

# On Designing a Learning Robot: Improving Morphology for Enhanced Task Performance and Learning

Maks Sorokin<sup>1\*,2</sup>, Chuyuan Fu<sup>1</sup>, Jie Tan<sup>3</sup>, C. Karen Liu<sup>4</sup>,  
Yunfei Bai<sup>5</sup>, Wenlong Lu<sup>1</sup>, Sehoon Ha<sup>2</sup>, Mohi Khansari<sup>1</sup>

**Abstract**—As robots become more prevalent, optimizing their design for better performance and efficiency is becoming increasingly important. However, current robot design practices overlook the impact of perception and design choices on a robot’s learning capabilities. To address this gap, we propose a comprehensive methodology that accounts for the interplay between the robot’s perception, hardware characteristics, and task requirements. Our approach optimizes the robot’s morphology holistically, leading to improved learning and task execution proficiency. To achieve this, we introduce a Morphology-AGnostIc Controller (MAGIC), which helps with the rapid assessment of different robot designs. The MAGIC policy is efficiently trained through a novel PRIVileged Single-stage learning via latent alignMent (PRISM) framework, which also encourages behaviors that are typical of robot onboard observation. Our simulation-based results demonstrate that morphologies optimized holistically improve the robot performance by 15-20% on various manipulation tasks, and require 25x less data to match human-expert made morphology performance. In summary, our work contributes to the growing trend of learning-based approaches in robotics and emphasizes the potential in designing robots that facilitate better learning. The project’s website can be found at [learning-robot.github.io](http://learning-robot.github.io).

## I. INTRODUCTION

Recent advances in hardware and software make autonomous robots more and more important in various environments, from manufacturing and warehouses to health-care and living spaces. Learning has been one of the most promising tools for operating robots in such unstructured environments, enabling them to acquire complex perception and reasoning capabilities. However, the current status quo of designing robots does not account for the impact of learning: rather, many robots are still designed based on human experts’ intuition or hand-crafted heuristics. Therefore, such designs can lead to a sub-optimal performance by causing unexpected visual occlusions. This is where the idea of guiding robot design to improve the robot learning capability comes in, inspired by the evolutionary process.

The evolution of physical attributes through natural selection has played a significant role in the emergence of advanced cognitive abilities among living beings [1]. By embracing the idea of evolution, robots also have the potential to evolve their designs for better real-world performance. However, it is extremely challenging to encapsulate all the perception, control, and hardware design into a single holistic evaluation pipeline due to the complexity of the components. For

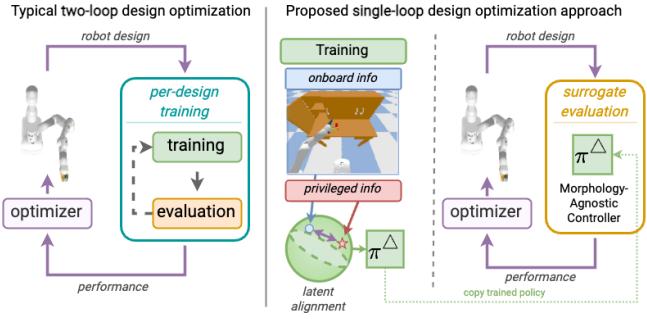


Fig. 1. A side-by-side comparison of existing vs proposed hardware design optimization approaches.

instance, it will be extremely expensive to formulate a typical two-loop design optimization process, which searches robot morphologies in the outer loop and trains a policy for each given design in the inner loop. As such, traditional design optimization techniques focus on enhancing certain attributes in isolation or exploring based on hand-designed heuristics.

This work particularly aims to discuss the design optimization for vision-based manipulation. In the general context of manipulation, visual sensors provide a rich stream of information that allow robots to perform tasks such as grasping, object manipulation, and assembly. The use of visual sensors in manipulation, however, inevitably poses challenges associated with complete or partial field-of-view occlusion, which can degrade performance due to limited perception. Whether an occlusion is harmless or fatal often depends on the task at hand and the stage of task execution. Such scenarios motivate us to explore hardware optimization without neglecting the interplay between the robot’s morphology, onboard perception abilities, and their interaction in different tasks.

We present a learning-oriented morphology optimization framework to improve the initial robot design crafted by a human expert. Our key idea is to develop a Morphology-AGnostIc Controller (MAGIC) that is capable of controlling a range of morphologies using onboard visual observations, greatly reducing the costs of traditional two-loop design optimization. The controller is trained using a novel PRIVileged Single-stage learning via latent alignMent (PRISM) formulation, which unifies the typical two-stage approach of teacher and student training [2]. In our experience, this novel formulation is essential for the student policy to learn behaviors characteristic of its own observation capability and not of a privileged agent. Once the morphology-agnostic controller is learned, we optimize robot design parameters

\* Correspondence to: [maks@gatech.edu](mailto:maks@gatech.edu)

<sup>1</sup>The research was conducted during Residency at Everyday Robots.

<sup>2</sup>Everyday Robots,<sup>2</sup> Georgia Institute of Technology,<sup>3</sup>Robotics at Google

<sup>4</sup>Stanford University, <sup>5</sup>Work done while at Everyday Robots

using Vizier [3] optimizer using the controller’s performance as a surrogate measure (Fig. 1). We leverage simulation throughout this work to accelerate the process of looking for an optimal configuration without actually building a physical robot, which is both costly and time-consuming.

We evaluate the proposed technique for optimizing the morphology of a mobile manipulator. We find that our framework can find an improved morphology that shows better performance on overall tasks and facilitates a more sample efficient learning. Specifically, an optimized design leads to a 15-20% success rate improvement on various manipulation tasks and is 25x more data efficient when trained from scratch. With this work, we would like to highlight the untapped potential of learning-based robot design optimization and show how robot designs can be tailored for better performance and learning with onboard sensing.

## II. RELATED WORK

### A. Morphology Optimization

Robot’s physical design is an important factor in its performance and ability to complete assigned tasks. Traditionally, robot designs are made by human experts who rely on heuristics to optimize the structure. A number of approaches have been developed to algorithmically assist designers in making better physical structures [4], [5], improving reachability [6], [7], [8], and observability of the workspace [9], [10], [11]. However, heuristic approaches don’t offer any insight on the design’s performance on actual tasks. This is particularly true for tasks that involve vision and learning, where the robot’s physical structure can significantly impact learning efficiency and performance.

