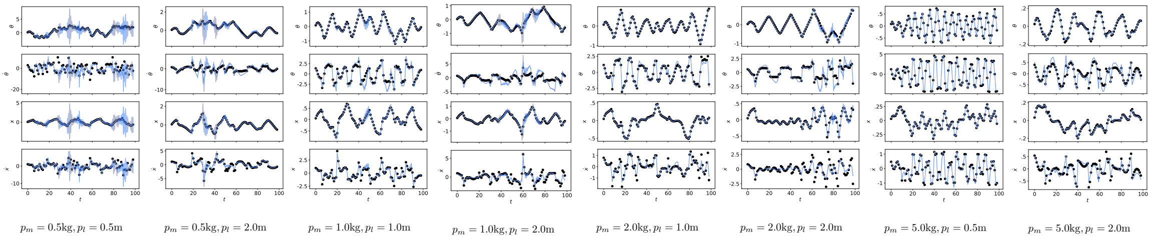
We highly appreciate the reviewers' time, efforts, and valuable suggestions! Recap: The contribution of our work is that we introduce task selection based on prior experience into a meta-learning algorithm by conceptualizing the learner and the active meta-learning setting using a probabilistic latent variable model. We are encouraged that reviewers find our approach intuitive without any overly restrictive assumptions [R1], effective for task selection [R2], relevant to our community with sound claims [R3], and novel as well as a valuable contribution to the field of meta-learning [R4].

R2, R3, R4 Comparisons with related work. Reviewers asked whether existing approaches, namely domain randomization (DR), active DR (ADR) [3] and active curriculum learning (ACL) [4,6,7], could be evaluated in the same setting. R3, R4 asked for further clarification on the differences between existing work and our approach. DR is equivalent to the UNI baseline in our experiments and we will add the appropriate citations. ADR is not applicable to our setup due to the kind of proxy that is used to rank the informativeness of a new task. ADR compares policy rollouts on potential tasks compared to a reference environment, dedicating more time to tasks that cause the agent difficulties. PAML learns a representation of the space of tasks and makes comparisons directly in that space. This way our approach does not require a) rollouts on new potential tasks, b) handpicked reference tasks and c) the task parameters to be observed directly. In comparison to ACL, we note that our key objective is data-efficient exploration of a task space from scratch. ACL performs unsupervised pre-training on environments to improve downstream RL tasks, making a direct comparison unsuitable. PAML and ACL can be seen as complimentary approaches, e.g., PAML might be used to select the tasks used for unsupervised pre-training in ACL.

R1 Empirical evaluation. We agree that it would be interesting to see how our approach would impact the quality of a policy in an RL setting, but see this as beyond the scope of the current work. We look at learning models, which can be used in a model-based RL setting, but additional (confounding) factors make policies successful, e.g., exploration and local optima. R1 also mentions that only one of the environments is learned from pixel data. With space constraints in mind, we chose to focus on evaluating PAML across the different task parameter scenarios where PAML is applicable. Lastly, we will add an analysis of the settings fully observed 4.1 and pixel-descriptor 4.4. to the paper: in 4.1, PAML consistently picks masses/lengths of the lower value end of the range but more diversely than in 4.2. In 4.4, it usually starts to select lengths at both ends of the range and then selects tasks towards the middle. VAE embedding. The VAE embedding is used directly as a latent task embedding and is jointly optimized with the meta-learning objective $\mathcal{L} = \mathcal{L}_{PAML} + p(\boldsymbol{\Psi}_{\text{candidates}}|\mathbf{H}_{\text{candidates}})$, where $\boldsymbol{\Psi}_{\text{candidates}}$ are candidate task pixel descriptors, see B.2 in the appendix. Discussion of meta-learning (ML). With space constraints in mind and since our work's goal is to incorporate active learning into ML rather than deriving a new ML method, we kept the part about prior work in ML short but detailed the ML approach used in this work in Section 2. Control signals. The number of actions is in Appendix Table 1.Plots of the model. In Figure 1, we show plots of the learnt model for 8 different task specifications. For better readability, we plot the trajectory of each task separately.



$p_m = 0.5$kg, $p_l = 0.5$m

$p_m = 0.5$kg, $p_l = 2.0$m

$p_m = 1.0$kg, $p_l = 1.0$m

$p_m = 1.0$kg, $p_l = 2.0$m

$p_m = 2.0$kg, $p_l = 1.0$m

$p_m = 2.0$kg, $p_l = 2.0$m

$p_m = 5.0$kg, $p_l = 0.5$m

$p_m = 5.0$kg, $p_l = 2.0$m

Figure 1: Plots of the learnt dynamics model for the cartpole environment on test tasks. $\theta, \dot{\theta}, x, \dot{x}$ denote the angle's position, angle's velocity, cart's position and cart's velocity, respectively. The figure shows the true data points (black discs) and the predictive distributions (blue) with ±2 standard deviations. The model generalizes well to the test tasks.

R2 Limited technical novelty. Although our work is built on existing ideas for probabilistic meta-learning and active learning, as far as we are aware, our algorithm addresses a gap in the literature when it comes to the combination of the two. Particularly in how we conceptualize task parameters in PAML's model. Our experiments show that this can be practically exploited in various scenarios. No introduction of related approaches. We introduced previous work (DR, ACL, ADR) that is similarly motivated and most close to ours in the third and fourth paragraph in Section 1.

R3 Domains outside robotic systems. Generally speaking, in situations where the meta-learning model is able to learn a useful task embedding our approach provides a way of taking advantage of that embedding to make informed task selections. Since data-efficient learning of task spaces is an important problem in robotics, we think that our selection of experiments does not detract from the evaluation but illustrates task conditions that are easy to interpret, such as the under-/ or over-specification.