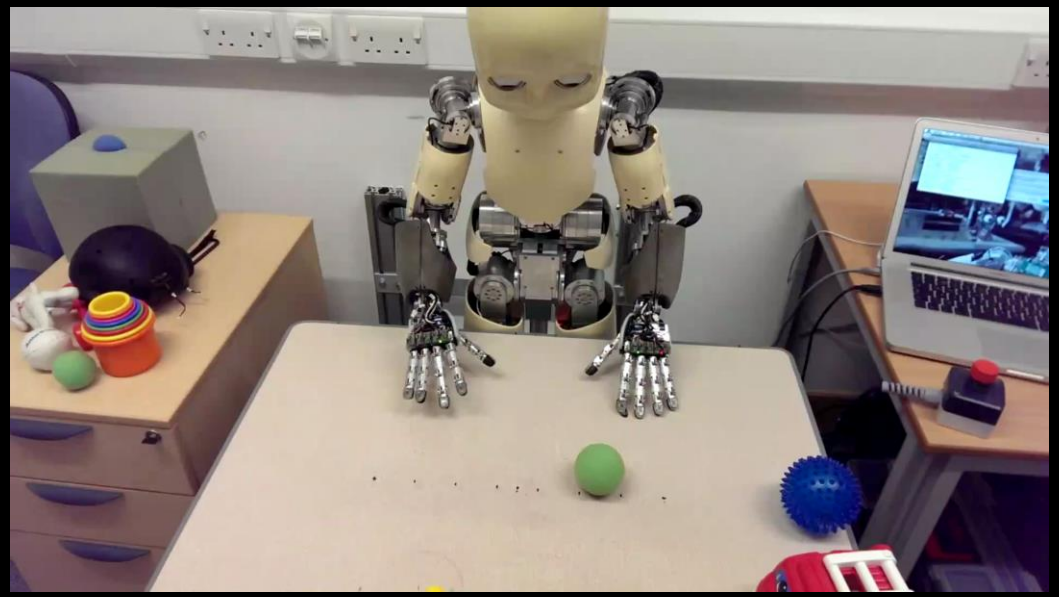
Multiple Time-scales Recurrent Neural Network

Scaling up the experiment on action-language grounding

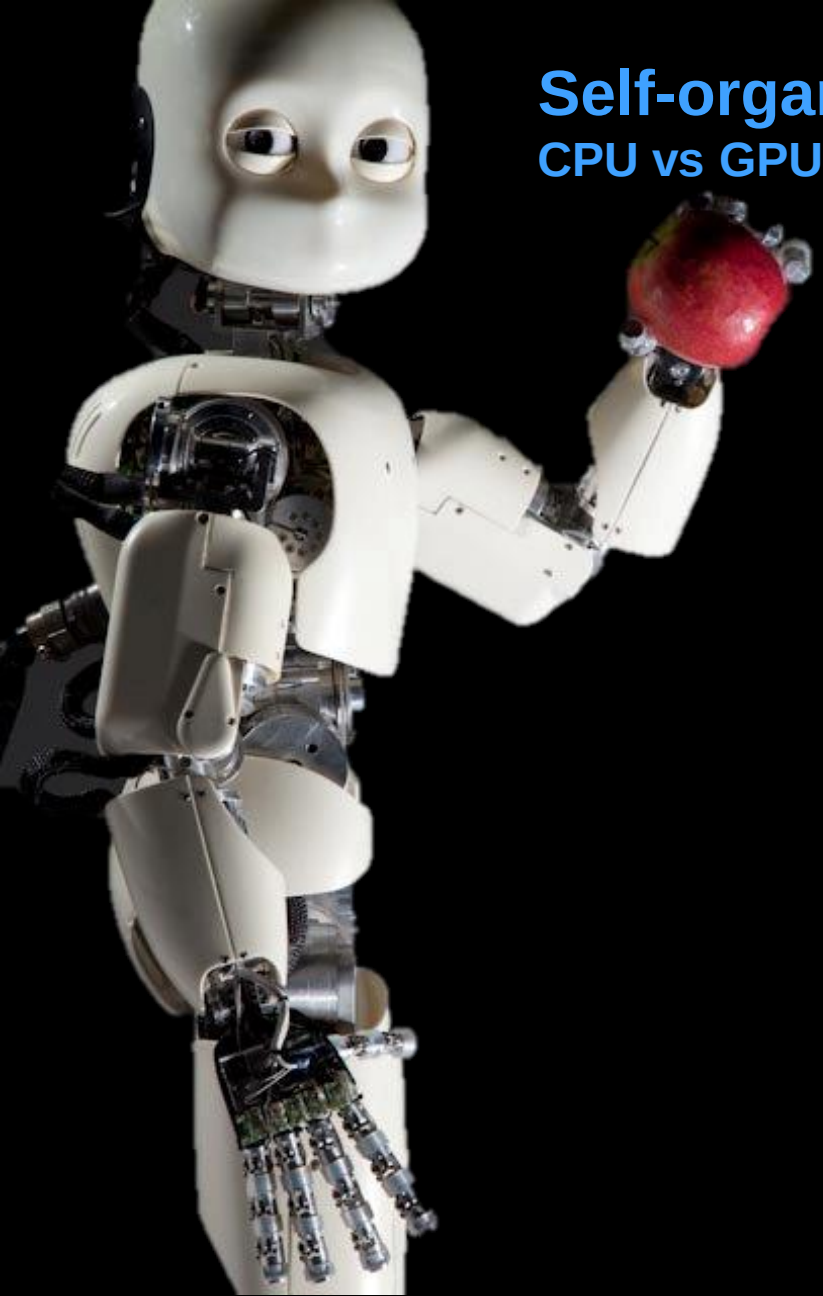
	Action 1	Action 2	Action 3	Action 4	Action 5	Action 6	Action N
Object 1	trained	trained	trained	trained	trained	trained	trained
Object 2	trained	trained	trained	trained	trained	trained	trained
Object 3	trained	trained	trained	trained	trained	trained	trained
Object 4	trained	trained	trained	trained	trained	trained	trained
Object 5	trained	trained	trained	trained	untrained	trained	trained
Object 6	trained	trained	trained	trained	trained	untrained	trained
Object N	trained	trained	trained	trained	trained	trained	untrained