Recently, researchers have aimed to address the limitations of robot designs by optimizing their physical structures in an informed and systematic way. A number of works have proposed computational approaches for co-optimizing the design parameters and motion trajectories of robotic systems [12], [13]. Typically these works model the relationship between form and function as solutions to an optimal control problem, often showing that such frameworks are effective in optimizing robotic design for a variety of tasks. Recently, a number of methods were proposed for the joint optimization of physical structures using policy optimization methods [14], [15], [16]. While these approaches bring us closer to optimizing the designs for performance on the actual tasks, so far they tackle cases with proprioceptive sensor information and do not take into account vision, which is often critical in deep learning approaches [17], [18]. In this work, we incorporate exteroceptive information obtained from the onboard camera sensor to account for an interplay between the design and the robot’s ability to sense task-relevant information.

### B. Vision-based Robot Learning

Visual sensing plays a crucial role in the field of robotics, enabling the robots’ perception and understanding of their surroundings. The field has come a long way in terms of explicitly understanding the world through pixels [19], [20], [21] and 3D [22], [23], [24], as well as, implicitly

through learned features that enhance downstream task performance [25], [26].

Utilizing the advances of learning, the field of robotics has been booming with pipelines that leverage Behavioral Cloning [27] and Reinforcement Learning [28] methods. Many systems that are designed to be deployed autonomously in the real-world leverage both exteroceptive sensing and learning. Notably, a number of projects that were deployed in the real-world leveraged learning and vision, some examples of such systems include robot locomotion [29], [30], [31], navigation [32], [33], [34], and manipulation [35], [36], [37], [38]. These works focus on improvements to the training pipelines and algorithms while keeping the robot characteristics intact. In this work, we leverage findings from vision-based robotics research and use manipulation problems as a testing ground for design optimization.

### C. Surrogate Evaluation

A surrogate function is a model that approximates the behavior of the true function that is costly to produce [39]. In the context of robot design optimization, a true function would measure the quality of the design through a full cycle of data collection, training, and evaluation. Alas, with a large number of designs to rate, it is infeasible to go through the whole process from collection to evaluation for each design. Thus, we explore the idea of a morphology-agnostic controller [40], [16], [41] which is capable of controlling a range of morphologies. Zhao et al. [16] used a model predictive control to evolve terrain traversal creatures. Such controllers can be used as a *proxy* to evaluate the performance of actual robot designs.

### D. Universal Policy

A universal policy can be defined as a controller which is able to operate in different environments while performing various tasks. In the context of learning, universal controllers are made possible through the exposure of the controller to a diverse set of environments during training [42]. Yu et al. [40] showed that it is possible to effectively adapt the controller to unknown task environments. Gupta et al. [41] leveraged learning and demonstrated cross-morphology policy transfer in the context of locomotion. In RT-1 [18], it was recently shown how a large-scale training could enable effective multi-task learning in different environments and help with cross-robot transfer. These approaches provide a strong foundation to enable generalization to unseen environments and tasks. In this work, we propose a recipe to enable the training of a universal controller that works across different morphologies while using the robot’s *onboard* sensing.

To better guide the controller during training, we leverage ideas of privileged learning [2] to bootstrap the training. A privileged pipeline enables a more capable and sample efficient policy training that can control different morphologies. Overall, using the morphology-agnostic controller as a surrogate function allows us to efficiently explore the design candidate space and find an improved morphology design.

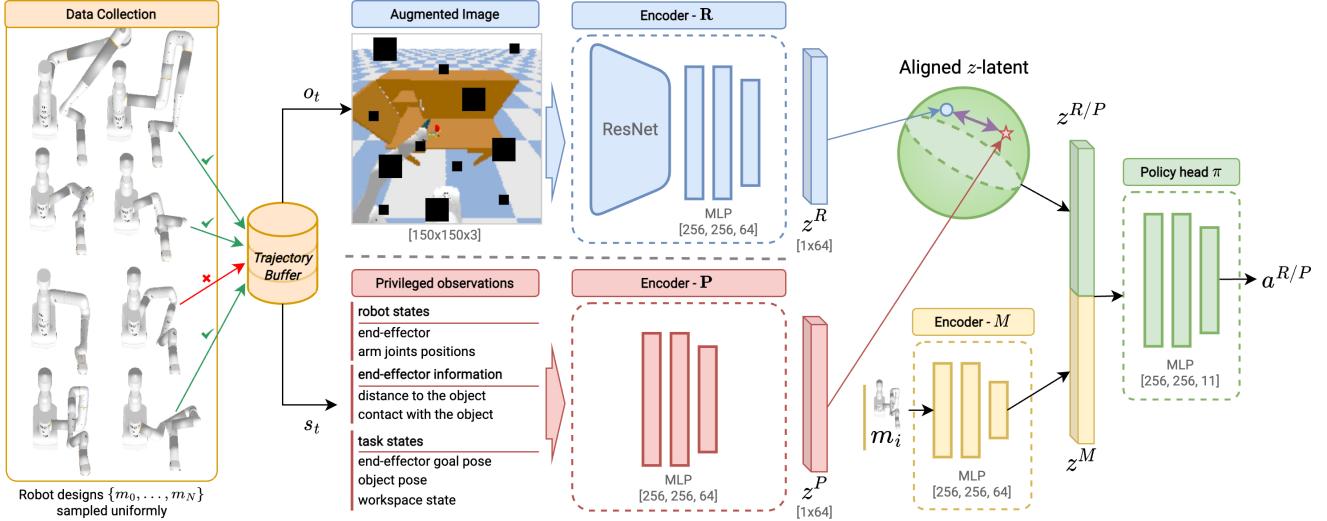


Fig. 2. **Overview of policy  $\pi^\Delta$  a Morphology-Agnostic Controller trained via PRIVileged Single-stage learning via latent alignMent (PRISM)**  
*Data Collection* - demonstrates a few robot designs randomly generated during the collection process. Robot trajectories are collected using motion planning and only successful trajectories are stored in the Trajectory Buffer. *Encoder - R* processes information obtained from onboard sensors (e.g., cameras) to generate embedding  $z^R$ . *Encoder - P* extracts  $z^P$  from privileged state which contains a comprehensive information about the robot and workspace.  $z^P$  and  $z^R$  are explicitly aligned through a loss function to produce mutual information. Then, *policy head π* generates onboard and privileged actions  $a^P$ / $a^R$  using morphology embedding  $z^M$ , and  $z^R/z^P$ . As such onboard ( $\pi_P^\Delta \leftarrow \pi(z^P, z^M)$ ) and privileged ( $\pi_R^\Delta \leftarrow \pi(z^R, z^M)$ ) policies are trained jointly via latent alignment and share *policy head π*. For details refer to Section III.

