SHOWCASE 5
Intermediate prototype, performing human-like object manipulation by robotic hand
(D3.5 – SGA3)

Simulation of robotic hand performing object manipulation.

Figure 1: A Simulated Anthropomorphic Robotic Hand Performing In-Hand Object Manipulation

The prototype solves the in-hand object manipulation task by learning to encode humanoid behaviour in a biologically constrained neural network model of the frontoparietal network. In its design, the network is constrained by structural connectivity data. In its input, the network is constrained by humanoid perceptual modalities. In its output, the network is constrained by controlling an anthropomorphic robotic hand.

Project Number:	945539	Project Title:	HBP SGA3

Document Title:	Showcase 5 – Intermediate prototype, performing human-like object manipulation by robotic hand
Document Filename:	D3.5 (D24) SGA3 M21 ACCEPTED 220520.docx
Deliverable Number:	SGA3 D3.5 (D20)
Deliverable Type:	Demonstrator
Dissemination Level:	PU = Public
Planned Delivery Date:	SGA3 M21 / 31 Dec 2021
Actual Delivery Date:	SGA3 M21 / 23 Dec 2021; accepted 20 May 2022
Author(s):	Mario SENDEN, UM (P117) Tonio WEIDLER, UM (P117)
Compiled by:	Mario SENDEN, UM (P117) Tonio WEIDLER, UM (P117)
Contributor(s):	Mario SENDEN, UM (P117) Tonio WEIDLER, UM (P117)
WP QC Review:	Yannick MOREL, UM (P117)
WP Leader / Deputy Leader Sign Off:	Rainer GOEBEL, UM (P117)
T7.4 QC Review:	N/A
Description in GA:	Simulation of robotic hand performing human-like in-hand object manipulation.
Abstract:	The demonstrator discussed is concerned with dexterity; coordination of complex hand movements by anthropomorphic robotic hand. To that end, a biologically constrained recurrent convolutional neural network (RCNN) is specified that provides a stochastic movement policy for each joint of the hand. Biological constraints are considered at a macroscopic scale. At the macroscopic scale, we identify the network of cortical regions involved in the coordination of complex hand movements, their functional specialisation and anatomical connections. We use this information to constrain the number of layers, the computational graph they form and their input modalities. The RCNN is trained through reinforcement learning on high-performance computing clusters.
Keywords:	Frontoparietal network, sensorimotor control, visuomotor control, deep learning, dexterity, manipulation
Target Users/Readers:	General scientific audience with some background in reinforcement learning.

Table of Contents

Table of Figures

1. Introduction

The work conducted within Work Package 3 (WP3) is chiefly concerned with development of architectures that implement either low- or high-level cognitive processes. Cognition includes several different abilities such as perception, attention, planning, problem solving, etc. Low-level cognitive processes are typically considered to occur in an automated manner with low levels of awareness, while high-level cognitive processes exercise voluntary mental activities with controlled awareness and are considered to involve abstract – symbolic - reasoning.

The demonstrator (Showcase 5 of WP3) discussed below addresses low-level cognition. Specifically, it is concerned with human dexterity; the control of complex hand movements such as in-hand object manipulation. This is cognitive in the sense that it involves coordination and planning of appropriate sequences of actions (joint movements) but at a low-level, as it does not require abstract reasoning and conscious planning of individual actions (only their goal). To achieve dexterity with an anthropomorphic robotic hand, a biologically constrained recurrent convolutional neural network (RCNN) is specified and trained with reinforcement learning to provide a stochastic movement policy for each joint of the hand. Biological constraints are considered at a macroscopic scale. Specifically, we identify the network of cortical regions involved in the coordination of complex hand movements, their functional specialisation and anatomical connections. We use this information to constrain the architecture of the RCNN in terms of the number of layers (reflecting the frontoparietal network and the visual system), the computational graph they form and their input modalities (including vision, touch and proprioception). Training of the RCNN occurs *in silico* on high-performance computing clusters (using SC6 HPC facilities). The demonstrator provides an end-to-end (deep reinforcement learning) approach towards developing a large-scale, embodied architecture subserving low-level cognitive functions. It thus complements other activities within WP3, which focus on hand-crafting architectures as well as individual modules subserving low- and high-level cognitive processes. While this demonstrator is in principle self-contained, we are exploring the manner in which it may be expanded upon (with additional layers and functionalities) to serve as a foundation upon which larger architectures requiring dexterity may build. Development of the demonstrator occurs in three stages. First, a proof-of-concept system was developed to establish the range of required infrastructure services and procedures (achieved by M9). Second, a first working prototype of in-hand object manipulation was trained *in silico*, which we present here (achieved by M21). The final stage involves training the RCNN using domain randomisation of parameters of the hand model and physics simulator. This ensures that the final demonstrator is generalisable towards other physics engines and potentially to physical systems as well and to subsequently analyse the solutions developed by the RCNN and compare these to solutions of (regions of) the biological FPN.