Multiple Time-scales Recurrent Neural Network

Generalisation testing

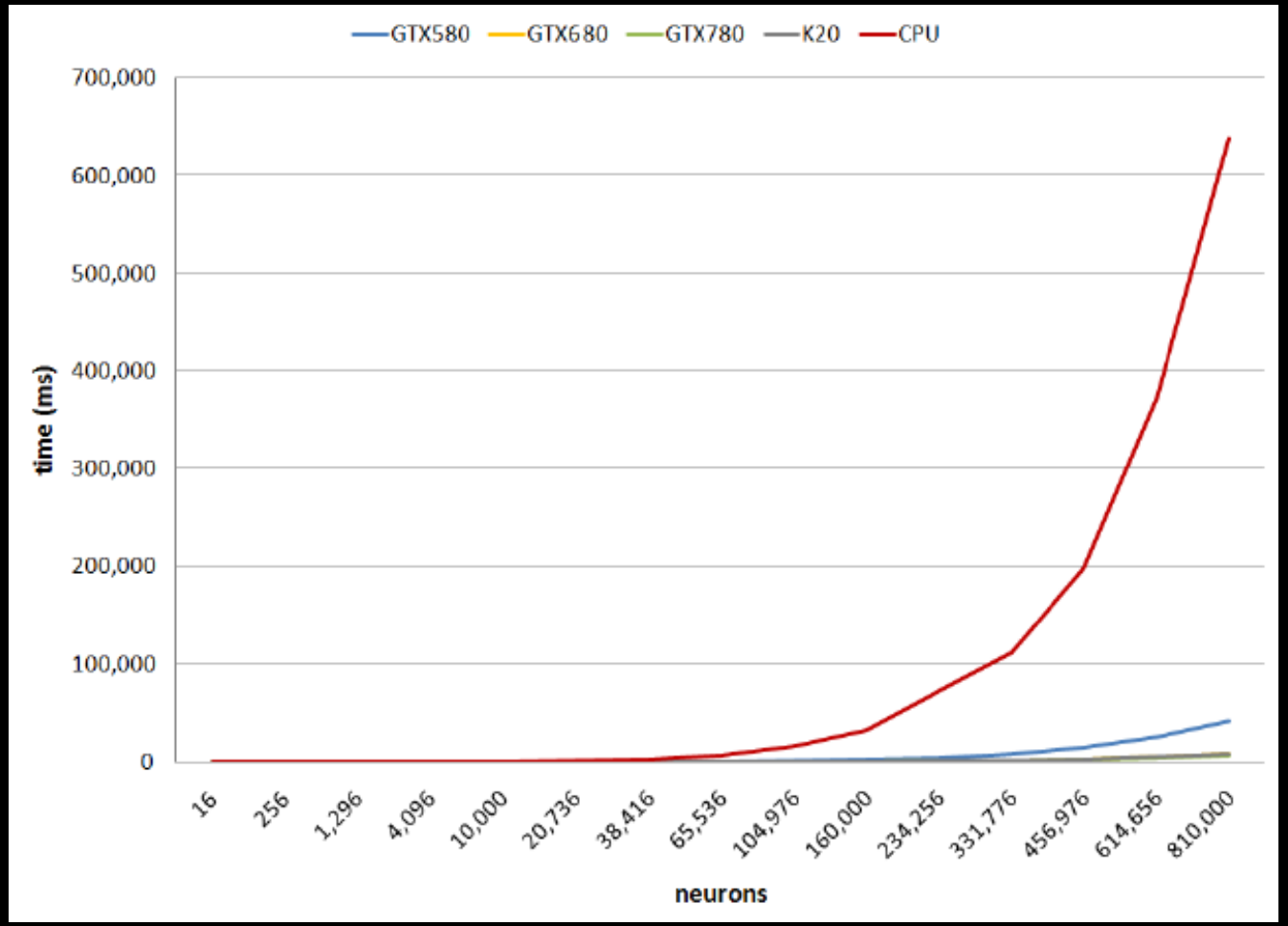


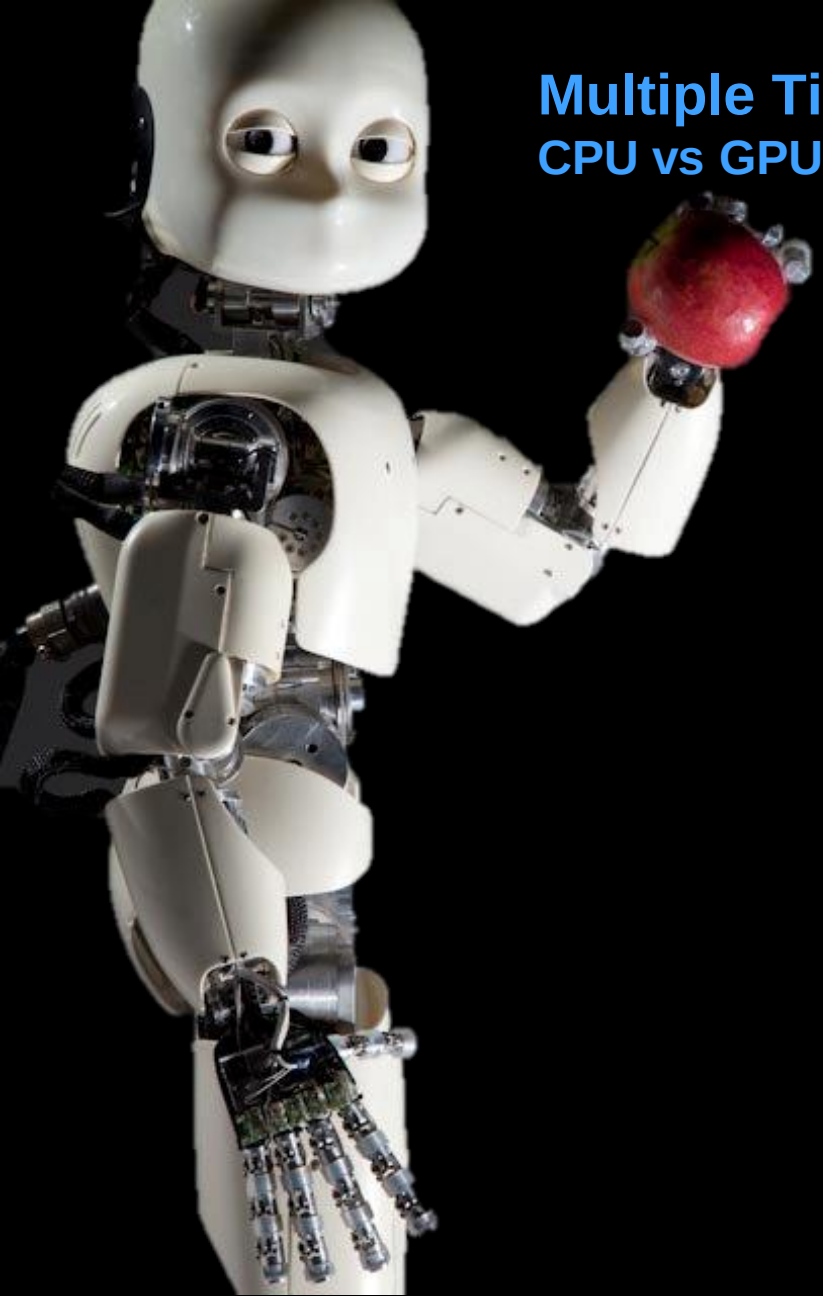
- **Experimental Setup**
  - **For each of the 9 objects, SOM and MTRNN was trained on 9 sequences each repeated 6x with different positions. Total of 478 sequences each with 100 41-wide vectors.**
  - **Left and Right hand used individually**
  - **MTRNN input: head, torso and arms (41 DOF)**
  - **Update rate: 50ms**



Self-organising maps

CPU vs GPU Performance





Multiple Time-scales Recurrent Neural Network

CPU vs GPU Performance

Neurons	2xGTX470	2xGTX580	2xC2075	2xK40
72	4.2x	5.6x	4.0x	1.7x
136	5.9x	7.8x	5.7x	4.0x
264	4.0x	5.8x	3.9x	4.9x
520	2.5x	3.5x	2.5x	4.8x
1,032	7.0x	9.0x	4.9x	3.6x
2,056	9.4x	14.7x	10.1x	7.8x
4,104	27.2x	51.4x	26.4x	32.6x
8,200	21.6x	33.3x	21.0x	37.9x

Biologically-inspired Active Vision system

Traditional Computer Vision

“Teaching a computer to classify objects has proved much harder than was originally anticipated”
Thomas Serre - Center for Biological and Computational Learning at MIT

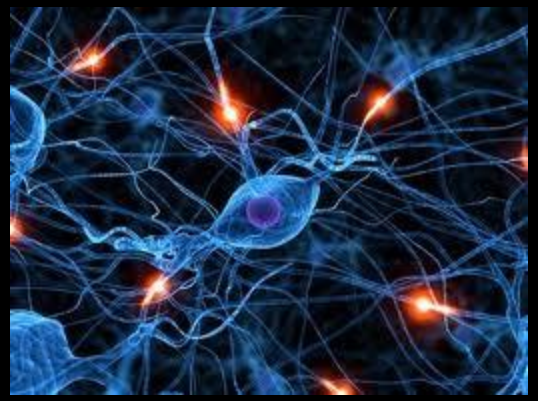
- **Specific template or computational representation is required to allow object recognition**

