COMPARING THE EFFICACY OF FINE-TUNING AND META-LEARNING FOR FEW-SHOT POLICY IMITATION

Massimiliano Patacchiola
University of Cambridge
mp2008@cam.ac.uk

Mingfei Sun
University of Manchester
mingfei.sun@manchester.ac.uk

Katja Hofmann
Microsoft Research
katja.hofmann@microsoft.com

Richard E. Turner
University of Cambridge
ret26@cam.ac.uk

ABSTRACT

In this paper we explore few-shot imitation learning for control problems, which involves learning to imitate a target policy by accessing a limited set of offline rollouts. This setting has been relatively under-explored despite its relevance to robotics and control applications. State-of-the-art methods developed to tackle few-shot imitation rely on meta-learning, which is expensive to train as it requires access to a distribution over tasks (rollouts from many target policies and variations of the base environment). Given this limitation we investigate an alternative approach, fine-tuning, a family of methods that pretrain on a single dataset and then fine-tune on unseen domain-specific data. Recent work has shown that fine-tuners outperform meta-learners in few-shot image classification tasks, especially when the data is out-of-domain. Here we evaluate to what extent this is true for control problems, proposing a simple yet effective baseline which relies on two stages: (i) training a base policy online via reinforcement learning (e.g. Soft Actor-Critic) on a single base environment, (ii) fine-tuning the base policy via behavioral cloning on a few offline rollouts of the target policy. Despite its simplicity this baseline is competitive with meta-learning methods on a variety of conditions and is able to imitate target policies trained on unseen variations of the original environment. Importantly, the proposed approach is practical and easy to implement, as it does not need any complex meta-training protocol. As a further contribution, we release an open source dataset called iMuJoCo (iMitation MuJoCo) consisting of 154 variants of popular OpenAI-Gym MuJoCo environments with associated pretrained target policies and rollouts, which can be used by the community to study few-shot imitation learning and offline reinforcement learning.

1 INTRODUCTION

The ability to learn quickly is critical for autonomous agents as it enables them to rapidly adapt to new environments and tasks. For example, in the field of robotics, agents may encounter changes in the environment or control dynamics, leading to a need for a policy update. These changes can occur in contexts such as manipulation (Fu et al., 2016), locomotion (Nagabandi et al., 2018), human-robot interaction (Khamassi et al., 2018), and may be caused by a variety of factors. There are many ways in which changes to the transition dynamics can arise, including faulty sensors, parts of the robot being replaced or upgraded over time, the design of the robot changing, parts of the robot corroding or wearing over time, environmental damage (sand in joints, lubricant running low, etc.). In all these cases, adaptation via imitation could be a possible way to adjust the policy on trajectories obtained via corrective feedback provided by humans or other agents.

Robotics and automation are not the only areas where imitation is important. For example, in multi-player games like DOTA (Berner et al., 2019), StarCraft (Vinyals et al., 2019), and Minecraft (Johnson et al., 2016), adapting a pretrained policy to the style of the team is crucial, and this can be accomplished through imitation. Additionally, in 1-vs-1 gameplay, a similar approach can be used to fine-tune a pretrained policy on the player's trajectories to create a fair opponent calibrated to the user's expertise (Paraschos & Koulouriotis, 2023).

While imitation can be a promising solution for adapting policies, it does have some limitations that need to be considered. One significant restriction is the potential scarcity of samples provided by a target policy, especially when relying on human-generated data. In some cases, obtaining a large number of diverse and informative trajectories can

be challenging. When humans provide the trajectories, the availability of samples may be limited due to factors such as time constraints, human resources, or the complexity of the task.

To address these limitations, researchers have been exploring the use of few-shot learning as a potential solution. Few-shot learning has gained significant attention in recent years, as it offers a promising approach for learning from a small amount of task-specific data. This setting is especially relevant in applications including personalization (Massiceti et al., 2021), object detection (Kang et al., 2019), and molecule design (Stanley et al., 2021). The majority of research has focused on image classification, where unlabeled query images are classified based on a support set of labeled images from the same classes. Historically, two approaches have been employed to tackle the few-shot image classification problem: fine-tuning and meta-learning. Fine-tuning involves taking a pretrained model, such as a convolutional neural network, and adjusting its parameters using a smaller dataset specific to the target task. Fine-tuning allows the model to learn task-specific features while still retaining the high-level features learned during pretraining, leading to improved accuracy and faster convergence on the target task. Meta-learning is based on learning-to-learn paradigms (Hospedales et al., 2020). The underlying idea of meta-learning is to improve the learning process of a model by leveraging previous learning experiences to quickly adapt to new tasks or environments with minimal additional data or training. Recent work has showed that fine-tuners are better than meta-learners in low-data regimes (Kolesnikov et al., 2020; Shysheya et al., 2023), reporting superior classification accuracy on large benchmarks such as Meta-Dataset (Triantafillou et al., 2019) and VTAB (Zhai et al., 2019). Specifically, pretraining a large backbone, such as ResNet50, on a generic dataset, like ImageNet, provides a robust starting point for fine-tuning routines that enable the model to adapt to specialized datasets. These results have led the community to question the effectiveness and utility of meta-learning in few-shot image classification.

In this paper, we take these considerations into account and assess to what extent they apply in the setting of few-shot imitation learning for control problems, which involves learning to imitate a target policy by accessing an offline set of its rollouts. Although it has received less attention compared to few-shot image classification, this setting holds high relevance due to the potential for developing techniques that can enhance policies in diverse applications, including robotics and autonomous driving. We focus on scenarios where only state-action pairs from a given trajectory over the target policy can be recorded and stored. This is the most challenging and practical scenario as it allows for learning from human demonstrations.

Meta-learning has been previously used to tackle few-shot imitation (Finn et al., 2017b). Using meta-learning in this context implies training a meta-policy over a task distribution that includes numerous target policies after which the model is adapted to imitate a new set of policies. Unlike the few-shot image classification case, fine-tuning baselines are difficult to define for few-shot imitation because there is no equivalent of ImageNet that can be used for model pretraining. We believe that this resulted in the underestimation of the potential of fine-tuning methods for few-shot imitation.

In this work, we demonstrate that it is possible to define a fine-tuning method for few-shot imitation by training a base model via reinforcement learning (e.g. Soft Actor-Critic) on a single environment and we then fine-tune the base model on offline rollouts, effectively generalizing to target policies trained on unseen variations of the base environment. One key advantage of our proposed baseline is that it eliminates the need for an expensive meta-training stage that relies on a task distribution (distribution over many target policies and variations of the environment). This advantage is crucial in many applications, since generating variations of the base environment (e.g. using domain randomization) may be difficult or costly in many real-world settings during the training phase. In addition, the use of policies trained via standard RL methods, open the door to the possibility of using open-source models that are publicly available and which can then be fine-tuned on new settings.

The *main contributions* of this paper can be summarized as follows:

1. We shed more light on few-shot imitation learning. This problem has been relatively under-explored despite its relevance to robotics and control applications. In particular, we conduct a comprehensive comparison of different methods across various environments and different numbers of shots, offering a broad perspective on each method's strengths and limitations in these settings.
2. We provide an effective fine-tuning baseline which is based on two stages: (i) pretraining a policy on a single environment via online reinforcement learning and (ii) fine-tuning the policy on offline trajectories from an unseen target policy. We empirically show that this baseline is competitive with meta-learning in the medium and high-shot conditions, achieving similar loss on a large number of tasks.
3. We release an open-source dataset called iMuJoCo (iMitation MuJoCo) which includes 154 environment variants obtained from the popular OpenAI-Gym MuJoCo suite with associated pretrained target policies and trajectories. This dataset can be used to study imitation-learning and offline RL. The dataset is available at: https://github.com/mpatacchiola/imujoco

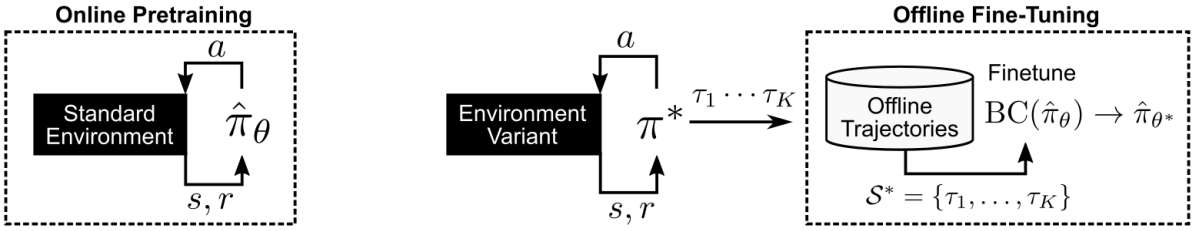

Figure 1: Online pretraining and offline fine-tuning of the proposed baseline. During pretraining (left schematic) a base policy $\hat{\pi}_\theta$ is trained online using standard RL methods (e.g. Soft Actor-Critic). In the adaptation phase (right schematic) we have access to an offline support set of trajectories $\mathcal{S}^* = \{\tau_1, \ldots \tau_K\}$ generated by a target policy π^* trained on a variant of the base environment. Offline fine-tuning exploits the offline support-set $\mathcal{S}^*$ to adjust the parameters of the base policy towards the target via Behavioral Cloning (BC).

2 Description of the Method

2.1 Problem Setting

The objective of our research is to develop a base policy that can quickly imitate a target policy given a few demonstrations. We consider a distribution over environments $p(\mathcal{E})$ where each environment $\mathcal{E}_i \sim p(\mathcal{E})$ is a specific Markov Decision Process (MDP). We make the assumption that the environments in $p(\mathcal{E})$ share a set of task-common parameters, but differ in terms of task-specific parameters. For instance, the environments can have the same transition dynamics but differ with respect to the agent's body structure (e.g. different body mass, limb length, range of motion, etc.). For each environment we have access to trajectories drawn from a target policy π_i that has been trained on that environment, where we define a trajectory as a sequence of state-action tuples

$$\tau \triangleq \{(s_1, a_1), (s_2, a_2), \ldots, (s_T, a_T)\}. \tag{1}$$

Trajectories are grouped into two disjoint sets: the *support set* $\mathcal{S}_i = \{\tau_1, \ldots, \tau_K\}$ and *query set* $\mathcal{Q}_i = \{\tau_1, \ldots, \tau_L\}$, forming a task $\mathcal{T}_i = \{\mathcal{S}_i, \mathcal{Q}_i\}$. The objective of a learner is to find the parameters $\boldsymbol{\theta}$ of a policy $\hat{\pi}_\theta$ that can be effectively adapted to imitate a target policy π^* (trained on the unseen environment $\mathcal{E}^*$) by conditioning on the support trajectories $\mathcal{S}^*$. Note that this setting is different from offline RL (Levine et al., 2020) since we assume that the rewards are not available (which is a less restrictive assumption), and that we are dealing with small amount of data (few-shot setting).

Meta-Learning Meta-learning methods used in this few-shot imitation learning setting (Finn et al., 2017b) assume that is possible to access a distribution over tasks $\mathcal{T}_i \sim p(\mathcal{T})$ during the training phase. The parameters $\boldsymbol{\theta}$ of the base policy $\hat{\pi}_\theta$ are optimized in two stages, first an inner-loss w.r.t. the support trajectories $\mathcal{S}_i$ is minimized to obtain a set of task-specific parameters, then an outer-loss w.r.t. the query trajectories $\mathcal{Q}_i$ is minimized by backpropagating the gradients through the inner-loop and updating the original set of parameters. At evaluation time, given the support set $\mathcal{S}^*$ from an unseen target policy π^*, the parameter of the imitation policy $\hat{\pi}_\theta$ are optimized through the inner-loop routine to obtain the task-specific parameters. Note that the initial assumption made in meta-learning, that we can access samples from a distribution over tasks $\mathcal{T}_i \sim p(\mathcal{T})$, may be unrealistic in many settings, as it implies sampling over a large number of trajectories generated by many different policies. This motivates our effort in finding a fine-tuning baseline that is not limited by such a constraint.

2.2 Method

The baseline presented in this paper consists of two phases: 1) online pretraining, and 2) offline fine-tuning. In the first phase, a policy is trained via standard reinforcement learning by accessing a single variation of the environment. In the second phase, the policy trained in the first phase is fine-tuned on the support set of an unseen task. Below we describe these two phases in more detail.

Online pretraining In the pretraining stage we find the parameters of a policy $\hat{\pi}_\theta$ using standard Reinforcement Learning (RL). In standard RL with infinite horizon, the objective to maximize is defined as $J(\pi) = \mathbb{E}_\pi\big[\sum_{t=0}^{\infty} \gamma^t r_t\big]$, where π is a stochastic policy $\pi : \mathcal{S} \mapsto \Delta(\mathcal{A})$ and $\gamma \in [0, 1)$ is the discount factor. We use Soft Actor-Critic (SAC, Haarnoja et al. 2018) to optimize the policy since it is a state-of-the-art off-policy algorithm for continuous control problems and enjoys good sample efficiency. SAC learns a policy $\hat{\pi}_\theta$ and a critic Q_ψ by maximizing a weighted

objective of the reward and the policy entropy term, $\mathbb{E}_{s_t,a_t\sim\hat{\pi}}\big[\sum_t r_t + \alpha\mathcal{H}\big(\hat{\pi}(\cdot|s_t)\big)\big]$. The critic parameters are learned by minimizing the squared Bellman error using transition tuples (s_t, a_t, s_{t+1}, r_t) from a replay buffer $\mathcal{D}$:

$$\mathcal{L}_Q(\psi) \triangleq \mathbb{E}_{(s_t,a_t,s_{t+1},r_t)\sim\mathcal{D}}\big[Q_\psi(s_t,a_t) - r_t - \gamma V(s_{t+1})\big]^2, \tag{2}$$

where the target value of the next state is obtained by drawing an action a' using the current policy:

$$V(s_{t+1}) \approx \bar{Q}_\psi(s_{t+1}, a') - \alpha\hat{\pi}_\theta(a|s_{t+1}). \tag{3}$$

The term $\bar{Q}_\psi(s_{t+1}, a')$ is usually a copy of the trained Q_ψ, with parameters that are updated more slowly than ψ. The policy is learned by minimizing its divergence from the exponential of the soft-Q function, which is equivalent to

$$\mathcal{L}_{\hat{\pi}}(\theta) \triangleq -\mathbb{E}_{a\sim\hat{\pi}}\big[Q_\psi(s_t,a) - \alpha\log\hat{\pi}_\theta(a|s_t)\big], \tag{4}$$

where α is the entropy regularization coefficient.

Note that, there is crucial difference between the online pretraining stage used in our method and the meta-training stage used in meta-learning (Finn et al., 2017a;b). Online pretraining only relies on a single base environment to train the imitation policy, whereas meta-learning needs a distribution over many environments and offline trajectories.

Offline Fine-Tuning After the imitation policy $\hat{\pi}_\theta$ has been pretrained on the base environment via SAC, we can use the policy at evaluation time. Given the support trajectories $\mathcal{S}^* = \{\tau_1, \ldots, \tau_K\}$ from an unseen target policy π^*, we fine-tune the parameters of the base policy by minimizing a loss function $\mathcal{L}$ over the support points via Behavioral Cloning (BC). This step provides a new set of parameters $\boldsymbol{\theta}^*$ of the base policy $\hat{\pi}_{\theta^*}$

$$\text{BC}(\hat{\pi}_\theta, \mathcal{S}^*, \mathcal{L}) \rightarrow \hat{\pi}_{\theta^*}, \tag{5}$$

where we have used the functional notation $\text{BC}(\cdot)$ as a shorthand to represent the iterative gradient-descent routine used in BC. After the offline fine-tuning stage we have access to a specialized policy $\hat{\pi}_{\theta^*}$ which is an approximation of the target policy π^*. Here it is important to stress the fact that offline fine-tuning is part of the evaluation phase, meaning that it has no impact on the training phase. A graphical representation of the online pretraining and offline fine-tuning stages is presented in Figure 1.

3 Previous Work

Meta-Learning vs. Fine-Tuning In the last few years, meta-learning has become a popular approach for tackling the few-shot setting in classification and regression (Hospedales et al., 2020). One of the most widely used meta-learning methods is Model-Agnostic Meta-Learning (MAML, Finn et al. 2017a). MAML trains a neural network to find a set of weights that can be easily adapted to new tasks in a few gradient steps. This is achieved through an outer loop that optimizes the neural network weights and an inner loop that updates the weights for a new task. In addition to MAML, other solutions that exploit the learning-to-learn paradigm have been proposed. For instance, metric learning (Snell et al., 2017; Bronskill et al., 2021), Bayesian methods (Gordon et al., 2018; Patacchiola et al., 2020; Sendera et al., 2021), and different adaptation mechanisms (Rebuffi et al., 2017; Patacchiola et al., 2022). A recent line of work has showed that fine-tuning methods are more effective than meta-learning while being easier to parallelize and train. In the context of few-shot image classification, Chen et al. (2019) were the first to demonstrate the potential of simple fine-tuning baselines for transfer learning. Triantafillou et al. (2019) have proposed MD-Transfer as an effective fine-tuning baseline for the MetaDataset benchmark. Kolesnikov et al. (2020) presented Big Transfer (BiT), which shows that large models pretrained on ILSVRC-2012 ImageNet and the full ImageNet-21k are highly effective in transfer learning. Fine-tuning only the last linear layer of a pretrained backbone (Bauer et al., 2017; Tian et al., 2020) or a subset of the parameters (Shysheya et al., 2023) has also been shown to be a valid approach to transfer learning and few-shot learning. In the context of reinforcement learning, Mandi et al. (2022) have compared meta-learning and fine-tuning for adaptation to new environments. This work differs from ours for a number of reasons: (i) it focuses on task-level online adaptation vs. we focus on offline policy imitation; (ii) it is based on tasks with non-overlapping state-action spaces vs. we use tasks with overlapping states; (iii) pretraining is performed using a multi-task schedule vs. pretraining on a single environment condition; (iv) discrete state spaces (vision tasks) vs. continuous state-action spaces (control tasks).

Few-shot imitation learning Finn et al. (2017b) are the first to tackle the imitation problem through a meta-learning approach. The authors adapt MAML to this setting, by assuming that support and query sets (containing state-action tuples) can be sampled from a distribution over many expert policies. Duan et al. (2017) consider a similar setting, assuming access to a very large (possibly infinite) set of tasks, the authors exploit soft attention mechanisms and recurrent architectures to perform adaptation to new tasks. James et al. (2018) departs from meta-learning and use a

metric learning approach for visual imitation in robotics applications. Yu et al. (2018) aim at performing imitation learning from human demonstrations. The support set consists of states (videos) of a human performing the task, while the query set is represented as a sequence of state-action pairs obtained from a robot that is performing the same task. The challenge here is that the human demonstration only include states (video frames) and there is no easy way to extract the actions (e.g. joint commands). To solve this problem the authors propose to meta-learn an adaptation objective that does not require actions. Mitchell et al. (2021) propose Meta-Actor Critic with Advantage Weighting (MACAW), an optimization-based meta-learning algorithm that uses simple, supervised regression objectives for both the inner and outer loop of meta-training. This work was the first to consider the setting where the state-action-reward tuples arrive as a stream, which is more aligned with the standard offline-RL setting. Xu et al. (2022) exploit prompting in conjunction with a pretrained decision transformer to achieve generalization across continuous control tasks. Few-shot demonstrations are passed as input prompt to a decision transformer, and concatenated with the state-action-reward tuple coming from the interaction with the environment. Like for Mitchell et al. (2021), the authors assume an offline-RL streaming containing state-action-reward tuples. Nagabandi et al. (2018) consider each timestep to potentially be a new task, and details or settings could have changed at any timestep. The authors present two versions of their method: recurrence-based adaptive learner (ReBAL) and gradient-based adaptive learner (GrBAL). Both methods rely on model predictive control (MPC), which is expensive to run.

Generalization in reinforcement learning Generalization in reinforcement dates back to early works which demonstrated generalization to new tasks in robotics (Sutton et al., 2011), and in video games (Schaul et al., 2015; Parisotto et al., 2015). In these applications, the tasks were defined as to reaching a set of goal states (Sutton et al., 2011), for which the reinforcement learning was leveraged to minimize the distance to the goal (Andrychowicz et al., 2017; Nasiriany et al., 2019), with possible combination of language (Jiang et al., 2019), and the learnt latent spaces (Eysenbach et al., 2018; Rakelly et al., 2019). Multi-task reinforcement learning (MTRL, Vithayathil Varghese & Mahmoud 2020) further generalizes the goal reaching to a full set of tasks, which features for example all tasks in Atari (Hafner et al., 2019), and various agents with different dynamics or morphology (Huang et al., 2020). Such generalization is challenging because different agents/tasks typically have incompatible observation and action spaces, precluding the direct application of classic deep reinforcement learning approaches. Recent work (Chang et al., 2021; Goyal et al., 2019) has shown that modularization simplifies learning over different state-action spaces (Huang et al., 2020). This family of approaches leverage Graph Neural Networks (Gori et al., 2005; Scarselli et al., 2005) to condition the policy on the morphological representation of the agent, and have demonstrated the ability to generalize to unseen morphologies (Huang et al., 2020; Wang et al., 2018; Sanchez-Gonzalez et al., 2018). Transformers (Vaswani et al., 2017), have also been adopted to further improve the task generalization in deep reinforcement learning (Kurin et al., 2020), mainly by exploiting the attention mechanism for structure modelling (Tenney et al., 2019; Vig & Belinkov, 2019).

4 Experiments

4.1 The iMuJoCo dataset

A few benchmarks have been proposed to address meta-learning and offline learning in RL, such as Meta-World (Yu et al., 2020), Procgen (Cobbe et al., 2020), and D4RL (Fu et al., 2020). However, differently from the standard meta-learning setting, in imitation learning we need a large variety of offline trajectories, collected from policies trained on heterogeneous environments. Existing benchmarks are not suited for this case as they: do not provide pretrained policies and their associated trajectories (e.g. Meta-World and Procgen), lack in diversity (Meta-World and D4RL), or do not support continuous control problems (e.g. Procgen). In order to satisfy these requirements, we created a variant of OpenAI-Gym MuJoCo that we called iMuJoCo (iMitation MuJoCo). The OpenAI-Gym MuJoCo suite (Todorov et al., 2012; Brockman et al., 2016) is a set of 10 control environments that has been widely used by the RL community. This suite is useful for training standard online agents, but is less useful for studying imitation learning, due to its lack in heterogeneity. The iMuJoCo dataset builds on top of MuJoCo providing a heterogeneous benchmark for training and testing imitation learning methods and offline RL methods. Heterogeneity is achieved by producing a large number of variants of three base environments: Hopper, Halfcheetah, and Walker2d. For each variant a policy has been trained via SAC, then the policy has been used to generate 100 offline trajectories. The user can access the environment variant (via the OpenAI-Gym API and a XML configuration file), the offline trajectories (via a Python data loader), and the underlying SAC policy network (using the Stable Baselines API, Raffin et al. 2021). Each environment variant falls into one of these four categories:

- *mass*: increase or decrease the mass of a limb by a percentage; e.g. if the mass is 2.5 and the percentage is 200% then the new mass for that limb will be 7.5.
- *joint*: limit the mobility of a joint by a percentage range, e.g. if the joint range is 180° and the percentage is -50% then the maximum range of motion becomes 90°.

- *length*: increase or decrease the length of a limb by a percentage; e.g. if the length of a limb is 1.5 and the percentage is 150% then the new length will be 3.75.
- *friction*: increase or decrease the friction by a percentage (only for body parts that are in contact with the floor); e.g. if the friction is 1.9 and the percentage is -50% then the new friction will be 0.95.

Note that each environment has unique dynamics and agent configurations, resulting in different numbers of variants. Specifically, we have 37 variants for Hopper, 53 for Halfcheetah, and 64 for Walker2d, making a total of 154 variants. For a comprehensive list of these variants and additional details about the dataset, please refer to Appendix A.

