Automated manipulation technology for industrial warehouse applications (GROW)
(D3.17 – SGA3)

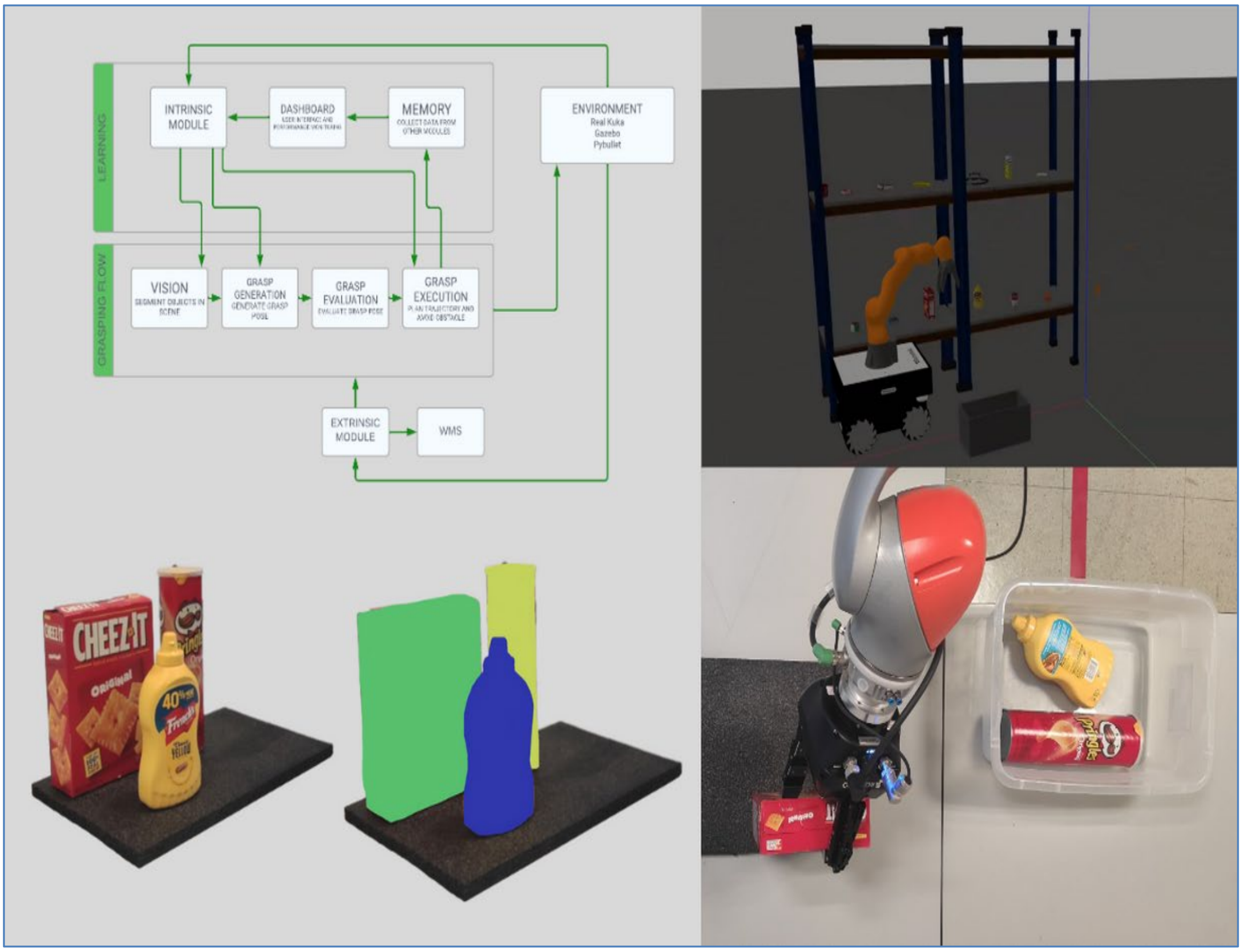

Figure 1: Integration of HBP Scene Understanding model

Within a robot architecture application and tested in an industrial grasping warehouse application