- **Must be flexible enough to account with all kinds of variations**

Biological Vision

"Researchers have been interested for years in trying to copy biological vision systems, simply because they are so good" ~ David Hogg - computer vision expert at Leeds University, UK

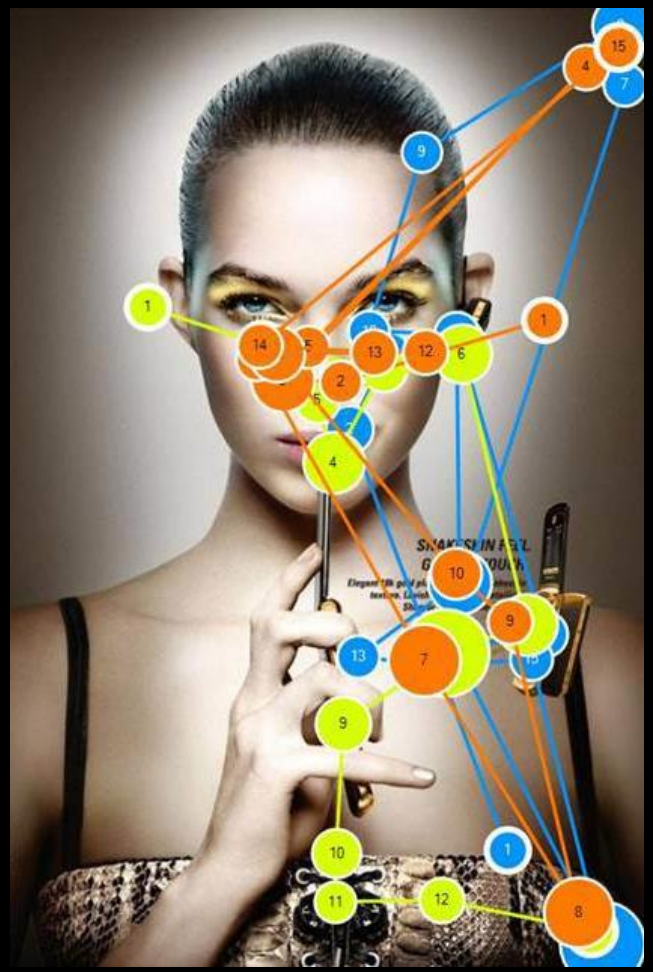
- **Highly optimized over millions of years of evolution, developing complex neural structures to represent and process stimuli**

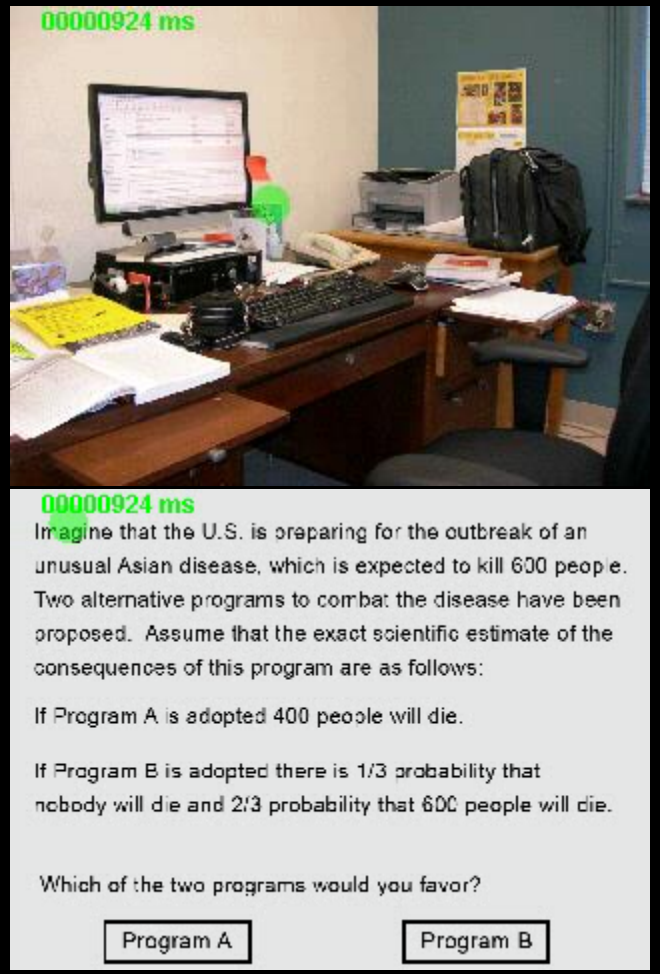


- **Superiority of biological vision systems is only partially understood**

- **Hardware architecture and the style of computation in nervous systems are fundamentally different**

Biological Vision





Seeing is a way of acting

Active Vision

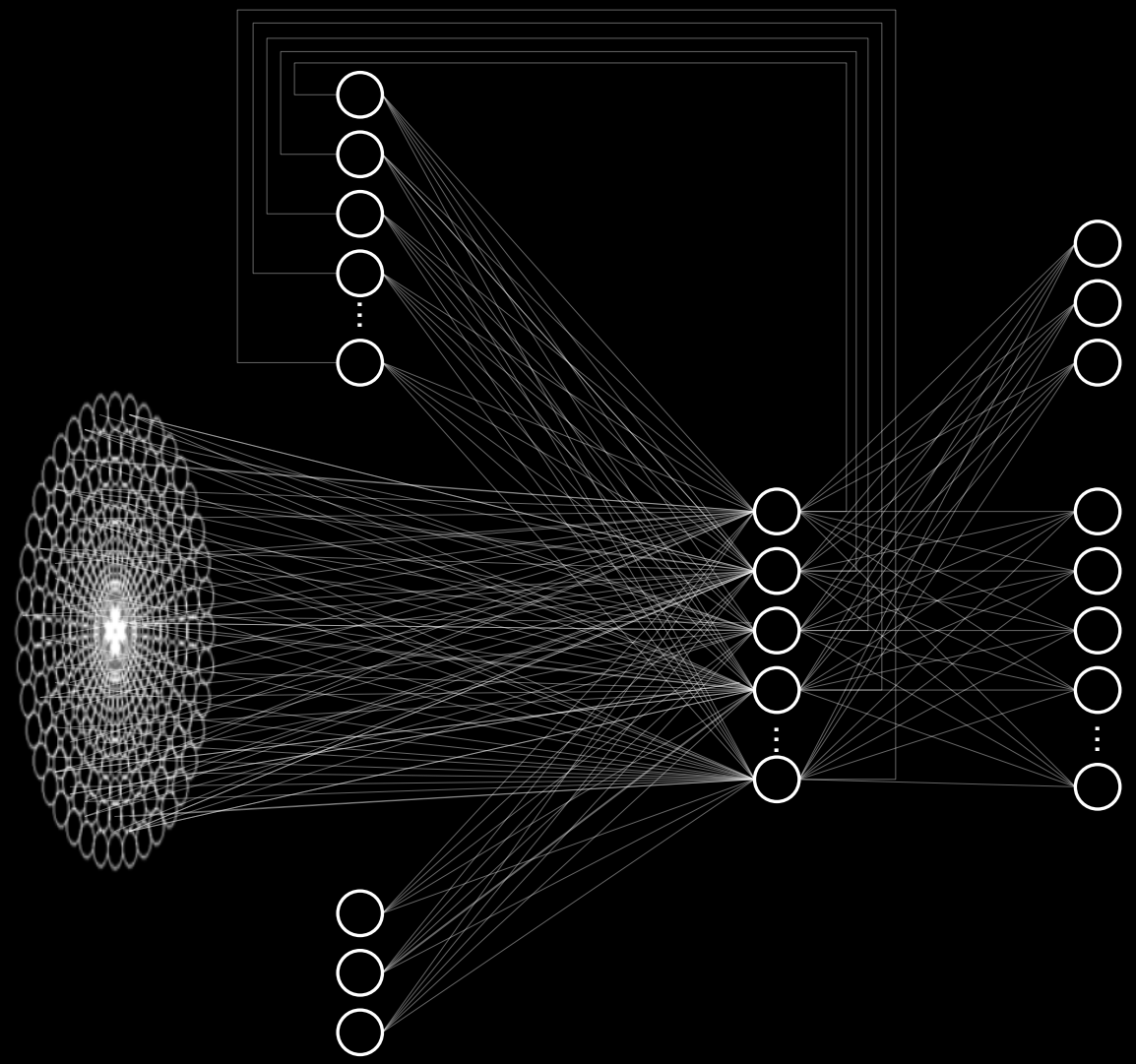
- Inspired by the vision systems of natural organisms that have been evolving for millions of years
- In contrast to standard computer vision systems, biological organisms actively interact with the world in order to make sense of it
- Humans and also other animals do not look at a scene in fixed steadiness. Instead, they actively explore interesting parts of the scene by rapid saccadic movements