### III. MORPHOLOGY-AGNOSTIC CONTROLLER

#### A. Overview

We present a Morphology-AGnostIc Controller (MAGIC) capable of operating across a variety of environments, tasks, and robot morphologies. The controller is a neural network policy  $\pi^\Delta$  trained to control a wide range of robot morphologies using the robot's onboard camera sensing. To achieve this challenging goal, we introduce a privileged single-stage learning framework (detailed in Section III-C) to bootstrap the behavioral learning from demonstrations. The data is collected across a spectrum of robot designs (detailed in Section V) to make the policy compatible with morphological variations. The  $\pi^\Delta$  training pipeline and policy architecture are summarized in Fig. 2. The  $\pi^\Delta$  will be later used as a surrogate function to rapidly assess the quality of different morphologies in Section IV. Below, we introduce the policy structure and training protocols used for obtaining such a policy.

#### B. States and Observation

We train our policy using Behavior Cloning, which has been proven effective for a variety of manipulation tasks [37], [38], [18]. In this paper, we use the following notation to describe the states and observations: *onboard* observations  $o_t$ , *privileged* states  $s_t$ , *morphology* configuration  $m_i$ . *Onboard* observations consist of sensor readings which are available during robot deployment. *Privileged* states provide a comprehensive representation that is only used to bootstrap the learning efficiency of the policy. These states do not suffer from sensor limitations or occlusions, notably, they are unobtainable during the policy deployment. *Morphology* configuration is used to inform the policy about the robot design parameters.

States and observations used in our tasks of robot manipulation are the following:

- $I \in R^{[150 \times 150 \times 3]}$  (onboard) - image observations rendered from the robot's onboard camera.
- $J \in R^8$  (privileged) - positions of the robot's arm joints.
- $E_{curr/goal} \in R^{11}$  (privileged) - current and goal states of the robot's end-effector, including Cartesian coordinates, quaternion orientation, and finger positions.
- $B \in R^7$  (privileged) - position and orientation of the object being manipulated during the task.
- $F \in R^2$  (privileged) - an indicator of contact between the fingers and the object, and the distance between them.
- $W \in R^2$  (privileged) - openness state of the cabinet, used to track the progress of the closing task.
- $m \in R^7$  (morphology) - configuration parameters that define the robot's physical structure and configuration.

These observations and states provide a comprehensive understanding of the robot's environment and actions, which are used to train the  $\pi^\Delta$ .

The policy output controls the displacement of robot end-effector position, orientation, and two-fingers.

#### C. Privileged Single-stage Learning via Latent Alignment

In this section, we describe the PRIVileged Single-stage learning via latent alignMent (PRISM) used for training of  $\pi^\Delta$ . Privileged learning has been shown to be effective for a number of applications [2], [34], [30], [31]. Generally in this paradigm, two policies are trained sequentially, a student policy with limited observations, and a privileged policy with full observations. The privileged policy's sole role is to guide the student during the transfer process, leading to improved learning efficiency and improved behavior due to a stronger learning signal.

We propose a novel training procedure, which unifies the traditional two-stage approach into one stage by using a latent space alignment loss during the optimization process. This unification allows us to discourage the student policy from learning *supernatural* behaviors which incorrectly exploit the information only present in privileged states, which is important to learn realistic vision-based policies. To make this approach possible, we use a combination of loss functions to train the policy network. We use Behavioral Cloning (BC) for action and Latent Alignment loss for the alignment of information in the encoder latent space. We utilize this framework for training  $\pi^\Delta$  that is agnostic of the robot morphology in both observation and control space.

**1) Network Architecture:** Our network architecture consists of four main components: image encoder  $R$ , a privileged encoder  $P$ , morphology encoder  $M$ , and the policy head  $\pi$ . Encoders produce unit-vector embeddings  $\mathbf{z}^R$ ,  $\mathbf{z}^P$ ,  $\mathbf{z}^M$  processing the respective inputs:  $\mathbf{o}_t^*$ ,  $\mathbf{s}_t$ , and  $\mathbf{m}_t$ . Onboard policy action  $\mathbf{a}^R$  and privileged policy action  $\mathbf{a}^P$  are produced by querying policy head with corresponding inputs:  $\pi(\mathbf{z}^R, \mathbf{z}^M)$  and  $\pi(\mathbf{z}^P, \mathbf{z}^M)$ . For the rest of the paper, we refer to onboard policy  $\pi(\mathbf{z}^R, \mathbf{z}^M)$  as  $\pi_R^\Delta$ , and to privileged policy  $\pi(\mathbf{z}^P, \mathbf{z}^M)$  as  $\pi_P^\Delta$ . For a detailed overview of the architecture refer to Fig. 2.

**2) Losses:** We use a combination of three loss function terms for training the policy network: two behavioral cloning terms for actions and one latent alignment term for the alignment of information in the encoder latent space.

*Behavioral Cloning loss.*  $L_{BC}$  consists of Mean Squared Error (MSE) for the Cartesian ( $xyz$ ) and finger ( $f$ ) actions, and the unsigned relative rotation angle for the quaternion orientation ( $q$ ) action:

$$L_{BC}(a, \hat{a}) = MSE(a_{xyz/f}, \hat{a}_{xyz/f}) + 2 \arccos(\hat{a}_q^T a_q).$$

$L_{BC}$  is calculated for both actions produced from privileged and onboard information encoder latents, against the demonstrated action  $\hat{a}$ .

