Neural End-to-End Self-learning of Visuomotor Skills by Environment Interaction

Matthias Kerzel and Stefan Wermter

Knowledge Technology Institute, Department of Informatics,
Universität Hamburg, Vogt-Kölln-Str. 30, 22527 Hamburg, Germany
{kerzel,wermter}@informatik.uni-hamburg.de

Abstract. Deep learning with neural networks is dependent on large amounts of annotated training data. For the development of robotic visuomotor skills in complex environments, generating suitable training data is time-consuming and depends on the availability of accurate robot models. Deep reinforcement learning alleviates this challenge by letting robots learn in an unsupervised manner through trial and error at the cost of long training times. In contrast, we present an approach for acquiring visuomotor skills for grasping through fast self-learning: The robot generates suitable training data through interaction with the environment based on initial motor abilities. Supervised end-to-end learning of visuomotor skills is realized with a deep convolutional neural architecture that combines two important subtasks of grasping: object localization and inverse kinematics.

Keywords: Visuomotor policies, Humanoid robot grasping, Deep learning, Self-learning

1 Introduction

Acquiring visuomotor skills is vital for robots that act as assistants and companions in complex domestic environments. Grasping is an essential capability for such a robot as it enables manipulation and multimodal inspection of objects. However, grasping is a challenging task. Even in a non-cluttered environment with an easy-to-grasp object, the robot must be able to localize its target in 3D-space and then use visuomotor skills to move its end-effector to this position.

Conventional frameworks solve this problem with modular approaches, which employ computer vision algorithms for determining the 3D-position of an object and inverse kinematics solvers to reach for the object, see [7] for example. These approaches rely on expert human knowledge and accurate data about the kinematic model of the robot.

Developmental robotics seeks to solve this challenge through learning and teaching. According to Cangelosi and Schlesinger [1], robots should autonomously develop increasingly sophisticated sensorimotor abilities. This paradigm opens up intuitive human-robot interaction scenarios, where a non-expert user can act

Citation: Kerzel M., Wermter S. (2017) Neural End-to-End Self-learning of Visuomotor Skills by Environment Interaction. In: Lintas A., Rovetta S., Verschure P., Villa A. (eds) Artificial Neural Networks and Machine Learning – ICANN 2017. ICANN 2017. Lecture Notes in Computer Science, vol 10613, pp 27-34. Springer, Cham.
The final publication is available at Springer via https://doi.org/10.1007/978-3-319-68600-4_4

as an instructor, much in the same way a human would teach a child. Moreover, developmental robotics benefits from the exchange with neuro-cognitive science [10], as robotic models can be inspired by and evaluate findings about ontogenetic development.

Following the idea of developmental robotics, we present a deep convolutional architecture for self-learning of robotic grasping skills through interaction with the environment. We perform supervised end-to-end training of a deep convolutional neural network architecture with training data that the robot acquired through interaction with the environment with minimal human involvement.

2 Related Work

Convolutional neural networks (CNNs), inspired by the visual system of mammals, have been successful in various visual and also nonvisual tasks including object localization [6]. Supervised training of CNNs relies on the availability of annotated training data. These annotations are represented as object bounding boxes [16] or coordinates of object centers [15]. Oquab et al. [12] have circumvented the necessity of spatial annotations to learn object localization by using labeled images.

To avoid the need for annotated data, deep reinforcement learning, introduced by Mnih et al. [11], combines a CNN architecture for visual processing with a neural realization of reinforcement learning to develop sensorimotor skills through trial and error. This and similar approaches [4] have been extended for continuous control problems. For instance, Lillicrap et al. [9] developed the Deep Deterministic Policy Gradient algorithm by coupling deep learning with a continuous actor-critic approach.