Creating Active Vision Systems

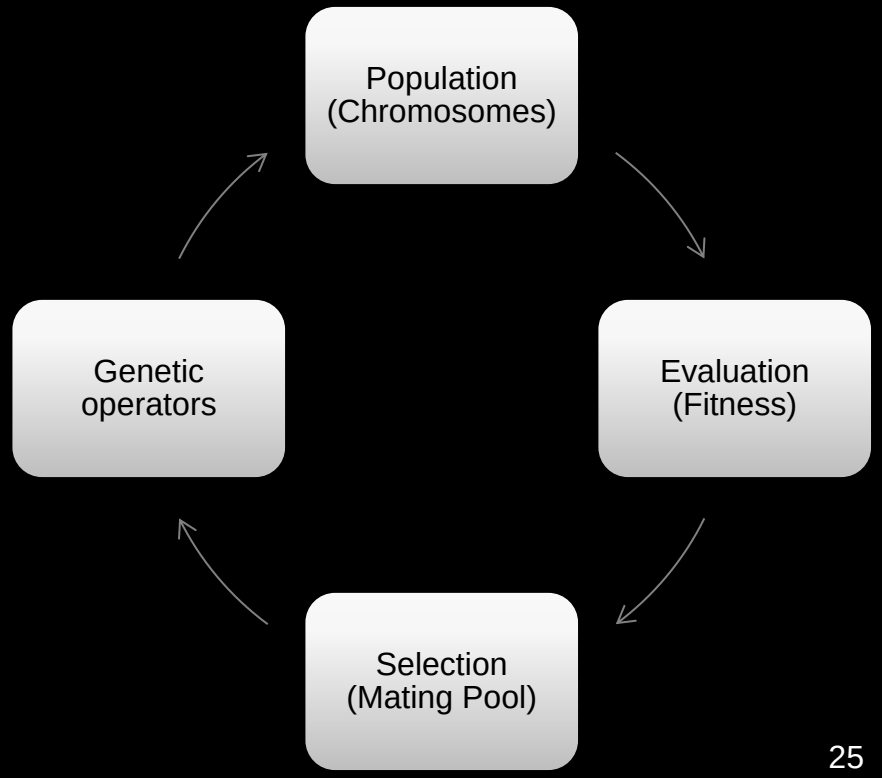
Evolutionary Robotics Approach

Evolutionary Robotics

- New technique for the automatic creation of autonomous robots
- Inspired by the Darwinian principle of selective reproduction of the fittest
- Views robots as autonomous artificial organisms that develop their own skills in close interaction with the environment and without human intervention
- Drawing heavily on biology and ethology, it uses the tools of neural networks, genetic algorithms, dynamic systems, and biomorphic engineering



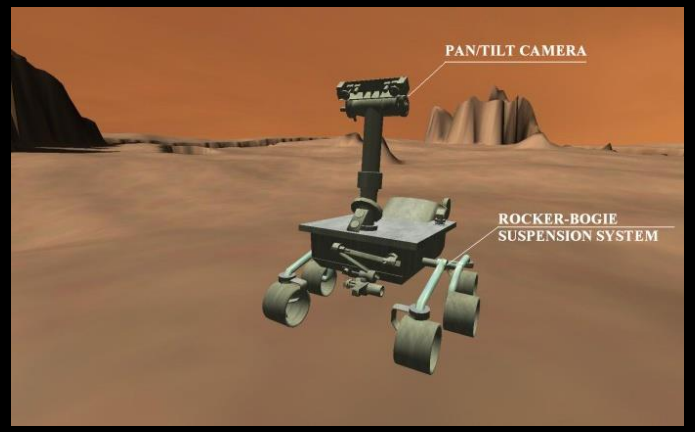
Genetic Algorithms (GAs) are adaptive heuristic search algorithm premised on the evolutionary ideas of natural selection and genetic. The basic concept of GAs is designed to simulate processes in natural system necessary for evolution.

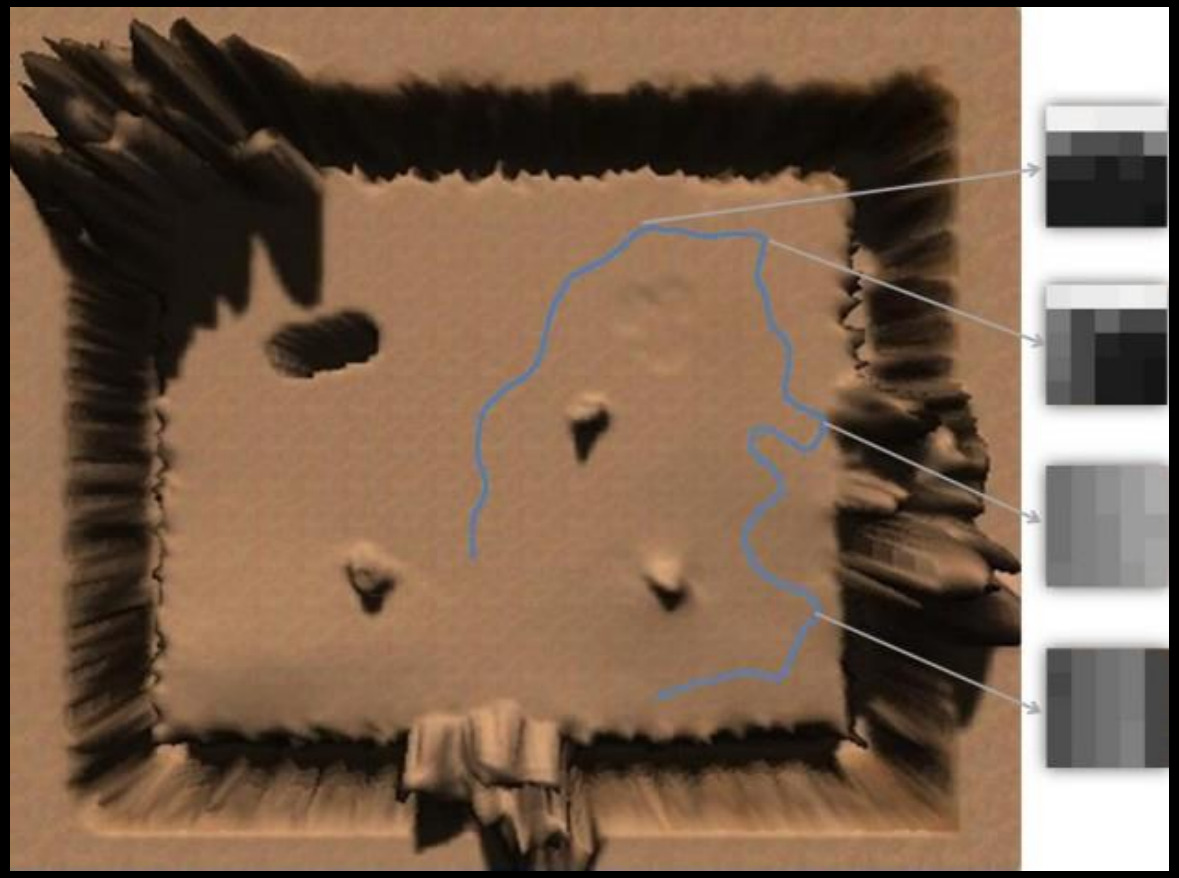


Artificial neural networks (ANNs) are very powerful brain-inspired computational models, which have been used in many different areas such as engineering, medicine, finance, and many others.