4.2 Setup

The policies used by all methods are parameterized via a Multi-Layer Perceptron (MLP) with two hidden layers (256 units) with ReLU activations. During the imitation phase we iterate over the support data for 20 epochs, minimizing the ℓ_2 loss between the prediction of the model and the true target. We use the Adam optimizer (Kingma & Ba, 2014) with weight decay of 10^{-5}, and an initial learning rate of 10^{-3} which is divided by 2 at 50% and 75% of the total learning epochs. These hyper-parameters were empirically derived from a grid search. At evaluation time, each method is adapted on the support data of each environment variation. There are a total of 100 trajectories from the target policy for each variant. Based on the number of shots, a random subset of trajectories is selected (e.g. 1 trajectory in 1-shot, 10 trajectories in 10-shot, etc) and used as support set. The remaining trajectories are used to estimate the test loss. For all methods we repeat the experiments using three different seeds and provide the aggregated results. We compare five different methods: Scratch, Fine-Tuning, Head Fine-Tuning, Meta-Learning, and Multi-Task. The Scratch baseline is a neural network initialized from scratch and directly fine-tuned on the support data (no pretraining). The Fine-Tuning baseline is the implementation of our method, as described in previous sections. The Head Fine-Tuning baseline is like the Fine-Tuning baseline, but only the parameters of the last linear layer are optimized, the remaining set of weights is kept equal to the pretrained values. The Meta-Learning baseline is trained using first-order MAML as described in Finn et al. (2017b;a). At evaluation time we adapt the model similarly to Fine-Tuning (20 epochs of gradient descent using Adam optimizer and initial learning rate of 10^{-3}); empirically this guarantees better results and ensures a fair comparison across methods. Finally, Multi-Task is a behavioral cloning baseline trained on the same data distribution used in Meta-Learning and adapted to a query task in the same way as Fine-Tuning. To test the generalization of the meta-learner and the multi-task learner we use a k-fold validation procedure: for each environment we split the variants into macro-categories (mass, friction, joint, length), and meta-train on all of them but one, the excluded category is then used for meta-evaluation. All experiments were carried out on a single NVIDIA A100 GPU with 80GB of memory.

Note that, in this section we compare all methods in terms of adaptation loss (loss over the query set of unseen tasks), reporting the reward scores in Appendix C. While in standard RL the maximum reward is often used to compare models, in few-shot imitation learning the reward is not a good proxy to asses the capabilities of a method. This is due to the fact that the target policy may have a reward that is lower/higher than the imitation policy. For instance if the imitation policy already performs well in the environment variant (high reward), but the target policy perform poorly (low reward), then the imitation policy has to lower its performance in order to approximate the target effectively.

4.3 Comparing individual results

In this section we examine the results reported in Figure 2 were we compare Fine-Tuning vs. other methods in terms of adaptation capabilities on all environments variants and shots reporting the individual losses for all conditions. The idea is to get a bird-eye view of the results, before further breaking them down in the other sections. We report all the results in tabular format in Appendix C.

Fine-Tuning vs. Scratch In order to asses whether the online pretraining stage is needed, we perform an ablation on the training protocol by removing the initial pretraining phase. We initialize a network from scratch for every task, and fine-tune it on the support set data. We refer to this variant as Scratch in the figures. The results are reported in the top row of Figure 2, and show that Fine-Tuning performs better than Scratch in a large majority of conditions clearly showing the advantage of the pretraining stage. However, in the one-shot condition the two methods perform similarly, showing that a severe data scarcity may reduce the effectiveness of online pretraining at lower-shots.

Fine-Tuning vs. Head Fine-Tuning The previous set of experiments confirms that pretraining is necessary to achieve good adaptation performance. However, fine-tuning a subset of the parameters may still be effective, as showed in the few-shot image classification setting (Bauer et al., 2017; Tian et al., 2020; Bronskill et al., 2021). Here we verify this hypothesis by fine-tuning just the last linear layer of the pretrained backbone, calling this method Head Fine-Tuning. The results are reported in the second row of Figure 2. Overall, fine-tuning the linear head significantly degrades the performance in all conditions. Interestingly, if we compare first and second row of Figure 2 we see that Head Fine-

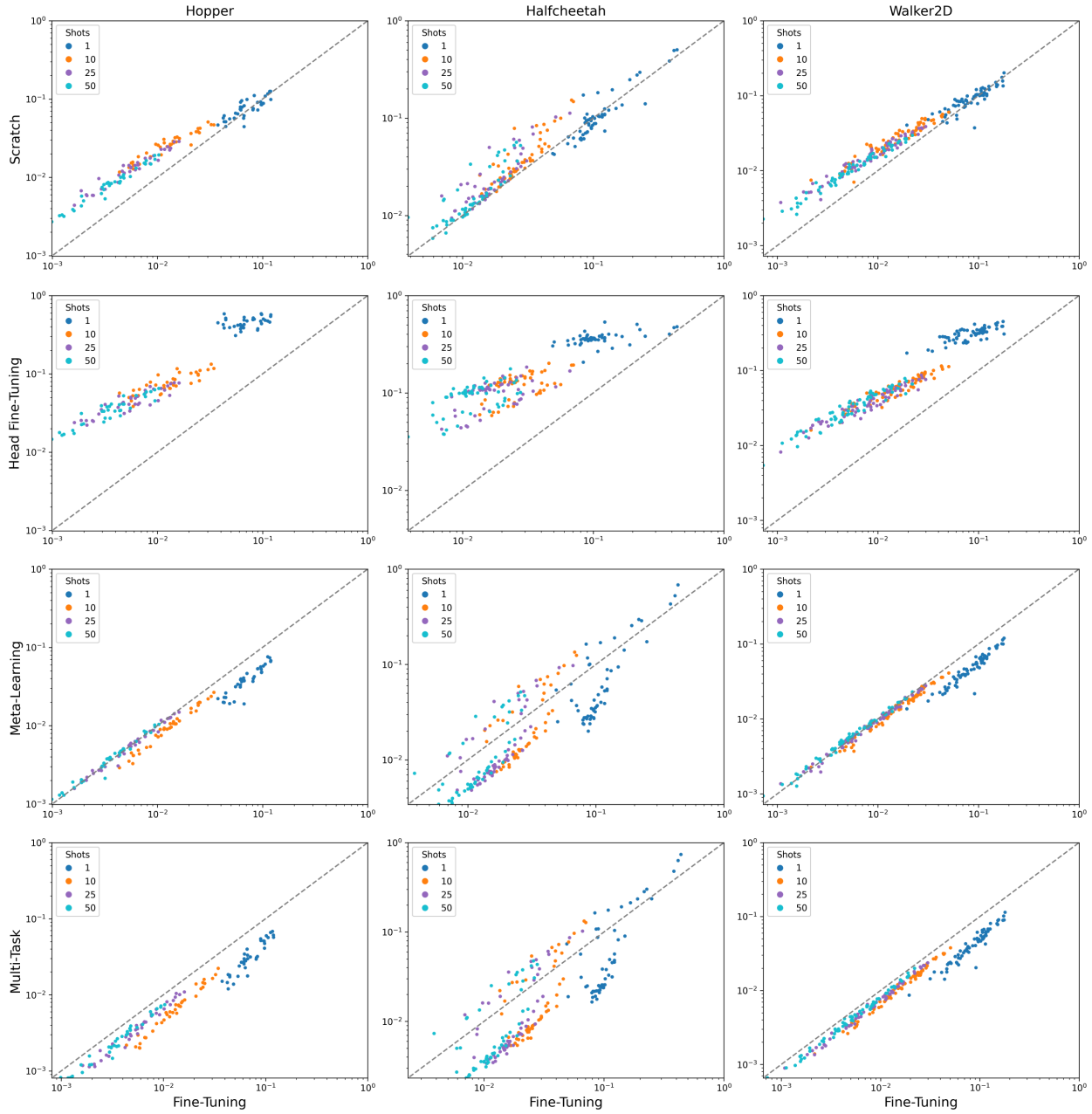

Figure 2: Comparing the average query-set loss (log-scale) of Fine-Tuning against Scratch, Head Fine-Tuning, Meta-Learning, and Multi-Task over all shots (1, 10, 25, 50) and environments (Hopper, Halfcheetah, and Walker2d). Lower loss means a better approximation of the target policy. Each dot is the average loss over three runs for a particular environment variation. The dotted identity line represents equal performance (equal loss). Points above the line mean better performance for Fine-Tuning (lower loss for Fine-Tuning), whereas points below the line mean better performance for the method noted on the y-axis (lower loss for the other method).

Tuning is less effective than Scratch. We hypothesize that this may be due to the fact that the neural networks used in RL are shallow, which reduce the reusability of the futures learned in early layers, this makes it necessary to adapt the whole set of parameters for obtaining the best performance.

Fine-Tuning vs. Meta-Learning We compare Fine-Tuning and Meta-Learning, reporting the results in the third row of Figure 2. The overall data show that the two methods perform very similarly on Hopper and Walker2d, while

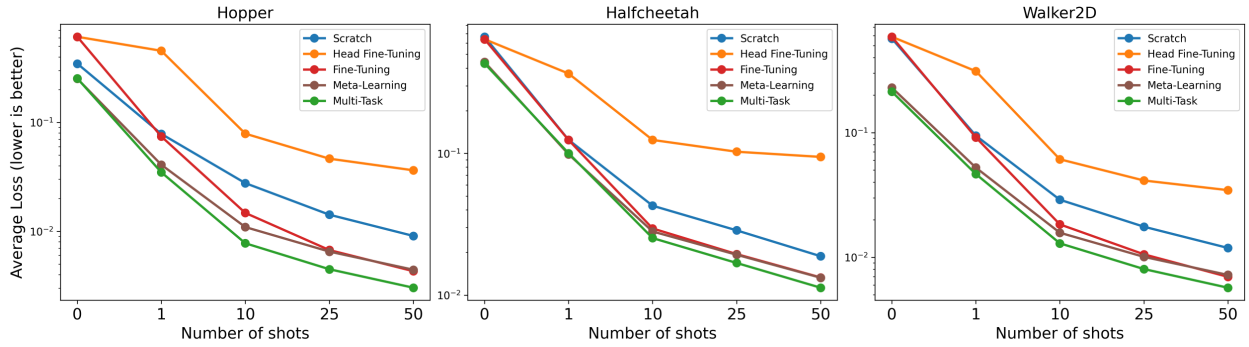

Figure 3: Comparing the average query-set loss per shots (log-scale) for all methods and environments. Lower loss is better as it implies a better approximation of the target policy. As a reference, we include the 0-shot condition, which shows the loss before adaptation. The general trend show that for all methods the loss decreases as the number of shots increases. The difference between Fine-Tuning (red line) and Meta-Learning (brown line) becomes marginal in the medium (10-shot) and high-shot regime (25 and 50-shot). Fine-Tuning outperforms Scratch (blue line) in all conditions, showing the importance of the pretraining stage.

having complementary strengths and weaknesses on Halfcheetah. In the low-shot regime, in particular in 1-shot, we observe an advantage for Meta-Learning but this advantage disappears as we move towards higher shots. At 10-shot the methods perform similarly on Walker2D and Hopper, while on Halfcheetah results are mixed, with one method outperforming the other based on the variant considered. On high shots (25 and 50-shot) we observe that the two methods convergence towards similar performance, with a significant decrease in loss, meaning that both methods are able to effectively imitate the target policy. These results suggest that, while Meta-Learning is more effective in the 1-shot case, the differences with Fine-Tuning tend to reduce substantially in the medium and high shot regimes. We conclude that at higher shots it may be more practical to use Fine-Tuning over Meta-Learning, bypassing expensive meta-training routines.

Fine-Tuning vs. Multi-Task Finally, we compare Fine-Tuning and Multi-Task, reporting results in the last row of Figure 2. Overall Multi-Task is better than Fine-Tuning in most conditions but the gap in performance is smaller in Walker2d and (on average) in Halfcheetah. The general trend observed with Multi-Task resemble the one observed with Meta-Lerning, which may be due to the use of the same k-fold pretraining schedule. Notably, there is a performance advantage of Multi-Task over Meta-Learning, even though both methods have access to the same type of data. This suggests that Multi-Task may offer a more convenient approach in terms of training ease and performance improvement compared to Meta-Learning. The performance advantage of Multi-Task over Fine-Tuning is likely due to the fact that Multi-Task has access to a larger variety of environments during the pretraining phase, while Fine-Tuning has only access to a single standard-environment. However, it is worth noting that assuming access to a distribution over tasks and policies may not be realistic in many practical applications, as this distribution may not be available.

4.4 Aggregated comparison

In this section we analyze the aggregated loss per shot over all variants and for all methods. Results are reported in Figure 3.

Analysis of the 0-shot setting The 0-shot condition is included as a baseline, showing how all methods perform *before adaptation*. The performance of Head Fine-Tuning is identical to Fine-Tuning in this condition, as both use the same pretrained backbone which are not adapted at 0-shot. Fine-Tuning performs similarly to or worse than Scratch, with an average loss of 0.61 vs. 0.35 in favor of Scratch for the Hopper environment, and similar scores on Halfcheetah (0.64 vs. 0.66) and Walker2d (0.59 vs. 0.57). These results confirm that the online pretraining stage does not provide an advantage for the fine-tuner. The situation is different for Meta-Learning, that does better at 0-shot in all three environments. The gap against Fine-Tuning is substantial in this condition with average losses of 0.35 vs. 0.25 in Hopper, 0.64 vs. 0.44 in Halfcheetah, and 0.57 vs. 0.22 in Walker2d. Similar results are also observed when comparing Fine-Tuning and Multi-Task. We hypothesize that this is due to the fact that Meta-Learning and Multi-Task have been exposed to a large number of task variants during training while the other methods have seen only one or none. Therefore, the set of weights obtained by those methods may be more general-purpose at 0-shot. This implicit advantage of Meta-Learning and Multi-Task at 0-shot could also explain why they do better at 1-shot.

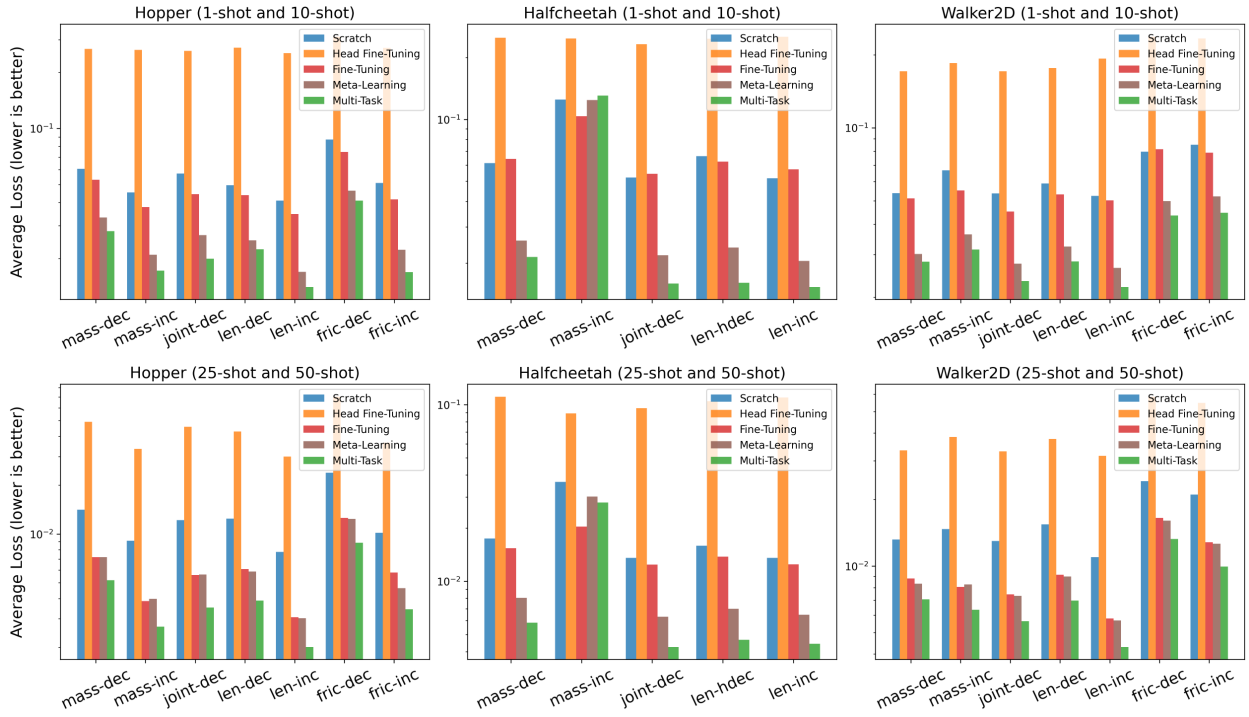

Figure 4: Comparing the average query-set loss (log-scale) on low-shot (top row) and high-shots (bottom row) for all methods over subset of environment variants that include: decreasing/increasing limb mass (mass-dec, mass-inc), decreasing joint mobility (joint-dec), increasing/decreasing length of a limb (len-dec, len-inc), and increasing/decreasing the friction with the floor (fric-dec, fric-inc). Lower loss is better as it implies a better approximation of the target policy.

Analysis of the low-shot setting Here we focus on the low-shot conditions: 1-shot and 10-shot. The highest loss for all methods (worst imitation capabilities) is observed in the 1-shot condition (1 trajectory in the support set). This is expected, as the amount of data available in 1-shot is limited. In 1-shot we observe the largest gap between Fine-Tuning and Meta-Learning, with an average loss of 0.07 vs. 0.04 on Hopper, 0.12 vs. 0.01 on Halfcheetah, and 0.09 vs. 0.05 on Walker2d. Multi-Task performs similarly to Meta-Learning or marginally better in this condition. Overall, Meta-learning and Multi-Task are more effective than Fine-Tuning in this scenario, but all methods suffer from poor approximation of the target policy due to the scarcity of data. In the 10-shot condition, we observe an overall decrease in loss (better approximation of the target policy) for all methods, showing the benefits of having more support data. The gap between Fine-Tuning and Meta-Learning decreases, with average losses of 0.014 vs. 0.011 on Hopper, 0.030 vs. 0.028 on Halfcheetah, and 0.018 vs. 0.016 on Walker2d. At 10-shot Fine-Tuning outperforms both Scratch and Head Fine-Tuning by a substantial margin. The gap between Fine-Tuning and Multi-Task remains approximately constant on Hopper but it decreases in Halfcheetah and Walker2d.

Analysis of the high-shot setting Here we analyze the high-shot conditions: 25-shot and 50-shot. Overall we see a better performance for all methods as the number of shots increases. At 25-shot and 50-shot there is no significant difference between Fine-Tuning and Meta-Learning, with average losses being equal up to the fourth decimal place. Both methods outperform the others by a substantial margin. Comparing Fine-Tuning with Scratch we see that the former does better in all conditions; e.g. in the 50-shot condition, we observe a loss of 0.004 vs. 0.009 on Hopper, 0.013 vs. 0.019 on Halfcheetah, and 0.007 vs. 0.012 on Walker2d. An important observation is that the gap between Scratch and Meta-Learning remains roughly constant as the number of shots increases, while the gap between Fine-Tuning an Meta-Learning gets smaller and eventually becomes marginal. This confirms the importance of online pretraining. The overall results for the Head Fine-Tuning method show that this approach has poor adaptation capabilities in all setting and environments. Finally, we observe that Multi-Task is the best performing method at higher shots. The gap with respect to Meta-Learning and Fine-Tuning increases pointing to the possibility that there is a qualitative difference in the optimum reached by this method.

4.5 Comparison over environment variants

In this section we breakdown the results by analyzing the performance over specific subsets of the iMuJoCo environment variants. In particular, we look at how methods perform with respect to modifications of the mass, joint mobility, length, and friction. Figure 4 provides a summary of the results.

Overall we observe that some variants are more challenging than others but this is environment-related. For instance, on Hopper decreasing the friction seems to be the most difficult variant, as showed by the higher loss for all methods in both low and high-shot. In Halfcheetah increasing the mass is the most difficult variant, while decreasing the range of motion of the joints is the less difficult. In Walker2d, modifying the friction is the most problematic variant, whereas increasing the length of the joint produces the less damage.

Meta-Learning outperforms Fine-Tuning in most of the low-shot conditions, but the difference is marginal in high-shot conditions on Hopper and Walker2d, while being more pronounced on Halfcheetah. We do not observe any sensible trend related to environment variants (e.g. differences in performance across environments correlated to the same set of modifications). The only notable difference seems to be the lower loss for Fine-Tuning on variants that increases the mass of the limbs. This is particularly evident for the Halfcheetah environment, where Fine-Tuning outperforms Meta-Learning and Multi-Task by a substantial margin.

Finally, Fine-Tuning performs similarly to Scratch in the low-shot case, but the situation is reversed in the high-shot condition. This shows once again that the pretraining stage becomes more important at high-shot.

5 Conclusions

In this paper we have investigated the few-shot imitation learning setting, proposing a simple yet effective fine-tuning baseline. We showed that by pretraining online via RL and fine-tuning offline on a set of support trajectories, it is possible to obtain results that are in line with meta-learning methods (in the high-shot regime) without the burden of meta-training on a distribution over tasks. In addition, we have released a dataset called iMuJoCo (iMitation MuJoCo) which includes 154 environment variants obtained from the popular OpenAI-Gym MuJoCo suite with associated pretrained target policies and trajectories. We believe that iMuJoCo can be a useful tool for investigating imitation learning and offline RL.

Comparison with few-shot image classification In the low-shot regime (1 and 10-shot), we have observed that meta-learning is a performant method able to provide solid results. However, the Multi-Task baseline is the best performing method in this regime, which is in line with what has been found in few-shot image classification (Chen et al., 2019; Kolesnikov et al., 2020; Shysheya et al., 2023). Therefore, when it is possible to generate data from many tasks (e.g. when domain randomization is available) we suggest to rely on meta-learning or multi-task learning instead of fine-tuning as those methods will likely lead to better performance. In the high-shot regime (25 and 50-shot) we did not see any significant difference between meta-learning and fine-tuning. The Multi-Task baseline is the best performing method overall in this condition but the gap with respect to Fine-Tuning and Meta-Learning is sometimes marginal (e.g. Halfcheetah and Walker2d). We conclude that fine-tuning is a viable alternative to meta-learning and multi-task learning in this setting, having the additional advantage of not requiring an expensive meta-training protocol and a distribution over tasks or policies, which may be more practical in many real-world applications. Another conclusion we have drawn from our experiments is that fine-tuning just the linear head does not work very well in few-shot imitation, this contrasts with results from the few-shot image classification literature (Bauer et al., 2017; Tian et al., 2020). Two factors may explain these results: differences in the neural network backbone and differences in the available data. Regarding the first factor, there is a substantial difference between the type backbones used in the two settings. Exploiting large convolutional networks in image classification makes it easier to reuse previously learned features, whereas the shallow networks used in RL may need more substantial adjustments in parameter space to be effective. Regarding the second factor, the amount of available data can also play a crucial role, since in few-shot image classification there is plenty of data that can be used to pretrain the model, making it easier to generalize, while few-shot imitation typically has limited data, making generalization more challenging.

Future work In future work we would like to investigate other fine-tuning protocols, and verify if the type of RL method used in the online pretraining stage has an impact on the overall performance during the evaluation phase. In particular, the use of on-policy vs. off-policy methods, parameters such as the entropy coefficient used in SAC, and the heterogeneity of the training environment, are factors that may play an important role in providing robust and general policies. Investigating these factors could shed more light on the differences between fine-tuning and meta-learning in the few-shot imitation setting.

ACKNOWLEDGMENTS

Funding in direct support of this work: Massimiliano Patacchiola and Richard E. Turner are supported by an EPSRC Prosperity Partnership EP/T005386/1 between the EPSRC, Microsoft Research, and the University of Cambridge.

REFERENCES

Marcin Andrychowicz, Filip Wolski, Alex Ray, Jonas Schneider, Rachel Fong, Peter Welinder, Bob McGrew, Josh Tobin, OpenAI Pieter Abbeel, and Wojciech Zaremba. Hindsight experience replay. *Advances in Neural Information Processing Systems*, 30, 2017.

Matthias Bauer, Mateo Rojas-Carulla, Jakub Bartlomiej Swiatkowski, Bernhard Scholkopf, and Richard E Turner. Discriminative k-shot learning using probabilistic models. *arXiv preprint arXiv:1706.00326*, 2017.

Christopher Berner, Greg Brockman, Brooke Chan, Vicki Cheung, Przemyslaw Debiak, Christy Dennison, David Farhi, Quirin Fischer, Shariq Hashme, Chris Hesse, et al. Dota 2 with large scale deep reinforcement learning. *arXiv preprint arXiv:1912.06680*, 2019.

Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. Openai gym. *arXiv preprint arXiv:1606.01540*, 2016.

John Bronskill, Daniela Massiceti, Massimiliano Patacchiola, Katja Hofmann, Sebastian Nowozin, and Richard Turner. Memory efficient meta-learning with large images. *Advances in Neural Information Processing Systems*, 2021.

Michael Chang, Sid Kaushik, Sergey Levine, and Tom Griffiths. Modularity in reinforcement learning via algorithmic independence in credit assignment. In *International Conference on Machine Learning*, pp. 1452–1462. PMLR, 2021.

Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Frank Wang, and Jia-Bin Huang. A closer look at few-shot classification. *arXiv preprint arXiv:1904.04232*, 2019.

Karl Cobbe, Chris Hesse, Jacob Hilton, and John Schulman. Leveraging procedural generation to benchmark reinforcement learning. In *International conference on machine learning*, pp. 2048–2056. PMLR, 2020.

Yan Duan, Marcin Andrychowicz, Bradly Stadie, OpenAI Jonathan Ho, Jonas Schneider, Ilya Sutskever, Pieter Abbeel, and Wojciech Zaremba. One-shot imitation learning. *Advances in Neural Information Processing Systems*, 30, 2017.

Benjamin Eysenbach, Abhishek Gupta, Julian Ibarz, and Sergey Levine. Diversity is all you need: Learning skills without a reward function. *arXiv preprint arXiv:1802.06070*, 2018.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*, 2017a.

Chelsea Finn, Tianhe Yu, Tianhao Zhang, Pieter Abbeel, and Sergey Levine. One-shot visual imitation learning via meta-learning. In *Conference on Robot Learning*, pp. 357–368. PMLR, 2017b.

Justin Fu, Sergey Levine, and Pieter Abbeel. One-shot learning of manipulation skills with online dynamics adaptation and neural network priors. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4019–4026. IEEE, 2016.

Justin Fu, Aviral Kumar, Ofir Nachum, George Tucker, and Sergey Levine. D4rl: Datasets for deep data-driven reinforcement learning. *arXiv preprint arXiv:2004.07219*, 2020.

Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *International Conference on Artificial Intelligence and Statistics*, pp. 249–256. JMLR Workshop and Conference Proceedings, 2010.

Jonathan Gordon, John Bronskill, Matthias Bauer, Sebastian Nowozin, and Richard E Turner. Meta-learning probabilistic inference for prediction. *arXiv preprint arXiv:1805.09921*, 2018.

Marco Gori, Gabriele Monfardini, and Franco Scarselli. A new model for learning in graph domains. In *IEEE International Joint Conference on Neural Networks*, volume 2, pp. 729–734. IEEE, 2005.

Anirudh Goyal, Alex Lamb, Jordan Hoffmann, Shagun Sodhani, Sergey Levine, Yoshua Bengio, and Bernhard Schölkopf. Recurrent independent mechanisms. *arXiv preprint arXiv:1909.10893*, 2019.

Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International conference on machine learning*, pp. 1861–1870. PMLR, 2018.

Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to control: Learning behaviors by latent imagination. *arXiv preprint arXiv:1912.01603*, 2019.

Timothy Hospedales, Antreas Antoniou, Paul Micaelli, and Amos Storkey. Meta-learning in neural networks: A survey. *arXiv preprint arXiv:2004.05439*, 2020.

Wenlong Huang, Igor Mordatch, and Deepak Pathak. One policy to control them all: Shared modular policies for agent-agnostic control. In *International Conference on Machine Learning*, pp. 4455–4464. PMLR, 2020.

Stephen James, Michael Bloesch, and Andrew J Davison. Task-embedded control networks for few-shot imitation learning. In *Conference on Robot Learning*, pp. 783–795. PMLR, 2018.

Yiding Jiang, Shixiang Shane Gu, Kevin P Murphy, and Chelsea Finn. Language as an abstraction for hierarchical deep reinforcement learning. *Advances in Neural Information Processing Systems*, 32, 2019.

Matthew Johnson, Katja Hofmann, Tim Hutton, and David Bignell. The malmo platform for artificial intelligence experimentation. In *Ijcai*, pp. 4246–4247, 2016.

Bingyi Kang, Zhuang Liu, Xin Wang, Fisher Yu, Jiashi Feng, and Trevor Darrell. Few-shot object detection via feature reweighting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8420–8429, 2019.

Mehdi Khamassi, George Velentzas, Theodore Tsitsimis, and Costas Tzafestas. Robot fast adaptation to changes in human engagement during simulated dynamic social interaction with active exploration in parameterized reinforcement learning. *IEEE Transactions on Cognitive and Developmental Systems*, 10(4):881–893, 2018.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Alexander Kolesnikov, Lucas Beyer, Xiaohua Zhai, Joan Puigcerver, Jessica Yung, Sylvain Gelly, and Neil Houlsby. Big transfer (bit): General visual representation learning. In *European Conference on Computer Vision*, 2020.

Vitaly Kurin, Maximilian Igl, Tim Rocktäschel, Wendelin Boehmer, and Shimon Whiteson. My body is a cage: the role of morphology in graph-based incompatible control. *arXiv preprint arXiv:2010.01856*, 2020.

Sergey Levine, Aviral Kumar, George Tucker, and Justin Fu. Offline reinforcement learning: Tutorial, review, and perspectives on open problems. *arXiv preprint arXiv:2005.01643*, 2020.

Zhao Mandi, Pieter Abbeel, and Stephen James. On the effectiveness of fine-tuning versus meta-reinforcement learning. *arXiv preprint arXiv:2206.03271*, 2022.

Daniela Massiceti, Luisa Zintgraf, John Bronskill, Lida Theodorou, Matthew Tobias Harris, Edward Cutrell, Cecily Morrison, Katja Hofmann, and Simone Stumpf. Orbit: A real-world few-shot dataset for teachable object recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10818–10828, 2021.

Eric Mitchell, Rafael Rafailov, Xue Bin Peng, Sergey Levine, and Chelsea Finn. Offline meta-reinforcement learning with advantage weighting. In *International Conference on Machine Learning*, pp. 7780–7791. PMLR, 2021.

Anusha Nagabandi, Ignasi Clavera, Simin Liu, Ronald S Fearing, Pieter Abbeel, Sergey Levine, and Chelsea Finn. Learning to adapt in dynamic, real-world environments through meta-reinforcement learning. *arXiv preprint arXiv:1803.11347*, 2018.

Soroush Nasiriany, Vitchyr Pong, Steven Lin, and Sergey Levine. Planning with goal-conditioned policies. *Advances in Neural Information Processing Systems*, 32, 2019.

Panagiotis D Paraschos and Dimitrios E Koulouriotis. Game difficulty adaptation and experience personalization: a literature review. *International Journal of Human-Computer Interaction*, 39(1):1–22, 2023.

Emilio Parisotto, Jimmy Lei Ba, and Ruslan Salakhutdinov. Actor-mimic: Deep multitask and transfer reinforcement learning. *arXiv preprint arXiv:1511.06342*, 2015.

Massimiliano Patacchiola, Jack Turner, Elliot J Crowley, Michael O'Boyle, and Amos J Storkey. Bayesian meta-learning for the few-shot setting via deep kernels. *Advances in Neural Information Processing Systems*, 2020.

Massimiliano Patacchiola, John Bronskill, Aliaksandra Shysheya, Katja Hofmann, Sebastian Nowozin, and Richard E Turner. Contextual squeeze-and-excitation for efficient few-shot image classification. *Advances in Neural Information Processing Systems*, 2022.

Antonin Raffin, Ashley Hill, Adam Gleave, Anssi Kanervisto, Maximilian Ernestus, and Noah Dormann. Stable-baselines3: Reliable reinforcement learning implementations. *Journal of Machine Learning Research*, 22(268): 1–8, 2021.

Kate Rakelly, Aurick Zhou, Chelsea Finn, Sergey Levine, and Deirdre Quillen. Efficient off-policy meta-reinforcement learning via probabilistic context variables. In *International Conference on Machine Learning*, pp. 5331–5340. PMLR, 2019.

Sylvestre-Alvise Rebuffi, Hakan Bilen, and Andrea Vedaldi. Learning multiple visual domains with residual adapters. *Advances in Neural Information Processing Systems*, 30, 2017.

Alvaro Sanchez-Gonzalez, Nicolas Heess, Jost Tobias Springenberg, Josh Merel, Martin Riedmiller, Raia Hadsell, and Peter Battaglia. Graph networks as learnable physics engines for inference and control. In *International Conference on Machine Learning*, pp. 4470–4479. PMLR, 2018.

Franco Scarselli, Sweah Liang Yong, Marco Gori, Markus Hagenbuchner, Ah Chung Tsoi, and Marco Maggini. Graph neural networks for ranking web pages. In *The IEEE/WIC/ACM International Conference on Web Intelligence*, pp. 666–672. IEEE, 2005.

Tom Schaul, Daniel Horgan, Karol Gregor, and David Silver. Universal value function approximators. In *International Conference on Machine Learning*, pp. 1312–1320. PMLR, 2015.

Marcin Sendera, Jacek Tabor, Aleksandra Nowak, Andrzej Bedychaj, Massimiliano Patacchiola, Tomasz Trzcinski, Przemysław Spurek, and Maciej Zieba. Non-gaussian gaussian processes for few-shot regression. *Advances in Neural Information Processing Systems*, 34:10285–10298, 2021.

Aliaksandra Shysheya, John Bronskill, Massimiliano Patacchiola, Sebastian Nowozin, and Richard E Turner. Fit: Parameter efficient few-shot transfer learning for personalized and federated image classification. In *International Conference on Machine Learning*, 2023.

Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. *Advances in Neural Information Processing Systems*, 2017.

Megan Stanley, John F Bronskill, Krzysztof Maziarz, Hubert Misztela, Jessica Lanini, Marwin Segler, Nadine Schneider, and Marc Brockschmidt. Fs-mol: A few-shot learning dataset of molecules. In *Advances in Neural Information Processing Systems Datasets and Benchmarks Track*, 2021.

Richard S Sutton, Joseph Modayil, Michael Delp, Thomas Degris, Patrick M Pilarski, Adam White, and Doina Precup. Horde: A scalable real-time architecture for learning knowledge from unsupervised sensorimotor interaction. In *International Conference on Autonomous Agents and Multiagent Systems*, pp. 761–768, 2011.

Ian Tenney, Patrick Xia, Berlin Chen, Alex Wang, Adam Poliak, R Thomas McCoy, Najoung Kim, Benjamin Van Durme, Samuel R Bowman, Dipanjan Das, et al. What do you learn from context? probing for sentence structure in contextualized word representations. *arXiv preprint arXiv:1905.06316*, 2019.

Yonglong Tian, Yue Wang, Dilip Krishnan, Joshua B Tenenbaum, and Phillip Isola. Rethinking few-shot image classification: a good embedding is all you need? In *European Conference on Computer Vision*, 2020.

Emanuel Todorov, Tom Erez, and Yuval Tassa. Mujoco: A physics engine for model-based control. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5026–5033. IEEE, 2012. doi: 10.1109/IROS.2012.6386109.

Eleni Triantafillou, Tyler Zhu, Vincent Dumoulin, Pascal Lamblin, Utku Evci, Kelvin Xu, Ross Goroshin, Carles Gelada, Kevin Swersky, Pierre-Antoine Manzagol, et al. Meta-dataset: A dataset of datasets for learning to learn from few examples. *arXiv preprint arXiv:1903.03096*, 2019.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.

Jesse Vig and Yonatan Belinkov. Analyzing the structure of attention in a transformer language model. *arXiv preprint arXiv:1906.04284*, 2019.

Oriol Vinyals, Igor Babuschkin, Wojciech M Czarnecki, Michaël Mathieu, Andrew Dudzik, Junyoung Chung, David H Choi, Richard Powell, Timo Ewalds, Petko Georgiev, et al. Grandmaster level in starcraft ii using multi-agent reinforcement learning. *Nature*, 575(7782):350–354, 2019.

Nelson Vithayathil Varghese and Qusay H Mahmoud. A survey of multi-task deep reinforcement learning. *Electronics*, 9(9):1363, 2020.

Tingwu Wang, Renjie Liao, Jimmy Ba, and Sanja Fidler. Nervenet: Learning structured policy with graph neural networks. In *International Conference on Learning Representations*, volume 30, 2018.

Mengdi Xu, Yikang Shen, Shun Zhang, Yuchen Lu, Ding Zhao, Joshua Tenenbaum, and Chuang Gan. Prompting decision transformer for few-shot policy generalization. In *International Conference on Machine Learning*, pp. 24631–24645. PMLR, 2022.

Tianhe Yu, Chelsea Finn, Annie Xie, Sudeep Dasari, Tianhao Zhang, Pieter Abbeel, and Sergey Levine. One-shot imitation from observing humans via domain-adaptive meta-learning. *arXiv preprint arXiv:1802.01557*, 2018.

Tianhe Yu, Deirdre Quillen, Zhanpeng He, Ryan Julian, Karol Hausman, Chelsea Finn, and Sergey Levine. Meta-world: A benchmark and evaluation for multi-task and meta reinforcement learning. In *Conference on robot learning*, pp. 1094–1100. PMLR, 2020.

Xiaohua Zhai, Joan Puigcerver, Alexander Kolesnikov, Pierre Ruyssen, Carlos Riquelme, Mario Lucic, Josip Djolonga, Andre Susano Pinto, Maxim Neumann, Alexey Dosovitskiy, et al. A large-scale study of representation learning with the visual task adaptation benchmark. *arXiv preprint arXiv:1910.04867*, 2019.

A ADDITIONAL DETAILS ON THE IMUJOCO DATASET

The iMuJoCo dataset is a dataset of offline trajectories and pretrained policies that can be used to study imitation learning and offline RL. The iMuJoCo dataset is based on a subset of the 10 environments of the OpenAI-Gym MuJoCo suite (Todorov et al., 2012; Brockman et al., 2016). The subset of environments used in iMuJoCo consists of: Hopper, Halfcheetah, and Walker2d. We modify each based environment to create multiple variants. For each variant we train a target policy via SAC. We use the Stable Baselines 3 library (Raffin et al., 2021), with default hyperparameters for the SAC algorithm. We train for 10^6 steps, using a clip range of 0.1. From each SAC target policy we generate 100 trajectories and store them offline. Each trajectory includes tuples (s_t, r_t, a_t) for all the steps taken (1000 steps per trajectory). Each variant is identified by a string that follows the convention *type-quantity-part* where:

- **type** refers to the variant type, it can be one of: *massinc, massdec, jointdec, lengthinc, lengthdec, frictioninc, frictiondec*, as explained in the paper.
- **quantity** is a percentage value identifying how much the modification is affecting the part, e.g. *massinc-100* means that the mass of that part has been increased by 100%.
- **part** refers to the body part affected by the modification; agents in each environment have a different number of body parts.

Note that each agent has a specific number of joints, and the dynamics of each environment are different. Therefore, the number of variants per each environment is different. We have 37 variants for Hopper, 53 for Halfcheetah, and 64 for Walker2d, for a total of 154. Below we report the ID strings identifying each variant:

Hopper massdec-50-thighgeom, jointdec-50-thighjoint, massdec-50-leggeom, frictioninc-25-footgeom, massinc-200-torsogeom, jointdec-25-legjoint, massinc-100-leggeom, massinc-100-footgeom, jointdec-50-footjoint, massinc-100-torsogeom, lengthinc-150-footgeom, jointdec-50-footjoint, massinc-200-thighgeom, massinc-300-footgeom, frictiondec-25-footgeom, massinc-100-thighgeom, jointdec-25-footjoint, massinc-300-thighgeom, massdec-25-torsogeom, lengthinc-100-torsogeom, jointdec-25-thighjoint, massinc-200-footgeom, massdec-25-footgeom, massinc-200-leggeom, massdec-50-footgeom, lengthdec-50-torsogeom, massdec-25-thighgeom, frictioninc-50-footgeom, massdec-25-leggeom, lengthdec-50-footgeom, lengthinc-100-footgeom, frictiondec-50-footgeom, jointdec-50-legjoint, massinc-300-leggeom, massdec-50-torsogeom, lengthinc-50-footgeom, massinc-300-torsogeom.

Halfcheetah massinc-300-bshin, massinc-200-ffoot, massinc-200-fthigh, massdec-50-head, jointdec-50-fshin, jointdec-25-bshin, jointdec-25-ffoot, massinc-100-fthigh, massdec-50-fthigh, massinc-300-fthigh, massdec-25-bthigh, massinc-300-fshin, jointdec-50-fthigh, massinc-300-head, jointdec-50-ffoot, lengthinc-100-ffoot, jointdec-25-fthigh, massinc-300-ffoot, massinc-200-head, massinc-300-bthigh, massinc-200-bthigh, massinc-300-bfoot, lengthinc-50-head, massinc-200-fshin, massdec-50-fshin, lengthinc-50-bfoot, jointdec-50-bthigh, massinc-100-fshin, massinc-100-head, lengthinc-50-ffoot, massdec-25-fthigh, massinc-200-bshin, massdec-50-bfoot, massinc-200-bfoot, massdec-50-ffoot, massdec-25-head, massdec-25-ffoot, massinc-100-bfoot, lengthdec-50-head, massinc-100-bthigh, massdec-25-bfoot, jointdec-50-bshin, jointdec-50-bfoot, massdec-50-bshin, jointdec-25-bfoot, massdec-50-bthigh, massinc-100-bshin, jointdec-25-fshin, lengthinc-100-bfoot, massdec-25-bshin, jointdec-25-bthigh, massdec-25-fshin, massinc-100-ffoot.

Walker2d massdec-25-footleftgeom, massinc-300-thighgeom, jointdec-25-thighleftjoint, massinc-200-thighleftgeom, massinc-200-leggeom, lengthinc-200-footgeom, jointdec-50-legleftjoint, lengthinc-200-torsogeom, massinc-200-torsogeom, massdec-25-torsogeom, jointdec-25-legjoint, massdec-25-leggeom, frictioninc-25-footleftgeom, massinc-100-thighgeom, lengthdec-50-torsogeom, massinc-200-thighgeom, massinc-100-thighleftgeom, massdec-50-legleftgeom, lengthinc-100-footgeom, frictioninc-50-footgeom, massdec-25-thighleftgeom, massdec-50-footgeom, lengthinc-200-footleftgeom, massdec-50-torsogeom, lengthinc-100-footleftgeom, jointdec-25-footjoint, massdec-50-thighleftgeom, massinc-300-leggeom, massdec-50-thighgeom, massinc-300-footgeom, lengthinc-100-torsogeom, massinc-200-footleftgeom, massinc-200-footgeom, massdec-50-footleftgeom, jointdec-50-footjoint, massdec-25-footgeom, jointdec-50-legjoint, frictiondec-25-footleftgeom, lengthdec-50-footleftgeom, lengthdec-50-footgeom, frictiondec-50-footgeom, massinc-100-torsogeom, massinc-100-footleftgeom, frictiondec-50-footleftgeom, jointdec-50-thighleftjoint, massdec-50-leggeom, massinc-200-legleftgeom, massinc-300-thighleftgeom, jointdec-25-legleftjoint, massinc-100-footgeom, massdec-25-thighgeom, massinc-100-leggeom, massdec-25-legleftgeom, massinc-100-legleftgeom, jointdec-50-footleftjoint, jointdec-50-thighjoint, massinc-300-footleftgeom, massinc-300-legleftgeom, massinc-300-torsogeom, frictiondec-25-footgeom, frictioninc-25-footgeom, frictioninc-50-footleftgeom, jointdec-25-thighjoint, jointdec-25-footleftjoint.

B Implementation Details

We used Pytorch for implementing all the experiments. The full code will be released with an open-source license.

Fine-Tuning We pretrain the Fine-Tuning baseline via SAC on the base environment (no variation applied). We use the Stable Baselines 3 library (Raffin et al., 2021), with default hyper-parameters for the SAC algorithm. We train for 10^6 steps, using a clip range of 0.1. At evaluation time we take the pretrained model for that environment, and optimize the ℓ_2 loss over the support points, obtaining a task-specific set of parameters. We use the same fine-tuning protocol described in the paper (20 epochs, Adam optimizer).

Head Fine-Tuning We use the exact same configuration used in Fine-Tuning, but during evaluation we only optimize the last linear layer, using same number of epochs and optimizer used for Fine-Tuning. We have tried different hyper-parameters (e.g. learning rate schedule) but the one used for Fine-Tuning lead to the best results.

Scratch This method initialize the weight of the policy randomly via Glorot initialization (Glorot & Bengio, 2010) for each evaluation task. The network is then fine-tuned using the protocol described in the paper (20 epochs, Adam optimizer).

Meta-Learning We use the 1-step MAML implementation used for regression problems as described in Finn et al. (2017a). We optimize over 100 tasks per meta-batch, using an inner learning rate of 0.01, and a meta learning rate of 0.001. We use Adam as meta-optimizer, and SGD as inner optimizer. During the evaluation phase, we adjust the weights of the policy via multiple optimization steps. We empirically found that applying the same adaptation schedule used for the other methods (20 epochs, Adam optimizer) lead to the best performance. Therefore, we used this adaptation scheduler in all experiments.

Multi-Task We pretrain the neural network for 50 epochs on the same k-fold data split used in Meta-Learning. We used Adap optimizer with a learning rate of 0.01 that is divided by half at 25 and 37 epochs. During the adaptation phase we use the same adaptation schedule used for the other methods (20 epochs, Adam optimizer).

C EXPERIMENTS: RESULTS IN TABULAR FORMAT

Table 1: Fine-Tuning in Hopper. Comparing average reward/loss per shot.