Project Number:	945539	Project Title:	HBP SGA3

Document Title:	Automated manipulation technology for industrial warehouse applications (GROW)
Document Filename:	D3.17 (D98) SGA3 M42 SUBMITTED 230929.docx
Deliverable Number:	SGA3 D3.17 (D98)
Deliverable Type:	Demonstrator
Dissemination Level:	PU
Planned Delivery Date:	SGA3 M42 / 30 SEP 2023
Actual Delivery Date:	SGA3 M42 / 29 SEP 2023
Author(s):	Emilio CARTONI, CNR (P12) Francesco MATTIOLI, AI2LIFE (P145) Gianluca BALDASSARRE, CNR (P12)
Compiled by:	Emilio, CARTONI, CNR (P12)
Contributor(s):	Michael ZECHMAIER, UM (P117), contributed to Sections 2 and 3 Yannick MOREL, UM (P117), contributed to Sections 2 and 3
WP QC Review:	Anita VAN OERS, UM (P 117)
WP Leader / Deputy Leader Sign Off:	Rainer GOEBEL, UM (P117)
T7.4 QC Review:	N/A
Description in GA:	The demonstrator includes simulation models, hosted on EBRAINS. It illustrates the developed cognitive architecture's ability, following a learning stage (used to acquire relevant knowledge of objects to be manipulated), to recognise desired sets of objects and manipulate them as required by the application. Relevant components are also demonstrated in a real-world setting.
Abstract:	The primary objective of the GROW project was to integrate a model, developed by the Human Brain Project (HBP) to study the brain, into a robot application with industrial relevance to assess its practical applicability. To achieve this, we opted for the integration of a scene-understanding model developed within the HBP, known as 'Seg-PredNet.' This model was seamlessly integrated into a robotic architecture named 'General-purpose Robot for Object Retrieval in Warehouses,' which has been independently developed by the GROW consortium for industrial purposes involving the manipulation of small objects. This application was driven by the recognition that automating the collection of small objects in warehouses holds significant commercial importance, particularly within the context of the thriving e-Commerce industry. The overarching goal was successfully realised through three key achievements: (a) the validation of an initial proof-of-concept integration of the system, conducted within a warehouse simulator developed in collaboration with the HBP UM partner; (b) the complete integration of the entire system into the Neuro Robotic Platform; (c) the benchmarking of the integrated system using a real robot and objects. These achievements align with the CEoI's mission of harnessing the potential of HBP models to address practical applications of industrial significance.

Keywords:	Scene understanding, visual segmentation, grasping, camera-arm-gripper robot, autonomous robotics, object manipulation applications, industrial relevance.
Target Users/Readers:	Neuroscience community, consortium members, warehouse stakeholders, platform users, students, general public.

Table of Contents

1. Introduction 4
2. HBP Scene understanding: Seg-PredNet 4
3. AI2LIFE grasping software 5
4. Gazebo simulation 6
5. Integration with EBRAINS NeuroRobotic Platform (NRP) 8
6. Test of the HPB Seg-PredNet in a real setting 9
7. General discussion 10
8. Looking forward 10

Table of Tables

Table 1: Test results in the real robotic scenario 9

Table of Figures

Figure 1: Integration of HBP Scene Understanding model 1
Figure 2: GROW system architecture 5
Figure 3: Gazebo simulation involving the whole Warehouse or grasping at a specific shelf 7
Figure 4: Integration of the GROW system within the NRP 8

1. Introduction

This Deliverable reports on the result of the GROW activity. GROW was selected as one of the two winning proposals of HBP SGA3 CEoI for the application of functional architectures supporting advanced cognitive functions to address AI and automation problems of industrial and commercial relevance. The mission of the activity was to apply HBP-developed scene understanding cognitive functions to a 'General-purpose Robot for Object retrieval in Warehouses' (from which the project acronym 'GROW') and also to integrate the whole system into the HBP software infrastructure, in particular the EBRAINS Neuro Robotics Platform (NRP). To accomplish this mission, an HBP neural-network model performing segmentation based on predictive coding accounts of the human vision system (Seg-PredNet) was selected to be integrated into a robotic grasping software for an industrial application. The Seg-PredNet model had been developed in T3.4 by EPFL (P134) and UM (P117). The model was integrated into a grasping application independently developed by the GROW consortium formed by the research institute coordinating the project, CNR (P12), and two companies, AI2LIFE (P145) and INGLOBE (P144).

After a first proof of concept demonstrating the information flow of the application in a simulated scenario, the network was fully integrated into the application and tested in a scenario involving real camera-arm-gripper robot picking-and-placing objects. The tests measured the performance of Seg-PredNet on several Key Performance Indicators relevant to the chosen industrial scenario. The software was also integrated with the NeuroRobotics Platform latest release, NRP 4.0; this enhances the compatibility with other possible bio-inspired or bio-constrained algorithms and modules developed within the HBP and other projects.

The GROW activity demonstrates how HBP scene understanding models can be profitably applied to problems with industrial and commercial relevance, such as the warehouse small-object manipulation and the post-harvesting agrifood sectors (See Figure 1). These types of applications, developed by AI2LIFE (P145), have recently received attention from various companies in the sector. In addition, the whole integration of the system showcases how the application potential of models developed to study the brain can be operationally measured with benchmarks having a relevant industrial value. In this respect, the system might become a service developed within the future EBRAINS research infrastructure.