*Latent Alignment loss.*  $L_{align}$  is used to align the unit vector outputs of the privileged and onboard information encoders. We empirically find that Huber loss [43] works better than MSE loss for the alignment on our tasks.

*Total loss.* Overall, single-stage privileged training loss looks as follows:

$$L_{total} = L_{BC}(\mathbf{a}^R, \hat{\mathbf{a}}) + L_{align}(\mathbf{z}^R, \mathbf{z}^P) + L_{BC}(\mathbf{a}^P, \hat{\mathbf{a}}).$$

*Training regimes.* The combination of all loss terms simulates a single-stage privileged co-training regime, where:

- 1)  $L_{BC}(\mathbf{a}^P, \hat{\mathbf{a}})$  trains the privileged  $\pi_P^\Delta$  (a.k.a. *Stage I*)
- 2)  $L_{align}(\mathbf{z}^R, \mathbf{z}^P)$  emulates the transfer (a.k.a. *Stage II*)
- 3)  $L_{BC}(\mathbf{a}^R, \hat{\mathbf{a}})$  adapts the behavior of  $\pi_R^\Delta$  for  $\mathbf{o}_t$ .

Through hyper-parameter search, we find that weighting all of the loss terms equally leads to the most efficient training regime of the onboard policy  $\pi_R^\Delta$ .

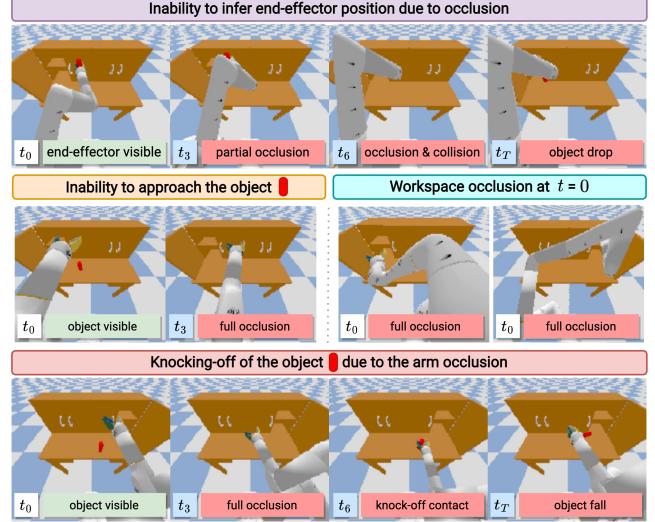


Fig. 3. **Visual Occlusion Examples** - a selected set of naturally occurring object and workspace occlusions due to the pose and configuration of the robot.

We also explore the variations of  $L_{align}(\text{sg}(\mathbf{z}^R), \mathbf{z}^P)$  and  $L_{align}(\mathbf{z}^R, \text{sg}(\mathbf{z}^P))$ , to only allow for alignment in one direction by stopping the gradients ( $\text{sg}()$ ), however, we find no significant difference in the performance of  $\pi_R^\Delta$ .

**3) Regularization:** To regularize the network and make it robust to occlusions from different robot morphologies, we use image augmentation during the training process. We use the *imgaug* library [44] to augment camera images during training. For each image, we apply either *Cutout* (random patch dropout) or *CoarseDropout* (smaller but more frequent patch dropout) augmentation with a 50% chance. We find these augmentations to be of high importance for training all our policies.

**4) Training Parameters:** For all experiments in this paper we use the Adam optimizer with a learning rate of  $3e-4$ . The total training time of the MAGIC policy  $\pi^\Delta$  takes five days on a single V100 GPU.

#### IV. HARDWARE OPTIMIZATION

The goal of the hardware optimization is to find a robot morphology that is best fitted for the tasks and workspaces considered. Unlike typical two-loop morphology optimization frameworks, our key idea is to leverage the morphology-agnostic policy  $\pi_R^\Delta$  trained only once. Note that, policy  $\pi_R^\Delta$  through learning is *primed* to disregard occlusions from onboard observation to generate suitable actions. Hence, if the  $\pi_R^\Delta$  is successful, then the information is sufficient, meaning that the morphology does not cause any major occlusions that prevent the task success (Fig. 3 shows common types of occlusions). Such an approach provides a more accurate evaluation of observability and reachability in terms of “learnability”, and it is substantially more feasible than per-design training.

\*For reaching  $\mathbf{o}_t$  consists on  $E^{goal}$  in addition to  $I$

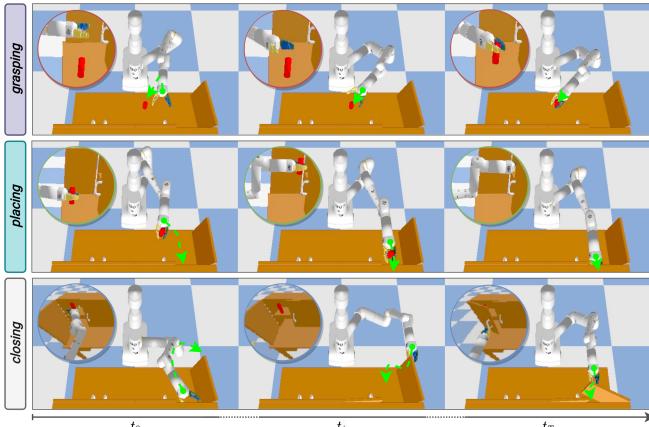


Fig. 4. **Manipulation Task** - Robot motion key-frame visualization for object grasping, object placing, and cabinet closing sub-tasks.

### A. Objective Function

The objective function for the optimization is the success rate of the morphology-agnostic policy  $\pi_R^\Delta$  on the task of reaching and/or manipulation. The success rate of  $\pi_R^\Delta$  serves as a *surrogate* measure of the robot morphology quality. Each morphology is evaluated 500 times on a seeded set of reaching tasks, and 200 times on each of the 3 manipulation tasks.