	Reward						Loss				
	target	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**
massdec-25-thigh	4533.1	4566.4	3369.0	4156.6	4384.6	4528.0	0.5721	0.1023	0.0086	0.0045	0.0034
massdec-25-legleft	4049.4	4661.4	3088.8	3358.1	3982.2	4118.8	0.7401	0.115	0.0236	0.0156	0.012
massdec-25-thighleft	5112.1	4554.8	4119.5	4186.3	4742.0	4847.3	0.4931	0.106	0.0363	0.0252	0.0189
massdec-25-footleft	4814.2	4610.9	3036.5	4400.5	4736.0	4788.7	0.737	0.0611	0.0057	0.0028	0.0017
massdec-25-leg	4820.6	4658.9	4082.1	4576.1	4796.3	4857.9	0.4416	0.0316	0.0042	0.0021	0.0016
massdec-25-foot	5863.6	4659.8	3342.1	4440.6	5498.1	5796.8	0.5954	0.0606	0.0112	0.006	0.0038
massdec-25-torso	4652.4	4488.7	2787.1	3914.4	4471.4	4632.3	0.5214	0.1209	0.0288	0.0167	0.0096
massdec-50-leg	5004.0	4663.3	3854.9	4517.5	4883.0	4851.6	0.4639	0.121	0.0266	0.0169	0.0139
massdec-50-thighleft	4117.1	4458.8	3556.3	3985.0	4120.1	4138.8	0.5321	0.0474	0.0055	0.0032	0.0019
massdec-50-foot	5664.3	4637.4	3028.2	5426.8	5587.2	5632.7	0.6323	0.0487	0.0084	0.0044	0.0027
massdec-50-torso	4491.8	4165.8	3750.1	4423.0	4620.7	4695.4	0.4547	0.0785	0.0149	0.0074	0.0043
massdec-50-footleft	4610.8	4464.8	3528.2	4534.3	4644.7	4659.7	0.7043	0.0686	0.0096	0.0056	0.0038
massdec-50-legleft	4346.1	4594.5	3541.1	4377.4	4558.3	4552.5	0.5063	0.0708	0.0117	0.0062	0.005
massdec-50-thigh	4407.7	4434.8	3764.5	4242.3	4426.3	4426.2	0.5747	0.1521	0.0507	0.0277	0.0193
massinc-100-legleft	4101.4	4242.4	2958.5	4019.5	4062.1	4099.7	0.7784	0.0635	0.0086	0.0044	0.0026
massinc-100-torso	4506.9	4191.0	3374.7	3758.3	4260.2	4441.9	0.6361	0.1712	0.0252	0.0134	0.0083
massinc-100-thigh	4067.2	4492.2	3182.4	3751.3	3882.9	4022.2	0.583	0.0953	0.0151	0.0087	0.0052
massinc-100-footleft	4524.0	4029.3	3588.4	4238.6	4317.9	4540.4	0.4841	0.0739	0.0202	0.0098	0.0055
massinc-100-thighleft	4062.8	4330.4	3532.5	3776.1	4004.8	4048.6	0.5307	0.0583	0.0078	0.0048	0.0024
massinc-100-leg	4305.0	4430.1	3591.9	3921.2	4232.0	4266.4	0.4507	0.0348	0.0048	0.0018	0.0011
massinc-100-foot	4145.7	4225.2	3460.3	3961.2	4266.9	4301.1	0.6526	0.1281	0.0285	0.0188	0.0136
massinc-200-torso	4261.9	2929.2	2869.6	3799.1	4092.3	4206.6	0.5231	0.0761	0.0114	0.0057	0.0035
massinc-200-footleft	4003.7	3586.1	3667.4	3742.3	3892.9	3851.6	0.4901	0.1797	0.0341	0.0218	0.0159
massinc-200-thighleft	4638.5	3440.0	3567.2	4593.6	4639.2	4665.1	0.4794	0.061	0.0084	0.0051	0.0037
massinc-200-legleft	3910.2	3847.5	3356.6	3799.6	3898.4	3920.6	0.5346	0.1102	0.0274	0.014	0.0085
massinc-200-leg	4107.1	4150.8	2889.1	3968.7	4082.9	4136.9	0.5822	0.1028	0.0158	0.0086	0.0057
massinc-200-foot	3899.0	3883.8	3112.6	3876.4	3843.4	3899.5	0.6198	0.072	0.0135	0.0085	0.0051
massinc-200-thigh	3936.4	3868.6	2982.6	3794.3	3876.0	3866.4	0.6411	0.1489	0.0295	0.017	0.0114
massinc-300-leg	4658.2	3908.4	2031.4	3978.7	4369.7	4441.4	0.5125	0.0943	0.0202	0.0111	0.0066
massinc-300-thigh	4472.5	2956.2	2943.5	4174.1	4423.5	4447.6	0.5486	0.094	0.0214	0.012	0.0073
massinc-300-thighleft	3625.1	2815.0	2733.9	3374.3	3602.9	3636.4	0.5038	0.0565	0.0085	0.0046	0.0035
massinc-300-legleft	3714.1	3183.3	2956.3	3497.2	3704.5	3703.1	0.5317	0.0413	0.0051	0.0022	0.0015
massinc-300-torso	5121.0	2421.3	2510.3	4033.6	4888.1	4970.2	0.5873	0.0982	0.0102	0.0055	0.0034
massinc-300-foot	3562.2	3690.7	3026.9	3452.1	3461.6	3496.5	0.5124	0.0636	0.0177	0.009	0.0045
massinc-300-footleft	3550.7	3421.3	2784.8	3362.8	3514.9	3567.4	0.6073	0.1269	0.0295	0.0187	0.0129
jointdec-25-footleft	4660.7	4235.0	3648.8	4318.6	4522.5	4607.2	0.5723	0.074	0.0169	0.0106	0.0073
jointdec-25-legleft	4102.6	4636.6	3245.2	3816.9	4090.2	4099.2	0.6767	0.0661	0.0119	0.0079	0.0045
jointdec-25-leg	4681.4	4639.5	3521.6	3788.4	4195.8	4535.8	0.6005	0.1066	0.0238	0.0149	0.0093
jointdec-25-foot	4864.2	4536.8	3513.4	4614.0	4837.3	4732.2	0.411	0.0194	0.0022	0.0011	0.0007
jointdec-25-thigh	4215.9	4640.8	3958.9	4017.4	4206.2	4231.9	0.4136	0.0748	0.0142	0.0083	0.0052
jointdec-25-thighleft	4414.2	4640.6	3105.7	3837.4	4226.4	4214.9	0.578	0.0848	0.0226	0.0114	0.0064
jointdec-50-foot	4510.1	4370.7	2909.0	3772.9	4148.0	4399.5	0.5006	0.0465	0.0059	0.0031	0.0019
jointdec-50-thigh	4917.7	4644.1	3598.1	4683.7	4842.0	4836.7	0.4531	0.0515	0.0099	0.0059	0.0042
jointdec-50-legleft	4158.8	4619.8	3430.6	3746.5	3986.6	4039.0	0.5896	0.1147	0.0266	0.0151	0.0116
jointdec-50-leg	4203.2	4645.6	3194.6	3688.2	4049.8	4105.6	0.6554	0.1074	0.0243	0.0125	0.0085
jointdec-50-thighleft	4545.9	4635.8	2989.5	4398.7	4576.3	4609.4	0.6719	0.0966	0.0201	0.0113	0.0079
jointdec-50-footleft	4590.2	3949.7	3401.4	4217.5	4538.4	4566.0	0.6628	0.0504	0.011	0.0068	0.0027
lengthdec-50-torso	3849.3	4085.4	3322.5	3623.8	3799.6	3845.4	0.5276	0.059	0.0125	0.0086	0.0044
lengthdec-50-footleft	4171.7	4282.4	2736.5	3763.3	4155.4	4179.3	0.6905	0.0883	0.0227	0.0129	0.0074
lengthdec-50-foot	4359.1	4563.5	3438.4	3734.6	3965.8	4177.3	0.6373	0.1118	0.0243	0.0132	0.0084
lengthinc-100-torso	5613.6	4282.0	3270.1	4699.1	5342.4	5562.7	0.5903	0.0442	0.0047	0.0022	0.0014
lengthinc-100-footleft	4485.5	4109.5	2719.8	3472.6	4141.5	4524.3	0.6619	0.114	0.0163	0.0091	0.0062
lengthinc-100-foot	4414.6	4319.0	3355.8	4187.9	4287.2	4328.9	0.5392	0.1158	0.0255	0.0132	0.0081
lengthinc-200-foot	4183.9	3738.8	3568.0	4026.8	4177.0	4197.1	0.555	0.1065	0.0257	0.0122	0.0089
lengthinc-200-footleft	3999.8	3517.2	2716.2	3584.6	3841.1	3965.1	0.6674	0.0465	0.0055	0.0026	0.0016
lengthinc-200-torso	4238.7	3906.4	2941.0	3522.9	3709.3	3894.1	0.8006	0.0915	0.0058	0.0027	0.0016
frictiondec-25-footleft	4104.4	4600.6	3825.8	3645.7	3949.1	4044.8	0.6611	0.0905	0.0163	0.0081	0.0051
frictiondec-25-foot	3841.3	4610.8	2692.6	3395.2	3973.2	3996.4	0.7064	0.1759	0.0429	0.0271	0.0186
frictiondec-50-foot	4118.6	4606.8	3149.1	3347.3	3790.2	3983.2	0.6848	0.1096	0.0225	0.013	0.0083
frictiondec-50-footleft	3919.6	4509.7	3115.5	3663.8	3679.8	3859.0	0.7702	0.1515	0.0434	0.0299	0.022
frictioninc-25-footleft	4338.9	4318.7	3147.3	4379.5	4540.2	4608.9	0.5891	0.1741	0.0357	0.0221	0.0139
frictioninc-25-foot	4393.1	4632.5	3394.3	3925.0	4325.1	4299.8	0.6602	0.1548	0.044	0.0256	0.0164
frictioninc-50-foot	4560.0	4612.9	3161.8	4247.2	4474.9	4565.8	0.641	0.1157	0.0164	0.0087	0.0053
frictioninc-50-footleft	4488.4	3835.6	3622.3	4104.3	4422.0	4447.3	0.6128	0.0774	0.0128	0.0063	0.0042

Table 2: Fine-Tuning in Halfcheetah. Comparing average reward/loss per shot.

	Reward						Loss				
	target	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**
massdec-25-bthigh	2617.1	1427.0	254.2	1840.1	2030.5	2324.6	0.6084	0.0966	0.0236	0.016	0.0113
massdec-25-ffoot	2607.3	2017.2	403.5	1699.7	1409.4	2239.9	0.6129	0.1096	0.0301	0.0201	0.014
massdec-25-fthigh	2647.0	1909.0	599.9	1853.5	2305.9	2477.1	0.6503	0.117	0.0286	0.0176	0.0122
massdec-25-bshin	2533.6	1025.7	1141.5	1819.2	1720.4	2186.0	0.5981	0.0899	0.021	0.0136	0.0095
massdec-25-bfoot	2511.3	1079.7	448.9	1686.8	1872.8	2367.8	0.5869	0.0873	0.0211	0.0124	0.0085
massdec-25-head	2986.3	4440.4	872.9	1368.7	2220.7	2632.5	0.6812	0.1171	0.034	0.0233	0.0165
massdec-25-fshin	2163.6	1846.2	1526.1	1853.5	1993.9	2074.6	0.6773	0.0503	0.0168	0.0107	0.0075
massdec-50-bfoot	2480.1	571.4	139.5	1768.3	2068.8	2232.0	0.637	0.1025	0.0296	0.0206	0.0152
massdec-50-fthigh	2341.8	2177.9	1349.0	1785.7	1956.2	2087.5	0.7176	0.1115	0.0404	0.0268	0.0188
massdec-50-head	3689.9	5988.2	3674.8	1556.5	2624.0	2925.1	0.613	0.1224	0.0328	0.0208	0.0153
massdec-50-fshin	2400.8	1748.6	335.7	1614.6	2027.3	2226.1	0.5824	0.097	0.0301	0.0196	0.0135
massdec-50-bthigh	2649.0	958.8	921.8	2021.1	2136.9	2161.3	0.6074	0.0963	0.0233	0.0152	0.0106
massdec-50-bshin	2412.3	549.5	1171.4	664.6	1396.2	1865.5	0.6463	0.0904	0.0232	0.0155	0.0105
massdec-50-ffoot	2406.6	2242.7	169.6	1448.6	1967.9	2151.5	0.5682	0.1223	0.0344	0.0203	0.0146
massinc-100-head	957.1	-9.3	95.1	892.7	912.6	922.4	0.6201	0.0727	0.016	0.0083	0.0059
massinc-100-fthigh	2224.8	1894.0	295.9	1611.9	1827.3	1970.6	0.6575	0.1224	0.0404	0.0259	0.0197
massinc-100-ffoot	1313.3	1379.5	1768.6	2927.4	2854.9	2999.4	0.5902	0.415	0.0446	0.029	0.0211
massinc-100-bshin	4852.2	5576.0	2781.0	3975.6	4267.1	4389.5	0.4742	0.14	0.0368	0.0247	0.0189
massinc-100-bfoot	34.8	4474.8	4345.0	3901.5	3959.9	3920.5	0.5303	0.438	0.042	0.0272	0.021
massinc-100-bthigh	3191.0	5068.5	3342.7	4438.4	4616.9	4622.9	0.3705	0.084	0.025	0.016	0.0115
massinc-100-fshin	1884.1	2402.9	765.9	1701.3	1789.1	1831.3	0.6872	0.0639	0.0149	0.0087	0.0059
massinc-200-fthigh	1631.2	2132.7	1187.8	1456.2	1530.4	1572.2	0.601	0.049	0.014	0.0088	0.0064
massinc-200-bfoot	3926.0	4191.2	2610.6	3561.9	3641.4	3640.5	0.7175	0.2275	0.0683	0.0342	0.0248
massinc-200-head	130.6	-71.4	8.6	94.3	90.6	87.4	0.539	0.25	0.0394	0.028	0.0263
massinc-200-ffoot	3284.7	1312.9	2254.4	3024.2	3125.1	2976.0	0.4649	0.1906	0.0508	0.0347	0.0278
massinc-200-fshin	34.3	2615.3	1337.2	2534.3	2743.4	2782.6	0.4512	0.3826	0.0394	0.0251	0.0162
massinc-200-bthigh	5637.0	6135.7	461.8	5465.7	5636.3	5733.6	1.035	0.0861	0.0172	0.0097	0.0072
massinc-200-bshin	4438.5	5340.1	3002.5	4133.0	4442.7	4445.0	0.4639	0.1086	0.0564	0.0517	0.0185
massinc-300-fshin	2896.7	2239.1	278.1	1527.3	2482.7	2630.7	0.6532	0.1315	0.0427	0.028	0.0211
massinc-300-head	-80.5	-101.2	-86.1	-60.6	-62.7	-63.5	0.6831	0.1664	0.0384	0.0261	0.0204
massinc-300-bthigh	5771.7	5359.7	570.6	5687.3	5838.5	5703.5	0.9935	0.0886	0.0136	0.0069	0.0038
massinc-300-fthigh	1351.2	2071.3	148.0	1254.2	1307.8	1362.5	0.8715	0.0897	0.0223	0.0137	0.0092
massinc-300-ffoot	2484.9	1362.5	355.7	1998.6	2236.3	2344.1	0.6179	0.1495	0.0458	0.031	0.0233
massinc-300-bshin	4637.7	4965.2	1327.7	4571.4	4825.2	4785.8	1.2693	0.1226	0.0188	0.0109	0.0072
massinc-300-bfoot	2352.2	3607.8	1961.5	2200.0	2372.9	2263.4	0.8817	0.2164	0.0706	0.0663	0.0259
jointdec-25-bshin	2530.5	2046.9	325.2	1685.5	2226.9	2391.9	0.5289	0.0794	0.019	0.0119	0.0076
jointdec-25-fthigh	2426.4	1932.3	960.1	1373.6	2123.3	2042.3	0.6154	0.0923	0.0237	0.0148	0.0106
jointdec-25-fshin	2291.3	1620.9	1545.4	1744.7	1943.9	2041.0	0.7088	0.0706	0.0226	0.0146	0.01
jointdec-25-ffoot	2621.1	1859.5	470.3	1925.8	2281.1	2433.3	0.5701	0.0915	0.0232	0.0155	0.0104
jointdec-25-bthigh	2513.6	2012.6	1222.8	1880.0	2120.2	2269.3	0.6018	0.0963	0.025	0.0175	0.0131
jointdec-25-bfoot	2525.1	1890.4	768.9	2025.0	2245.0	2403.7	0.5938	0.0805	0.0187	0.012	0.0082
jointdec-50-fthigh	2175.1	2169.6	372.3	1705.3	1897.6	2023.2	0.6059	0.1044	0.0331	0.0224	0.0143
jointdec-50-bshin	2457.2	1828.1	92.7	831.2	1651.1	2178.3	0.5364	0.081	0.0182	0.0105	0.0075
jointdec-50-fshin	2621.3	1351.9	431.7	1687.5	2174.1	2323.5	0.6531	0.095	0.0253	0.0162	0.0119
jointdec-50-bfoot	2559.5	1966.6	535.4	1059.8	1383.5	1539.1	0.6008	0.0835	0.024	0.0151	0.0107
jointdec-50-bthigh	2118.8	2123.4	976.6	1757.2	1909.0	1967.6	0.6196	0.0645	0.0246	0.0125	0.0071
jointdec-50-ffoot	2604.9	1798.4	248.2	1601.3	2131.9	2345.3	0.6176	0.0867	0.023	0.0143	0.0105
lengthdec-50-head	2536.7	798.2	289.6	1704.0	1797.1	2217.8	0.62	0.0992	0.0253	0.0162	0.0115
lengthinc-50-bfoot	2522.1	2005.3	297.0	1929.8	2233.8	2184.1	0.6005	0.0849	0.0194	0.0111	0.0082
lengthinc-50-ffoot	2532.5	1401.6	745.7	1546.0	1522.6	2269.5	0.6292	0.1019	0.0263	0.0158	0.0118
lengthinc-50-head	2566.2	2958.7	1594.5	1635.2	2060.5	2263.8	0.6222	0.0793	0.0194	0.0135	0.01
lengthinc-100-bfoot	2586.0	1814.9	388.4	1827.8	1781.4	2369.0	0.6071	0.0988	0.0256	0.0165	0.0122
lengthinc-100-ffoot	2627.8	1257.8	217.6	1443.8	2094.8	2263.0	0.5887	0.0926	0.0231	0.0149	0.0109

Table 3: Fine-Tuning in Walker2d. Comparing average reward/loss per shot.

	Reward						Loss				
	target	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**
massdec-25-torso	4652.4	4488.7	2787.1	3914.4	4471.4	4632.3	0.5214	0.1209	0.0288	0.0167	0.0096
massdec-25-footleft	4814.2	4610.9	3036.5	4400.5	4736.0	4788.7	0.737	0.0611	0.0057	0.0028	0.0017
massdec-25-thighleft	5112.1	4554.8	4119.5	4186.3	4742.0	4847.3	0.4931	0.106	0.0363	0.0252	0.0189
massdec-25-leg	4820.6	4658.9	4082.1	4576.1	4796.3	4857.9	0.4416	0.0316	0.0042	0.0021	0.0016
massdec-25-legleft	4049.4	4661.4	3088.8	3358.1	3982.2	4118.8	0.7401	0.115	0.0236	0.0156	0.012
massdec-25-foot	5863.6	4659.8	3342.1	4440.6	5498.1	5796.8	0.5954	0.0606	0.0112	0.006	0.0038
massdec-25-thigh	4533.1	4566.4	3369.0	4156.6	4384.6	4528.0	0.5721	0.1023	0.0086	0.0045	0.0034
massdec-50-thigh	4407.7	4434.8	3764.5	4242.3	4426.3	4426.2	0.5747	0.1521	0.0507	0.0277	0.0193
massdec-50-legleft	4346.1	4594.5	3541.1	4377.4	4558.3	4552.5	0.5063	0.0708	0.0117	0.0062	0.005
massdec-50-leg	5004.0	4663.3	3854.9	4517.5	4883.0	4851.6	0.4639	0.121	0.0266	0.0169	0.0139
massdec-50-foot	5664.3	4637.4	3028.2	5426.8	5587.2	5632.7	0.6323	0.0487	0.0084	0.0044	0.0027
massdec-50-torso	4491.8	4165.8	3750.1	4423.0	4620.7	4695.4	0.4547	0.0785	0.0149	0.0074	0.0043
massdec-50-thighleft	4117.1	4458.8	3556.3	3985.0	4120.1	4138.8	0.5321	0.0474	0.0055	0.0032	0.0019
massdec-50-footleft	4610.8	4464.8	3528.2	4534.3	4644.7	4659.7	0.7043	0.0686	0.0096	0.0056	0.0038
massinc-100-leg	4305.0	4430.1	3591.9	3921.2	4232.0	4266.4	0.4507	0.0348	0.0048	0.0018	0.0011
massinc-100-thigh	4067.2	4492.2	3182.4	3751.3	3882.9	4022.2	0.583	0.0953	0.0151	0.0087	0.0052
massinc-100-thighleft	4062.8	4330.4	3532.5	3776.1	4004.8	4048.6	0.5307	0.0583	0.0078	0.0048	0.0024
massinc-100-torso	4506.9	4191.0	3374.7	3758.3	4260.2	4441.9	0.6361	0.1712	0.0252	0.0134	0.0083
massinc-100-foot	4145.7	4225.2	3460.3	3961.2	4266.9	4301.1	0.6526	0.1281	0.0285	0.0188	0.0136
massinc-100-footleft	4524.0	4029.3	3588.4	4238.6	4317.9	4540.4	0.4841	0.0739	0.0202	0.0098	0.0055
massinc-100-legleft	4101.4	4242.4	2958.5	4019.5	4062.1	4099.7	0.7784	0.0635	0.0086	0.0044	0.0026
massinc-200-leg	4107.1	4150.8	2889.1	3968.7	4082.9	4136.9	0.5822	0.1028	0.0158	0.0086	0.0057
massinc-200-footleft	4003.7	3586.1	3667.4	3742.3	3892.9	3851.6	0.4901	0.1797	0.0341	0.0218	0.0159
massinc-200-thigh	3936.4	3868.6	2982.6	3794.3	3876.0	3866.4	0.6411	0.1489	0.0295	0.017	0.0114
massinc-200-foot	3899.0	3883.8	3112.6	3876.4	3843.4	3899.5	0.6198	0.072	0.0135	0.0085	0.0051
massinc-200-thighleft	4638.5	3440.0	3567.2	4593.6	4639.2	4665.1	0.4794	0.061	0.0084	0.0051	0.0037
massinc-200-legleft	3910.2	3847.5	3356.6	3799.6	3898.4	3920.6	0.5346	0.1102	0.0274	0.014	0.0085
massinc-200-torso	4261.9	2929.2	2869.6	3799.1	4092.3	4206.6	0.5231	0.0761	0.0114	0.0057	0.0035
massinc-300-thigh	4472.5	2956.2	2943.5	4174.1	4423.5	4447.6	0.5486	0.094	0.0214	0.012	0.0073
massinc-300-footleft	3550.7	3421.3	2784.8	3362.8	3514.9	3567.4	0.6073	0.1269	0.0295	0.0187	0.0129
massinc-300-legleft	3714.1	3183.3	2956.3	3497.2	3704.5	3703.1	0.5317	0.0413	0.0051	0.0022	0.0015
massinc-300-foot	3562.2	3690.7	3026.9	3452.1	3461.6	3496.5	0.5124	0.0636	0.0177	0.009	0.0045
massinc-300-thighleft	3625.1	2815.0	2733.9	3374.3	3602.9	3636.4	0.5038	0.0565	0.0085	0.0046	0.0035
massinc-300-leg	4658.2	3908.4	2031.4	3978.7	4369.7	4441.4	0.5125	0.0943	0.0202	0.0111	0.0066
massinc-300-torso	5121.0	2421.3	2510.3	4033.6	4888.1	4970.2	0.5873	0.0982	0.0102	0.0055	0.0034
jointdec-25-thighleft	4414.2	4640.6	3105.7	3837.4	4226.4	4214.9	0.578	0.0848	0.0226	0.0114	0.0064
jointdec-25-legleft	4102.6	4636.6	3245.2	3816.9	4090.2	4099.2	0.6767	0.0661	0.0119	0.0079	0.0045
jointdec-25-thigh	4215.9	4640.8	3958.9	4017.4	4206.2	4231.9	0.4136	0.0748	0.0142	0.0083	0.0052
jointdec-25-leg	4681.4	4639.5	3521.6	3788.4	4195.8	4535.8	0.6005	0.1066	0.0238	0.0149	0.0093
jointdec-25-footleft	4660.7	4235.0	3648.8	4318.6	4522.5	4607.2	0.5723	0.074	0.0169	0.0106	0.0073
jointdec-25-foot	4864.2	4536.8	3513.4	4614.0	4837.3	4732.2	0.411	0.0194	0.0022	0.0011	0.0007
jointdec-50-thigh	4917.7	4644.1	3598.1	4683.7	4842.0	4836.7	0.4531	0.0515	0.0099	0.0059	0.0042
jointdec-50-thighleft	4545.9	4635.8	2989.5	4398.7	4576.3	4609.4	0.6719	0.0966	0.0201	0.0113	0.0079
jointdec-50-foot	4510.1	4370.7	2909.0	3772.9	4148.0	4399.5	0.5006	0.0465	0.0059	0.0031	0.0019
jointdec-50-legleft	4158.8	4619.8	3430.6	3746.5	3986.6	4039.0	0.5896	0.1147	0.0266	0.0151	0.0116
jointdec-50-footleft	4590.2	3949.7	3401.4	4217.5	4538.4	4566.0	0.6628	0.0504	0.011	0.0068	0.0027
jointdec-50-leg	4203.2	4645.6	3194.6	3688.2	4049.8	4105.6	0.6554	0.1074	0.0243	0.0125	0.0085
lengthdec-50-foot	4359.1	4563.5	3438.4	3734.6	3965.8	4177.3	0.6373	0.1118	0.0243	0.0132	0.0084
lengthdec-50-torso	3849.3	4085.4	3322.5	3623.8	3799.6	3845.4	0.5276	0.059	0.0125	0.0086	0.0044
lengthdec-50-footleft	4171.7	4282.4	2736.5	3763.3	4155.4	4179.3	0.6905	0.0883	0.0227	0.0129	0.0074
lengthinc-100-foot	4414.6	4319.0	3355.8	4187.9	4287.2	4328.9	0.5392	0.1158	0.0255	0.0132	0.0081
lengthinc-100-footleft	4485.5	4109.5	2719.8	3472.6	4141.5	4524.3	0.6619	0.114	0.0163	0.0091	0.0062
lengthinc-100-torso	5613.6	4282.0	3270.1	4699.1	5342.4	5562.7	0.5903	0.0442	0.0047	0.0022	0.0014
lengthinc-200-foot	4183.9	3738.8	3568.0	4026.8	4177.0	4197.1	0.555	0.1065	0.0257	0.0122	0.0089
lengthinc-200-torso	4238.7	3906.4	2941.0	3522.9	3709.3	3894.1	0.8006	0.0915	0.0058	0.0027	0.0016
lengthinc-200-footleft	3999.8	3517.2	2716.2	3584.6	3841.1	3965.1	0.6674	0.0465	0.0055	0.0026	0.0016
frictiondec-25-footleft	4104.4	4600.6	3825.8	3645.7	3949.1	4044.8	0.6611	0.0905	0.0163	0.0081	0.0051
frictiondec-25-foot	3841.3	4610.8	2692.6	3395.2	3973.2	3996.4	0.7064	0.1759	0.0429	0.0271	0.0186
frictiondec-50-footleft	3919.6	4509.7	3115.5	3663.8	3679.8	3859.0	0.7702	0.1515	0.0434	0.0299	0.022
frictiondec-50-foot	4118.6	4606.8	3149.1	3347.3	3790.2	3983.2	0.6848	0.1096	0.0225	0.013	0.0083
frictioninc-25-footleft	4338.9	4318.7	3147.3	4379.5	4540.2	4608.9	0.5891	0.1741	0.0357	0.0221	0.0139
frictioninc-25-foot	4393.1	4632.5	3394.3	3925.0	4325.1	4299.8	0.6602	0.1548	0.044	0.0256	0.0164
frictioninc-50-foot	4560.0	4612.9	3161.8	4247.2	4474.9	4565.8	0.641	0.1157	0.0164	0.0087	0.0053
frictioninc-50-footleft	4488.4	3835.6	3622.3	4104.3	4422.0	4447.3	0.6128	0.0774	0.0128	0.0063	0.0042

Table 4: Head Fine-Tuning in Hopper. Comparing average reward/loss per shot.