Related Research

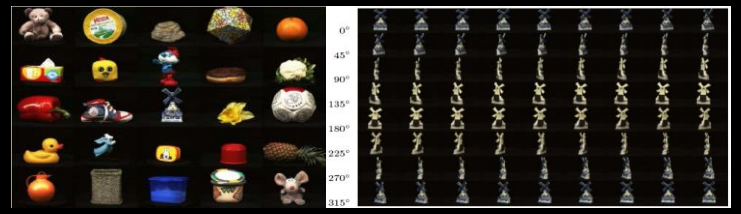
Mars Rover obstacle avoidance (Peniak et al.)





Method

- **Evolution of the active vision system for real-world object recognition**
  - training the system in a parallel manner on multiple objects viewed from many different angles and under different lighting conditions
- **Amsterdam Library of Object Images (ALOI)**
  - provides a color image collection of one-thousand small objects
  - recorded for scientific purposes
  - systematically varied viewing angle, illumination angle, and illumination color



- **Active Vision Training**
  - trained on a set of objects from the ALOI library
  - each genotype is evaluated during multiple trials with different randomly rotated objects and under varying lighting conditions
  - evolutionary pressure provided by a fitness function that evaluates overall success or failure of the object classification
  - trained on increasingly larger number of objects
- **Active Vision Testing**
  - robustness and resiliency of recognition of the dataset
  - generalization to previously unseen instances of the learned objects

Experimental Setup

- Recurrent Neural Network
  - Inputs: 8x8 neurons for retina, 2 neurons for proprioception (x,y pos)
  - No hidden neurons
  - Outputs: 5 object recognition neurons, 2 neurons to move retina (16px max)
- Genetic Algorithm
  - Generations: 10000
  - Number of individuals: 100
  - Number of trials: 36+16 (object rotations + varying lighting conditions)
  - Mutation probability: 10%
  - Reproduction: best 20% of individuals create new population
  - Elitism used (best individual is preserved)

Experimental Setup

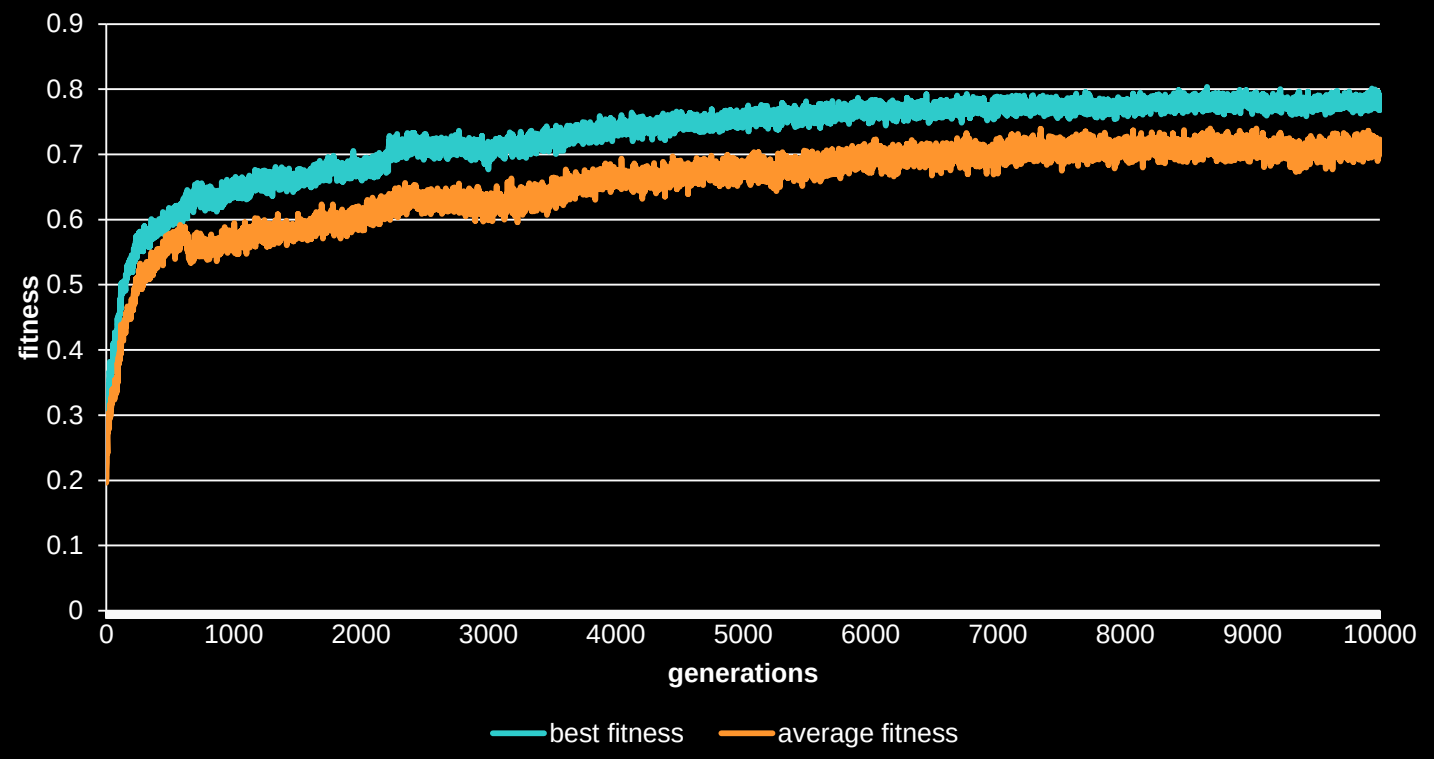
- Each individual (neural network) could freely move the retina and read the input from the source image (128x128) for 20 steps
- At each step, neural network controlled the behavior of the system (retina position) and provide recognition output
- The recognition output neuron with the highest activation was considered the network's guess about what the object was
  - Fitness function = number of correct answers / number of total steps

GPU Accelerating GA and ANN

- GPUs were used to accelerate:
  - Evolutionary process – parallel execution of trials
  - Neural Network – parallel calculation of neural activities