Overall, these activities thus contribute substantially to the HBP OBJECTIVE 3.1 WPO3.1 'Enhanced real-world task performance through biologically plausible adaptive Cognitive architectures running on neuromorphic hardware and closed-loop neurorobotics Platforms'.

A video summary of the GROW activity is available at the following link:
https://youtu.be/nO_H1VAMFVw

2. HBP Scene understanding: Seg-PredNet

In agreement with the T3.4 management, it was decided to use the Seg-PredNet developed by EPFL (P134) and UM (P117) as the scene understanding technology developed within HBP to be integrated into an industrial application. Seg-PredNet was part of the output of OP3.11 Functional closed-loop demonstrators addressing sensorimotor functions with modular cognitive architecture [T3.4] at M18.

Seg-PredNet is a neural network inspired by predictive coding accounts of human vision, which performs segmentation and object identification. In the field of computer vision, segmentation is a fundamental process that involves partitioning or categorising an image into distinct and meaningful regions or segments. These segments correspond to objects, shapes, or regions of interest within the image. The segmentation process is pivotal for enabling robots to interpret and understand the environment (e.g. localise an object to be grasped).

Seg-PredNet combines the layers of PredNet, which can be used for next-frame video prediction with further layers that use the internal representations to predict segmentation masks. The network is based on a recurrent architecture, so it can process a continuous information flow, such as video frames, to update its predictions through time.

3. AI2LIFE grasping software

The mission of GROW was to integrate and validate a model from HBP within a ‘General-purpose Robot for Object-Retrieval in Warehouses’. This scenario involves a robot that gets an ‘order’ (a list of objects) and executes the order by moving among the shelves of the warehouse and by picking-and-placing the objects into a container it brings with it. This scenario in particular requires the robot to autonomously navigate among the shelves and collect objects. The navigation functions are for now abstracted, as out-of-the-shelves solutions are available for this (e.g. based on SLAM models). Instead, the focus is on the grasping challenges since partially unstructured industrial settings, such as those of warehouses, require the robot to be able to grasp a large variety of (small) objects independently of their shape, rigidity, position, and orientation on shelves.

Functional to this mission, the GROW consortium (CNR (P12), AI2LIFE (P145), INGLOBE (P144)) has independently developed the Architecture described in Figure 2 and below. As aforementioned, the mission of the activity was to integrate into this architecture some brain-based modules from HBP and also cast the whole system into the HBP software infrastructure, in particular the NRP.

An important feature of this architecture is that its composing modules can be implemented with different specific models, such as models coming from machine learning (ML), as commonly done by the consortium, or brain-based models, as those coming from the HBP project. Functionally, this is made possible as the whole software is divided into modules whose input and output are standardised.