	Reward						Loss				
	target	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**
massdec-25-thigh	3672.3	3345.5	3395.0	2605.5	2450.0	2187.7	0.6338	0.4992	0.097	0.052	0.0392
massdec-25-leg	3429.9	3473.5	3525.5	2673.2	2616.5	1655.0	0.7654	0.5671	0.1053	0.0689	0.0525
massdec-25-foot	3835.8	3600.7	3677.1	2869.3	2801.0	2292.6	0.565	0.4181	0.0653	0.0423	0.0339
massdec-25-torso	3337.4	3326.6	3345.3	3007.8	2750.3	2668.0	0.4914	0.3767	0.0817	0.0531	0.0453
massdec-50-foot	3892.0	3623.5	3696.1	3255.8	2661.0	2849.5	0.5965	0.4733	0.1172	0.075	0.0594
massdec-50-thigh	3330.1	3298.0	3305.0	2076.6	1830.8	1847.3	0.6131	0.4703	0.0991	0.0652	0.0549
massdec-50-leg	3487.1	3529.3	3284.8	2726.5	2548.7	2705.8	0.5016	0.3868	0.0756	0.0401	0.0292
massdec-50-torso	3362.1	3162.9	3206.4	2791.9	2806.7	2697.8	0.5259	0.3756	0.0719	0.0418	0.0333
massinc-100-thigh	3209.6	3208.9	3296.0	2449.2	1974.2	2218.0	0.6549	0.4921	0.0848	0.0516	0.0404
massinc-100-foot	3051.3	2907.7	3022.8	2516.6	2694.3	2563.6	0.5451	0.417	0.0784	0.0407	0.0312
massinc-100-leg	3247.4	3089.7	3328.1	2889.2	2740.7	2568.2	0.4331	0.3083	0.0541	0.0335	0.0271
massinc-100-torso	3208.2	3164.5	2931.7	2399.6	2691.6	1945.4	0.6585	0.4872	0.0717	0.0411	0.0334
massinc-200-leg	3256.6	3159.0	3203.6	2784.3	3060.1	3057.5	0.4896	0.3618	0.0652	0.0407	0.0342
massinc-200-thigh	3375.3	2871.5	2826.3	2168.8	2009.9	1989.7	0.7446	0.5274	0.0709	0.0433	0.0333
massinc-200-torso	3518.0	2807.0	2814.3	1210.8	1834.6	1127.0	0.7135	0.5138	0.0555	0.0312	0.0243
massinc-200-foot	2775.1	2531.3	2546.5	2423.9	2512.5	2649.6	0.7297	0.5862	0.1045	0.0564	0.041
massinc-300-thigh	3583.3	2687.3	2655.7	2408.1	1534.4	1759.8	0.6637	0.491	0.0452	0.0253	0.019
massinc-300-torso	3295.9	2615.7	2603.0	2029.2	2176.1	2191.1	0.7712	0.584	0.0525	0.0254	0.0178
massinc-300-foot	2502.3	2151.3	2310.4	2248.9	2366.5	2279.2	0.6304	0.4271	0.0568	0.0239	0.0146
massinc-300-leg	3174.2	3106.5	3092.6	2702.7	2351.3	1417.2	0.4912	0.3448	0.0655	0.0408	0.0345
jointdec-25-foot	3287.7	3294.3	3300.1	2362.6	2133.1	1954.0	0.5414	0.4088	0.076	0.0503	0.0411
jointdec-25-leg	3395.4	3442.8	3509.2	2825.1	2604.7	2616.0	0.5669	0.3961	0.0781	0.0489	0.0384
jointdec-25-thigh	3645.5	3444.1	3042.1	1819.6	1708.7	1697.0	0.6491	0.4739	0.0863	0.0487	0.0379
jointdec-50-leg	3356.1	3445.7	3481.9	1930.8	2065.6	2266.3	0.6146	0.4829	0.1105	0.0702	0.0588
jointdec-50-thigh	3269.8	3440.7	3380.2	2630.9	2266.1	2374.3	0.5664	0.4374	0.1019	0.0639	0.0529
jointdec-50-foot	3553.6	2730.0	2753.2	2883.1	2852.5	2625.5	0.5814	0.449	0.0384	0.0221	0.0165
lengthdec-50-foot	3532.6	3190.0	3149.8	1647.0	2310.5	2614.0	0.5951	0.4264	0.0634	0.0349	0.023
lengthdec-50-torso	3404.1	3541.9	3443.6	2433.1	2520.7	2713.0	0.7456	0.5015	0.0963	0.0621	0.0511
lengthinc-50-foot	3674.2	2692.4	3183.6	2726.1	2761.6	2705.3	0.554	0.3851	0.0528	0.0307	0.0225
lengthinc-100-torso	3451.8	3391.3	3296.6	2465.0	1923.2	1886.6	0.5528	0.3901	0.0617	0.0247	0.0173
lengthinc-100-foot	3698.2	2391.9	2256.0	2551.5	2899.7	2722.8	0.557	0.4038	0.0499	0.0253	0.0169
lengthinc-150-foot	3740.8	1825.6	1819.1	1808.2	2537.0	3130.5	0.7806	0.5734	0.1164	0.0617	0.0411
frictiondec-25-foot	3412.5	3191.5	3175.8	2193.4	1690.2	1629.1	0.5875	0.4584	0.1132	0.0765	0.0651
frictiondec-50-foot	3600.0	2941.7	2936.3	2498.9	2257.4	2093.1	0.6827	0.5422	0.1329	0.077	0.0626
frictioninc-25-foot	3735.7	3319.9	3360.4	2615.6	2526.7	2309.2	0.5638	0.426	0.07	0.0455	0.0364
frictioninc-50-foot	3212.6	2821.8	2812.4	2135.3	2024.0	2103.9	0.6739	0.5175	0.0718	0.0373	0.0272

Table 5: Head Fine-Tuning in Halfcheetah. Comparing average reward/loss per shot.

	Reward						Loss				
	target	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**
massdec-25-fshin	2163.6	1844.3	929.0	42.3	515.0	1057.5	0.6773	0.3337	0.0586	0.0458	0.0416
massdec-25-ffoot	2607.3	2010.5	1178.9	-121.6	318.7	350.4	0.6129	0.351	0.1463	0.1231	0.1144
massdec-25-bthigh	2617.1	1451.1	684.7	-120.3	227.1	289.9	0.6084	0.3596	0.1311	0.1087	0.0999
massdec-25-fthigh	2647.0	1852.3	797.5	-128.0	220.9	314.3	0.6503	0.3695	0.15	0.1261	0.1177
massdec-25-bshin	2533.6	1021.2	1268.7	-67.5	153.0	118.4	0.5981	0.3575	0.1293	0.1067	0.0991
massdec-25-head	2986.3	4484.2	1734.3	-69.8	-86.7	168.2	0.6812	0.3946	0.1566	0.1293	0.1196
massdec-25-bfoot	2511.3	1069.6	176.6	-18.4	46.0	59.3	0.5869	0.3563	0.1397	0.1186	0.1106
massdec-50-bfoot	2480.1	611.9	134.3	223.6	324.8	504.8	0.637	0.3547	0.122	0.1047	0.0987
massdec-50-head	3689.9	6038.8	4641.0	-88.7	285.3	365.9	0.613	0.3636	0.1681	0.1415	0.1314
massdec-50-fshin	2400.8	1755.6	362.3	-128.2	-160.4	-71.4	0.5824	0.3496	0.1425	0.1213	0.114
massdec-50-bshin	2412.3	541.7	817.3	2.6	44.4	17.7	0.6463	0.3944	0.1398	0.1147	0.1065
massdec-50-fthigh	2341.8	2001.5	786.4	264.9	909.6	1101.3	0.7176	0.3826	0.126	0.1083	0.1025
massdec-50-ffoot	2406.6	2303.1	814.4	112.2	101.5	83.3	0.5682	0.3421	0.1665	0.1487	0.1424
massdec-50-bthigh	2649.0	1037.8	1018.9	168.2	509.8	556.2	0.6074	0.3694	0.1332	0.1095	0.1019
massinc-100-bthigh	3191.0	5126.2	2639.3	2886.3	2278.7	2449.7	0.3705	0.2066	0.0794	0.0655	0.0611
massinc-100-bshin	4852.2	5569.9	4396.6	3333.3	3756.1	3698.4	0.4742	0.2874	0.0966	0.0775	0.0718
massinc-100-ffoot	1313.3	1336.7	1455.1	1661.8	1525.6	1875.0	0.5902	0.4687	0.1203	0.097	0.088
massinc-100-bfoot	34.8	4297.7	4534.2	2471.1	2553.5	2437.4	0.5303	0.4785	0.1067	0.0847	0.0827
massinc-100-head	957.1	-8.6	-8.8	68.4	487.8	549.3	0.6201	0.3207	0.1198	0.0912	0.0795
massinc-100-fshin	1884.1	2394.4	478.3	237.1	380.3	489.9	0.6872	0.3441	0.0848	0.0664	0.0592
massinc-100-fthigh	2224.8	1935.9	575.4	-117.3	46.2	38.2	0.6575	0.3693	0.1642	0.1431	0.1351
massinc-200-bthigh	5637.0	6138.0	1681.7	1983.6	3637.3	4007.8	1.035	0.4302	0.0632	0.0438	0.0376
massinc-200-fshin	34.3	2556.4	2383.1	2206.7	2311.7	2196.4	0.4512	0.4017	0.0996	0.0774	0.0643
massinc-200-bshin	4438.5	5350.6	3434.0	3460.4	3267.0	3169.0	0.4639	0.2669	0.1218	0.1155	0.0765
massinc-200-fthigh	1631.2	2211.3	1192.5	295.5	332.9	474.4	0.601	0.3036	0.0767	0.0578	0.0498
massinc-200-ffoot	3284.7	1437.2	1177.6	1770.1	2054.5	2337.7	0.4649	0.3117	0.1251	0.1043	0.1011
massinc-200-bfoot	3926.0	4206.5	3390.8	2720.8	2359.2	2538.3	0.7175	0.4471	0.1928	0.1336	0.1272
massinc-200-head	130.6	-73.5	-67.5	113.0	154.5	154.9	0.539	0.382	0.1054	0.0867	0.0985
massinc-300-fshin	2896.7	2155.6	1469.3	-79.8	10.3	27.2	0.6532	0.4012	0.17	0.1495	0.143
massinc-300-head	-80.5	-98.0	-132.3	-68.6	-53.3	-44.6	0.6831	0.3795	0.1003	0.0766	0.0746
massinc-300-bthigh	5771.7	5341.8	2579.9	997.2	3974.8	4080.4	0.9935	0.369	0.0597	0.0424	0.0354
massinc-300-fthigh	1351.2	2068.5	204.3	-113.8	567.8	956.3	0.8715	0.3866	0.0727	0.0523	0.0456
massinc-300-ffoot	2484.9	1408.6	783.8	65.8	181.6	209.8	0.6179	0.3936	0.2016	0.1842	0.1767
massinc-300-bshin	4637.7	4965.6	2513.6	2327.3	3508.3	4291.1	1.2693	0.5342	0.0634	0.0449	0.0379
massinc-300-bfoot	2352.2	3610.4	1619.0	1835.9	1714.2	1630.8	0.8817	0.5064	0.1934	0.168	0.135
jointdec-25-fshin	2291.3	1799.2	551.6	187.6	571.9	1025.6	0.7088	0.3597	0.085	0.0687	0.0635
jointdec-25-bfoot	2525.1	1950.9	1141.3	17.8	215.5	201.0	0.5938	0.3518	0.1284	0.1062	0.0977
jointdec-25-bthigh	2513.6	1887.2	1707.6	79.6	455.2	420.1	0.6018	0.3474	0.1291	0.1068	0.0999
jointdec-25-fthigh	2426.4	1914.2	1436.1	-167.3	46.6	459.6	0.6154	0.3509	0.1319	0.1085	0.1004
jointdec-25-bshin	2530.5	2011.6	1288.3	94.6	349.6	314.5	0.5289	0.3119	0.1166	0.0976	0.0904
jointdec-25-ffoot	2621.1	1915.5	1371.0	-50.5	-48.9	37.3	0.5701	0.3429	0.1346	0.1125	0.104
jointdec-50-fthigh	2175.1	2211.0	824.9	-164.8	125.4	165.6	0.6059	0.3529	0.1426	0.1213	0.1111
jointdec-50-fshin	2621.3	1263.0	1132.8	-69.2	-91.6	105.6	0.6531	0.3646	0.1304	0.1077	0.0993
jointdec-50-ffoot	2604.9	1462.8	920.5	72.6	193.2	326.9	0.6176	0.3508	0.1286	0.1072	0.0969
jointdec-50-bshin	2457.2	1992.9	819.9	-63.3	17.3	-40.4	0.5364	0.3242	0.1208	0.0993	0.0904
jointdec-50-bfoot	2559.5	1958.7	1491.5	-43.8	277.0	319.4	0.6008	0.3604	0.1244	0.1015	0.0926
jointdec-50-bthigh	2118.8	1934.1	334.1	124.4	796.5	988.9	0.6196	0.31	0.0728	0.0582	0.0538
lengthdec-50-head	2536.7	815.6	225.1	209.8	758.8	687.6	0.62	0.3664	0.1279	0.1085	0.101
lengthinc-50-bfoot	2522.1	2082.4	1271.2	-110.3	480.7	634.6	0.6005	0.3601	0.1265	0.1039	0.0949
lengthinc-50-head	2566.2	2866.9	1891.1	97.6	284.1	308.5	0.6222	0.3861	0.1398	0.1103	0.0997
lengthinc-50-ffoot	2532.5	1246.7	810.3	-27.8	-39.7	185.2	0.6292	0.3649	0.1331	0.1127	0.1042
lengthinc-100-ffoot	2627.8	1251.5	743.9	-100.7	-19.3	-49.8	0.5887	0.356	0.1521	0.126	0.1143
lengthinc-100-bfoot	2586.0	1660.5	1262.7	164.5	229.7	85.0	0.6071	0.3633	0.1453	0.121	0.1125

Table 6: Head Fine-Tuning in Walker2d. Comparing average reward/loss per shot.

	Reward						Loss				
	target	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**
massdec-25-thighleft	5112.1	4556.5	3851.8	3158.8	3068.4	3083.3	0.4931	0.2564	0.0786	0.0552	0.049
massdec-25-footleft	4814.2	4612.0	4252.7	2683.3	3319.1	3146.1	0.737	0.335	0.0333	0.0209	0.0156
massdec-25-torso	4652.4	4490.0	4271.8	2861.7	2422.9	2380.5	0.5214	0.3267	0.0875	0.0568	0.0479
massdec-25-legleft	4049.4	4653.1	3165.4	2135.6	2636.9	2805.1	0.7401	0.3991	0.0682	0.0461	0.0403
massdec-25-leg	4820.6	4659.2	4129.1	3500.0	3084.6	2991.6	0.4416	0.1876	0.0201	0.0126	0.0096
massdec-25-thigh	4533.1	4566.9	4433.3	2518.6	3368.1	3373.9	0.5721	0.3137	0.046	0.0266	0.0223
massdec-25-foot	5863.6	4661.9	4015.6	2216.0	1899.9	2227.1	0.5954	0.272	0.0437	0.0278	0.0228
massdec-50-footleft	4610.8	4462.6	3989.8	2861.8	2976.5	3362.2	0.7043	0.3553	0.0513	0.0295	0.0221
massdec-50-torso	4491.8	4150.5	4200.2	1456.6	1941.4	3398.4	0.4547	0.2396	0.0439	0.0277	0.0216
massdec-50-thigh	4407.7	4428.4	4209.2	2728.2	3149.1	2852.4	0.5747	0.3597	0.1127	0.08	0.0722
massdec-50-thighleft	4117.1	4450.7	4410.6	3750.3	2961.7	3250.9	0.5321	0.2008	0.0283	0.0181	0.0131
massdec-50-legleft	4346.1	4602.4	4393.7	3840.0	3623.6	3585.2	0.5063	0.2342	0.0499	0.0311	0.0255
massdec-50-leg	5004.0	4660.2	4484.5	2032.5	2195.2	2626.8	0.4639	0.2646	0.0724	0.0537	0.0514
massdec-50-foot	5664.3	4640.1	4195.8	3148.8	3619.9	3422.5	0.6323	0.274	0.0343	0.0203	0.0146
massinc-100-torso	4506.9	4214.1	3457.5	2823.7	2738.8	2566.6	0.6361	0.3878	0.0876	0.0592	0.0504
massinc-100-footleft	4524.0	4042.0	4017.0	3243.6	2691.2	2981.0	0.4841	0.2431	0.054	0.0339	0.0265
massinc-100-legleft	4101.4	4249.1	3493.7	2627.1	2648.7	2562.6	0.7784	0.314	0.047	0.0305	0.0225
massinc-100-foot	4145.7	4227.7	4043.4	2559.5	3282.7	3371.1	0.6526	0.3784	0.0805	0.0628	0.0575
massinc-100-thigh	4067.2	4503.8	3462.2	2589.2	2435.5	2336.1	0.583	0.308	0.061	0.0449	0.037
massinc-100-leg	4305.0	4414.3	3955.9	2278.5	2718.9	2841.0	0.4507	0.1809	0.0262	0.0149	0.0107
massinc-100-thighleft	4062.8	4356.5	3616.5	2994.3	2693.0	2506.1	0.5307	0.3065	0.05	0.0298	0.022
massinc-200-footleft	4003.7	3590.9	3528.9	3476.9	3131.5	3168.3	0.4901	0.3068	0.0894	0.0664	0.0579
massinc-200-thigh	3936.4	3799.3	3311.5	3119.2	3234.4	3408.9	0.6411	0.42	0.0951	0.0644	0.0561
massinc-200-legleft	3910.2	3837.3	3884.7	2401.8	1346.2	2538.1	0.5346	0.3142	0.092	0.0625	0.053
massinc-200-torso	4261.9	2907.3	3144.2	3019.0	2972.2	2817.2	0.5231	0.2327	0.0483	0.034	0.0276
massinc-200-thighleft	4638.5	3444.8	3353.5	2664.3	2975.7	3316.0	0.4794	0.2662	0.046	0.0288	0.0241
massinc-200-leg	4107.1	4180.4	3469.1	3373.9	3293.8	3292.5	0.5822	0.3445	0.0621	0.0421	0.0361
massinc-200-foot	3899.0	3904.1	3634.0	2948.3	2746.7	2721.6	0.6198	0.3466	0.0475	0.0326	0.0262
massinc-300-thighleft	3625.1	2825.8	2949.6	2821.1	2819.7	2758.7	0.5038	0.2555	0.04	0.0242	0.0187
massinc-300-foot	3562.2	3692.8	3295.8	3261.2	3104.7	3085.6	0.5124	0.2938	0.0547	0.0364	0.0274
massinc-300-torso	5121.0	2430.1	2376.1	1897.2	1659.8	2109.0	0.5873	0.3117	0.0497	0.0333	0.0271
massinc-300-footleft	3550.7	3423.7	3018.3	2731.0	2199.7	2696.1	0.6073	0.3805	0.0915	0.0693	0.0608
massinc-300-legleft	3714.1	3204.7	2992.9	2828.1	2955.6	3158.8	0.5317	0.282	0.0304	0.0186	0.0135
massinc-300-leg	4658.2	3905.1	3145.4	1695.8	2041.3	2079.9	0.5125	0.3102	0.0717	0.0481	0.0407
massinc-300-thigh	4472.5	2970.4	2778.0	2563.2	2374.3	2430.4	0.5486	0.2967	0.061	0.0441	0.0373
jointdec-25-thighleft	4414.2	4637.2	3790.0	2920.7	3052.9	3058.0	0.578	0.2922	0.0595	0.0385	0.029
jointdec-25-foot	4864.2	4541.7	4267.7	2780.9	3547.9	3519.7	0.411	0.1709	0.016	0.0082	0.0054
jointdec-25-thigh	4215.9	4637.9	4299.0	2983.0	3765.1	3772.0	0.4136	0.2346	0.0571	0.0383	0.0327
jointdec-25-legleft	4102.6	4638.2	4172.5	2791.7	3138.7	3070.8	0.6767	0.3469	0.0536	0.0336	0.0256
jointdec-25-footleft	4660.7	4247.9	4112.4	3109.9	3038.9	2950.5	0.5723	0.251	0.0469	0.035	0.0293
jointdec-25-leg	4681.4	4640.8	4059.4	2341.5	2543.8	2827.7	0.6005	0.3542	0.0762	0.0541	0.0472
jointdec-50-leg	4203.2	4641.1	4068.4	2755.2	2937.1	2711.6	0.6554	0.3266	0.0758	0.0514	0.0437
jointdec-50-legleft	4158.8	4621.8	4295.1	2179.3	2962.0	3216.7	0.5896	0.3756	0.0808	0.0593	0.0541
jointdec-50-thighleft	4545.9	4638.7	3436.7	2972.7	2819.3	2672.7	0.6719	0.3497	0.0703	0.0475	0.0416
jointdec-50-thigh	4917.7	4637.5	4223.9	3341.9	3501.2	3711.5	0.4531	0.2502	0.0396	0.0251	0.0204
jointdec-50-foot	4510.1	4387.4	4063.5	2690.5	2767.2	2519.6	0.5006	0.2366	0.0333	0.0205	0.0155
jointdec-50-footleft	4590.2	3955.8	3546.5	3127.1	3190.3	3334.3	0.6628	0.2731	0.0321	0.0228	0.0148
lengthdec-50-footleft	4171.7	4314.0	3268.5	3271.0	3137.9	3138.5	0.6905	0.3316	0.0599	0.0434	0.0347
lengthdec-50-torso	3849.3	4097.6	3946.2	3097.4	2383.9	2383.2	0.5276	0.2332	0.0396	0.0297	0.021
lengthdec-50-foot	4359.1	4560.4	3640.4	2922.5	2926.6	2811.2	0.6373	0.321	0.0743	0.0521	0.0447
lengthinc-100-torso	5613.6	4274.5	3792.2	2427.5	2517.2	2649.7	0.5903	0.2834	0.0295	0.017	0.0122
lengthinc-100-foot	4414.6	4321.9	4034.5	2135.8	3140.7	3518.4	0.5392	0.3288	0.0854	0.054	0.0446
lengthinc-100-footleft	4485.5	4086.9	3636.5	2658.1	2601.6	2321.6	0.6619	0.3442	0.066	0.0422	0.0357
lengthinc-200-torso	4238.7	3860.6	3131.3	2885.3	2602.4	2877.3	0.8006	0.4147	0.0367	0.0194	0.0138
lengthinc-200-footleft	3999.8	3537.5	2957.1	2619.3	2445.7	2443.9	0.6674	0.2807	0.0362	0.0208	0.015
lengthinc-200-foot	4183.9	3738.6	3557.1	2631.7	3488.4	3535.8	0.555	0.3265	0.0828	0.0552	0.0494
frictiondec-25-footleft	4104.4	4596.4	4418.9	2445.3	2483.7	2346.7	0.6611	0.318	0.0598	0.0355	0.0291
frictiondec-25-foot	3841.3	4605.1	4133.9	2259.1	2921.8	1792.4	0.7064	0.4506	0.115	0.0865	0.0782
frictiondec-50-footleft	3919.6	4520.2	3914.5	3039.3	2948.2	2434.2	0.7702	0.4016	0.1111	0.0758	0.0659
frictiondec-50-foot	4118.6	4612.2	4004.0	1825.5	2393.0	1932.6	0.6848	0.3673	0.0788	0.0519	0.0448
frictioninc-25-foot	4393.1	4617.9	4349.4	3428.1	3313.6	3341.6	0.6602	0.4354	0.104	0.0757	0.0658
frictioninc-25-footleft	4338.9	4314.3	4118.9	3253.4	2316.5	2648.2	0.5891	0.3885	0.1001	0.0734	0.0635
frictioninc-50-footleft	4488.4	3827.9	3874.0	3280.3	3355.7	3572.0	0.6128	0.3434	0.0619	0.0401	0.0334
frictioninc-50-foot	4560.0	4619.7	3813.9	3275.5	3123.7	2346.6	0.641	0.3655	0.0709	0.0477	0.0389

Table 7: Scratch in Hopper. Comparing average reward/loss per shot.