1.1 Scientific Motivation

This project combines deep learning, robotics and neuroscientific knowledge to understand how the brain coordinates visually guided object manipulation. Although tool use is common among animals, the significance that tools have acquired for humans is unique. Indeed, skilful object manipulation is an essential component of our everyday life, and our manual dexterity is unmatched by even our closest simian relatives. Given that in-hand object manipulation is a highly complex task that engages a large-scale network encompassing sensory, association and motor regions, it is unsurprising that it remains insufficiently understood how the brain achieves this. It has been established that the frontoparietal network performs sensorimotor transformations necessary for goal-oriented action execution. For instance, the intraparietal sulcus (IPS) contains “visual”, “visuomotor”, and “motor” neurons and sends input to the rostroventral premotor cortex (PMVr). The PMVr stores motor synergies, a low-dimensional vocabulary of motor actions, suggesting that the IPS-PMVr circuit transforms visual information into motor commands. Transformations follow a hierarchical progression, with brain regions storing different intermediate representations. Much can be learned with respect to representations from decoding studies. For instance, recent studies conducted in macaque monkeys using both intracortical electrophysiological recordings and fMRI have shown that object-specific hand configurations can be reliably decoded from monkey homologues of the IPS, PMVr and primary motor cortex. While decoding studies are important to identify representations, they cannot provide information on transformations. However, in order to truly understand visually guided actions, it is pivotal to identify, not only representations stored in individual brain regions, but also the computations occurring between brain regions. The central contribution of this Showcase is to shed new light on these transformations by transferring a uniquely human skill to an artificial agent. To that end, we specify an ecologically valid set of visually guided actions and use deep learning to train a recurrent convolutional neural network (RCNN) to perform these actions with an anthropomorphic robotic hand. The architecture of the RCNN is based on functional and anatomical knowledge of the frontoparietal network in the primate brain involved in visually guided hand and arm movements. Training data will be gathered from experience obtained by the agent itself within a simulated environment. The transformations and computations emerging in the trained architecture can subsequently be investigated using tools from neuroscience (e.g. representational similarity analysis), machine learning (e.g. feature visualisation) and data-driven science (e.g. sparse identification of nonlinear dynamics) to provide new insights and hypotheses as to how the brain coordinates complex hand movements.

1.2 Relation to and use of EBRAINS

The involved work greatly relies on the HBP’s EBRAINS research infrastructure. Training in-hand object manipulation as reported here requires approximately 3-5 years of simulated experience. Domain randomisation as tackled in the third stage will scale this up to 100-150 years. Generating this experience and optimising upon it is computationally expensive and requires extensive parallelisation, for which the use of high-performance computing (HPC platforms, SC6) clusters is essential. Furthermore, the design of the RCNN is guided the human brain atlas (SC2) to identify brain regions involved in complex hand movements, as well as their interconnectivity and local cell profile. This benefits from the development of the “Brainscapes” atlas client, which gives direct programmatic access to data organised within the EBRAINS human brain atlas. It supports the multilevel character of this atlas, which defines cytoarchitectonic maps in multiple reference template spaces at different spatial resolutions (namely the MNI Colin, ICBM152 asymmetric as well as the BigBrain microscopic space). It links them with complementary maps related to brain function, connectivity and fibre architecture.