### B. Optimization Algorithm

We use a black-box optimization algorithm, Vizier [3], to optimize robot design parameters. Vizier is a sampling-based optimization algorithm that allows us to efficiently search for the optimal solution among a large number of possible morphologies. We evaluate 50 morphologies in parallel capping the total number of evaluations at 1000.

### C. Optimization Space

As an initial design configuration, we use Everyday Robots' manipulator robot [37], [17], [18], which features a seven degree-of-freedom arm and a two-finger gripper. We use this design as a baseline design and refer to it as a *human-expert* design in the upcoming sections.

The optimizer is configured to propose link length deltas to the initial robot configuration. We allow for modification of the following robot links for respective ranges (in meters): *torso* ( $-0.3, 0.3$ ), *shoulder* ( $0.0, 0.3$ ), *bicep* ( $-0.05, 0.4$ ), *elbow* ( $-0.05, 0.3$ ), *forearm* ( $-0.2, 0.3$ ), *wrist* ( $0.0, 0.2$ ), *gripper* ( $-0.05, 0.2$ ). We show a few random robot morphology configurations on the left of Fig. 2.

## V. EXPERIMENTAL SETUP

In this section, we describe the tasks, environment, and data collection process used for training the policy  $\pi^\Delta$  and optimization of the robot morphology. We also describe the training process of the targeted policy, which serves as a true performance evaluator for a given robot morphology.

### A. Tasks

We use two tasks for training the policy: reaching and manipulation. The reaching task involves moving the end-effector to a specific goal position, while the manipulation task involves picking an object, placing it in a cabinet, and closing the cabinet door (Fig. 4).

**1) Reaching task:** The reaching task involves moving the robot's end-effector to a specific goal position within the workspace. The goal position is randomly sampled at the beginning of each trial, and the robot's joint positions are also randomly placed in a different initial configuration. The robot's end-effector is required to finish in a fixed orientation used across all trials. The task is considered successful if the end-effector is within 3cm and 10 degrees from the goal pose for more than 20 time steps. This task is used to evaluate the robot's ability to reach specific points within the workspace and helps test the onboard policies robot's arm/end-effector pose inference capabilities.

**2) Manipulation task:** The manipulation task is a harder task, which is broken down into three sub-tasks: picking, placing, and closing. The goal of this task is to evaluate the robot's ability to perform more complex manipulation actions, such as picking, manipulating objects, and interacting with the environment. Unlike reaching, this task additionally requires that the network infers the object position and the cabinet states.

*Picking.* For this task, the robot is required to pick up an object off the table surface. The position of the object is randomized at the beginning of each trial, and the robot arm is reset into a randomized location in near proximity to the object. The task is considered successful if the object is raised 20 cm above the table for more than 20 time steps.

*Placing.* The robot must place the object in one of two cabinets. One of the cabinets is randomly selected as the goal at the beginning of each trial with its door being set as open. The robot arm is initialized in the object-holding pose, with the goal to bring and release that object inside the target cabinet.

*Closing.* The robot must close the door of the cabinet that the object was placed in. The goal is to retract the arm from the post object placing pose and fully close the door of the cabinet. The task is considered successful if the door is closed with a threshold of five degrees.

### B. Robot Data Collection

Next, we describe the robot control and data collection procedures used for training the policy  $\pi^\Delta$  and optimizing the robot morphology. We collect trajectories using motion planning, which uses ground-truth simulation information, for various robot designs. During training (Section III),  $\pi_R^\Delta$  is trained to produce these motions using only onboard observations.

**1) Robot Control:** For our tasks, we use a robot manipulator with a seven degrees of freedom arm and a fixed base. The robot arm is actuated through end-effector displacement control and Inverse Kinematics (IK) with collision avoidance. During data collection, we use a set of predefined end-effector poses, which, IK guides the robot arm through avoiding any collisions.

**2) Data Collection:** The data collection process involves initializing a morphology with uniformly sampled link lengths and collecting synthetic trajectories using the steps outlined in Algorithm 1.

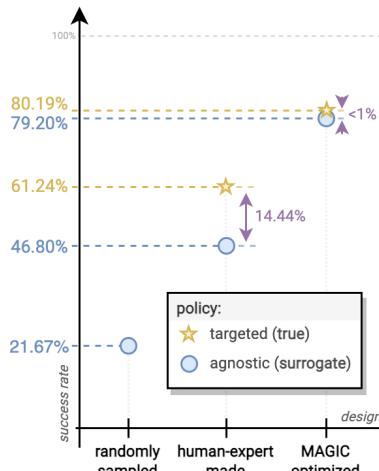


Fig. 5. **Targeted vs. Morphology-Agnostic Policies.** Policies  $\pi_R^\Delta$  and  $\pi_R^T$  have a 14.44% performance gap on the *human-expert* design, which shrinks to 1% on the *MAGIC-optimized* design. We omit the targeted performance evaluation on random morphologies due to computational cost.

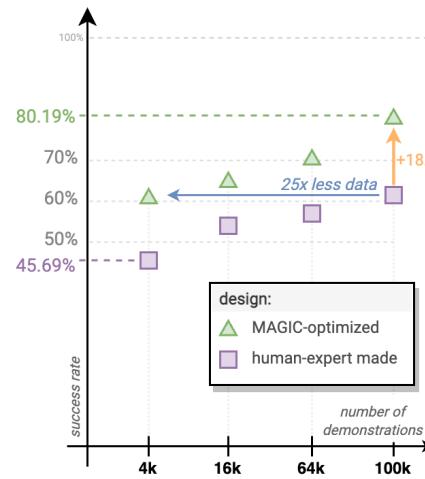


Fig. 6. **Robot Performance & Learning Efficiency.** (a) *Learning oriented* morphology design outperforms the *human-expert* design by 18.95% when trained with the same amount of data. (b) *Learning oriented* design can achieve the same performance as *human-expert* design using 25x less data.