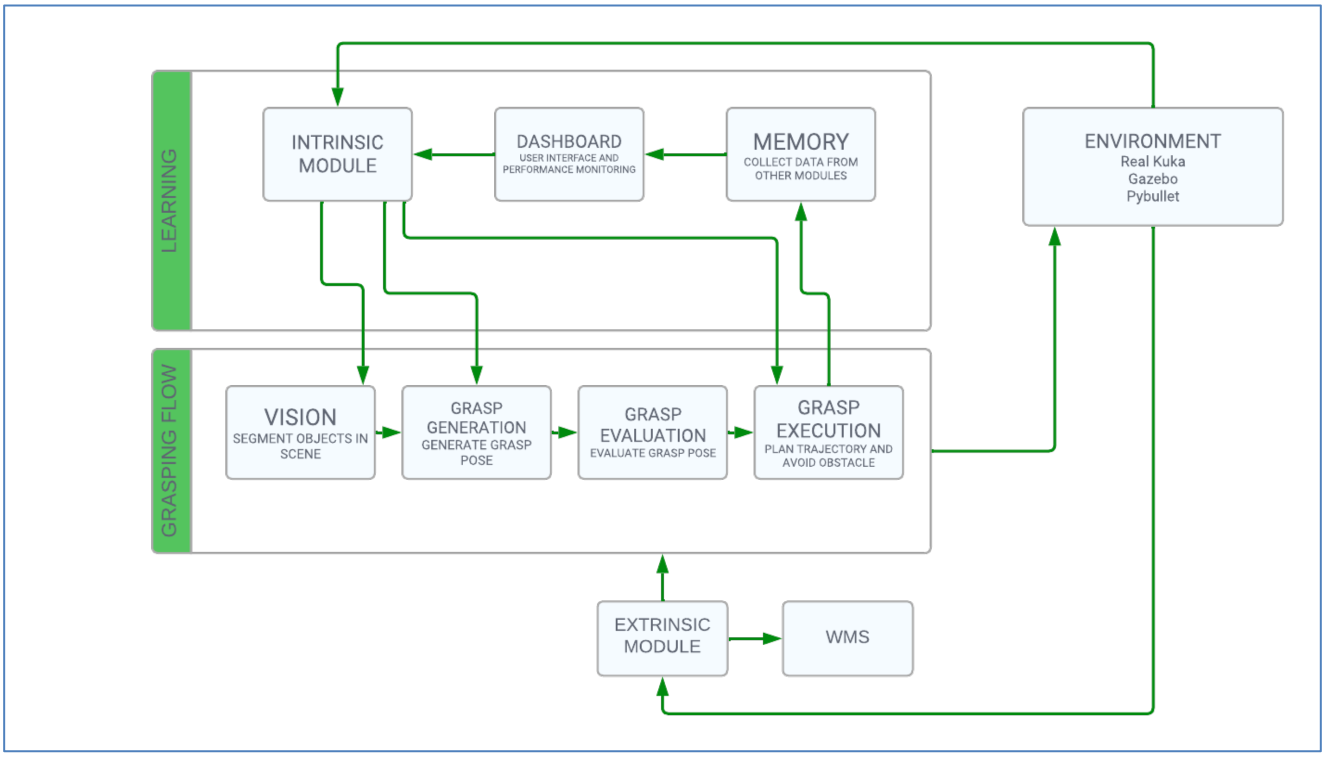

Figure 2: GROW system architecture

As shown in Figure 2, the architecture is split into these modules:

- 3 Learning modules: Intrinsic Module, Dashboard, and Memory
- 4 modules for the Grasping Flow: Vision, Grasp generation, Grasp evaluation, and Grasp execution
- 3 modules to interact with the external environment: Warehouse Management System (WMS), Extrinsic module, Environment

During the warehouse operations, the Robot controller gets orders (sets of objects to be collected) from the *WMS*. These orders are processed by the *Extrinsic module* which sets the specific goals (the ‘extrinsic goals’, in the words of the open-ended learning framework within which the architecture was developed) which moves the robot around the warehouse to reach the relevant shelves and asks the *Grasping Flow modules* to collect the required objects from each shelf, one after the other, until the order is complete.

When robot is in front of the shelf and the *Grasping Flow* is interrogated to grasp a certain class of object (e.g. a 'bottle of mustard'), the *Vision* module acquires RGBD images of the environment, **segments the scene and detects where the desired object is, returning a point cloud of the target object** (the function implemented by the HBP scene understanding component) to the Grasp Generation module. By analysing the point cloud the *Grasp Generation* module generates possible ways to grasp the object and sends the relevant parameters (grasp position and orientation in 6-dof space) to the *Grasp Evaluator*. The *Grasp Evaluator* evaluates the generated grasp poses and ranks them, from the one, which is predicted to have the higher chance of success to the lowest.

Finally, the *Grasp Executor* plans and executes the poses, starting from the highest ranked ones. If a plan is found, the *Grasp Executor* moves the gripper to the desired pose, performs the grasp, and retrieves the object by moving it to the container.

All experience collected when executing the *Grasping Flow* is sent to the *Memory* module inside the *Learning process* (images acquired, grasp poses tried and whether the object was grasped or not). A dashboard analyses all the grasps performed in Memory and returns to the user the current performance of the robot, measured by both global KPI and object-specific KPI (e.g. how many times it successfully grasped the bottles of mustard or how frequently it required a human intervention when doing so). If the performance on some class of objects is not satisfactory, the user can trigger new learning for that class of objects by engaging the *Intrinsic Module* (within the open-ended learning framework, 'intrinsic' refers to the autonomous learning of the robot driven by internally generated needs for particular pieces of knowledge).

In the current implementation, the learning processes of the *Intrinsic Module* **extract object-specific experience to train the HBP Vision module.**

Both the *Learning Flow* and *Grasping Flow* interact with an *Environment* module, which is a common interface to control the robot arm, in both the simulated and real robot, and to get information on the robot state and to access the RGBD cameras needed for *Vision*. The *Environment* module can be used in the same way to control simulated robots (simulations based on Gazebo or Pybullet) or to control the real robot (in this case a KUKA iiwa R800 arm, a Robotiq 3-finger gripper, and an Intel d435 RGBD camera).

4. Gazebo simulation

A simulated Warehouse environment was constructed in collaboration with the UM (P117)to make an initial test of a proof-of-concept implementation of the system before porting the application to a real robot (see Figure 3).