	Reward						Loss				
	target	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**
massdec-25-foot	3835.8	994.6	1683.9	1938.6	3059.5	3182.0	0.2974	0.0647	0.0243	0.0129	0.0085
massdec-25-torso	3337.4	1162.9	1875.4	1802.8	2512.6	2518.5	0.299	0.0725	0.0371	0.0223	0.017
massdec-25-leg	3429.9	839.0	1784.7	1916.5	2039.7	2431.8	0.3666	0.0987	0.0383	0.0176	0.0108
massdec-25-thigh	3672.3	491.0	1788.4	2567.3	2657.0	3398.2	0.3664	0.1108	0.0258	0.0107	0.0062
massdec-50-thigh	3330.1	855.9	2213.2	3350.3	2743.6	3098.6	0.2504	0.092	0.042	0.0228	0.0152
massdec-50-foot	3892.0	1041.4	1683.0	2162.9	3451.7	3381.4	0.3393	0.1166	0.0468	0.0235	0.0148
massdec-50-torso	3362.1	1012.0	1784.7	3098.3	2892.8	3280.9	0.3919	0.0912	0.0265	0.012	0.0075
massdec-50-leg	3487.1	805.6	1616.5	1627.5	2689.8	2637.7	0.2409	0.0569	0.0259	0.015	0.0089
massinc-100-thigh	3209.6	1025.6	1666.0	2776.7	3116.2	3160.5	0.3542	0.0729	0.0321	0.0154	0.0087
massinc-100-leg	3247.4	1042.1	1995.8	2959.0	3156.1	3193.9	0.3254	0.0659	0.0244	0.0118	0.0083
massinc-100-foot	3051.3	977.4	2003.7	2568.4	2596.7	2649.2	0.3437	0.0742	0.0183	0.0071	0.0043
massinc-100-torso	3208.2	1008.6	1818.7	1607.3	1835.9	2533.9	0.2987	0.0612	0.0275	0.0142	0.0088
massinc-200-thigh	3375.3	831.3	1569.7	1623.8	2455.0	3105.6	0.4515	0.0821	0.0306	0.0142	0.0081
massinc-200-torso	3518.0	990.7	1682.3	1373.9	2047.6	2782.1	0.3435	0.0513	0.0194	0.0114	0.007
massinc-200-foot	2775.1	728.4	1780.5	2433.1	2441.3	2642.8	0.5458	0.0848	0.0314	0.015	0.0098
massinc-200-leg	3256.6	950.8	1761.1	2135.0	2429.2	3212.9	0.2862	0.0908	0.0264	0.0144	0.01
massinc-300-foot	2502.3	876.5	1989.4	2118.3	2553.9	2532.2	0.318	0.0444	0.0117	0.0044	0.0027
massinc-300-thigh	3583.3	695.5	1721.6	2109.9	2377.1	2262.2	0.3763	0.0446	0.0153	0.0059	0.0039
massinc-300-leg	3174.2	1216.0	2194.2	2400.2	3134.0	3181.2	0.3909	0.0725	0.0276	0.0142	0.0086
massinc-300-torso	3295.9	935.4	1432.1	2187.2	2754.3	2967.3	0.2848	0.0612	0.0156	0.0067	0.0032
jointdec-25-foot	3287.7	814.6	2098.2	2545.0	3217.7	3115.9	0.2669	0.0787	0.0264	0.0132	0.0083
jointdec-25-leg	3395.4	597.0	1692.2	2676.8	2953.2	3177.5	0.4156	0.0953	0.0247	0.0125	0.0079
jointdec-25-thigh	3645.5	971.1	1947.4	2711.4	2735.9	3130.6	0.3878	0.0988	0.0289	0.0135	0.0084
jointdec-50-thigh	3269.8	674.6	1645.2	1915.4	2523.1	3230.1	0.2824	0.0874	0.0363	0.0189	0.0131
jointdec-50-leg	3356.1	982.2	1864.8	2685.9	3057.0	2969.6	0.3237	0.1089	0.0423	0.0246	0.0164
jointdec-50-foot	3553.6	919.5	1873.2	2203.2	2376.3	2254.5	0.4316	0.0466	0.0128	0.0058	0.0033
lengthdec-50-foot	3532.6	1059.1	1704.2	2236.7	3108.9	3079.9	0.3607	0.0573	0.0147	0.0071	0.005
lengthdec-50-torso	3404.1	705.2	1851.5	2605.5	3230.1	3240.1	0.3481	0.088	0.0382	0.0228	0.0148
lengthinc-50-foot	3674.2	782.6	2045.5	3108.3	3260.1	3518.6	0.3435	0.0514	0.0184	0.0096	0.0058
lengthinc-100-foot	3698.2	1071.5	1726.0	2527.9	3170.7	3205.1	0.3	0.057	0.0134	0.006	0.0032
lengthinc-100-torso	3451.8	832.5	2222.1	1534.4	2557.2	2858.3	0.2417	0.0558	0.0142	0.0059	0.0037
lengthinc-150-foot	3740.8	818.4	1774.5	3273.9	3538.7	3619.1	0.4391	0.0846	0.0327	0.0175	0.0103
frictiondec-25-foot	3412.5	1169.4	1929.5	2079.3	2783.1	2921.6	0.3253	0.1255	0.0509	0.0287	0.0193
frictiondec-50-foot	3600.0	735.1	1868.6	2323.5	2730.2	2962.9	0.3731	0.1255	0.0471	0.0284	0.0191
frictioninc-25-foot	3735.7	863.0	1755.8	2578.3	3305.4	3282.9	0.3556	0.0791	0.0283	0.0165	0.0102
frictioninc-50-foot	3212.6	818.6	2327.4	2598.5	2842.9	3144.6	0.4227	0.0755	0.0205	0.0086	0.0053

Table 8: Scratch in Halfcheetah. Comparing average reward/loss per shot.

	Reward						Loss				
	target	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**
massdec-25-fthigh	2647.0	-3.3	-87.8	198.9	672.8	1901.8	0.6504	0.1091	0.035	0.022	0.0149
massdec-25-bthigh	2617.1	-3.3	361.2	863.8	1265.6	2286.0	0.6731	0.08	0.0264	0.0166	0.0115
massdec-25-bfoot	2511.3	-2.9	393.9	1246.0	1896.6	2295.1	0.6291	0.0701	0.0241	0.0135	0.0089
massdec-25-bshin	2533.6	-3.5	819.9	1450.6	1573.5	2200.3	0.605	0.0715	0.0243	0.0157	0.0104
massdec-25-head	2986.3	-2.4	-126.7	1203.2	1599.2	1858.4	0.6798	0.1027	0.0422	0.028	0.019
massdec-25-ffoot	2607.3	-2.3	-28.2	166.6	1618.6	1984.5	0.7169	0.105	0.0373	0.0235	0.0157
massdec-25-fshin	2163.6	-3.2	1594.2	1222.6	1861.4	2018.0	0.7417	0.0425	0.021	0.0138	0.0086
massdec-50-fshin	2400.8	-2.6	21.1	298.4	1627.9	1973.3	0.5567	0.0783	0.0306	0.0219	0.0138
massdec-50-bshin	2412.3	-3.2	808.4	1317.0	2023.4	2225.1	0.6869	0.0723	0.0273	0.0164	0.0111
massdec-50-head	3689.9	-1.7	-62.5	845.6	1315.4	1873.3	0.6539	0.1081	0.0387	0.0247	0.0169
massdec-50-bfoot	2480.1	-3.3	467.4	724.5	1645.2	2136.1	0.7193	0.0994	0.0352	0.0227	0.0159
massdec-50-ffoot	2406.6	-2.5	64.8	942.2	1650.5	1953.4	0.6116	0.1243	0.0367	0.0227	0.0149
massdec-50-bthigh	2649.0	-3.0	561.4	1259.7	1765.2	2166.6	0.6211	0.0941	0.0274	0.0171	0.0118
massdec-50-fthigh	2341.8	-3.3	1034.6	1274.0	1424.5	1921.0	0.7891	0.1038	0.0481	0.0342	0.0231
massinc-100-head	957.1	-3.1	282.3	498.8	657.2	683.5	0.5379	0.0511	0.0163	0.0094	0.0058
massinc-100-bthigh	3191.0	-2.4	-8.2	257.9	329.5	287.9	0.7164	0.1725	0.0785	0.05	0.0336
massinc-100-bshin	4852.2	-2.2	-354.3	657.6	2612.1	3087.0	0.7518	0.1955	0.0838	0.0607	0.0447
massinc-100-fshin	1884.1	-2.9	897.6	1520.6	1721.6	1783.7	0.6936	0.0546	0.0186	0.0111	0.0075
massinc-100-ffoot	1313.3	-3.1	140.5	824.6	2803.1	2848.8	0.7943	0.4948	0.0751	0.0491	0.0364
massinc-100-bfoot	34.8	-2.3	-9.0	289.0	469.0	1133.7	0.6579	0.5032	0.0862	0.0593	0.0471
massinc-100-fthigh	2224.8	-3.2	-18.1	147.2	1609.7	1929.4	0.6156	0.1095	0.0449	0.0303	0.0208
massinc-200-fthigh	1631.2	-3.7	1002.8	227.3	1431.2	1511.3	0.6595	0.0433	0.0167	0.0112	0.0078
massinc-200-fshin	34.3	-4.1	-90.7	158.3	1343.3	2586.6	0.545	0.3856	0.0709	0.0496	0.0318
massinc-200-head	130.6	-6.2	-147.2	-27.7	88.4	95.9	0.4123	0.1402	0.0605	0.0429	0.0362
massinc-200-bfoot	3926.0	-3.2	-5.8	372.7	1634.5	3130.2	0.7709	0.2953	0.1531	0.081	0.0525
massinc-200-bshin	4438.5	-2.0	98.3	2062.2	2156.1	2605.1	0.6696	0.1824	0.0997	0.1027	0.0414
massinc-200-ffoot	3284.7	-2.5	-49.9	2236.5	2635.1	2794.9	0.6525	0.2479	0.0933	0.0686	0.0523
massinc-200-bthigh	5637.0	-2.1	-26.5	513.2	2962.4	4616.0	0.8046	0.1002	0.0336	0.0208	0.0147
massinc-300-bthigh	5771.7	-2.1	133.6	2472.5	4178.1	4912.8	0.8225	0.0941	0.026	0.0159	0.0096
massinc-300-bshin	4637.7	-1.8	775.5	1034.0	2553.5	3975.9	0.8385	0.0736	0.0322	0.0212	0.0143
massinc-300-fshin	2896.7	-4.0	4.9	669.9	1089.1	1822.1	0.6425	0.1207	0.0544	0.0319	0.022
massinc-300-ffoot	2484.9	-3.1	7.4	1153.4	2145.5	2207.3	0.5584	0.1258	0.0511	0.0364	0.0255
massinc-300-head	-80.5	-5.2	-56.7	-53.5	-76.7	-73.2	0.6075	0.1364	0.0513	0.0363	0.0281
massinc-300-fthigh	1351.2	-3.0	-537.6	1010.5	1283.2	1244.3	0.7933	0.1099	0.0426	0.0266	0.0183
massinc-300-bfoot	2352.2	-2.4	369.7	1079.6	1583.0	1761.1	0.7153	0.277	0.1478	0.1124	0.0565
jointdec-25-fshin	2291.3	-2.8	1533.6	449.9	1744.0	1996.7	0.7799	0.0624	0.0305	0.0188	0.0116
jointdec-25-ffoot	2621.1	-2.8	112.9	491.8	1466.9	2247.9	0.6532	0.096	0.0255	0.0157	0.01
jointdec-25-bfoot	2525.1	-2.7	784.1	850.0	1822.1	2227.2	0.622	0.0623	0.0229	0.0155	0.0089
jointdec-25-bshin	2530.5	-3.0	233.5	713.1	1728.7	2242.7	0.5541	0.0612	0.0196	0.0117	0.008
jointdec-25-fthigh	2426.4	-3.0	-118.9	231.8	1582.6	2043.8	0.607	0.0876	0.0279	0.0158	0.0106
jointdec-25-bthigh	2513.6	-3.4	-47.1	1234.0	1371.7	1615.7	0.6977	0.0968	0.0282	0.0183	0.0133
jointdec-50-bshin	2457.2	-3.8	237.4	734.9	1745.3	2213.5	0.5477	0.0597	0.0176	0.0103	0.0066
jointdec-50-fshin	2621.3	-2.3	70.0	316.0	1368.5	1778.5	0.6086	0.0985	0.0295	0.0188	0.0128
jointdec-50-bfoot	2559.5	-3.1	751.3	487.6	1559.8	2165.6	0.6447	0.0675	0.0251	0.0159	0.0102
jointdec-50-bthigh	2118.8	-2.9	1281.9	1479.7	1807.8	1905.5	0.8661	0.0641	0.0301	0.0199	0.0089
jointdec-50-fthigh	2175.1	-2.7	-11.9	567.2	1657.0	1928.9	0.6905	0.088	0.0362	0.0227	0.0147
jointdec-50-ffoot	2604.9	-3.6	463.2	1115.0	1792.4	2029.4	0.6432	0.0896	0.0269	0.0159	0.0107
lengthdec-50-head	2536.7	-2.6	-86.2	884.7	1133.6	1791.3	0.6395	0.1034	0.0292	0.019	0.0129
lengthinc-50-bfoot	2522.1	-1.8	376.2	1394.3	1760.7	2004.4	0.6117	0.0613	0.0243	0.015	0.0091
lengthinc-50-head	2566.2	-3.2	4.9	276.7	1656.9	2051.8	0.6666	0.0732	0.0197	0.0132	0.0096
lengthinc-50-ffoot	2532.5	-3.8	4.0	858.8	1265.3	1964.4	0.5847	0.0878	0.0296	0.0186	0.013
lengthinc-100-ffoot	2627.8	-3.1	-8.2	507.1	1546.3	2028.3	0.5952	0.0849	0.0237	0.0145	0.0103
lengthinc-100-bfoot	2586.0	2.1	199.9	1568.3	1873.6	2151.1	0.6462	0.0844	0.0285	0.0188	0.0133

Table 9: Scratch in Walker2d. Comparing average reward/loss per shot.

	Reward						Loss				
	target	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**
massdec-25-foot	5863.6	175.6	308.4	1249.0	2802.8	4164.1	0.6355	0.0547	0.0164	0.0099	0.0064
massdec-25-torso	4652.4	171.1	1382.2	343.7	2502.3	3167.6	0.4405	0.1002	0.0379	0.024	0.0151
massdec-25-legleft	4049.4	339.1	-106.1	1417.4	2173.5	3659.9	0.4278	0.0899	0.0302	0.0201	0.0154
massdec-25-leg	4820.6	411.0	-29.0	383.4	2860.4	3210.7	0.5622	0.0479	0.0088	0.0051	0.0037
massdec-25-thigh	4533.1	87.8	1601.2	1424.3	2208.8	3764.6	0.6173	0.1033	0.0167	0.0089	0.0063
massdec-25-thighleft	5112.1	1029.8	1505.0	1547.9	2728.3	3480.2	0.612	0.1172	0.0428	0.0304	0.0239
massdec-25-footleft	4814.2	461.4	1320.7	2145.4	2881.8	3593.7	0.4417	0.0487	0.0115	0.0061	0.0036
massdec-50-footleft	4610.8	254.7	267.9	1825.5	2099.2	3854.0	0.6406	0.0897	0.0199	0.0115	0.0076
massdec-50-legleft	4346.1	-59.3	1149.1	1642.7	1959.5	2955.1	0.6263	0.0819	0.0196	0.0118	0.0091
massdec-50-thigh	4407.7	244.3	1605.1	2288.2	3634.4	3837.5	0.5999	0.1397	0.0603	0.0365	0.0266
massdec-50-foot	5664.3	-79.9	634.0	1602.0	2921.3	4621.8	0.6021	0.0434	0.0139	0.0083	0.005
massdec-50-thighleft	4117.1	280.4	278.2	566.2	3055.3	3720.7	0.6859	0.0555	0.0125	0.0069	0.0042
massdec-50-torso	4491.8	-458.8	1214.3	1644.1	2143.9	3901.2	0.6083	0.0795	0.0226	0.0131	0.0087
massdec-50-leg	5004.0	458.3	181.7	2013.4	3126.1	4292.7	0.5276	0.1049	0.0344	0.0224	0.0188
massinc-100-leg	4305.0	54.5	1863.3	1068.4	2016.8	3820.8	0.6284	0.0571	0.0111	0.0052	0.0029
massinc-100-thighleft	4062.8	215.5	1313.5	1006.6	2115.6	3037.0	0.592	0.089	0.0225	0.0124	0.0069
massinc-100-torso	4506.9	730.7	520.6	592.5	1592.6	3525.4	0.5378	0.1755	0.0435	0.024	0.0156
massinc-100-footleft	4524.0	195.6	490.5	-112.3	1730.8	3169.9	0.5771	0.0751	0.0268	0.015	0.0098
massinc-100-foot	4145.7	172.8	523.0	2630.9	2919.9	3875.4	0.56	0.1207	0.0422	0.0285	0.0209
massinc-100-thigh	4067.2	-27.1	1695.5	291.2	400.7	1601.7	0.6904	0.1206	0.0313	0.0194	0.0121
massinc-100-legleft	4101.4	31.5	-315.5	-70.4	2569.8	3340.0	0.7039	0.0743	0.0204	0.0096	0.0055
massinc-200-footleft	4003.7	110.8	1387.7	2171.3	1313.2	3092.9	0.6095	0.2016	0.0466	0.0314	0.0235
massinc-200-thigh	3936.4	258.3	1120.6	2923.2	2923.1	3101.0	0.5831	0.1503	0.0496	0.0291	0.0201
massinc-200-legleft	3910.2	513.9	569.4	2454.6	2825.8	2971.1	0.5663	0.1275	0.0494	0.0266	0.0164
massinc-200-torso	4261.9	603.2	299.9	1014.0	2376.2	3197.9	0.6059	0.0859	0.0255	0.0132	0.0079
massinc-200-leg	4107.1	240.1	1262.4	2260.1	2972.3	3542.4	0.601	0.1001	0.0311	0.0173	0.012
massinc-200-thighleft	4638.5	18.5	211.7	1192.9	2456.3	3766.6	0.5499	0.0851	0.0214	0.011	0.0069
massinc-200-foot	3899.0	233.1	1357.1	2564.3	3209.6	3415.4	0.4528	0.0699	0.021	0.0134	0.0091
massinc-300-foot	3562.2	301.4	1786.5	2169.9	2710.2	3212.9	0.5232	0.084	0.0331	0.0169	0.009
massinc-300-leg	4658.2	-4.1	1757.4	1741.5	2229.5	3399.6	0.5319	0.1025	0.0342	0.0203	0.0133
massinc-300-thighleft	3625.1	371.7	74.8	1595.1	2789.4	3254.8	0.5812	0.0744	0.0218	0.0118	0.0073
massinc-300-thigh	4472.5	525.0	860.4	1726.6	1954.5	3659.9	0.5802	0.0999	0.0296	0.0188	0.0125
massinc-300-legleft	3714.1	146.4	975.8	1131.9	2586.7	3291.7	0.5181	0.0607	0.0125	0.0067	0.0043
massinc-300-footleft	3550.7	466.4	1096.1	1868.2	2340.5	2798.3	0.5402	0.1331	0.0478	0.0308	0.0228
massinc-300-torso	5121.0	797.2	1648.5	2865.4	3280.6	4324.1	0.5066	0.0784	0.0191	0.0109	0.0065
jointdec-25-thighleft	4414.2	270.0	1191.1	1345.2	3288.0	3559.1	0.49	0.0802	0.0301	0.0173	0.0103
jointdec-25-leg	4681.4	26.2	583.5	2128.9	2926.4	3514.1	0.4869	0.1016	0.0342	0.0219	0.015
jointdec-25-footleft	4660.7	115.0	1850.0	1561.1	1883.1	3203.0	0.5496	0.0718	0.0231	0.0164	0.0119
jointdec-25-foot	4864.2	456.8	1689.9	2445.9	3007.1	3147.6	0.563	0.0404	0.0075	0.0038	0.0023
jointdec-25-legleft	4102.6	406.0	1840.4	2678.7	3496.7	3680.2	0.615	0.0836	0.0218	0.0141	0.009
jointdec-25-thigh	4215.9	651.1	329.5	1808.3	3582.9	3669.4	0.4748	0.0879	0.0274	0.0165	0.0106
jointdec-50-foot	4510.1	266.8	1685.3	1671.4	2658.4	3079.1	0.63	0.0669	0.0155	0.0084	0.0051
jointdec-50-legleft	4158.8	384.7	1322.0	382.8	3355.6	3863.1	0.4749	0.1101	0.0441	0.0256	0.0184
jointdec-50-leg	4203.2	233.0	1146.0	2118.3	2891.0	3561.7	0.5597	0.1081	0.0355	0.0214	0.0144
jointdec-50-footleft	4590.2	968.8	993.0	1264.6	3218.8	3784.9	0.5679	0.0645	0.0178	0.0115	0.006
jointdec-50-thigh	4917.7	506.1	1051.5	893.1	2056.7	2123.2	0.5562	0.06	0.018	0.011	0.0081
jointdec-50-thighleft	4545.9	37.4	779.6	930.2	2817.7	3615.3	0.4896	0.1037	0.0334	0.0196	0.0138
lengthdec-50-footleft	4171.7	210.3	-99.3	-142.6	1723.2	3382.5	0.5279	0.0907	0.0318	0.019	0.0121
lengthdec-50-foot	4359.1	608.6	-248.2	2175.3	2761.2	3335.6	0.5628	0.1009	0.0376	0.0228	0.015
lengthdec-50-torso	3849.3	289.0	875.6	2502.1	2830.9	2782.4	0.5964	0.0727	0.0194	0.0151	0.0086
lengthinc-100-foot	4414.6	302.7	540.2	1722.2	3397.0	3774.6	0.6395	0.1231	0.0411	0.0244	0.0151
lengthinc-100-footleft	4485.5	-2.4	1769.4	1236.6	1865.2	2463.0	0.5628	0.1075	0.0291	0.0167	0.0117
lengthinc-100-torso	5613.6	728.6	593.7	2836.1	3645.3	3651.7	0.5895	0.0451	0.0098	0.0052	0.0031
lengthinc-200-foot	4183.9	179.1	1240.7	1850.6	3484.5	3987.1	0.5721	0.1245	0.041	0.0234	0.017
lengthinc-200-footleft	3999.8	421.7	1936.2	2036.1	2547.7	3099.0	0.4675	0.052	0.0103	0.0053	0.0033
lengthinc-200-torso	4238.7	1178.2	444.2	3353.4	3513.4	3554.4	0.4433	0.0372	0.007	0.0041	0.0026
frictiondec-25-foot	3841.3	-90.2	1393.1	1377.3	2157.3	3403.8	0.5191	0.1349	0.0537	0.0372	0.0266
frictiondec-25-footleft	4104.4	202.6	1648.9	1037.1	2579.1	3144.5	0.5834	0.097	0.0256	0.015	0.0096
frictiondec-50-foot	4118.6	192.4	1717.4	642.7	915.2	3247.1	0.6722	0.1087	0.0335	0.0223	0.0153
frictiondec-50-footleft	3919.6	79.4	1706.2	1373.5	2338.8	2887.8	0.6829	0.1323	0.0522	0.0387	0.0296
frictioninc-25-foot	4393.1	86.7	16.8	1903.3	3646.2	3871.9	0.583	0.1391	0.0565	0.0367	0.0248
frictioninc-25-footleft	4338.9	511.0	1039.1	1672.6	3235.2	3977.2	0.614	0.1569	0.0476	0.0328	0.0218
frictioninc-50-foot	4560.0	300.6	1955.4	2774.5	3585.7	4285.4	0.598	0.1237	0.0344	0.0188	0.0111
frictioninc-50-footleft	4488.4	273.5	1726.6	2560.6	2656.0	3549.8	0.609	0.099	0.0246	0.0137	0.0091

Table 10: Meta-Learning in Hopper. Comparing average reward/loss per shot.