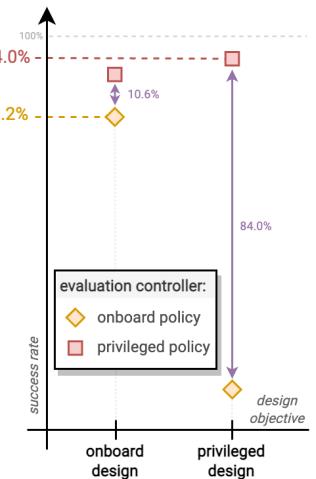


Fig. 7. **Onboard and privileged design objective comparison.** Morphology optimized using privileged objective has a major performance drop when evaluated using onboard information, emphasizing the importance of evaluation with onboard sensing.

### Algorithm 1 Cross-morphology data collection

```

1: Initialize buffer  $\mathcal{B} \leftarrow \{\}$ 
2: while  $\mathcal{B}.size() < N$  do
3:   Sample  $m_i \sim \mathcal{M}$   $\triangleright$  Uniformly Sample Morphology
4:   workspace.reset()  $\triangleright$  Randomize Task/Workspace
5:   plan  $\leftarrow$  planner(workspace,  $m_i$ )  $\triangleright$  Plan the Motion
6:    $s_{0...T} \leftarrow$  execute(plan, workspace,  $m_i$ )  $\triangleright$  Execute
7:   if  $s_T.is\_success$  then  $\triangleright$  Filter out failures
8:     Store  $s_{0...T}$  in  $\mathcal{B}$ 
9: Return  $\mathcal{B}$ 

```

All of the sub-tasks are timed, and the trajectory lengths are capped at 200 time steps ( $\sim 10$  seconds). The data is collected using a planner which follows through a designated set of end-effector poses. We use PyBullet [45] simulator to collect the demonstration data. The data collection process is conducted in parallel by hundreds of workers to efficiently collect a large amount of data. In total, we collect  $N = 500k$  successful trajectories for training the policy  $\pi^\Delta$ .

### C. Targeted Policy

A targeted policy  $\pi_R^T$  is a controller designed to control a specific robot, unlike a MAGIC policy  $\pi^\Delta$ , which is designed to perform well across multiple robots. We use targeted policy as an ablation to evaluate how much the performance of our proposed MAGIC policy  $\pi^\Delta$  is compromised to accommodate for multi-morphology capability.

To evaluate the true performance without multi-morphology compromise, we train the *targeted* policy  $\pi_R^T$  using the procedure described in Section III but with a fixed robot design. For a particular robot, we collect 100k successful targeted demonstrations while keeping morphology  $m_i$  intact (skip Algorithm 1:line 3) during the data collection process. We use targeted demonstration to train the targeted policy  $\pi^T$  from scratch until convergence.

## VI. EXPERIMENTS

The results of our experimentation are presented in this section and structured to answer the following questions:

- Does a morphology-agnostic policy provide a good surrogate of a true performance?
- How does the *MAGIC-optimized* design compare to a *human-expert* design?
- How does incorporating onboard sensing affects the morphology optimization process?
- What effects do additional robot tasks have on the optimized robot morphology?

### A. Agnostic vs Targeted Policies

First, we show how a surrogate policy performance compares to a true performance on a set of selected robot designs (see Fig. 5). We report the performance of the morphology-agnostic controller  $\pi_R^\Delta$  (Section. III) and targeted controller  $\pi_R^T$  (Section. V-C) on a *human-expert* robot design  $m^H$  and the best *MAGIC-optimized* solution candidate  $m^*$  (described in Section IV).

We find that the reaching task performance of the  $\pi_R^\Delta$  and  $\pi_R^T$  is nearly identical for  $m^*$  (difference:  $< 1\%$ ). On the other hand, the performance on  $m^H$  design has a gap of 14.44%. We hypothesize that  $m^*$  causes less occlusion and hence is easier to control using  $\pi_R^\Delta$  compared to  $m^H$ . However, policy  $\pi_R^T$  can learn to exploit the structure of occluding arm to infer the missing information, such as a rough pose of an end-effector, through targeted re-training.

### B. Comparison to a human-expert design

Next, we analyze the difference between *MAGIC-optimized*  $m^*$  and *human-expert*  $m^H$  designs in terms of task performance and learning efficiency (Fig. 6). We highlight the absolute performance improvement of *targeted*

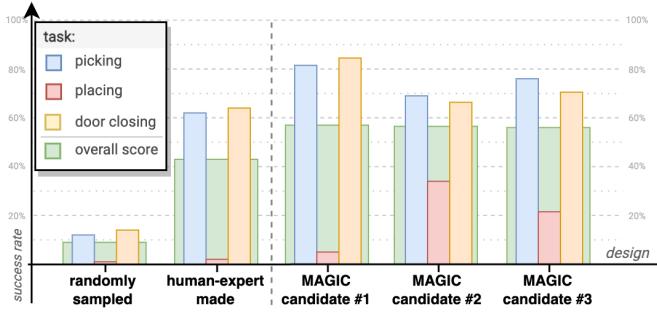


Fig. 8. **Various design performance comparison on the manipulation task.** Overall *MAGIC*-optimized top solution candidates perform  $\sim 15\%$  better than *human-expert* design. Different optimization candidates show varying levels of performance at different sub-tasks, allowing a human to make a final decision about the performance trade-offs.

policies when evaluated on  $m^*$  compared to  $m^H$  on the task of reaching. Specifically,  $m^*$  achieves success rate of 80.19% compared to 61.24% with  $m^H$ .

The performance gain on  $m^*$  could be attributed to a reduced frequency of sensor occlusions. We hypothesize that the reduced distortion of onboard information can also result in a higher quality of data during collection, which naturally leads to improved robot learning capabilities. To measure the learning efficiency, we compare the success rates of the policies trained with a varying number of demonstrations. Fig. 6 shows that  $m^*$  is a better-suited robot for learning compared to  $m^H$ , as it requires 25x less data than  $m^H$  to reach the same performance when training the controller from scratch.

### C. Significance of onboard sensing

Next, we seek to investigate the significance of onboard sensing during design optimization. If the robot policy is given access to all of the information present in the environment, a robot may be able to reach its upper-bound performance, which is mostly constrained by kinematic and dynamic capabilities. However, it might perform suboptimally when tested with onboard sensing because the morphology can limit its sensing capabilities via visual occlusions.