The simulated environment relied on the Gazebo simulator and it included a KUKA LBR-iiwa-arm mounted on a mobile platform and equipped with a Robotiq 3-finger adaptive gripper. Two versions of the environment were made:

1) A simulation encompassing a larger part of the warehouse (to be used for simulation of robot navigation by the concurrent HBP PROMENAID activity)
2) A simulation reproducing a reduced environment focusing on a rack with multiple shelves and objects for testing grasping

The Gazebo simulation utilised ROS for communication and its MoveIt! package for controlling the robot. These were also used to control the real robot.

While the ROS interface was almost identical to the setup with the real robot, thus facilitating a sim to real setup transfer, simulating grasps correctly inside Gazebo proved particularly challenging. In most simulations, the objects seemed to 'actively escape the grasp' of the robot, with no apparent fix to this behaviour. Looking at the experience of other Gazebo/ROS users, we found this to be a widespread problem, which no one seemed to have solved. Indeed, most users and other researchers we have been in contact with, in the end resolved with a 'hack': creating a virtual fixed joint linking the gripper and the object when they contact. As this 'fix' was not viable for simulations where we wanted the robot to autonomously practise to learn to grasp properly, a notable additional effort was spent looking for a proper solution. In the end, we found out the issue to be caused by multiple different causes, which, beyond setting correct parameters for friction, mass, moment of inertias of objects and robot links, also involved the way the robot is controlled. The default way of controlling the robot through ROS via a position interface is incompatible with correctly simulating the grasp and that a force-based interface has to be used, not only for the gripper but also for the whole robot — wheels included.

In the end, this solution was finally found and we started to see proper simulations of grasps. Overall, the testing of the proof of concept was successful as planned, thus enabling the successive steps of testing on a real robot on which we started to focus.

