# Open-Ended Learning of Grasp Strategies using Intrinsically Motivated Self-Supervision*

**Quentin Delfosse[1], Svenja Stark[1], Daniel Tanneberg[1], Vieri Giuliano Santucci[2], Jan Peters[1,3]**
[1]Intelligent Autonomous Systems, Technische Universität Darmstadt, Germany
[2]Institute of Cognitive Science and Technologies, National Research Council, Italy
[3]Robot Learning Group, Max Plank Institute for Intelligent Systems, Germany

***Abstract*—Despite its apparent ease, grasping is one major unsolved task of robotics. Equipping robots with dexterous manipulation skills is a crucial step towards autonomous and assistive robotics. This paper presents a task space controlled robotic architecture for open-ended self-supervised learning of grasp strategies, using two types of intrinsic motivation signals. By using the robot-independent concept of object offsets, we are able to learn grasp strategies in a simulated environment, and to directly transfer the knowledge to a different 3D printed robot.**

## I. Introduction

Despite a lot of recent significant progress in robotics and computer vision, robots are still far from being autonomous, as they lack lifelong learning skills. Therefore, one of the biggest remaining challenges is to turn the assistants of tomorrow into lifelong learners, able to adapt to their changing environment, and to transfer learned knowledge [1]. As a step towards this goal, we propose an architecture that autonomously learns grasp positions for different object shapes, enabling a robotic arm to grasp previously unknown objects. The learning process is self-organized through the usage of performance improvement, a type of competence-based intrinsic motivation.

## II. Approach

The Goal-Discovering Robotic Architecture for Intrinsically-Motivated Learning (GRAIL) [2] allows a robot to discover abstract goals in its environment, create internal representations of those, and use Competence-Based Intrinsic Motivation to self-supervise its learning. The architecture is equipped with a predefined number of goals and experts. Experts compete for solving the goals and each expert will eventually be assigned to one goal. Both goals and experts are assigned scores, representing their recent performance improvement. At timestep $t$, the goal to train on, and the expert to train are chosen through a softmax on the respective scores. After execution of the expert, the score of the selected goal ($G^t$) is updated with the performance improvement, i.e., the difference between the current performance $p^t$ and the previous performance $p^{t-1}$, smoothed with a moving average given by

$$G^t = (1-\alpha)G^{t-1} + \alpha(p^t - p^{t-1}) \quad ,$$

*This project has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement #713010 (GOAL-Robots) and #640554 (SKILLS4ROBOTS).

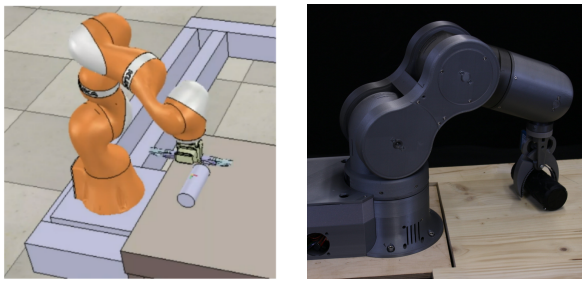

Fig. 1. *Left:* Training environment with a 7DOF Kuka arm simulated in V-REP. *Right:* Real 3D printed robot environment for the evaluation of the knowledge transfer.

with $\alpha$ as smoothing factor. The score of the selected expert is updated analogously.

### A. Extending the Architecture

Acquiring knowledge about previously unknown objects on-the-job is crucial for lifelong learning [3]. We thus adapted GRAIL s.t. at any point, the architecture can incorporate new goals. When a new goal is discovered, the architecture instantiates random experts for it. Existing experts of other goals are also tested. If they perform well, i.e., succeed in grasping at least one object, they are duplicated and the duplicate is added to the new goal (expert transfer). We thus bootstrap the learning process on newly discovered goals, and allow a not predetermined thus unlimited number of goals and experts. This extension of the architecture is depicted in Fig. 2.

GRAIL uses *Maximizing competence progress motivation* [4] to self-organize the learning order of different goals, as goals and experts making the most progress have higher

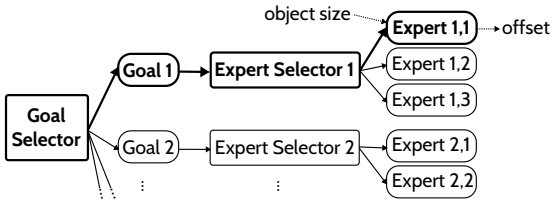


Fig. 2. A schematic depiction of the goal-expert relation within the architecture. The goal selector selects a goal to train on (e.g. Goal 1) by sampling from the softmax of the goal scores. The expert selector of the selected goal selects one of its experts (e.g. Expert 1) analogously. The selected expert predicts the offset pose according to the object size.

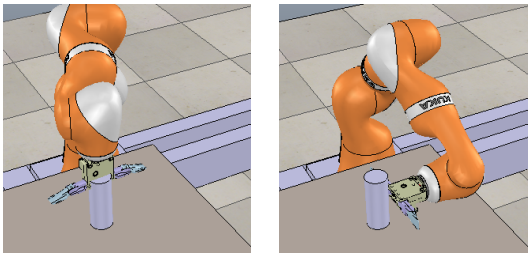

Fig. 3. *Left:* The end effector is sent to the center of the object which is permeable for demonstration purpose. *Right:* The predicted offset (5cm along y-axis, and 90 degrees around x- and around y-axis) is added.

chances to be selected. We use the performance, which is part of this intrinsic motivation model, to initialize the score of the experts, by setting it to the initially observed performance ($p^0$). The initial goal score is the maximum score of its experts. We thus encourage the robot to first train on goals with experts that show promising performances, hence maximizing the chances to have stable grasping positions in a short training period. We use the average score as temperature in the softmax, to balance between exploration and exploitation.