Deep reinforcement learning proved to be strong in virtual environments where collecting training data can happen fast, without human assistance and without danger of damaging the learning agent, e.g. [13]. But when applying trial-and-error methods to real robots in complex environments, the time needed for training increases dramatically. Pinto and Gupta [14] showed how a robot can learn grasping positions and angles in 700 hours. They employed staged learning where continuously improving neural networks were used to collect samples for the next iteration of training. The large number of required trials makes this approach unsuitable for non-industrial, domestic robots. To solve this challenge, Levine et al. [8] proposed a guided policy search method that transforms motor policy search into supervised learning. To achieve this transformation, Levine et al. used visuomotor training setups in which the state of the environment was fully observable at training time, thus generating sufficiently annotated data for supervised learning by exploiting the known forward kinematics of the robot.

In summary, existing approaches rely either on human-annotated training data, extensive learning phases that are not suitable for domestic robots, or information about the kinematic model of the robot. We extend the state of the art by presenting an approach that circumvents all of these necessities by letting

the robot randomly generate annotated training samples through autonomous interaction with its environment.

3 Self-Learning to Grasp by Randomly Placing Objects

For the supervised learning of visuomotor grasping skills with a neural architecture, training samples are required that associate the state of the environment, represented by an image, with the desired action, represented as a joint configuration that places the robot's end-effector in a grasp position. With extensive human effort, these samples could be generated by repeatedly placing an object in front of the robot at different positions and manually guiding the robot's end-effector into a grasping position.

We present a novel approach to acquiring training samples through self-learning with minimal human intervention by inverting the grasping task. The robot autonomously places objects at random positions in front of itself, memorizes the joint configuration that led to this state of the environment and then associates an image with the joint configuration. The rationale behind this approach is the reversibility of the grasping action: When the robot places an object on a surface from a given joint configuration, the same configuration is likely to be suitable for grasping the object.

3.1 Experimental Setup

Experiments were carried out on NICO (Neuro-Inspired COmpanion) [5], a child-sized developmental robot that we designed as a multimodal research platform for neural architectures, see Figure 1. NICO's arms have six degrees of freedom; its Seed Robotics[1] hands have three segmented fingers. All joints provide proprioceptive information. The head is equipped with two cameras and is able to perform pitch and yaw movements. We used a 3D-printed grasping object that has a broad base for stability and is rotation-invariant along its upright axis for easy grasping, see Figure 1.

3.2 Initial Motor Skills

To enable the robot to place objects on the table during its self-learning phase, it requires initial motor skills. The robot needs to explore the surface of the table to learn joint configurations that bring its hand close to the table surface for placing the object for grasping. We used human demonstration to train this skill: For easy handling, the robot closes its hand around the grasping object. A human demonstrator then moves the object around on the table surface. The motors of the robot's arm have no torque during this time and are only used to record joint angles. Only 30 seconds of demonstration time were needed to record 600 samples.

[1] http://www.seedrobotics.com

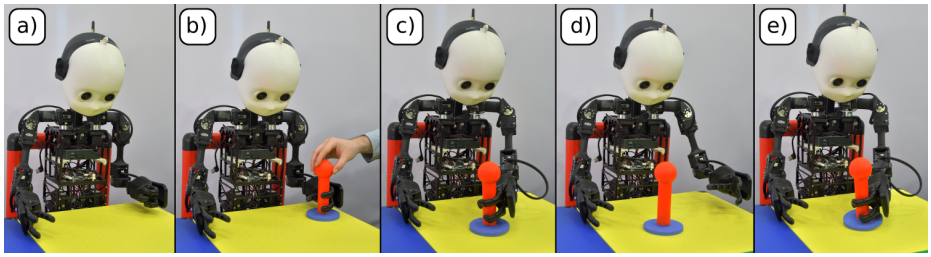

Fig. 1. NICO (Neuro-Inspired COmpanion) performing a self-learning cycle with the grasping object. a) The robot begins with its hand in the home position. b) A human experimenter places the grasping object in the robot's hand. c) The robot moves to a random position from the initial motor training. d) The robot releases the object, removes the hand and records an image. e) The robot moves back to the last joint configuration to grasp the object again. The self-learning cycle repeats steps c) to e).

