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).

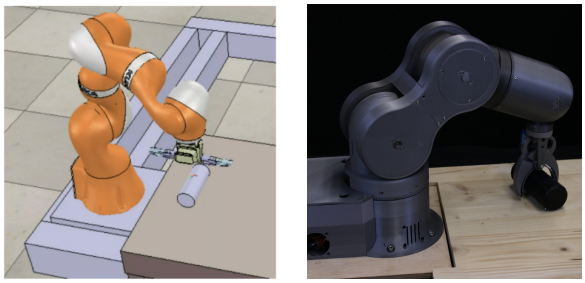

Fig. 1. *Left:* Training environment with a 7DOF Kuka arm simulated in V-REP. *Right:* Real 3D printed robot environment for the evaluation of the knowledge transfer.

with α 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

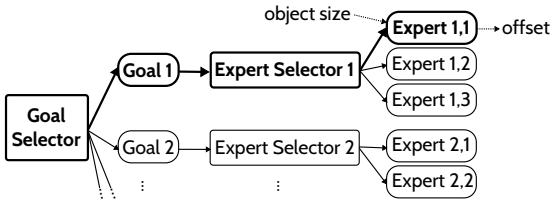

Fig. 2. A schematic depiction of the goal-expert relation within the architecture. The goal selector selects a goal to train on (e.g. Goal 1) by sampling from the softmax of the goal scores. The expert selector of the selected goal selects one of its experts (e.g. Expert 1) analogously. The selected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