	Reward						Loss				
	target	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**
massdec-25-leg	3429.9	2313.5	2306.2	2940.4	3197.5	3237.6	0.295	0.0676	0.0154	0.0085	0.0051
massdec-25-torso	3337.4	2535.1	2164.5	2962.3	3238.2	3304.9	0.1951	0.0534	0.0199	0.0125	0.0097
massdec-25-thigh	3672.3	2540.1	2630.6	3157.2	3373.8	3530.7	0.2727	0.0423	0.0072	0.0041	0.0029
massdec-25-foot	3835.8	2140.4	2675.5	3207.6	3574.2	3785.1	0.2089	0.0337	0.0093	0.0061	0.0042
massdec-50-foot	3892.0	1939.9	2452.0	3618.7	3705.9	3755.1	0.2996	0.0758	0.0266	0.0128	0.0082
massdec-50-leg	3487.1	2320.0	2813.4	3064.0	2999.5	3156.2	0.1889	0.0406	0.0166	0.0085	0.0046
massdec-50-thigh	3330.1	2491.2	2667.0	3296.4	3386.8	3389.6	0.2133	0.0564	0.0193	0.0117	0.0085
massdec-50-torso	3362.1	2741.1	2841.4	3264.9	3285.7	3367.9	0.1888	0.0387	0.0075	0.0046	0.003
massinc-100-leg	3247.4	2478.4	2368.9	3250.4	3226.3	3278.5	0.171	0.0314	0.009	0.0057	0.0039
massinc-100-foot	3051.3	2520.6	2474.2	3016.8	3029.5	3056.2	0.2499	0.0312	0.0054	0.003	0.002
massinc-100-thigh	3209.6	2169.3	2657.1	3261.2	3214.3	3207.1	0.1904	0.0368	0.0105	0.0065	0.0045
massinc-100-torso	3208.2	1994.0	1615.4	3255.2	3068.8	3052.5	0.2985	0.0379	0.0113	0.0063	0.0044
massinc-200-thigh	3375.3	2125.7	1964.6	3217.0	3366.6	3364.6	0.3019	0.0373	0.009	0.0051	0.0033
massinc-200-foot	2775.1	2362.7	2394.4	2695.9	2712.5	2815.5	0.3696	0.0526	0.0115	0.0073	0.0049
massinc-200-leg	3256.6	2449.1	2534.6	3310.4	3313.6	3312.6	0.2176	0.0533	0.0106	0.0068	0.0047
massinc-200-torso	3518.0	1748.4	2141.1	2791.8	3008.6	3331.6	0.2006	0.0229	0.0075	0.0044	0.0031
massinc-300-foot	2502.3	2302.4	2241.4	2514.4	2528.9	2508.5	0.2574	0.019	0.0029	0.0016	0.0011
massinc-300-leg	3174.2	2564.9	2534.8	3142.0	3155.9	3168.7	0.2353	0.0396	0.0093	0.0056	0.0038
massinc-300-torso	3295.9	1661.2	2374.6	3134.9	3096.8	3126.3	0.2524	0.0197	0.0033	0.0019	0.0013
massinc-300-thigh	3583.3	2060.0	2003.9	2515.2	2965.4	2876.5	0.4286	0.0261	0.0045	0.0027	0.0019
jointdec-25-foot	3287.7	2052.3	3185.7	3233.5	3152.0	3102.2	0.2122	0.0386	0.0096	0.0054	0.0037
jointdec-25-thigh	3645.5	2188.7	2773.1	3420.2	3333.0	3337.7	0.3004	0.0464	0.0107	0.006	0.0038
jointdec-25-leg	3395.4	2184.2	2793.0	3338.7	3362.8	3370.0	0.2384	0.0365	0.009	0.0057	0.0037
jointdec-50-thigh	3269.8	2197.8	2890.8	2899.7	3030.6	3122.2	0.1694	0.0477	0.0147	0.0087	0.0063
jointdec-50-foot	3553.6	2028.9	2743.3	3343.8	3493.2	3509.4	0.2838	0.0221	0.0042	0.0022	0.0014
jointdec-50-leg	3356.1	2183.7	2931.3	3214.6	3206.0	3216.8	0.2175	0.0611	0.0198	0.0123	0.0083
lengthdec-50-torso	3404.1	2120.1	2770.8	3319.0	3336.8	3351.7	0.3997	0.0577	0.0182	0.011	0.0078
lengthdec-50-foot	3532.6	2111.3	2100.8	3395.4	3511.5	3485.8	0.2283	0.0199	0.0044	0.0027	0.002
lengthinc-50-foot	3674.2	1981.0	2752.2	3431.3	3647.0	3574.8	0.2454	0.023	0.0049	0.0027	0.0019
lengthinc-100-foot	3698.2	1977.1	3448.4	3628.0	3694.6	3639.4	0.216	0.0201	0.0035	0.002	0.0013
lengthinc-100-torso	3451.8	2132.3	2652.2	3027.7	3342.1	3448.3	0.1687	0.0191	0.0042	0.0025	0.0016
lengthinc-150-foot	3740.8	1623.5	2341.2	3691.0	3758.2	3746.0	0.3692	0.0484	0.0123	0.0073	0.005
frictiondec-25-foot	3412.5	2148.8	2446.4	2996.2	3110.9	3257.7	0.2048	0.0736	0.022	0.0149	0.0104
frictiondec-50-foot	3600.0	2208.0	2670.5	3081.3	3230.9	3301.1	0.2952	0.066	0.0237	0.0142	0.0102
frictioninc-25-foot	3735.7	1971.7	3214.5	3514.6	3537.3	3737.3	0.2366	0.0332	0.0106	0.0083	0.005
frictioninc-50-foot	3212.6	2235.6	2594.8	3145.7	3147.9	3171.5	0.2571	0.0399	0.0053	0.0031	0.0021

Table 11: Meta-Learning in Halfcheetah. Comparing average reward/loss per shot.

	Reward						Loss				
	target	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**
massdec-25-ffoot	2607.3	2091.4	1969.0	2299.4	2397.5	2450.3	0.1772	0.0401	0.0149	0.0103	0.0077
massdec-25-bshin	2533.6	2239.9	2292.2	2397.7	2446.6	2471.4	0.079	0.0239	0.0089	0.006	0.0047
massdec-25-bthigh	2617.1	2338.8	2430.1	2523.0	2563.6	2593.9	0.0996	0.0271	0.0109	0.0075	0.0057
massdec-25-bfoot	2511.3	2217.2	2266.7	2407.0	2449.3	2456.1	0.0683	0.02	0.0084	0.0057	0.0043
massdec-25-fthigh	2647.0	2297.7	2056.5	2323.2	2379.5	2436.5	0.1812	0.039	0.0138	0.0097	0.0074
massdec-25-head	2986.3	2430.8	1873.7	2393.2	2510.5	2651.2	0.242	0.0529	0.0212	0.0144	0.0107
massdec-25-fshin	2163.6	2190.9	1877.1	2076.2	2120.2	2141.3	0.2117	0.0251	0.0075	0.0048	0.0035
massdec-50-bfoot	2480.1	2093.7	2049.4	2250.8	2331.0	2371.8	0.1723	0.0371	0.0146	0.0099	0.0073
massdec-50-head	3689.9	2051.4	1791.3	2213.1	2527.8	2728.5	0.3432	0.0552	0.0193	0.0128	0.0096
massdec-50-bthigh	2649.0	2381.0	2383.4	2504.2	2553.6	2574.7	0.1232	0.0277	0.0109	0.0073	0.0055
massdec-50-ffoot	2406.6	1582.6	1542.4	2011.4	2134.5	2232.3	0.2853	0.0533	0.0169	0.0111	0.0085
massdec-50-fshin	2400.8	2017.9	1825.6	2142.2	2243.6	2302.1	0.2621	0.0423	0.0149	0.01	0.0072
massdec-50-fthigh	2341.8	2258.1	1807.9	2066.3	2170.3	2227.5	0.2586	0.058	0.0201	0.0126	0.0094
massdec-50-bshin	2412.3	2126.5	2057.6	2303.5	2310.9	2355.2	0.1067	0.0267	0.0104	0.0068	0.0052
massinc-100-ffoot	1313.3	2008.4	77.3	2523.7	2770.6	2916.5	0.8706	0.5251	0.0609	0.0425	0.0318
massinc-100-fshin	1884.1	2320.3	1260.0	1761.3	1802.8	1840.2	0.8468	0.0624	0.0228	0.0105	0.0048
massinc-100-bshin	4852.2	2161.8	377.6	2058.4	2524.8	2597.0	0.8641	0.19	0.0696	0.0498	0.0374
massinc-100-bthigh	3191.0	1995.3	-207.9	228.5	851.0	1245.6	1.0144	0.1633	0.0589	0.0409	0.0286
massinc-100-bfoot	34.8	2305.1	-266.0	924.4	1323.4	1998.9	0.9084	0.6825	0.0744	0.0522	0.0415
massinc-100-head	957.1	1033.5	855.7	759.1	860.0	904.7	0.3247	0.0328	0.0085	0.0048	0.0034
massinc-100-fthigh	2224.8	2057.8	1584.0	1931.5	2010.1	2044.1	0.3478	0.0708	0.0246	0.0169	0.0123
massinc-200-bshin	4438.5	1417.5	56.0	1858.7	2534.5	3021.4	0.9815	0.1693	0.0974	0.0927	0.036
massinc-200-ffoot	3284.7	1161.1	-217.7	1809.2	2587.6	2659.9	0.8368	0.2557	0.0807	0.0596	0.0471
massinc-200-bthigh	5637.0	1376.2	-95.7	888.0	2838.2	4799.0	1.099	0.117	0.0261	0.0164	0.0118
massinc-200-fthigh	1631.2	1602.3	915.0	1413.4	1504.5	1564.7	0.8639	0.054	0.0155	0.0076	0.0056
massinc-200-fshin	34.3	2035.7	-260.8	712.7	2042.4	2405.6	0.6256	0.4306	0.0623	0.0427	0.0282
massinc-200-bfoot	3926.0	1696.3	-195.9	1287.8	2297.1	3169.9	1.0048	0.2874	0.1347	0.0681	0.0471
massinc-200-head	130.6	490.9	-7.0	106.4	7.3	93.3	0.6566	0.1731	0.05	0.0361	0.0331
massinc-300-head	-80.5	181.9	-182.0	-63.7	-67.7	-73.2	0.8858	0.1414	0.0425	0.0305	0.0244
massinc-300-fthigh	1351.2	1281.3	64.3	755.3	1150.1	1315.5	1.2373	0.1008	0.0317	0.021	0.0149
massinc-300-fshin	2896.7	1678.7	1821.7	2395.2	2614.0	2682.3	0.4753	0.0859	0.0304	0.0202	0.0152
massinc-300-bfoot	2352.2	801.2	1108.5	674.6	1240.6	1943.2	0.9057	0.2962	0.1246	0.0974	0.0514
massinc-300-bshin	4637.7	195.5	-230.2	872.7	2384.7	3984.5	0.9805	0.0884	0.0258	0.0167	0.0116
massinc-300-bthigh	5771.7	324.5	126.7	2142.6	4490.1	4972.7	1.1136	0.0992	0.0204	0.011	0.0072
massinc-300-ffoot	2484.9	651.6	1889.4	2212.0	2274.5	2363.6	0.4766	0.0942	0.0328	0.0222	0.0164
jointdec-25-bshin	2530.5	528.1	653.0	2288.6	2427.5	2471.4	0.1563	0.024	0.0085	0.0055	0.0039
jointdec-25-bfoot	2525.1	511.5	1614.8	2246.4	2387.4	2478.5	0.1522	0.0237	0.0092	0.0062	0.0046
jointdec-25-fshin	2291.3	529.6	576.4	1987.2	2126.1	2161.2	0.3085	0.0371	0.0121	0.0076	0.0056
jointdec-25-ffoot	2621.1	676.9	1856.7	2406.5	2536.4	2565.4	0.1007	0.0288	0.0106	0.0071	0.0051
jointdec-25-bthigh	2513.6	345.1	-52.9	1526.8	2254.9	2345.8	0.1984	0.0328	0.0128	0.0084	0.0062
jointdec-25-fthigh	2426.4	237.7	239.0	2075.1	2283.9	2356.1	0.2311	0.0346	0.0115	0.0075	0.0052
jointdec-50-fthigh	2175.1	8.9	1045.6	1939.5	2056.5	2086.6	0.3103	0.0482	0.0174	0.0114	0.0082
jointdec-50-bshin	2457.2	243.1	472.6	2253.9	2285.3	2416.2	0.2125	0.0277	0.0079	0.005	0.0037
jointdec-50-ffoot	2604.9	561.9	1196.3	2256.1	2409.0	2468.5	0.14	0.0287	0.0109	0.0073	0.0054
jointdec-50-fshin	2621.3	444.8	1842.5	1988.5	2269.4	2386.4	0.1735	0.0339	0.0128	0.0089	0.0066
jointdec-50-bthigh	2118.8	343.3	1064.3	1929.2	2008.3	2032.7	0.3423	0.0419	0.0105	0.0056	0.0037
jointdec-50-bfoot	2559.5	484.9	1863.7	2264.5	2427.7	2497.5	0.1272	0.0277	0.0113	0.0076	0.0054
lengthdec-50-head	2536.7	792.2	1139.4	1968.8	2250.3	2372.5	0.2166	0.0352	0.0125	0.0081	0.0059
lengthinc-50-bfoot	2522.1	833.3	216.9	2076.6	2451.1	2510.4	0.1523	0.027	0.0095	0.0063	0.0045
lengthinc-50-ffoot	2532.5	617.6	212.5	2117.1	2380.2	2423.5	0.1597	0.0365	0.0128	0.0087	0.0065
lengthinc-50-head	2566.2	796.4	1534.4	1999.6	2446.0	2494.8	0.1768	0.0262	0.0099	0.0066	0.0049
lengthinc-100-ffoot	2627.8	610.5	275.3	2137.6	2434.9	2471.2	0.1635	0.0282	0.0107	0.0071	0.0052
lengthinc-100-bfoot	2586.0	877.2	1471.4	2133.0	2402.6	2467.7	0.1438	0.0319	0.0125	0.0085	0.0064

Table 12: Meta-Learning in Walker2d. Comparing average reward/loss per shot.

	Reward						Loss				
	target	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**
massdec-25-leg	4820.6	2867.9	1974.4	2831.6	4732.3	4861.4	0.1734	0.0173	0.0037	0.0022	0.0018
massdec-25-thighleft	5112.1	2870.7	1887.8	3636.5	4373.1	4567.1	0.1842	0.0671	0.03	0.0226	0.0185
massdec-25-torso	4652.4	2790.7	1792.2	3769.5	4073.7	4496.5	0.2022	0.0638	0.0225	0.0145	0.0094
massdec-25-foot	5863.6	2923.0	1836.8	3677.2	5049.9	5490.7	0.2058	0.0308	0.0083	0.0051	0.0036
massdec-25-footleft	4814.2	2869.7	1343.7	4123.9	4489.7	4714.5	0.3206	0.0267	0.0045	0.0026	0.0017
massdec-25-legleft	4049.4	2874.2	1962.7	3187.8	3884.8	4122.9	0.2983	0.0659	0.0198	0.0144	0.0115
massdec-25-thigh	4533.1	2836.6	3165.5	3865.3	4182.2	4337.1	0.1819	0.0481	0.0072	0.0042	0.0033
massdec-50-thigh	4407.7	2774.6	2671.1	3979.3	4331.7	4354.6	0.2126	0.0894	0.0409	0.0251	0.0188
massdec-50-torso	4491.8	2796.9	2420.4	3827.3	4185.1	4729.4	0.1796	0.0401	0.0135	0.008	0.0052
massdec-50-footleft	4610.8	2853.8	2011.0	3968.2	4460.6	4575.7	0.1942	0.0387	0.0079	0.0055	0.004
massdec-50-foot	5664.3	2879.8	2632.7	4194.5	5235.6	5540.5	0.2631	0.0239	0.0069	0.0041	0.0027
massdec-50-leg	5004.0	2853.1	2211.5	4227.8	4710.6	4809.7	0.1915	0.0675	0.0212	0.0145	0.0126
massdec-50-legleft	4346.1	2940.5	2082.7	4309.5	4517.7	4558.7	0.189	0.0394	0.0103	0.0064	0.0055
massdec-50-thighleft	4117.1	2965.6	1572.9	3926.8	4086.9	4130.4	0.2214	0.0248	0.0049	0.0033	0.0022
massinc-100-thighleft	4062.8	2621.1	443.4	3459.2	3699.4	3986.9	0.2057	0.0344	0.0079	0.0052	0.003
massinc-100-leg	4305.0	2656.2	1134.9	3888.3	4160.4	4267.4	0.1946	0.0214	0.0043	0.002	0.0013
massinc-100-foot	4145.7	2697.7	2297.3	3600.0	4045.2	4214.0	0.331	0.0869	0.0257	0.0182	0.0139
massinc-100-thigh	4067.2	2591.9	1649.3	2918.6	3907.0	3987.2	0.2258	0.0567	0.0151	0.0095	0.0063
massinc-100-legleft	4101.4	2682.4	1761.4	3642.9	3929.9	3899.6	0.2958	0.0382	0.008	0.0046	0.0029
massinc-100-torso	4506.9	2446.0	1790.1	3029.8	3922.7	4216.7	0.2231	0.1006	0.0209	0.0128	0.0086
massinc-100-footleft	4524.0	2670.2	2602.5	3475.4	4103.4	4442.8	0.1798	0.0382	0.0161	0.009	0.0058
massinc-200-legleft	3910.2	2580.2	2257.7	3616.6	3682.4	3927.6	0.1959	0.0727	0.0272	0.014	0.0092
massinc-200-torso	4261.9	2242.0	1065.9	3503.3	3914.3	4098.5	0.2275	0.0461	0.0103	0.0059	0.004
massinc-200-footleft	4003.7	2591.5	2321.2	2503.5	3504.8	3731.4	0.2298	0.1196	0.0298	0.021	0.0166
massinc-200-leg	4107.1	2450.6	2752.6	3136.1	3555.6	3490.8	0.2211	0.0644	0.0147	0.0089	0.0063
massinc-200-thighleft	4638.5	2500.6	1271.4	3979.0	4381.2	4549.0	0.1638	0.0335	0.0075	0.0048	0.0038
massinc-200-thigh	3936.4	2103.3	2864.1	3409.4	3737.7	3808.4	0.2304	0.0929	0.0273	0.0169	0.012
massinc-200-foot	3899.0	2603.8	1226.2	2978.4	3819.3	3904.4	0.3242	0.0464	0.0117	0.008	0.0055
massinc-300-thighleft	3625.1	2153.9	2632.9	3279.8	3603.1	3618.7	0.1757	0.0364	0.0084	0.0048	0.0039
massinc-300-legleft	3714.1	2504.5	2774.2	3525.5	3677.9	3704.6	0.1771	0.0226	0.0049	0.0025	0.0018
massinc-300-foot	3562.2	2403.1	2829.0	3133.7	3244.9	3426.6	0.1664	0.0423	0.0164	0.0093	0.005
massinc-300-footleft	3550.7	2590.2	2589.2	3111.9	3410.7	3496.5	0.2289	0.073	0.0277	0.0194	0.0142
massinc-300-leg	4658.2	2098.8	2489.2	3262.3	3669.2	3965.3	0.2769	0.0675	0.0184	0.0116	0.0075
massinc-300-thigh	4472.5	1416.5	1666.5	3769.3	4334.1	4211.6	0.2745	0.0541	0.018	0.011	0.0075
massinc-300-torso	5121.0	1999.3	1648.1	4293.8	4669.1	4932.6	0.2265	0.0495	0.0086	0.005	0.0033
jointdec-25-leg	4681.4	2319.4	2032.0	3284.2	3594.4	4340.8	0.2708	0.0649	0.0203	0.0136	0.0096
jointdec-25-thighleft	4414.2	2331.6	2369.2	3389.5	3915.8	4089.8	0.2157	0.0453	0.0174	0.0101	0.0062
jointdec-25-legleft	4102.6	2448.5	2852.4	3820.7	4000.2	4136.8	0.2345	0.0358	0.0104	0.0075	0.0048
jointdec-25-thigh	4215.9	2303.5	3156.6	4078.1	4075.6	4116.2	0.163	0.0392	0.0124	0.0083	0.0057
jointdec-25-footleft	4660.7	2556.3	2875.6	3497.6	3965.4	4476.5	0.1838	0.0379	0.0136	0.0099	0.0075
jointdec-25-foot	4864.2	2507.5	2755.1	2976.2	3925.5	4369.8	0.2098	0.0136	0.0023	0.0014	0.0009
jointdec-50-footleft	4590.2	2068.6	3004.4	3797.7	4242.0	4523.2	0.1695	0.0262	0.0088	0.0062	0.0029
jointdec-50-foot	4510.1	2317.8	2593.3	3504.5	3798.1	4032.3	0.2182	0.0267	0.0057	0.0034	0.0023
jointdec-50-thighleft	4545.9	2304.5	2629.2	3270.7	4069.7	4546.1	0.2343	0.0545	0.0171	0.0105	0.0078
jointdec-50-thigh	4917.7	2356.9	2152.1	2823.6	4286.5	4599.3	0.2201	0.0295	0.0089	0.0062	0.0049
jointdec-50-leg	4203.2	2361.1	2529.0	3612.9	3857.1	4098.6	0.2231	0.0557	0.0199	0.0115	0.0084
jointdec-50-legleft	4158.8	2522.2	2271.0	3575.7	3787.0	3905.9	0.2195	0.0697	0.0238	0.0146	0.0115
lengthdec-50-foot	4359.1	2615.8	2407.5	3181.7	3998.8	4037.7	0.2644	0.0635	0.021	0.0128	0.0088
lengthdec-50-torso	3849.3	2321.7	2095.0	3786.9	3855.2	3858.0	0.1279	0.0279	0.0101	0.0079	0.0047
lengthdec-50-footleft	4171.7	2636.5	2055.0	3730.9	3976.4	4106.0	0.3254	0.0528	0.0186	0.0119	0.0077
lengthinc-100-foot	4414.6	2143.4	2178.2	3695.6	4052.4	4270.5	0.2197	0.068	0.0222	0.0128	0.008
lengthinc-100-torso	5613.6	2379.5	2192.8	3922.6	4341.5	5151.5	0.235	0.0212	0.0038	0.002	0.0014
lengthinc-100-footleft	4485.5	2489.1	1923.2	2509.2	3464.3	4265.8	0.2273	0.0552	0.0137	0.0088	0.0067
lengthinc-200-torso	4238.7	2281.1	1692.4	3374.5	3528.6	3879.6	0.3018	0.0217	0.0037	0.002	0.0013
lengthinc-200-footleft	3999.8	2449.9	1572.6	3192.7	3551.4	3811.1	0.2449	0.0221	0.0041	0.0023	0.0016
lengthinc-200-foot	4183.9	2225.1	3021.0	3870.1	4162.8	4182.5	0.2702	0.0599	0.0216	0.012	0.0094
frictiondec-25-footleft	4104.4	2132.3	2155.7	3391.3	3699.7	3965.7	0.2231	0.0463	0.0133	0.0075	0.0052
frictiondec-25-foot	3841.3	1993.7	1502.5	3162.0	3901.2	3864.1	0.3065	0.0999	0.0358	0.0252	0.0187
frictiondec-50-footleft	3919.6	2238.3	2567.6	2821.7	3343.2	3748.8	0.3127	0.0878	0.036	0.0278	0.0229
frictiondec-50-foot	4118.6	1985.3	1953.3	3153.1	3522.0	3921.0	0.2884	0.0601	0.0192	0.0127	0.0087
frictioninc-25-footleft	4338.9	2627.1	1772.8	3578.0	3948.2	4501.4	0.2376	0.1136	0.0308	0.0207	0.0142
frictioninc-25-foot	4393.1	2068.4	2485.9	3836.4	4167.6	4199.4	0.2888	0.0965	0.0359	0.0239	0.0167
frictioninc-50-footleft	4488.4	2599.0	2749.8	3780.4	4167.2	4318.3	0.2174	0.0439	0.0108	0.0063	0.0046
frictioninc-50-foot	4560.0	2118.6	2170.6	3117.3	4105.0	4513.6	0.2325	0.0697	0.0154	0.0089	0.0057

Table 13: Multi-Task in Hopper. Comparing average reward/loss per shot.