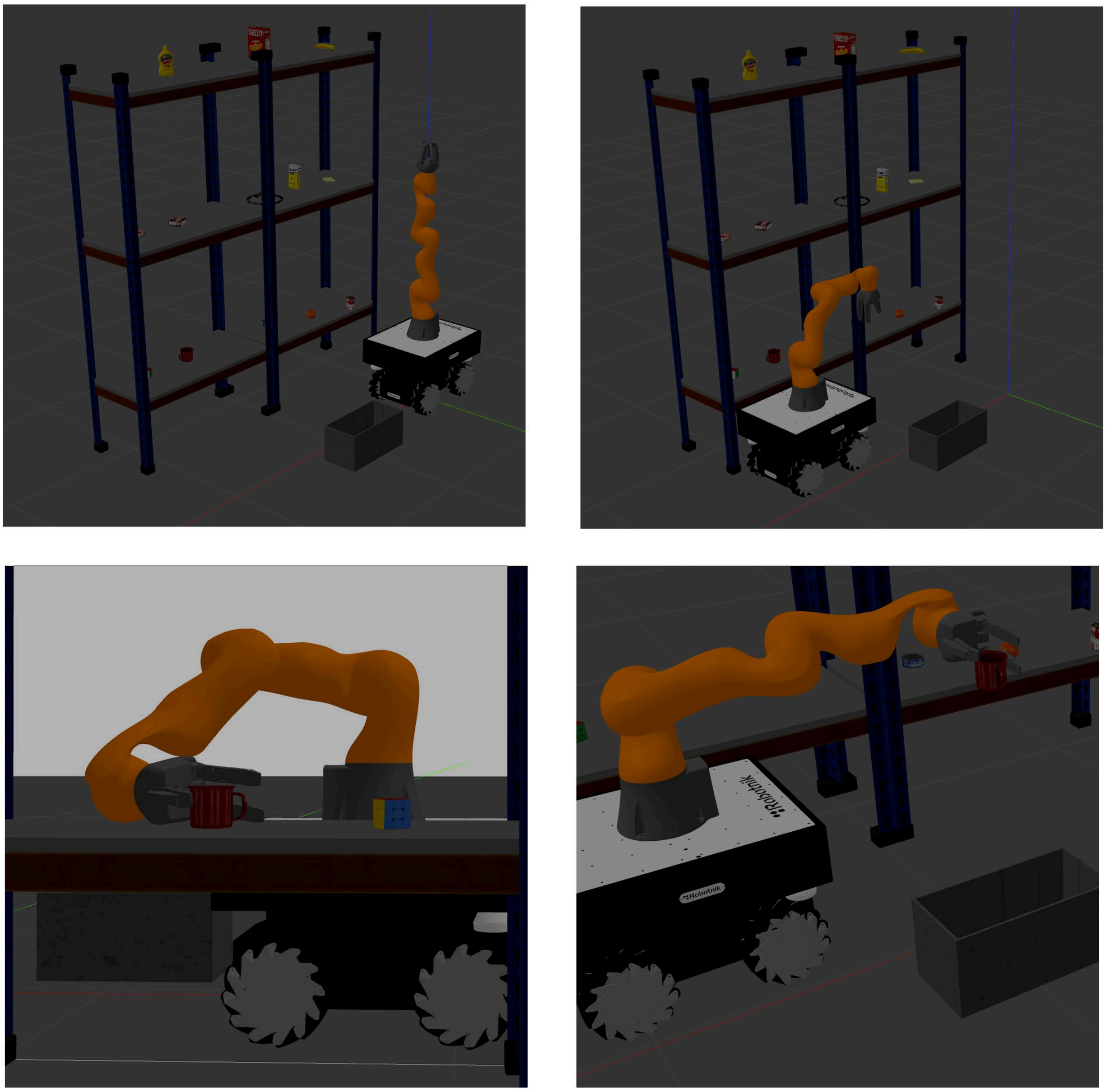

Figure 3: Gazebo simulation involving the whole Warehouse or grasping at a specific shelf

5. Integration with EBRAINS NeuroRobotic Platform (NRP)

The GROW software components were then integrated with the EBRAINS NeuroRobotic Platform (NRP), in collaboration with the HBP NRP developers. The Neurorobotics Platform is a simulation platform that facilitates the integration of different brain models to different simulators and robots. The fundamental structural principle underlying NRP 4.0 is modularity. Within the NRP, simulations adopt a hub-and-spoke architecture, wherein a central component, referred to as NRP-core, manages the execution of all other components. These spoke modules can exhibit heterogeneity in both their nature and function, and there is no fixed limit on their quantity. This flexibility empowers the NRP to leverage a wide range of simulators (e.g. Gazebo, NEST, Unity, MuJoCo, PyBullet, EDLUT, OpenSim) and facilitates the creation of complex control architectures by combining various components. For our activity, we used the latest available version of NRP, version 4.0. The GROW system architecture was broken down into three NRP components (or 'engines' – see Figure 4):

- A vision component containing the object segmentation and identification functions (Seg-PredNet).
- A "grasping" component, containing the algorithm to generate and rank grasps.
- An "agent" component, containing the rest of the architecture.

NRP 4.0 provides two communication styles for components:

- "FTILoop" enables synchronous communication between components
- "EventLoop" allows components to exchange messages asynchronously