### B. Application to Grasping

To learn grasp strategies in a time efficient manner, we let the architecture predict the robot end effector pose which is necessary for grasping a target object. Therefor, we categorize different object shapes as goals. This representation enables the architecture to learn experts which, for every object shape and size, determine the offset (position and orientation) which is added to the pose of the target object. An example is shown in Fig. 3. Hence, for any given object pose, the architecture predicts an end effector pose enabling a stable grasp.

## III. Experiments and Results

The learning setup is simulated in V-REP and consists of a Kuka LBR R820 manipulator and a table on which objects are placed. During learning, the robot encounters objects of certain shapes, which differ in their size and pose. In the following, we exemplary describe the self-supervised learning procedure of one run, for which the results are also shown in Fig. 4. We initially place a cube and a cylinder in the simulated environment. The architecture successfully discovers these two goals by checking if the observed shapes have already been seen. For the cube shape (Goal 1), four experts are instantiated (i.e. they successfully grasped at least once), and two for cylinders (Goal 2). Fig. 4 shows the evolution of these experts, how they are selected and updated during the 19 first episodes. One episode corresponds to the selected expert performing stochastic hill climbing to improve its performance with 10 trials. For each goal, experts with high score are first selected, and if no (further) progress is made, their selection probability decreases. Hence, the worse performing cylinder expert (Goal 2) is selected after the better one fails to improve. Grasps are considered successful when the object is lifted without slipping, i.e., a stable grasp achieves maximum performance. After Episode 10, we place a new object in the

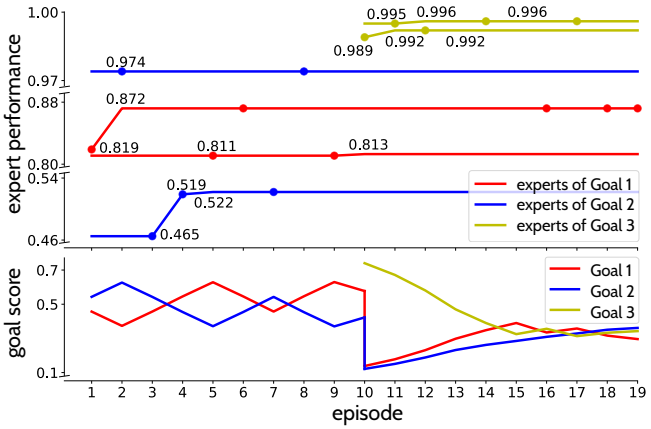


Fig. 4. An exemplary Development of the goal probabilities and the expert performances during 19 episodes. In each episode, first a goal, and then one of its experts is selected. The dots mark the selected expert. Goal 3 is detected at the beginning of Episode 10. *Top:* The precise performance values are marked for the initial scores and after each competence progress. *Bottom:* The plot shows the softmax of the goal scores which are based on their learning progress. It is also their probability of getting selected.

environment and the architecture autonomously discovers it. Four new experts are instantiated for the new goal, but only two are selected during training (see Fig 4). After few training iterations on this new goal, performances and progress are similar to the other goals, and the architecture automatically trains on all of them.

Working with offsets in task space allows an easy transfer of the learned grasp strategies. We show this by transferring the strategies learned on the Kuka arm in simulation to an open source 3D printed robot, Thor [5]. The robot has performed successful grasps for several object positions despite its different kinematics and shape[1], as shown in Fig. 1.

## IV. Conclusion and Discussion

To enable real open-ended learning, we extended GRAIL's self-supervised learning to handle infinitely many goals and experts and to transfer knowledge between experts. Further, it is possible to transfer strategies from one robot to another, which we show by training on a simulated robot manipulator and a successful transfer to a different 3D printed robot. We are currently training the architecture on a wider range of goals, and we are planning to integrate an object decomposition framework to learn grasping of more complex objects.

## References


[1] Thrun, S. Mitchell, T. Lifelong robot learning. *Robotics And Autonomous Systems*. (1995)

[2] Santucci, V., Baldassarre, G. Mirolli, M. GRAIL: A goal-discovering robotic architecture for intrinsically-motivated learning. *Transactions On Cognitive And Developmental Systems*. (2016)

[3] Young, J., Basile, V., Kunze, L., Cabrio, E. Hawes, N. Towards Lifelong Object Learning by Integrating Situated Robot Perception and Semantic Web Mining. *Proceedings of the Twenty-second European Conference on Artificial Intelligence*. (2016)

[4] Oudeyer, P. Kaplan, F. How can we define intrinsic motivation? *Proc. of the 8th Conf. on Epigenetic Robotics*. (2008)

[5] Thor: Open Source 3D printed Robot. https://github.com/AngelLM/Thor. (19 09 2019)


[1] Video of a successful grasp: http://quentindelfosse.me/index.php/thor