Conversely, the demonstrator work informs the design of the supporting framework, in terms of specifications. We continuously communicate with the development team of the Neurorobotics Platform (NRP, SC4) to provide insights on the MuJoCo physics engine and its relevance for such training tasks, as well as essential requirements for transferring simulation of the RCNN onto the NRP, such as accelerated simulation time and the possibility to programmatically launch several instances of the NRP for deployment on high performance computing clusters. In fact, future versions of the NRP are now projected to include MuJoCo as an optional native simulator, which is further facilitated by the simulator’s recent move to a free-to-use model. Additionally, our recent efforts, allow the agent to control a hand simulated in the NRP’s native simulator Gazebo. Here, the major challenge is to train policies that generalise over differences in the physical parameters and governing equations of different simulation software, which we tackle in the Showcase’s final stage (M36). Aforementioned progress in the NRP paired with simulator-agnostic policies then constitute models that can subserve other research relying on arbitrary simulators, all within the NRP.

Together with the technical coordination team, the Showcase is currently being prepared as an EBRAINS service. Standardised as a modular workflow, other researchers will be able to leverage the software without understanding in detail the underlying mechanisms.

Crucially, this demonstrator motivates the emergence of infrastructure workflows that span a meaningful range of complementary services, from data services (SC2) to embodiment and function-related services (SC4), with direct support from dedicated compute resources (SC6). No such combination of interoperable services is to be found outside of EBRAINS. Demonstrating the manner in which EBRAINS thus empowers us to break new ground in cognitive computational neuroscience - and its practical application - contributes to establishing the infrastructure’s Unique Value Proposition.

2. Technical Specification

The behavioural policies adopted by our model of the frontoparietal network (depicted in Figure 2) are trained in a reinforcement learning setup using proximal policy optimisation (PPO). PPO alternates between two phases: the gathering of experience based on its current policy and the updating of the policy based on the previously collected data. Experience is represented by the state of the environment, the action the agent decided to execute in this state and the reward it received. Based on an estimate of the average value of every state and the actual reward received, advantageous actions are fostered, whereas disadvantageous decisions are suppressed. Our implementation of this setup efficiently trains arbitrary neural network models with potentially multimodal input (vision, proprioception and touch). Since M9, this framework has been further improved to meet the challenge of in-hand object manipulation. On the technical side, we most notably developed a new parallelisation system for both the gathering of experience and the optimisation (Figure 3). This new, native MPI implementation integrates smoothly with the remainder of the software without requiring additional third-party packages. Additionally, it is tailored towards the setup of EBRAINS' HPC cluster CSCS, on which we train the policies. Together, this enables substantially faster training on an arbitrary number of nodes, a crucial requirement for the complexity of the tasks tackled in Showcase 5. To integrate the visual modality, the system now communicates all modalities of states in separate streams, allowing any possible architecture to process them flexibly. Furthermore, policy and value functions are now trained asynchronously. Whereas the policy only receives humanoid sensory inputs, the value function can access additional high-level information such as the object's rotational and positional velocity. Thereby, the value function can make more precise predictions, which is crucial for effective learning, but the model retains its plausibility, since the dedicated parts of the network are omitted after training. Lastly, we shifted from a continuous, beta-distributed to a discrete, multi-categorical action space. While this substantially increases the dimensionality of the output space the model projects to, the predicted distribution is less complex. One should note that even using continuous actions, the agent learns reasonable policies. However, the performance in a multi-categorical distribution cannot be matched in reasonable training times. Regardless, since we directly predict joint angles and thereby already omit a biologically plausible translation from motor commands to actions, this does not harm the validity of our approach.