	Reward						Loss				
	target	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**
massdec-25-torso	3337.4	2967.8	2879.6	3241.3	3281.9	3074.5	0.2117	0.0517	0.016	0.0093	0.0071
massdec-25-leg	3429.9	3344.1	2743.1	3295.9	3270.3	3291.5	0.2608	0.057	0.012	0.0057	0.0034
massdec-25-foot	3835.8	3204.8	2700.5	3539.8	3665.0	3798.9	0.1791	0.0268	0.0064	0.0044	0.003
massdec-25-thigh	3672.3	3187.3	2846.4	3418.9	3359.9	3462.6	0.3035	0.0319	0.0047	0.0028	0.0021
massdec-50-leg	3487.1	3390.7	2757.8	2812.1	3225.4	3301.7	0.1792	0.0398	0.0139	0.0065	0.0033
massdec-50-torso	3362.1	2711.8	2962.0	3341.3	3364.0	3358.1	0.1899	0.0297	0.0043	0.0026	0.0018
massdec-50-foot	3892.0	3169.0	2019.4	3437.0	3750.9	3790.5	0.2441	0.0659	0.0223	0.0104	0.0066
massdec-50-thigh	3330.1	2952.0	2391.0	3377.0	3368.0	3377.7	0.2135	0.0511	0.0145	0.0081	0.006
massinc-100-torso	3208.2	2895.0	2145.8	3166.3	3137.8	3202.3	0.3389	0.0332	0.0072	0.004	0.0028
massinc-100-foot	3051.3	2628.2	2584.3	3049.4	3059.2	3059.8	0.2638	0.0271	0.0042	0.0023	0.0016
massinc-100-thigh	3209.6	3004.9	2914.9	3195.9	3171.4	3211.8	0.2207	0.0294	0.007	0.0043	0.003
massinc-100-leg	3247.4	3019.7	2769.7	3203.1	3250.4	3268.8	0.1827	0.0239	0.0058	0.004	0.0027
massinc-200-foot	2775.1	2129.8	2419.8	2679.5	2777.2	2812.1	0.4498	0.0507	0.0086	0.0055	0.0039
massinc-200-thigh	3375.3	2628.9	2154.6	3314.8	3355.6	3360.3	0.3163	0.0298	0.0052	0.003	0.0021
massinc-200-torso	3518.0	2369.4	2483.4	3077.7	3166.3	3384.9	0.1749	0.0181	0.0055	0.0029	0.002
massinc-200-leg	3256.6	2833.6	2545.4	3303.3	3285.0	3294.1	0.2173	0.0544	0.0074	0.0044	0.0031
massinc-300-foot	2502.3	2212.9	2565.5	2513.8	2501.7	2515.0	0.3016	0.0177	0.0022	0.0012	0.0008
massinc-300-thigh	3583.3	2463.2	2014.7	2811.7	2967.4	3048.7	0.4246	0.0172	0.003	0.0018	0.0012
massinc-300-torso	3295.9	1994.6	2238.9	3111.2	3195.1	3064.5	0.246	0.012	0.002	0.0011	0.0008
massinc-300-leg	3174.2	2741.1	2838.8	3181.3	3164.3	3187.5	0.2435	0.0344	0.006	0.0035	0.0025
jointdec-25-thigh	3645.5	3107.3	2540.0	3292.4	3384.1	3511.7	0.2484	0.0315	0.0073	0.0036	0.0022
jointdec-25-leg	3395.4	3105.9	2672.5	3284.1	3335.3	3370.5	0.208	0.0227	0.0059	0.0037	0.0021
jointdec-25-foot	3287.7	2930.8	2846.1	3132.4	3303.2	3310.1	0.2046	0.0279	0.0057	0.0031	0.0021
jointdec-50-leg	3356.1	3102.1	2964.9	3297.5	3264.6	3210.5	0.2208	0.0566	0.0141	0.0084	0.0055
jointdec-50-thigh	3269.8	3118.9	2788.2	3093.1	3214.3	3221.5	0.1978	0.041	0.0096	0.0055	0.0041
jointdec-50-foot	3553.6	2025.9	2790.7	3405.9	3530.2	3547.9	0.2698	0.0151	0.002	0.0011	0.0008
lengthdec-50-torso	3404.1	3284.2	2809.8	3326.5	3372.7	3372.4	0.3054	0.0597	0.0126	0.0074	0.0053
lengthdec-50-foot	3532.6	2885.4	2430.2	3296.6	3479.8	3443.9	0.2093	0.0145	0.0027	0.0016	0.0012
lengthinc-50-foot	3674.2	2652.9	2321.2	3518.0	3639.1	3667.9	0.2645	0.0184	0.0027	0.0016	0.0012
lengthinc-100-foot	3698.2	2283.9	2918.6	3595.5	3684.2	3669.4	0.3034	0.0175	0.0021	0.0012	0.0009
lengthinc-100-torso	3451.8	3211.9	2014.4	3052.4	3366.4	3050.5	0.1734	0.0138	0.0023	0.0014	0.001
lengthinc-150-foot	3740.8	1907.6	2629.5	3742.6	3764.6	3766.5	0.5187	0.0473	0.0082	0.0053	0.0036
frictiondec-25-foot	3412.5	2956.5	2176.4	2960.8	3312.0	3195.4	0.19	0.0682	0.0164	0.0109	0.0074
frictiondec-50-foot	3600.0	2722.5	2695.2	3164.1	3222.2	3382.5	0.2787	0.0609	0.0184	0.0101	0.0069
frictioninc-25-foot	3735.7	3138.1	3036.4	3542.8	3708.6	3714.0	0.1714	0.0255	0.0079	0.0065	0.0037
frictioninc-50-foot	3212.6	2631.1	2543.1	3027.0	3136.0	3159.6	0.2098	0.0308	0.0033	0.002	0.0015

Table 14: Multi-Task in Halfcheetah. Comparing average reward/loss per shot.

	Reward						Loss				
	target	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**
massdec-25-bshin.xml	2533.6	2141.7	2180.9	2406.0	2448.2	2486.2	0.0786	0.022	0.007	0.0046	0.0036
massdec-25-ffoot.xml	2607.3	1926.3	1778.5	2370.2	2466.2	2477.4	0.1407	0.0329	0.011	0.0073	0.0054
massdec-25-head.xml	2986.3	2237.0	1963.4	2423.8	2666.6	2822.5	0.2249	0.0454	0.016	0.0104	0.0075
massdec-25-bthigh.xml	2617.1	2142.1	2342.8	2544.1	2598.1	2594.2	0.086	0.024	0.0082	0.0054	0.004
massdec-25-fshin.xml	2163.6	1887.7	1867.3	2118.4	2145.4	2155.5	0.0971	0.0189	0.0054	0.0034	0.0025
massdec-25-fthigh.xml	2647.0	2105.7	1945.0	2443.9	2532.2	2575.6	0.1561	0.0314	0.0104	0.0068	0.0052
massdec-25-bfoot.xml	2511.3	2068.6	2216.3	2416.9	2472.2	2486.3	0.0635	0.0177	0.0061	0.0039	0.0029
massdec-50-fthigh.xml	2341.8	2102.0	1726.1	2133.4	2246.6	2278.3	0.2136	0.0492	0.0147	0.0093	0.0069
massdec-50-head.xml	3689.9	1300.9	1600.2	2047.0	2304.5	3043.3	0.2993	0.0482	0.0148	0.0096	0.0069
massdec-50-bfoot.xml	2480.1	2080.4	2114.3	2272.2	2372.5	2405.3	0.1179	0.03	0.0106	0.0071	0.0053
massdec-50-ffoot.xml	2406.6	1309.9	1516.5	2101.8	2219.4	2275.6	0.2831	0.046	0.0133	0.0084	0.006
massdec-50-bshin.xml	2412.3	2169.4	1843.6	2290.2	2325.1	2385.2	0.1088	0.0252	0.0083	0.0051	0.0039
massdec-50-fshin.xml	2400.8	1756.9	1760.3	1945.0	2170.5	2328.7	0.2622	0.0385	0.0114	0.0072	0.005
massdec-50-bthigh.xml	2649.0	2299.5	2384.9	2535.4	2485.4	2570.2	0.0988	0.0244	0.0087	0.0055	0.0041
massinc-100-fthigh.xml	2224.8	1609.2	1244.8	1881.4	2006.8	1988.1	0.3459	0.0567	0.0197	0.013	0.0094
massinc-100-head.xml	957.1	627.5	629.1	698.3	841.6	819.3	0.3653	0.0249	0.0066	0.0039	0.0027
massinc-100-bthigh.xml	3191.0	1192.6	-205.5	163.9	1000.4	1296.8	0.9885	0.1636	0.0538	0.0356	0.0251
massinc-100-bshin.xml	4852.2	1073.0	755.6	1595.4	1995.8	2390.0	1.0102	0.1914	0.0672	0.0469	0.0353
massinc-100-fshin.xml	1884.1	1657.9	1165.2	1723.1	1838.8	1850.6	0.8016	0.0566	0.0169	0.0071	0.005
massinc-100-ffoot.xml	1313.3	1611.1	326.5	2531.9	2927.0	2930.1	0.9712	0.6305	0.0581	0.0387	0.0283
massinc-100-bfoot.xml	34.8	1194.0	-262.6	1212.4	1099.7	2135.5	0.9677	0.7375	0.0726	0.049	0.0377
massinc-200-fthigh.xml	1631.2	956.4	889.2	1473.6	1563.5	1608.0	0.7948	0.0736	0.0121	0.0082	0.0051
massinc-200-head.xml	130.6	298.0	-80.2	54.5	58.5	112.3	0.7218	0.2351	0.0475	0.0334	0.0307
massinc-200-bfoot.xml	3926.0	131.1	377.3	1829.9	2579.9	3322.2	1.0497	0.3014	0.1325	0.0608	0.0418
massinc-200-bthigh.xml	5637.0	392.2	-132.1	634.4	3361.0	4985.0	1.1786	0.1078	0.0272	0.0161	0.011
massinc-200-bshin.xml	4438.5	-57.0	331.2	983.7	2254.3	2984.3	1.013	0.1757	0.0967	0.0864	0.0333
massinc-200-fshin.xml	34.3	1335.5	-254.3	901.1	1866.0	2774.7	0.8602	0.4783	0.0613	0.0402	0.0262
massinc-200-ffoot.xml	3284.7	1035.4	674.5	1989.1	2678.1	2815.0	0.887	0.2345	0.0773	0.0562	0.0434
massinc-300-bthigh.xml	5771.7	104.7	-148.6	1281.4	3700.0	5270.2	1.1868	0.0871	0.0221	0.0118	0.0073
massinc-300-bshin.xml	4637.7	-274.0	-145.5	630.5	2872.4	4198.3	1.1516	0.1041	0.0276	0.016	0.0108
massinc-300-bfoot.xml	2352.2	-201.4	1014.6	769.9	1151.2	1874.1	0.9461	0.283	0.1276	0.1023	0.0475
massinc-300-fshin.xml	2896.7	794.3	1625.6	1975.9	2512.1	2685.7	0.4376	0.0818	0.0261	0.0169	0.013
massinc-300-fthigh.xml	1351.2	423.4	-141.6	841.4	1182.9	1298.3	1.3	0.1084	0.0291	0.019	0.0134
massinc-300-head.xml	-80.5	92.0	-223.0	-122.7	-74.7	-101.3	0.9513	0.2124	0.0414	0.029	0.0226
massinc-300-ffoot.xml	2484.9	443.1	1663.9	2125.3	2322.1	2399.8	0.5282	0.09	0.0298	0.019	0.0137
jointdec-25-fthigh.xml	2426.4	787.1	798.7	2228.4	2343.8	2400.9	0.1331	0.0245	0.0076	0.0048	0.0034
jointdec-25-bthigh.xml	2513.6	1289.1	288.9	2028.6	2476.6	2500.4	0.1206	0.0256	0.0085	0.0054	0.0041
jointdec-25-fshin.xml	2291.3	1058.2	1626.9	2097.1	2180.3	2217.5	0.1178	0.0243	0.008	0.005	0.0035
jointdec-25-bshin.xml	2530.5	1094.1	-15.2	1861.6	2415.2	2510.6	0.0981	0.0165	0.0054	0.0035	0.0025
jointdec-25-bfoot.xml	2525.1	1103.9	956.7	2412.4	2477.1	2509.0	0.0862	0.0186	0.0064	0.0042	0.003
jointdec-25-ffoot.xml	2621.1	1286.3	1430.0	2509.6	2590.6	2575.9	0.0717	0.021	0.0065	0.0043	0.0031
jointdec-50-fshin.xml	2621.3	1067.7	1477.3	2124.2	2495.9	2544.5	0.125	0.0247	0.0089	0.0058	0.0043
jointdec-50-ffoot.xml	2604.9	1204.2	1128.0	2408.5	2508.2	2555.9	0.0837	0.0219	0.0078	0.0053	0.0038
jointdec-50-bfoot.xml	2559.5	1215.5	1323.3	2224.2	2491.2	2528.7	0.0804	0.0217	0.0081	0.0053	0.0037
jointdec-50-bshin.xml	2457.2	585.0	200.0	1825.3	2263.2	2427.2	0.1733	0.0217	0.0057	0.0034	0.0025
jointdec-50-fthigh.xml	2175.1	688.1	1219.8	2021.4	2095.4	2132.0	0.2339	0.038	0.0129	0.0088	0.0066
jointdec-50-bthigh.xml	2118.8	1164.3	773.8	1954.7	2057.6	2069.4	0.1947	0.0311	0.0065	0.0036	0.0023
lengthdec-50-head.xml	2536.7	1571.3	1948.4	2338.1	2429.2	2488.9	0.1114	0.0238	0.0083	0.0054	0.0039
lengthinc-50-ffoot.xml	2532.5	1075.0	1410.8	2411.7	2503.7	2519.8	0.0968	0.0257	0.0092	0.006	0.0044
lengthinc-50-head.xml	2566.2	2178.2	1667.4	2473.6	2530.5	2552.3	0.1173	0.0209	0.0075	0.005	0.0036
lengthinc-50-bfoot.xml	2522.1	1442.1	1492.2	2301.6	2490.0	2531.5	0.0942	0.0191	0.0063	0.0038	0.0028
lengthinc-100-bfoot.xml	2586.0	1589.9	1375.6	2273.1	2503.6	2548.2	0.1168	0.0255	0.009	0.006	0.0044
lengthinc-100-ffoot.xml	2627.8	1081.6	1320.3	2452.2	2542.5	2558.8	0.1174	0.0225	0.0076	0.0048	0.0035

Table 15: Multi-Task in Walker2d. Comparing average reward/loss per shot.

	Reward						Loss				
	target	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**	**0-shot**	**1-shot**	**10-shot**	**25-shot**	**50-shot**
massdec-25-foot	5863.6	2375.7	2604.5	3995.7	5663.6	5797.3	0.2178	0.0304	0.0073	0.0043	0.0029
massdec-25-footleft	4814.2	2473.8	2003.9	4300.5	4775.4	4799.9	0.3992	0.0222	0.0033	0.0018	0.0012
massdec-25-leg	4820.6	2579.2	2464.5	4149.1	4776.0	4850.4	0.1617	0.0144	0.0028	0.0015	0.0013
massdec-25-thighleft	5112.1	2775.7	1324.3	4168.4	4668.5	4808.9	0.1777	0.0654	0.0281	0.0209	0.0173
massdec-25-legleft	4049.4	2457.4	1289.4	3439.4	3978.5	4030.6	0.285	0.0597	0.0172	0.0127	0.0101
massdec-25-thigh	4533.1	2793.6	3101.6	4267.2	4481.7	4523.1	0.1605	0.0466	0.0056	0.0031	0.0026
massdec-25-torso	4652.4	2714.2	1550.8	4261.2	4483.7	4619.8	0.2093	0.0631	0.0202	0.0124	0.0077
massdec-50-thighleft	4117.1	3135.3	2116.2	3937.7	4063.9	4125.7	0.2289	0.0225	0.0039	0.0027	0.0017
massdec-50-torso	4491.8	3037.4	1976.2	4161.5	4530.0	4679.7	0.1854	0.0377	0.012	0.0065	0.0042
massdec-50-thigh	4407.7	3039.3	2705.4	4121.9	4350.1	4443.0	0.227	0.0884	0.0378	0.0218	0.0158
massdec-50-foot	5664.3	2606.2	1685.0	4937.8	5502.8	5610.8	0.2813	0.0234	0.006	0.0034	0.0022
massdec-50-leg	5004.0	2911.2	1666.0	4067.0	4655.6	4896.2	0.2021	0.0678	0.0186	0.0127	0.0109
massdec-50-footleft	4610.8	2484.2	3112.2	4122.0	4512.3	4652.1	0.1748	0.0324	0.0059	0.0042	0.003
massdec-50-legleft	4346.1	2701.1	1710.2	4497.8	4557.9	4546.6	0.2043	0.0343	0.0082	0.005	0.0044
massinc-100-leg	4305.0	1597.8	1795.5	4129.8	4261.9	4261.3	0.244	0.0178	0.0033	0.0014	0.0009
massinc-100-legleft	4101.4	2337.1	2124.8	3986.5	4046.2	4112.0	0.2509	0.0272	0.0057	0.0033	0.0021
massinc-100-foot	4145.7	1807.4	2122.0	3850.3	4152.1	4324.9	0.3182	0.0807	0.0219	0.0151	0.0114
massinc-100-footleft	4524.0	2528.2	2480.9	3822.5	4303.9	4507.8	0.1686	0.0365	0.0147	0.0076	0.0047
massinc-100-thigh	4067.2	1731.4	1824.8	3363.9	3538.3	3958.0	0.22	0.048	0.0107	0.0066	0.0044
massinc-100-thighleft	4062.8	1721.6	1867.2	3808.4	4037.8	4050.7	0.2017	0.024	0.0048	0.0033	0.0019
massinc-100-torso	4506.9	1970.6	1817.1	3770.8	4342.4	4407.7	0.208	0.0891	0.0161	0.0094	0.0063
massinc-200-torso	4261.9	1702.5	2571.6	3988.0	4102.9	4208.7	0.2391	0.0393	0.0076	0.0042	0.0028
massinc-200-leg	4107.1	1787.8	2890.7	3690.0	4041.3	4080.8	0.2086	0.0552	0.0109	0.0063	0.0045
massinc-200-footleft	4003.7	2336.5	2121.4	3415.6	3795.4	3826.8	0.2164	0.1141	0.0268	0.0181	0.0139
massinc-200-thigh	3936.4	1330.7	2989.1	3807.7	3844.7	3857.4	0.2385	0.0802	0.0216	0.0131	0.0092
massinc-200-legleft	3910.2	2054.4	2797.0	3589.9	3880.3	3894.2	0.2113	0.067	0.0219	0.0107	0.007
massinc-200-foot	3899.0	1926.1	2290.9	3649.1	3902.7	3947.6	0.3363	0.0429	0.0101	0.0068	0.0044
massinc-200-thighleft	4638.5	1625.6	3208.2	4213.4	4618.6	4666.1	0.1566	0.0275	0.005	0.0034	0.0028
massinc-300-foot	3562.2	2072.9	2723.3	3428.8	3514.8	3470.1	0.1608	0.0343	0.0124	0.007	0.0034
massinc-300-thighleft	3625.1	1551.4	2986.9	3527.3	3626.9	3649.2	0.1856	0.0288	0.0061	0.0034	0.003
massinc-300-leg	4658.2	1831.5	1695.6	3640.4	4194.5	4322.2	0.2718	0.0613	0.0146	0.0087	0.0055
massinc-300-legleft	3714.1	1988.0	3186.0	3671.7	3704.5	3664.8	0.1805	0.0176	0.0036	0.0017	0.0013
massinc-300-torso	5121.0	1493.8	2161.1	4658.9	4958.9	5005.7	0.2362	0.0492	0.0061	0.0035	0.0023
massinc-300-footleft	3550.7	2150.7	2423.3	3293.8	3514.7	3537.4	0.2352	0.0678	0.0231	0.0153	0.0111
massinc-300-thigh	4472.5	1326.0	2538.2	4208.5	4359.6	4430.2	0.2689	0.0508	0.0156	0.0091	0.006
jointdec-25-foot	4864.2	3447.3	3074.9	4023.4	4375.5	4614.7	0.1836	0.0086	0.0014	0.0009	0.0007
jointdec-25-footleft	4660.7	3128.1	2572.2	3835.9	4272.6	4598.6	0.1566	0.0345	0.0121	0.0082	0.0062
jointdec-25-thighleft	4414.2	3382.6	1409.5	3954.9	4216.4	4236.8	0.187	0.0403	0.0142	0.008	0.0048
jointdec-25-leg	4681.4	3343.1	2654.9	3555.5	4361.2	4471.5	0.2643	0.0613	0.0176	0.0113	0.0077
jointdec-25-legleft	4102.6	3333.4	3048.7	3949.3	4117.6	4125.2	0.1925	0.0294	0.0079	0.0056	0.0035
jointdec-25-thigh	4215.9	3285.9	3376.2	4111.7	4192.6	4222.5	0.1152	0.0341	0.0092	0.0059	0.004
jointdec-50-thigh	4917.7	3313.7	2571.0	4345.5	4643.8	4754.0	0.1469	0.0233	0.0071	0.0048	0.0037
jointdec-50-thighleft	4545.9	3290.6	2507.9	4114.2	4567.5	4593.0	0.1996	0.0458	0.013	0.0076	0.0056
jointdec-50-legleft	4158.8	3215.0	2012.9	3703.3	3968.2	4035.9	0.2123	0.0605	0.0191	0.0114	0.0092
jointdec-50-foot	4510.1	2744.6	2842.8	4028.5	4257.0	4443.2	0.1791	0.0196	0.0038	0.0022	0.0015
jointdec-50-leg	4203.2	3322.9	2102.6	3874.7	4083.8	4155.2	0.2131	0.0496	0.0161	0.0089	0.0067
jointdec-50-footleft	4590.2	2519.4	3066.0	4433.5	4565.2	4567.5	0.1642	0.024	0.0071	0.0051	0.002
lengthdec-50-foot	4359.1	2426.3	2761.1	3291.1	4004.1	4238.6	0.2174	0.0547	0.0163	0.0096	0.0063
lengthdec-50-torso	3849.3	2291.8	2861.1	3794.2	3847.2	3860.2	0.1005	0.0234	0.0085	0.0065	0.0036
lengthdec-50-footleft	4171.7	2327.6	1320.8	3921.9	3964.5	4059.9	0.2671	0.0497	0.0158	0.0098	0.0062
lengthinc-100-torso	5613.6	3149.4	2651.3	3984.0	5063.3	5140.5	0.1799	0.0142	0.0026	0.0014	0.001
lengthinc-100-foot	4414.6	2412.4	3127.8	4228.5	4306.6	4359.3	0.1675	0.0623	0.017	0.0096	0.0061
lengthinc-100-footleft	4485.5	2354.3	1455.8	3172.0	4258.3	4500.7	0.1932	0.0425	0.0108	0.0069	0.0052
lengthinc-200-foot	4183.9	2088.0	3415.0	4037.3	4213.5	4201.2	0.194	0.0524	0.018	0.0089	0.0069
lengthinc-200-torso	4238.7	2685.3	2809.5	3668.9	3911.6	4040.3	0.2775	0.0203	0.0035	0.0019	0.0011
lengthinc-200-footleft	3999.8	2262.3	2502.1	3287.6	3795.4	3963.2	0.2006	0.0178	0.0031	0.0017	0.0012
frictiondec-25-footleft	4104.4	3075.8	2639.1	3619.0	4022.1	4057.8	0.1671	0.0405	0.0106	0.0058	0.0039
frictiondec-25-foot	3841.3	3220.2	1928.6	3515.0	3904.3	4035.8	0.3082	0.0899	0.0317	0.0218	0.0151
frictiondec-50-foot	4118.6	3243.4	1338.4	3591.2	3867.2	4009.3	0.2518	0.0516	0.0146	0.0095	0.0065
frictiondec-50-footleft	3919.6	2975.4	2709.9	3503.7	3706.6	3798.1	0.2184	0.078	0.0309	0.0237	0.02
frictioninc-25-footleft	4338.9	3537.0	2400.2	3704.3	4545.9	4615.7	0.1882	0.1017	0.0267	0.0171	0.0113
frictioninc-25-foot	4393.1	3350.7	2571.6	4059.9	4239.2	4360.9	0.2711	0.0905	0.0316	0.0201	0.0134
frictioninc-50-foot	4560.0	3245.2	2249.5	4119.7	4499.4	4580.6	0.1653	0.0553	0.0106	0.0061	0.0039
frictioninc-50-footleft	4488.4	3380.3	3117.6	3822.6	4360.7	4435.2	0.17	0.0329	0.0079	0.0044	0.0033