3.3 Self-Learning by Gathering Training Samples

In the main training phase, the robot repeats a training cycle to gather pairs of images of the object on the table and matching joint configurations. The training begins with the robot's hand in the home position, see Figure 1 a). A human experimenter puts the grasping object into the robot's hand, Figure 1 b). The robot moves to a randomly selected joint configuration from the initial motor training, thus moving its hand and the object on the table, Figure 1 c). The robot releases the object, and moves the hand to a predefined position outside its field of vision to record an image along with the previously selected joint configuration, Figure 1 d). The hand is moved away to avoid interference with the training of the visual system. After recording the image, the robot moves its hand back to the previously selected joint configuration, thus attempting to grasp the previously placed object, Figure 1 e). After closing its hand, the robot repeats its training cycle with a new random position from the initial motor training.

The robot uses proprioceptive haptic information from its hand motors to determine if a grasping attempt during the self-learning phase has been successful. The attempt can fail if the grasping object tumbles during release or is accidentally moved while the robot retracts its hand. In this case, the training cycle is automatically stopped and the last collected sample is deleted, as it might lead to learning wrong visuomotor policies. The robot moves its hand back to the home position and requests human assistance. Once the object is placed back into the robot's hand the self-learning is resumed.

3.4 Neural Architecture

A deep neural architecture is used for end-to-end learning of grasping skills from the collected samples. The architecture consists of convolutional layers for object localization and dense layers for learning motor policies, see Figure 2. The input

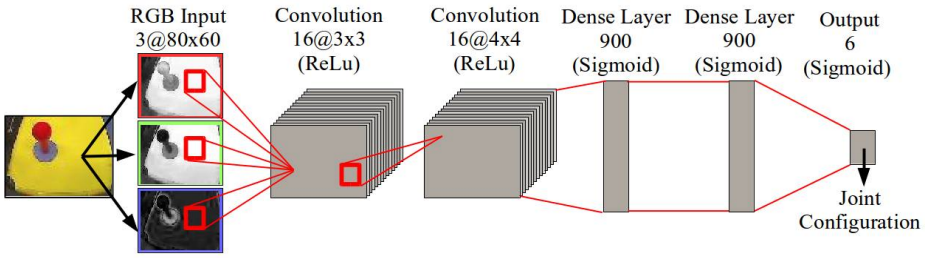

Fig. 2. Neural architecture for end-to-end learning of grasping.

layer of the architecture takes a downsampled cutout of the RGB image that shows the area directly in front of the robot with a dimensionality of 3*80*60. The input is processed by two convolutional layers with 16 filters of size 3x3 and 4x4 that are moved over the input with a stride of 1. The filter size and number is empirically determined. The convolutional layers are not followed by pooling, as these layers would realize translational invariance, which is detrimental for localization tasks (compare [11]). Two dense layers with 900 neurons each follow the convolution layers. Finally, the output layer consists of 6 neurons, one for each degree of freedom of the robot arm. The neurons in the convolutional layers use the rectified linear activation function introduced by Hahnloser et al. [3] to prevent the exploding or vanishing-gradient effect. The dense layers and output layer use sigmoid activation function to ensure an output in the [0..1] interval. Joint configurations are accordingly normalized to the same interval, with 0 being the minimum and 1 being the maximum joint value from the set of training samples. The network architecture and its hyper-parameters, as shown in Figure 2, were empirically determined. We chose a set of parameters that performed best on the average of the different experimental conditions described below.

4 Experimental Results

We conducted a full training and evaluation of the system that encompassed the initial motor training, a self-learning phase, neural network learning, grasping evaluation and an experiment on autonomous recovery during self-learning.

4.1 Self-Learning

During the self-learning phase, the robot autonomously collects training samples. Each training cycle takes ∼30 seconds to perform. 500 samples were collected. We analyzed three factors during the self-learning phase: Did collisions occur that required urgent intervention by the experimenters, did the robot recognize failed grasping attempts, and how often did grasping attempts fail?

During the 500 self-learning cycles, no self-collision of the robot or a forceful collision of the robot with the environment occurred, because all joint configurations selected during self-learning stem from the initial motor training which

is inherently collision free. Compared to trial-and-error methods like Deep Deterministic Policy Gradient [9] or multi-staged learning [14] the robot performs no explorative actions that could lead to harmful states.