Due to these improvements to our system, we are now able to solve the manipulation task. The learned behaviour is human-like with few exceptions. The agent effortlessly utilises the thumb to perform most rotations along the x and y-axis (that is, those rotations leading to a new side of the cube lying on the palm). Rotations around the z-axis are guided by the fingertips and knuckles of the remaining digits. Occasionally, these occur by sliding the cube towards the wrist and back. Here, the robot exploits the lack of skin-evoked friction on the robotic hand, allowing for the cube to gently slide over the palm. In contrast, humans would likely leverage the friction of their skin alongside the ability to bend their palms. Despite this limitation, it becomes clear that under anthropomorphic constraints, the agent achieves human-like performance and where the constraints are precise, finds humanoid solutions. Likely, we can even further fine-tune these solutions towards their human example by implementing the missing constraints/flexibility into the robot. For instance, we observe that integrating a squared-force minimisation term into the reward function results in less chaotic, more human-like behaviour.

Moreover, the model's architecture has been refined to more closely match current knowledge about the frontoparietal network. A depiction of this network architecture is given in Figure 2. While subcortical regions (e.g., basal ganglia) and cerebellum participate in motor control, we deliberately narrow down the focus of the current modelling on cortical regions. These comprise visual and somatosensory cortex as input regions, prefrontal cortex (PFC) as an input and processing region for goal descriptions, posterior parietal cortex (PPC) as an integrator of perceptual representations from VC and SSC, and motor cortex (MC) which determines the output of the network, a motor command. Note that the motor commands issued by our model are relative joint angles. Thus, we do not explicitly model individual muscles in the hand, nor the efferent projections through the spinal cord. Connectivity between regions and subregions is based on the brain atlases available through EBRAINS, as well as a comprehensive review of the neuroimaging literature.

Crucially, we can train our policies on this new model. In fact, our preliminary experiments show that this architecture learns impressive policies substantially faster than networks proposed in state-of-the-art AI research on the same task.