Results

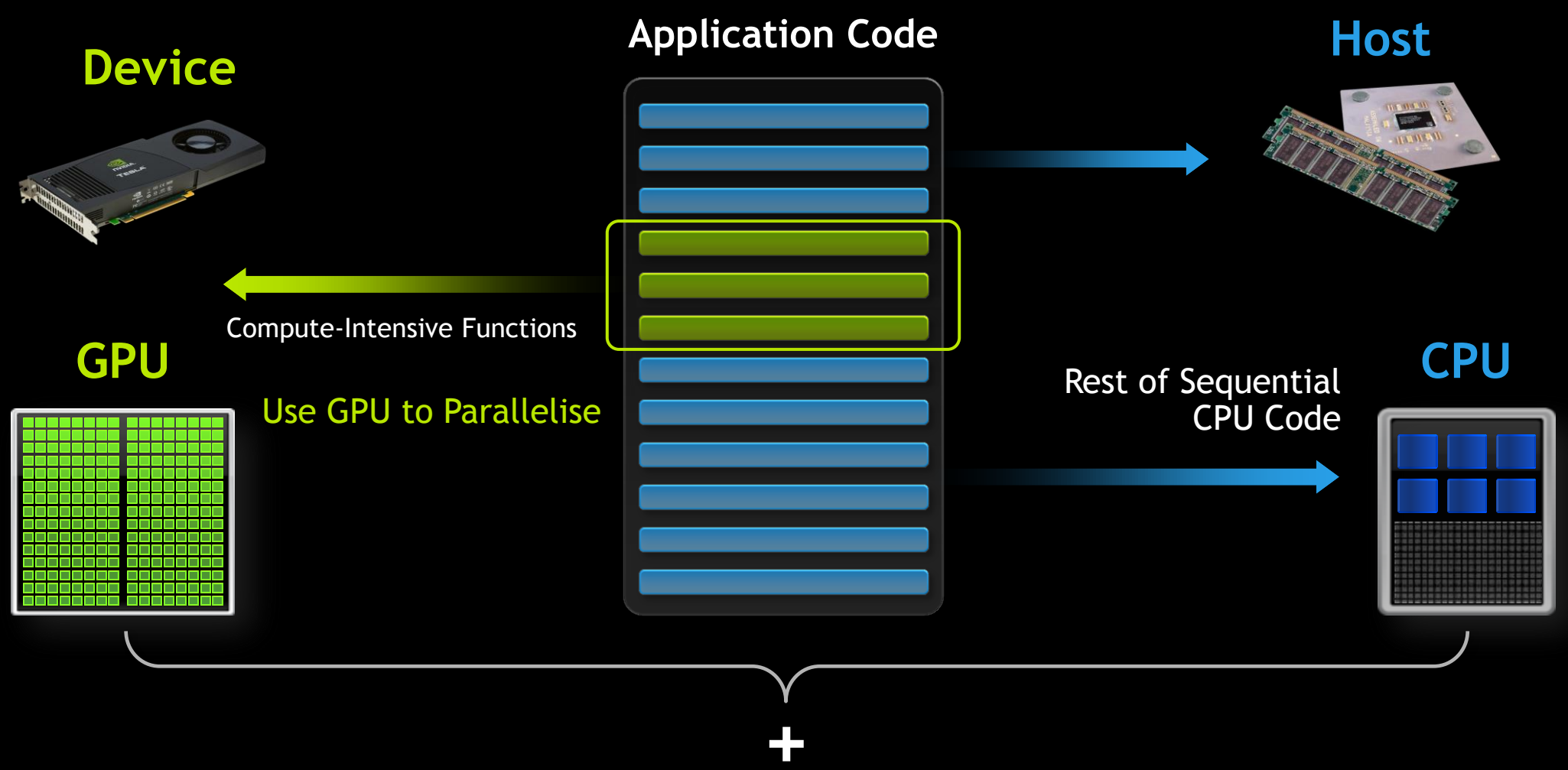
- Fitness can not reach 1.0 since it takes few time-steps to recognize an object
- All objects are correctly classified at the end of the each test



Evolved Behavior

Software development

Heterogeneous computing



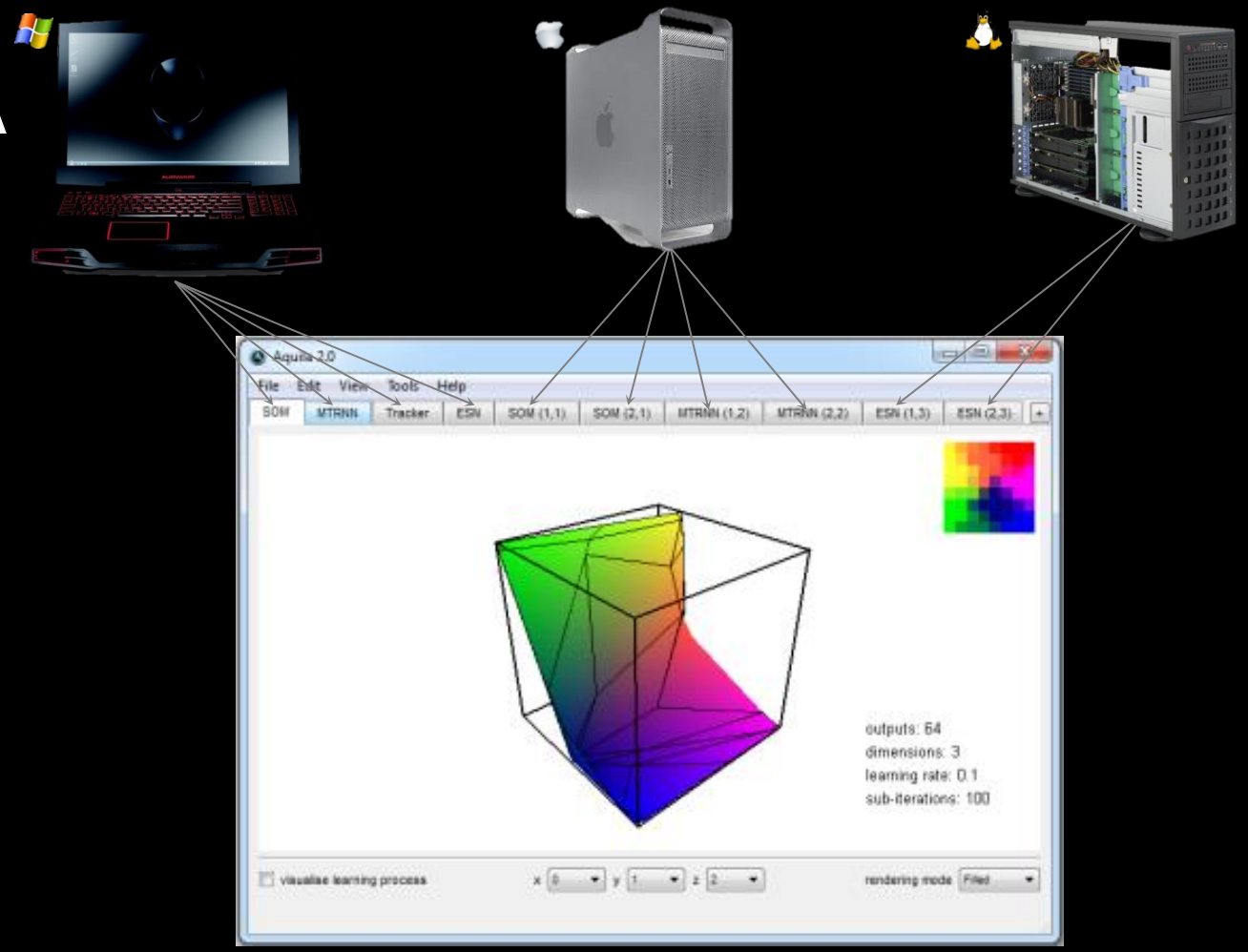
What is Aquila?