Also, the robot reliably detected failed grasps and stopped its self-learning in such cases. The average number of consecutive, error-free self-learning cycles was 31.24 with a high fluctuation between 2 and 106 sequential cycles.

4.2 Neural Network Learning and Grasping Success

We evaluated the neural network's ability to learn visuomotor grasping skills from training sets of 10, 25, 50, 100, 200 and 400 samples. 2000 epochs of training were performed with stochastic gradient descent with Nesterov momentum [17] (learning rate = 0.01, momentum = 0.9). The batch size depended on the size of the training set. We used a batch size of 10 for the 10-sample condition, a batch size of 20 for the 25-sample condition and a batch size of 40 for all other conditions. The squared error was used as a loss function. Glorot uniform initialization, also known as Xavier initialization, was used for all layers to stabilize the strength of the input signal throughout the deep network [2].

Each experimental condition was repeated ten times to minimize the influence of randomization. For each trial, 50 validation and 50 test samples were randomly chosen from the sample set along with the desired number of training samples. Figure 3 (left) shows the results of training. As expected, the squared error in the test set decreased steadily with increasing training set size.

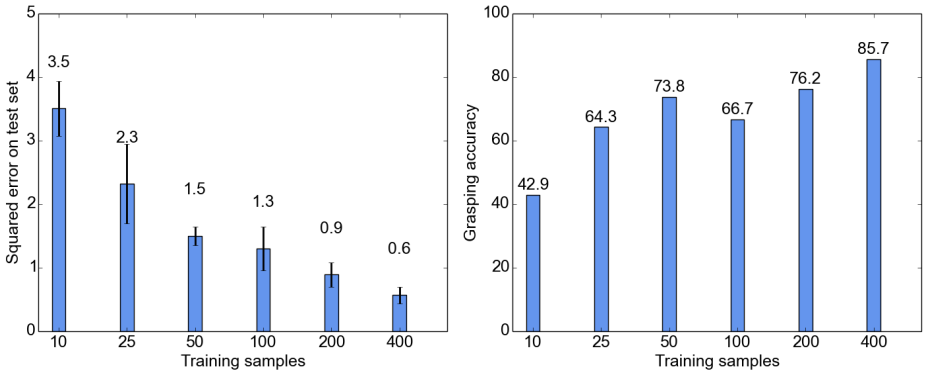

Fig. 3. Results of training with different numbers of samples. Left: Squared error of learned joint configurations from ten trails per experimental condition. Right: Percentage of successful grasps from 42 grasping tasks using the neural model that performed best on its test set.

As different joints contribute differently to the grasping action, it is difficult to predict grasping success of the neural models from the squared error. Therefore, we performed standardized grasping experiments to evaluate the best model

for each training condition. We subtly marked 42 positions in a six by seven grid on the table in front of the robot and manually placed the object in these positions for the robot to grasp. The results are shown in Figure 3 (right). After 400 samples, the robot is reliably grasping with over 85% success rate. The most common error was that the robot's hand was slightly too low, pushing the object away instead of grasping it.

4.3 Autonomous Recovery during Self-Learning

The results indicate that a low number of training samples are sufficient to achieve a reasonable grasp chance. We adapted the idea of multi-staged learning [14] to automate the self-learning phase further. In the case of a failed grasp, the already collected samples are used to train the neural architecture which then controls up to ten further grasp attempts before human assistance is requested.

We repeated the collection of 500 grasping samples. Slightly exceeding the result reported above, the average number of continuous grasping trials was 45.5. Ten grasping attempts failed, mostly due to displacing the object while removing the hand. Except for the first two failed grasps, the robot could recover from the error state. On average, 1.6 grasping attempts using the trained neural architecture were needed.