We use a privileged controller  $\pi_P^\Delta$  during the morphology optimization phase, which acts as an optimal motion planner with the ground-truth robot and task information. Once the morphology is found, we evaluate its performance with the policy which relies on onboard information  $\pi_R^\Delta$ . In Fig. 7, we compare the performance of privileged information morphology and onboard sensing morphology  $m^*$ . We observe that if we only use the privileged policy during the morphology design, there is a significant 84% drop in performance when the robot is evaluated with onboard sensing. In contrast,  $m^*$  performs well in both onboard and privileged settings. This result demonstrates the significance of using onboard sensing during the robot design process.

### D. Multitask-based regularization

Finally, we investigate the direction of task-based regularization, through the exposure of the robot to a wider set of manipulation tasks. There exists a large number of

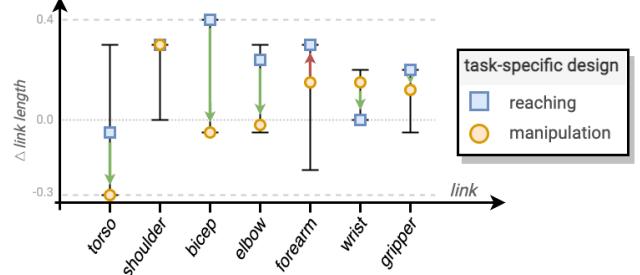


Fig. 9. **Task Complexity Effects on Robot Design** - Optimized link lengths of the robot morphologies tend to converge to more plausible solutions when optimized on more complex tasks.

regularization approaches applied during the optimization to improve the practicality of the design such as energy, or material penalties [4], [5]. However, we intend to avoid an explicit regularization that can directly impact the final morphology. Instead, we seek implicit regularization via learning on multiple tasks.

We repeat data collection, training, and optimization procedures with manipulation tasks and report the performance of multiple solution candidates in Fig. 8. In addition to seeing similar trends in performance improvements of *MAGIC*-optimized design  $m^*$  in manipulation similar to reaching task, we also observe overall improvements in the robot design solutions. Fig. 9 compares two robot morphologies: one optimized for the reaching task, and one that is optimized for the manipulation task. Not surprisingly, the type and complexity of the tasks being considered could significantly impact the optimal morphology design. Hence, including multiple tasks during the optimization process can lead to a more regularized morphology, with shorter and more manageable links that are likely easier to manufacture.

## VII. CONCLUSION

In this work, we explore the potential of optimizing robot morphology for improved learning and demonstrate that it can be achieved through a holistic task-oriented optimization process. We propose a cost-effective method to improve the robot hardware by training a morphology-agnostic surrogate controller and demonstrate that a robot designed with learning considerations can excel at learning compared to a human-expert design. We introduce a single-stage privileged learning framework that enables rapid acquisition of an onboard policy without artifacts on two-stage privileged learning transfer. Through our experiments, we show that robot designs with enhanced learning capabilities can improve performance by 15% on manipulation and by 20% on complex manipulation tasks compared to a human-expert robot design. Moreover, an optimized robot reaches a human-expert design performance level with 25x less demonstration data. We hope this work contributes to the growing trend of learning-based robots and sheds light on opportunities in hardware designs that facilitate better learning.

## VIII. FUTURE WORK

While using simulation was necessary for making it tractable to evaluate hundreds of different morphologies, a natural next step is to build the final optimized robot design in the real world and test its performance against the simulation results. Another extension of our work is to include sensor placement during the optimization as well as evaluating on more complex task distributions. Investigations in those directions could help us discover some unconventional designs that are overlooked by human engineers.