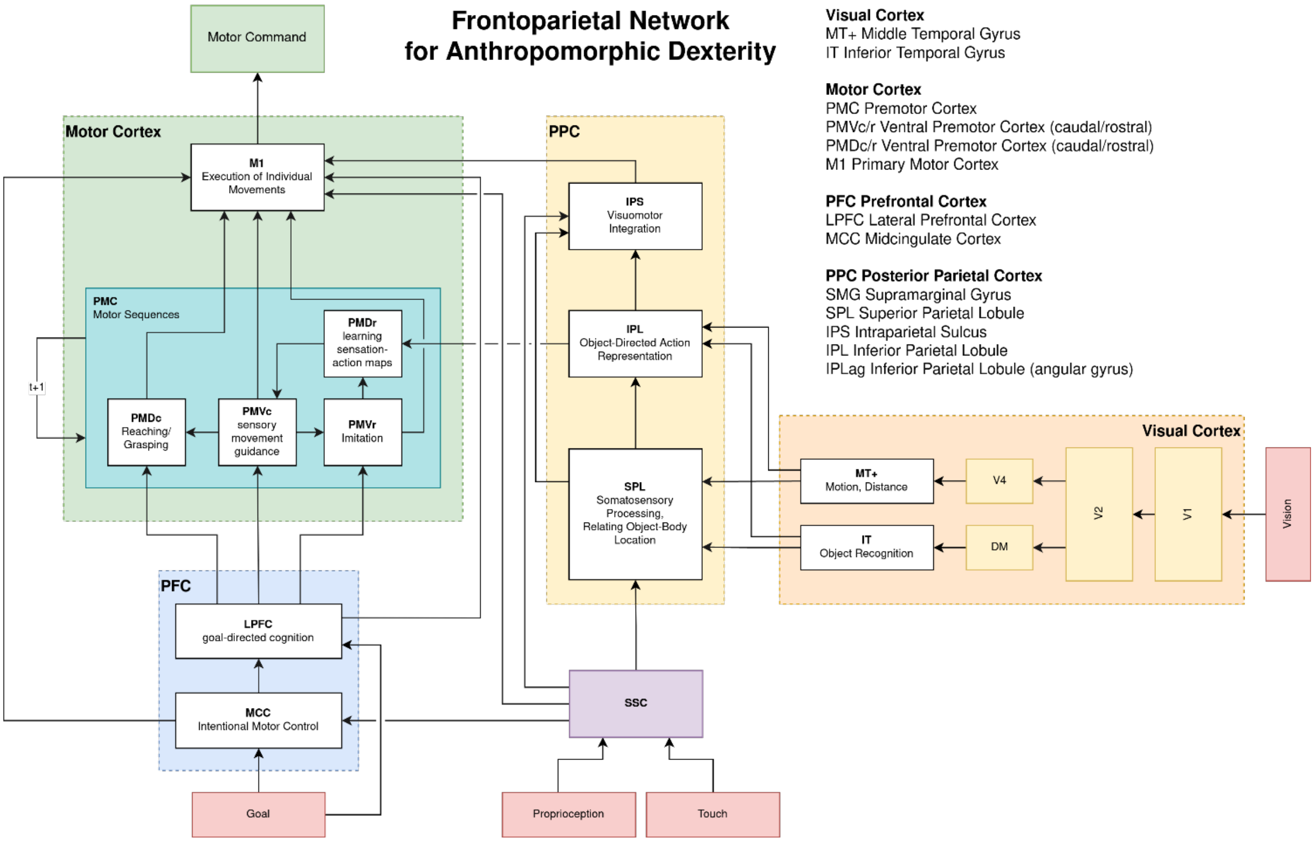

Figure 2: Model of the Frontoparietal Network Aiming for Macro-Scale Plausibility

Our model of the frontoparietal network aims for macroscale plausibility. Participating regions and their intended functions are derived from the neuroimaging literature whereas their connectivity is based both on literature and the EBRAINS Human Brain Atlas. The model receives multimodal input: vision from a camera placed above the hand, proprioception from the angles and velocities of the hand's 24 joints as well as touch sensation from 92 sensors distributed all over the hand's surface. The output is a 20-dimensional vector representing relative joint angle changes.