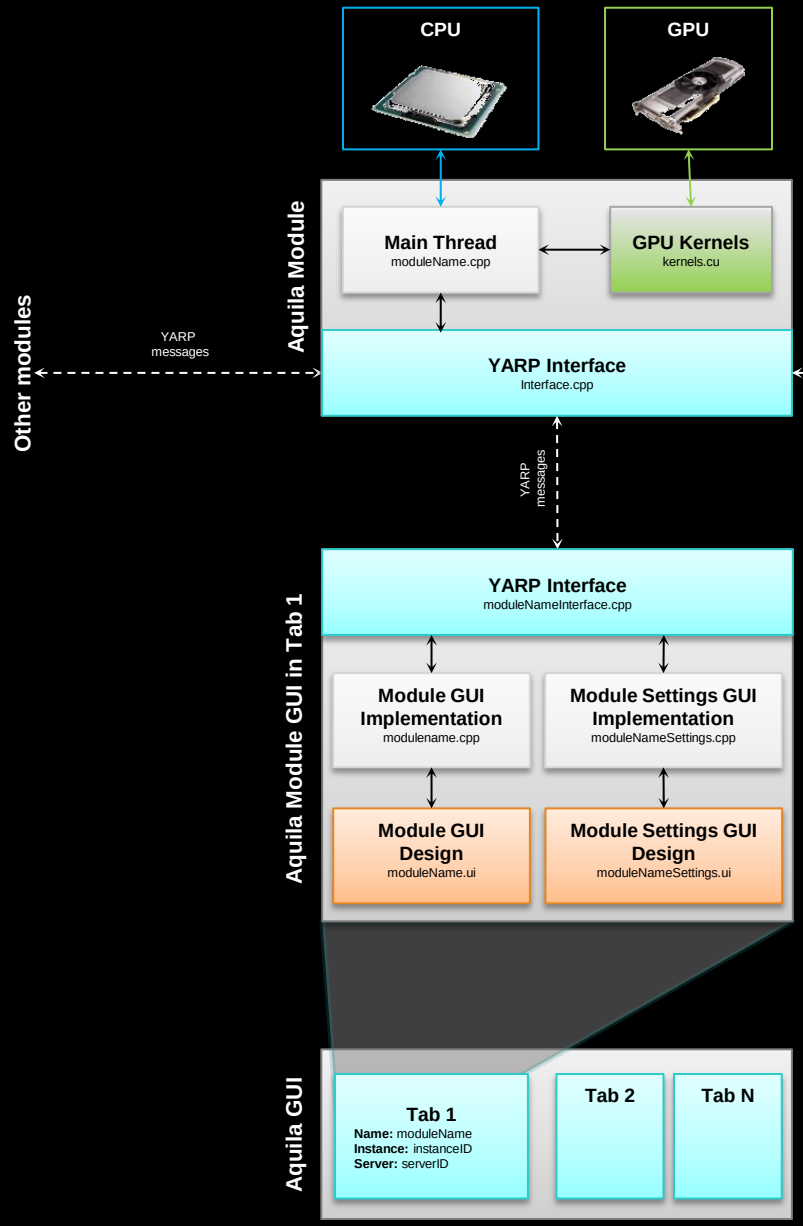
DISCOVER WITH PLYMOUTH UNIVERSITY

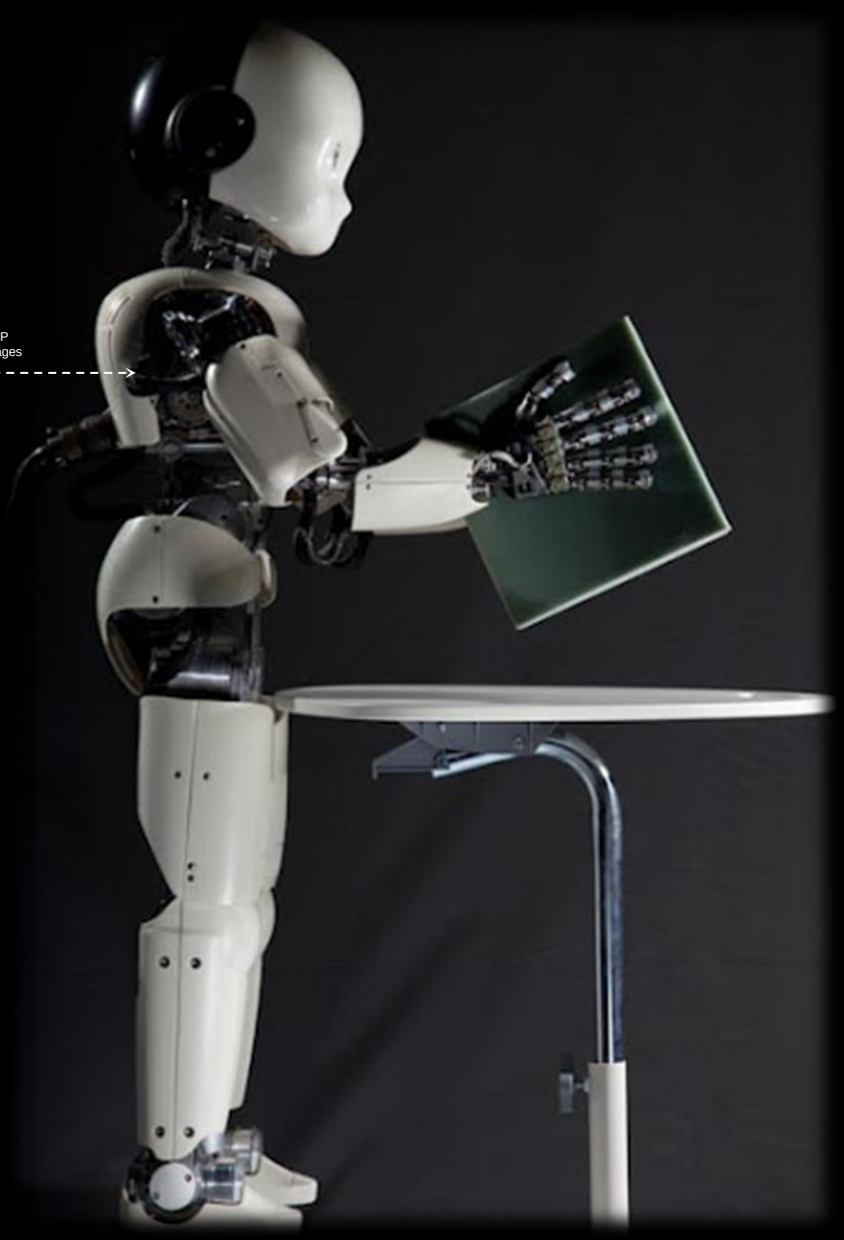
- Heterogeneous software architecture for the development of modules loosely coupled to their graphical user interfaces
- Provides simple and user friendly GUI client
  - Distribute, control and visualise existing modules
  - Generate new modules
  - Monitor connected server
  - Tools
- Modules
  - Run heterogeneous CPU-GPU code doing the actual work

What is Aquila?

- Developed in C++ and CUDA
- Cross-platform
  - Linux
  - OSX
  - Windows
- Dependencies
  - Qt
  - YARP
  - CUDA