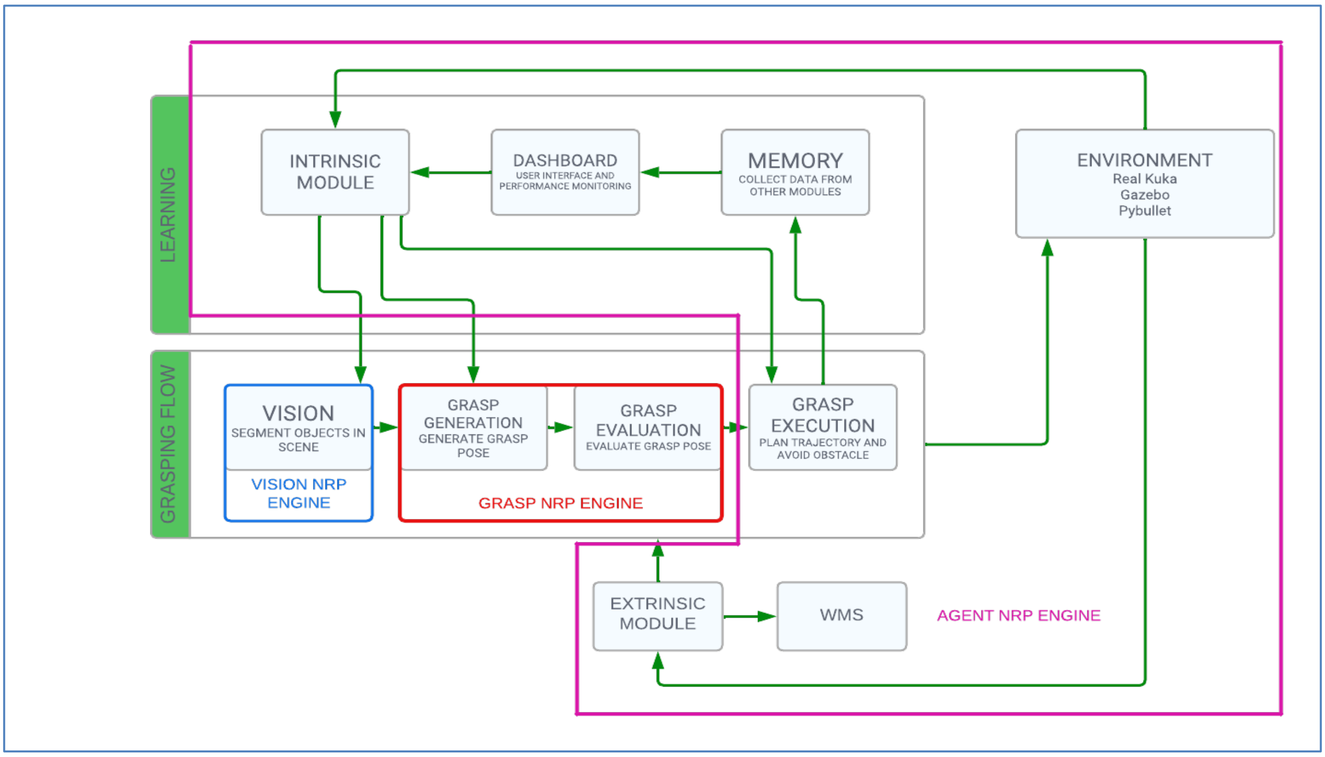

Figure 4: Integration of the GROW system within the NRP

Initially, for our specific requirements, the utilisation of an EventLoop appeared to be the preferable choice. However, it is worth noting that this mode of communication is relatively novel in version 4.0 and has not achieved full stability yet. Consequently, the NRP team recommended opting for the more established FTILoop for the current integration. The process of implementing an NRP simulation with the actual robot was relatively straightforward, given that our system was already divided into distinct modules. Undoubtedly, the most challenging aspect of this work was the implementation of the Transreceiver Functions, which facilitate communication among the various

NRP engines. Future improvements of the integration might switch to the EventLoop simulation style to greatly enhance communication efficiency and improve system response times.

6. Test of the HPB Seg-PredNet in a real setting

To evaluate the performance of the new GROW system, which has been integrated into the NRP and incorporates the Seg-PredNet vision network developed by HBP, we established a testing environment. This environment is set up within the workspace of the IIWA LBR robot, which includes an 88 cm by 125 cm table. The scene is captured by three RGBD cameras (Intel d435), with two positioned near the robot's base and one near the hand effector. The hand effector used is a Robotiq 3-finger gripper.

The objects used for testing are selected from a subset of the YCB Benchmarks - Object and Model Set, the same subset has been used for training the Seg-PredNet vision network. During the testing procedure, each object is randomly placed within the robot's workspace; the robot must pick up the object and move it to the desired position (a bucket to the right of the robot). This process is repeated 40 times. For each repetition, the robot has three attempts to recognise and grasp it. If the maximum number of attempts is reached without success, the repetition is considered a failure.

To better assess the performance of Seg-PredNet, we compared its results with a different GROW system using AI2LIFE's (P145) proprietary vision component and without the NRP integration. In contrast to Seg-PredNet, AI2LIFE's (P145) algorithm relies on a geometry-based solution for object segmentation and identification.

To evaluate the efficacy of the system, we employed six key performance indicators (KPIs):

- *PER - Production Error Rate*: this metric is the frequency at which the robot fails to fulfil an order. Such failures may arise from various sources, including instances where the vision system fails to recognise the object, the robot's grasp attempt is unsuccessful, or the object is misplaced en route to its destination.
- *PSR - Pick Success Rate:* this indicator quantifies the success rate of pick attempts.
- *IR - Intervention Rate:* this KPI gauges the frequency with which the system necessitates human intervention. Human intervention becomes necessary when a robot encounters an impasse, primarily due to inaccurate handling of the target object.
- *MAR - Maximum Attempts Reached:* in a single iteration, the robot is permitted up to three attempts to secure the requested object. Exceeding this threshold results in the trial being classified as a failure. The MAR metric reflects this occurrence.
- *TIME - Time to Complete Pick:* the TIME indicator records the duration required to execute each pick operation, encompassing the entire process from the acquisition of RGBD images to the moment the object is deposited in the designated container.
- *NP - No Plan:* this metric tracks the instances when the robot is unable to formulate a viable trajectory to reach the target object. This error can be attributed to a multitude of factors, such as segmentation failures in the vision system resulting in incomplete object recognition and subsequent collision issues during the planning phase. It is one of the most prevalent sources of errors encountered.