## REFERENCES

- [1] A. Manning and M. S. Dawkins, *An introduction to animal behaviour*. Cambridge University Press, 1998.
- [2] D. Chen, B. Zhou, V. Koltun, and P. Krähenbühl, “Learning by cheating,” in *Conference on Robot Learning*. PMLR, 2020.
- [3] D. Golovin, B. Solnik, S. Moitra, G. Kochanski, J. Karro, and D. Sculley, “Google vizier: A service for black-box optimization,” in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, August 13 - 17, 2017*. ACM, 2017, pp. 1487–1495. [Online]. Available: <https://doi.org/10.1145/3097983.3098043>
- [4] C. A. C. Coello, A. D. Christiansen, and A. H. Aguirre, “Using a new ga-based multiobjective optimization technique for the design of robot arms,” *Robotica*, vol. 16, no. 4, pp. 401–414, 1998.
- [5] S. A. Kouritem, M. I. Abouheaf, N. Nahas, and M. Hassan, “A multi-objective optimization design of industrial robot arms,” *Alexandria Engineering Journal*, vol. 61, no. 12, pp. 12 847–12 867, 2022.
- [6] H. Seraji, “Reachability analysis for base placement in mobile manipulators,” *Journal of Robotic Systems*, 1995.
- [7] N. Vahrenkamp, D. Berenson, T. Asfour, J. Kuffner, and R. Dillmann, “Humanoid motion planning for dual-arm manipulation and re-grasping tasks,” in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2009, pp. 2464–2470.
- [8] A. Makhal and A. K. Goins, “Reuleaux: Robot base placement by reachability analysis,” in *2018 Second IEEE International Conference on Robotic Computing (IRC)*. IEEE, 2018, pp. 137–142.
- [9] B. Triggs and C. Laugier, “Automatic camera placement for robot vision tasks,” in *Proceedings of 1995 Ieee International Conference on Robotics and Automation*, vol. 2. IEEE, 1995, pp. 1732–1737.
- [10] S. Chen and Y. Li, “Automatic sensor placement for model-based robot vision,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 34, no. 1, pp. 393–408, 2004.
- [11] S. Nikolaidis, R. Ueda, A. Hayashi, and T. Arai, “Optimal camera placement considering mobile robot trajectory,” in *2008 IEEE International Conference on Robotics and Biomimetics*. IEEE, 2009.
- [12] S. Ha, S. Coros, A. Alspach, J. Kim, and K. Yamane, “Computational co-optimization of design parameters and motion trajectories for robotic systems,” *The International Journal of Robotics Research*, vol. 37, no. 13–14, pp. 1521–1536, 2018.
- [13] S. Ha, S. Coros, A. Alspach, J. M. Bern, J. Kim, and K. Yamane, “Computational design of robotic devices from high-level motion specifications,” *IEEE Transactions on Robotics*, vol. 34, no. 5, 2018.
- [14] K. Wampler and Z. Popović, “Optimal gait and form for animal locomotion,” *ACM Transactions on Graphics*, vol. 28, no. 3, 2009.
- [15] D. Ha, “Reinforcement learning for improving agent design,” *Artificial life*, vol. 25, no. 4, pp. 352–365, 2019.
- [16] A. Zhao, J. Xu, M. Konaković Luković, J. Hughes, A. Speilberg, D. Rus, and W. Matusik, “Robogrammar: Graph grammar for terrain-optimized robot design,” *ACM Transactions on Graphics (TOG)*, 2020.
- [17] M. Ahn and et.al., “Do as i can and not as i say: Grounding language in robotic affordances,” in *arXiv preprint arXiv:2204.01691*, 2022.
- [18] A. Brohan and et.al., “Rt-1: Robotics transformer for real-world control at scale,” in *arXiv preprint arXiv:2212.06817*, 2022.
- [19] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [20] K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask r-cnn,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.
- [21] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, “An image is worth 16x16 words: Transformers for image recognition at scale,” *arXiv preprint arXiv:2010.11929*, 2020.
- [22] C. Godard, O. Mac Aodha, M. Firman, and G. J. Brostow, “Digging into self-supervised monocular depth prediction,” *The International Conference on Computer Vision (ICCV)*, October 2019.
- [23] J. Philion and S. Fidler, “Lift, splat, shoot: Encoding images from arbitrary camera rigs by implicitly unprojecting to 3d,” in *Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIV 16*. Springer, 2020, pp. 194–210.
- [24] V. Guizilini, R. Ambrus, S. Pillai, A. Raventos, and A. Gaidon, “3d packing for self-supervised monocular depth estimation,” in *Proceedings of the IEEE/CVF CVPR conference*, 2020.
- [25] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. Avila Pires, Z. Guo, M. Gheshlaghi Azar *et al.*, “Bootstrap your own latent-a new approach to self-supervised learning,” *Advances in neural information processing systems*, vol. 33, pp. 21 271–21 284, 2020.
- [26] R. M. Shah and V. Kumar, “Rrl: Resnet as representation for reinforcement learning,” in *International Conference on Machine Learning*. PMLR, 2021, pp. 9465–9476.
- [27] D. A. Pomerleau, “Alvinn: An autonomous land vehicle in a neural network,” *Advances in neural information processing systems*, 1988.
- [28] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.
- [29] W. Yu, D. Jain, A. Escontrela, A. Iscen, P. Xu, E. Coumans, S. Ha, J. Tan, and T. Zhang, “Visual-locomotion: Learning to walk on complex terrains with vision,” in *CORL*, 2021.
- [30] T. Miki, J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, “Learning robust perceptive locomotion for quadrupedal robots in the wild,” *Science Robotics*, vol. 7, no. 62, p. eabk2822, 2022.
- [31] A. Agarwal, A. Kumar, J. Malik, and D. Pathak, “Legged locomotion in challenging terrains using egocentric vision,” in *6th Annual Conference on Robot Learning*, 2022.
- [32] V. Tolani, S. Bansal, A. Faust, and C. Tomlin, “Visual navigation among humans with optimal control as a supervisor,” *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 2288–2295, 2021.
- [33] D. Hoeller, L. Wellhausen, F. Farshidian, and M. Hutter, “Learning a state representation and navigation in cluttered and dynamic environments,” *IEEE Robotics and Automation Letters*, vol. 6, no. 3, 2021.
- [34] M. Sorokin, J. Tan, C. K. Liu, and S. Ha, “Learning to navigate sidewalks in outdoor environments,” *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 3906–3913, 2022.
- [35] D. Kalashnikov, A. Irpan, P. Pastor, J. Ibarz, A. Herzog, E. Jang, D. Quillen, E. Holly, M. Kalakrishnan, V. Vanhoucke *et al.*, “Qt-opt: Scalable deep reinforcement learning for vision-based robotic manipulation,” *arXiv preprint arXiv:1806.10293*, 2018.
- [36] F. Xia, C. Li, R. Martín-Martín, O. Litany, A. Toshev, and S. Savarese, “Relmogen: Leveraging motion generation in reinforcement learning for mobile manipulation,” *arXiv preprint arXiv:2008.07792*, 2020.
- [37] E. Jang, A. Irpan, M. Khansari, D. Kappler, F. Ebert, C. Lynch, S. Levine, and C. Finn, “Bc-z: Zero-shot task generalization with robotic imitation learning,” in *CORL*. PMLR, 2022.
- [38] M. Khansari, D. Ho, Y. Du, A. Fuentes, M. Bennice, N. Sievers, S. Kirmani, Y. Bai, and E. Jang, “Practical imitation learning in the real world via task consistency loss,” *arXiv preprint :2202.01862*, 2022.
- [39] A. Sobester, A. Forrester, and A. Keane, *Engineering design via surrogate modelling: a practical guide*. John Wiley & Sons, 2008.
- [40] W. Yu, J. Tan, C. K. Liu, and G. Turk, “Preparing for the unknown: Learning a universal policy with online system identification,” in *Proceedings of Robotics: Science and Systems*, 2017.
- [41] A. Gupta, L. Fan, S. Ganguli, and L. Fei-Fei, “Metamorph: Learning universal controllers with transformers,” in *International Conference on Learning Representations*, 2022.
- [42] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, “Domain randomization for transferring deep neural networks from simulation to the real world,” in *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2017.
- [43] P. J. Huber, “Robust estimation of a location parameter,” *Breakthroughs in statistics: Methodology and distribution*, 1992.
- [44] A. B. Jung and et.al., “imgaug,” <https://github.com/aleju/imgaug>.
- [45] E. Coumans and Y. Bai, “Pybullet, a python module for physics simulation for games, robotics and machine learning,” <http://pybullet.org>, 2016–2021.