5 Conclusion

The presented approach extends the state of the art by facilitating self-learning of robotic visuomotor skills for grasping without the need for annotated training data or information about the kinematics of the robot. In contrast to trial and error methods that can take hundreds of training hours, the presented approach continuously improves its performance in a time span of several minutes to a few hours. Human assistance is only needed for 30 seconds of initial motor training and occasionally during the self-learning phase. The self-learning phase does not require monitoring, as there is no danger of harmful collisions. The robot will request help if needed.[2]

Our approach enables a domestic developmental robot to learn to grasp in a modest amount of time. We have shown how complex visuomotor skills can develop through interaction with the environment based on more simple initial motor skills. End-to-end visuomotor learning offers opportunities to research the emergence of spatial representations in hidden layers as well as the benefits of end-to-end learning compared to the separate training of components. In future work, our approach will be extended to multiple objects with different grasping geometries and bimanual grasping. Self-organization will be used to select samples from the initial motor training in a more principled way. Also, we will evaluate how well our self-learning method is suited to generate pre-trained networks for continuous deep reinforcement learning.

[2] Visit nico.knowledge-technology.info for further information and video material.

Acknowledgments. This work was partially funded by the German Research Foundation (DFG) in project Crossmodal Learning (TRR-169) and the Hamburg Landesforschungsförderungsprojekt.

References

1. Cangelosi, A., Schlesinger, M.: Developmental Robotics; from Babies to Robots. MA: MIT Press/Bradford Books, Cambridge (2014)
2. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proc. of Aistats, 9, pp. 249–256 (2010)
3. Hahnloser, R. H., Sarpeshkar, R., Mahowald, M. A., Douglas, R. J., Seung, H. S.: Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. Nature, 405(6789), 947–951 (2000)
4. van Hasselt, H., Guez, A., Silver, D.: Deep Reinforcement Learning with Double Q-Learning. arXiv preprint arXiv:1509.06461, 2015.
5. Kerzel, M., Strahl, E., Magg, S., Navarro-Guerro, N., Heinrich, S., Wermter, S.: NICO - Neuro-Inspired COmpanion: A Developmental Humanoid Robot Platform for Multimodal Interaction (RO-MAN 2017 accepted)
6. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature, 521(7553), 436–444 (2015)
7. Leitner, J., Harding, S., Förster, A., Corke, P.: A Modular software Framework for eyehand coordination in humanoid robots. Frontiers in Robotics and AI, 3 (2016)
8. Levine, S., Finn, C., Darrell, T., Abbeel, P.: End-to-end training of deep visuomotor policies. Journal of Machine Learning Research, 17(39), 1–40 (2016)
9. Lillicrap, T.P., Hunt, J.J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., Wierstra, D.: Continuous control with deep reinforcement learning. arXiv preprint arXiv:1509.02971 (2015)
10. Lungarella, M., Metta, G., Pfeifer, R., Sandini, G.: Developmental robotics: a survey. Connection Science, 15(4), 151–190 (2003)
11. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., Petersen, S., others: Human-level control through deep reinforcement learning. Nature, 518(7540), 529–533 (2015)
12. Oquab, M., Bottou, L., Laptev, I., Sivic, J.: Is object localization for free?-weakly-supervised learning with convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 685-694 (2015)
13. Peng, X. B., Berseth, G., v. d. Panne, M.: Terrain-adaptive locomotion skills using deep reinforcement learning. ACM Transactions on Graphics, 35(4), 81 (2016)
14. Pinto, L., Gupta, A.: Supersizing self-supervision: Learning to grasp from 50k tries and 700 robot hours. In: 2016 IEEE International Conference on Robotics and Automation (ICRA), pp. 3406–3413. IEEE Press (2016)
15. Speck, D., Barros, P., Weber, C., Wermter, S.: Ball localization for robocup soccer using convolutional neural networks. RoboCup Symposium, Leipzig, Germany (2016)
16. Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y.: Overfeat: Integrated recognition, localization and detection using convolutional networks. arXiv preprint arXiv:1312.6229 (2013)
17. Sutskever, I., Martens, J., Dahl, G. E., Hinton, G. E.: On the importance of initialization and momentum in deep learning. In: Proceedings of The 30th International Conference on Machine Learning, pp. 1139-1147 (2013)