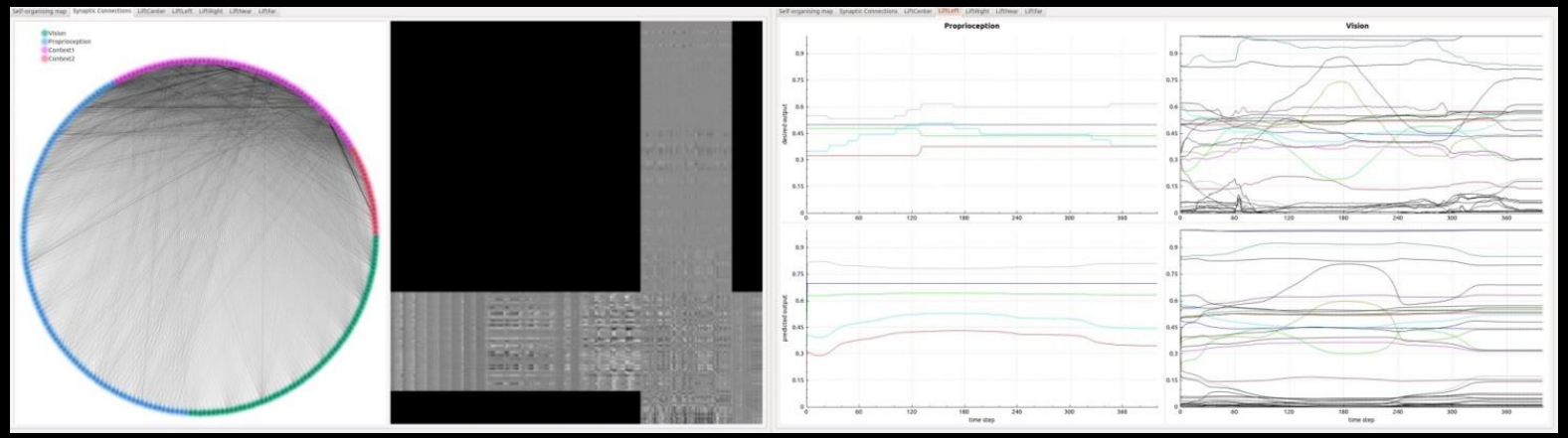
DISCOVER WITH PLYMOUTH UNIVERSITY

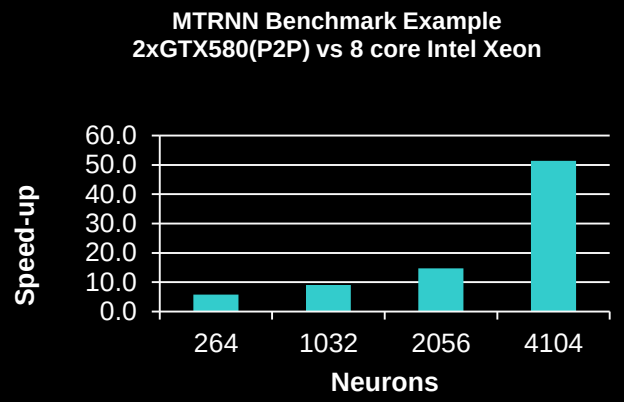
NVIDIA CUDA TEACHING CENTER

Existing Aquila Ecosystem

MTRNN

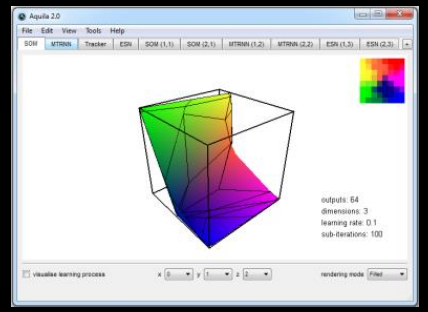
Multiple Time-scales Recurrent Neural Network





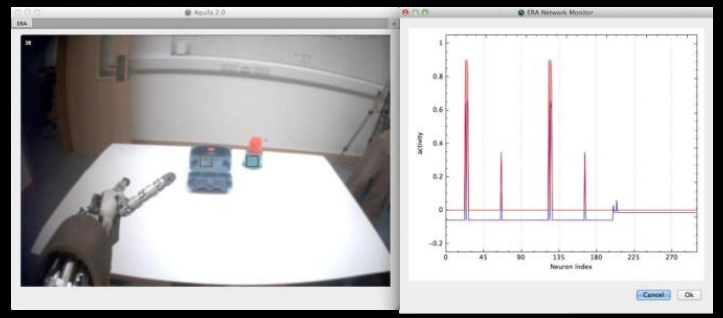
SOM

Self-organising Map



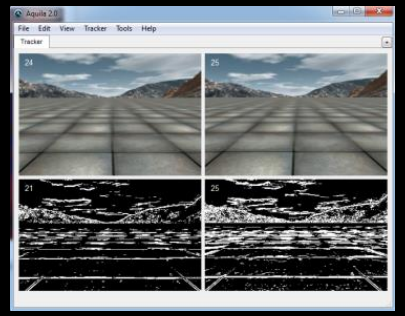
ERA

Epigenetic Robotics Architecture



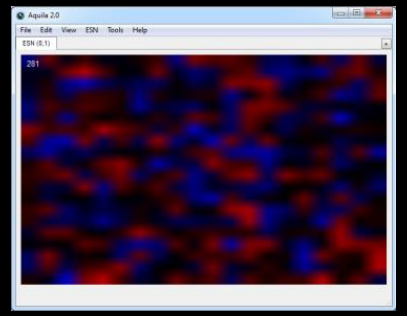
Tracker

Object tracking

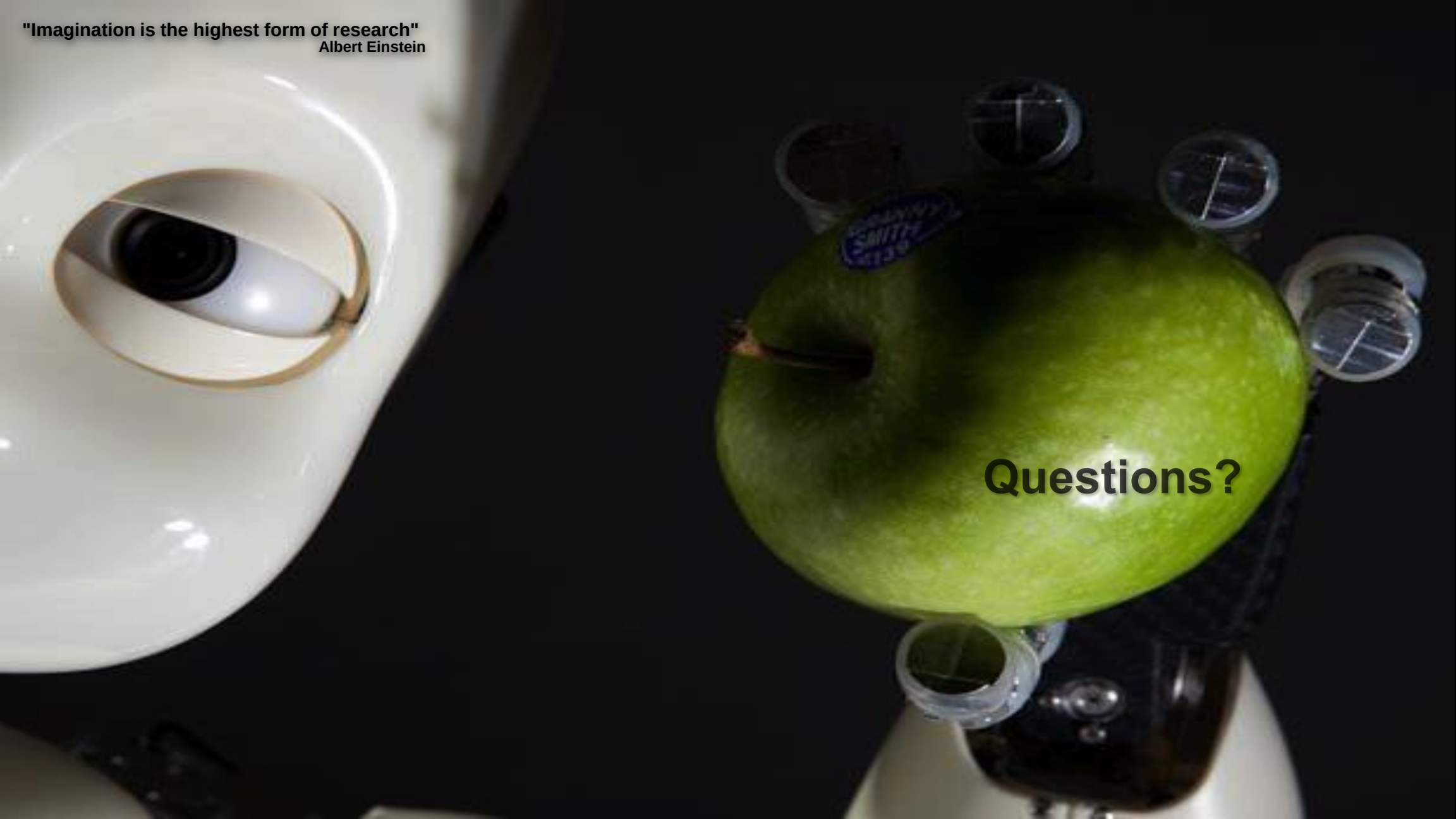


ESN

Echo State Networks



"Imagination is the highest form of research"
Albert Einstein



Questions?