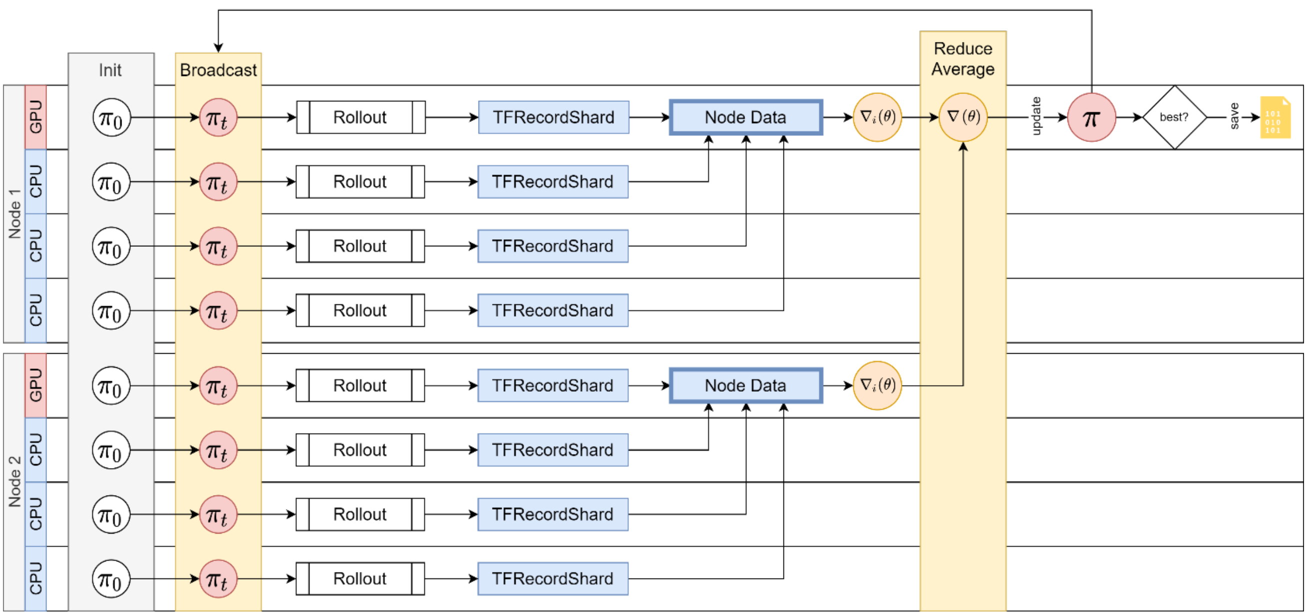

Figure 3: Integrated Parallel Setup of Both Gathering and Optimisation During Training

During training, both the gathering and the optimisation phases are parallelised in a native MPI implementation. This setup allows for fast training on an arbitrary number of machines, as hundreds of workers collect experience in parallel. Every node, comprising multiple workers, combines the data produced by its own workers on the GPU process. The GPU calculates a partial gradient, followed by the root process (usually the GPU process on the first node) averaging all these parts into one gradient constituting the update to the policy. The new policy is then broadcasted to all workers, where new experience is gathered.

3. How to access the Showcase

The demonstrator is implemented in TensorFlow 2.4, supported by Python 3.8 and can be accessed from GitHub: https://github.com/ccnmaastricht/dexterous-robot-hand. Simulations are run in MuJoCo, which is now freely available at https://mujoco.org/.

4. Looking Forward

So far, a flexible framework for training biologically inspired models for visuomotor control has been built and its ability to learn the complex task of in-hand object manipulation was showcased. From here on, the focus shifts towards building new models, training generalisable policies, and analysing the neurocomputations that emerge under humanoid constraints.

Analysing Neurocomputations Based on the prototype at hand, the Showcase enters the next phase in which we will analyse the neurocomputations emerging in the proposed model and its future iterations. The analysis will encompass a variety of methods from deep learning (e.g., filter visualisations and saliency maps), neuroscience (e.g., representational similarity analysis) and dynamical systems theory / data-driven science (e.g., stability analysis and sparse identification of nonlinear dynamics; SINDy).

Modular and End-to-End Systems Constructing the RCNN models comprises two stages. Hand-engineered pathways through cortex model the macro-scale level of the frontoparietal network. Weights emerging during reinforcement learning constitute the policy and thereby model behaviour and its neurocomputational implementation. In its simplest form, the latter stage occurs end-to-end. That is, all network weights are initialised randomly and trained on the target task (e.g., object manipulation). While this approach is conceptually simple, learning itself is a highly complex procedure. Being randomly initialised, the network has no prior knowledge. Thus, every aspect from the simplest detection of edges in the visual input to the difficult storage of useful memory needs to be learned simultaneously. Alternative to this approach, individual (or groups of) regions will be pre-trained on auxiliary tasks facilitating the primary task. For instance, the visual cortex model will be pre-trained on pose estimation in the dorsal stream and object recognition in the ventral stream. Moreover, the model will be split into executive and planning modules, comprised of both motor and frontal areas. The former will first learn to execute specific movements, whereas the latter learns to command the executive module to achieve a more complex task, like finger-tapping sequences. If successful, such modular training procedures can shed light on transferrable skill acquisition.

Micro-Scale Plausibility Structural connectivity data from the EBRAINS multilevel human brain atlas guided the network design on a macro-scale. From here, the biological realism of the RCNN will be improved by replacing activation functions in its layers with input-output relations (firing-current curves) of local circuits within their corresponding cortical regions. This is an important first step to facilitate the transformation of the RCNN into a biophysical model.

Domain Randomisation Within the final stage of the demonstrator, the agent will be additionally trained under domain randomisation. This paradigm randomises the parameters of the simulator to train policies that can adapt to any simulation, including a physical robot.

In conclusion, the M36 demonstration will feature networks trained to adapt to arbitrary physical parameters. These networks will have improved micro-scale plausibility, and a variety of methods will be employed to analyse their neurocomputations.