Table 1: Test results in the real robotic scenario

	PER	PSR	IR	MAR	TIME	NP
Seg-PredNet	7%	69%	5%	7%	65.6s	22%
AI2LIFE (P145) proprietary algorithm	2%	> 90%	<1%	1%	15-20s	4%

The system using Seg-PredNet achieved a 7% Production Error Rate, which means it successfully picked the requested objects 93% of the time (see Table 1). In the remaining 7% of cases, the system gave up after reaching the maximum number of attempts (3 attempts). Failures of the attempts were mainly due to failures either during the planning phase (22% planning attempts rejected) or during grasping execution, where the execution led to the necessity of human intervention.

In general, these failures can only partly be attributed to the general system (e.g. the planning algorithm) as the comparison shows much lower MAR, IR and NP indices for the AI2LIFE (P145) proprietary algorithm. Seg-PredNet masks seemed to be of lower quality in the case of the failures, resulting in incomplete object segmentation and attempts to plan a grasp inside the object. In turn these led to failures of planning or plans that pushed the object against the table (due to not accounting for some parts of the object), thus triggering safety mechanisms and human intervention. TIME index for Seg-PredNet was also higher due to the necessity of multiple attempts and some overhead from the NRP middleware.

7. General discussion

The findings of the test indicate that the HBP brain-inspired visual component performs less effectively than the ML component it was tested against. However, it is important to note that the performance achieved by the HBP component, though somewhat lower, remains significantly valuable. This is particularly noteworthy considering that the component was primarily designed to address computational constraints inspired by the brain's functioning, rather than solely prioritising performance, as done for the ML components. This suggests relevant possibilities for enhancing performance if this becomes an explicit research target.

At a more general level, the comparison performed remains highly useful as it underscores the potential advantages of models inspired by the brain that must adhere to bio-inspired constraints. A prime example of this is the feature of the HBP visual component for which it utilises a recurrent neural network resembling an important feature of the brain regions corresponding to the model. This feature enables the component to perform a real-time processing of information streaming from the environment. In contrast, the ML components used for comparison operate only on static images and are ill-suited for more dynamic, real-world scenarios where a robot has to deal with a continuous stream of visual data from the environment (e.g. for closed-loop grasping of moving objects).

Finally, it is worth highlighting that the GROW system realised and presented here is one of the first instances of systems that allows a rigorous benchmarking of brain-inspired model components against state-of-the-art ML models within an industrially relevant scenario. As shown, this type of comparison is ideal to indicate the weaknesses of brain-inspired models when a real-life application is addressed, but also to highlight their potentialities deriving from the brain-inspired features. If further developed, these features might have a relevant application value.

A last additional note involves the quality of the specific industrially relevant scenario addressed here. The whole system in particular involves the manipulation of small objects in warehouses. This is a very important area of application for AI models. The importance of this sector and applications is demonstrated by the fact that the AI2LIFE (P145) technology has received the expression of interest by other companies of the warehouse management sector.

8. Looking forward

Tests of the Seg-PredNet in the real setting highlighted that its performance is for now significant but still lower than competitive ML approaches. This opens up different possibilities for the future regarding its usability in applications. One could be that the teams developing the component will succeed to increase its performance, for example in terms of the quality of the object segmentation and typology of the usable objects. Another interesting possibility is that the component is developed and tested in applications requiring a continuous streaming of input, for example to support close-loop manipulation actions of the robots. Indeed, the model might have an interesting potential for this challenge due its brain-inspired recurrent nature.

In the period following the Human Brain Project (HBP) termination, AI2LIFE (P145) will continue to work on robotic applications involving the Warehouse and post-harvesting agrifood sectors, also seeking to involve other stakeholders operating within them. These will involve in particular two types of stakeholders. The first are potential client SME companies needing customised robotic applications for unstructured or semi-unstructured applications where robot adaptation is important. The second are supply companies that might contribute with specific complementary solutions, possibly within joint ventures, such as customised robot hardware solutions. Now that the GROW consortium has developed an important tool to test the application relevance of brain-inspired components, this paves the way to easily test further brain-inspired models that might be proposed to AI2LIFE (P145) to evaluate their possible use within its target applications.

In addition, CNR (P12) has become a partner of the 'EBRAINS-Italy' project, which started in November 2022 and will last 3 years. This is a project funded by the Italian Government through the EU-funded post-covid National Recovery and Resilience Plan ('PNRR - Piano Nazionale di Ripresa e Resilienza'). This project is working in synergy with the nascent EU-funded EBRAINS project aiming to support the realisation of a European Infrastructure starting from the HBP EBRAINS infrastructure legacy. This infrastructure is planned to support both research and innovation involving brain-related models as those developed within the HBP. Within this context, CNR (P12) is currently evaluating if the GROW system could become a service within the new EBRAINS infrastructure. This service would allow other research teams to test if their models, possibly initially developed to study the brain, do have an application potential if used to implement some of the components of the GROW system.