RSS 2020 Reviews

Reviewer #3

1. Summary of Contributions: Please briefly list three things this paper contributes (e.g., system demonstration or development, theoretical, methodological, algorithmic, empirical; bridging fields; or providing an important critical analysis). For each contribution, briefly state the level of significance (i.e., how much impact will this work have on researchers and practitioners in the future?). If you cannot think of three things, please explain why. Not all good papers will have three contributions.

1. An algorithmic approach to model-based meta-reinforcement learning based on variational inference of latent dynamics parameters, and online variational inference via gradient ascent on an ELBO for online model identification. (Unclear significance)

2. Experimental comparison of proposed approach on hardware against naive model based RL approaches as well as simple PID feedback control. (Low significance)

3. Demonstration of utility of approach on a full payload transportation task. (Medium significance)

The authors propose a novel algorithm for model-based meta-reinforcement learning which performs online gradient-based optimization on the variational parameters on a posterior distribution over latent system parameters of a deep neural network dynamics model. However, there are several algorithmic choices that are not adequately justified or experimentally validated, which make it hard to judge the significance of the proposed approach to future research in model-based meta-RL for robotic control in changing and uncertain environments.

2. Detailed comments: Please provide a thorough review of the submission, including its originality, quality, clarity, and significance.

The paper proposes a novel algorithm for system identification which is to be used in a MPC based model-identification adaptive control framework. As a target application, the authors consider a the problem of a quadrotor picking up an unknown mass payload with an unknown length cable, and transporting it along a desired trajectory. The paper is well written, and the authors present an impressive systems demo showing that their approach is able to perform this task with lower deviation from the desired trajectory than baseline approaches. While the hardware demo is well executed and impressive, I have concerns with the proposed approach, the lack of ablations to existing model-based meta-RL algorithms and lack of comparisons in the plots made it hard to judge the utility of the proposed approach against prior art and simpler alternatives.

Concerns regarding the algorithm:

- The MBRL baseline operated on states that were a concatenation of the past 8 states. Was this the same state representation used for the latent variable approach? How was the choice of a sliding window length of 8 chosen? Is the performance of the MBRL baseline sensitive to this choice?

- A key motivation of this work is that quick adaptation is necessary to prevent instability for a dynamic, potentially unstable system such as a quadrotor with a suspended payload. However, the online adaption is performed via gradient descent on variational parameters, which is a non-convex optimization with no convergence guarantees, and may in fact converge to suboptimal local minima. Furthermore, unlike MAML and related meta-learning algorithms, the meta-training phase does not optimize for the performance of this online gradient descent.

- From my reading, the MPC using the "current estimate of z^{test}," implies that the MPC only uses the mean of $q_{\phi^{test}}(z)$. If this is the case, a simpler baseline would be to use gradient descent to only estimate the mean, rather than both the mean and variance (reducing the number of parameter that are learned online and potentially improving online learning efficiency). This is an important ablation that should be compared against, and is closely related to the meta-learning approach taken in [1]. Alternatively, if modeling the posterior uncertainty over z is important, the authors should demonstrate this by comparing against a controller which acts with knowledge of the model uncertainty by, e.g., sampling models with $z \sim q_{\phi^{test}}(z)$

Concerns regarding the experimental results:
- The authors do not compare to other existing model-based meta-RL approaches, such as [2], [3]. Both of these approaches seem applicable to the proposed scenario, and thus should be compared against.

- The authors perform no hyperparameter selection comparisons (apart from whether on not z is known at train time). The paper would be strengthened with sensitivity of their approach with, e.g. the dimension chosen for the latent variable z, the number of samples of z used to compute the ELBO (9) used for the test time adaptation. Furthermore, the authors should compare against simplifications of their method, such only performing point-estimation of z at test time, as discussed above.

- It is important for any learning algorithm to demonstrate generalization, the algorithm should at least be able to interpolate between training examples, and ideally, should extrapolate as well. The authors only report performance on tether lengths that exactly match a subset of the training set, and thus it is impossible to tell if the proposed approach is able to generalize to novel test-time cable lengths and masses.

- As the authors argue in the motivation of this paper, learning a black box dynamics model is particularly relevant when dealing with dynamics that are otherwise hard to model, such as those involving friction or aerodynamical effects. However, the scenario considered in the experiments does not seem to have these criteria, and is similar to the tasks performed with model-based approaches such as [4]. The paper would be strengthened with an experimental demonstration which highlights the utility of a learning based approach, which could be possible if the suspended payload had non-negligible, hard-to-model drag.

Overall, the paper presents a novel approach to meta-learning which is validated in a impressive systems demo of a quadrotor performing a payload transportation task. However, there are concerns with the algorithm and experimental validation that make it difficult for me to judge the contribution of this new approach to the field. If the authors perform the suggested experimental comparisons and the results demonstrate the efficacy of their approach, the contribution would be quite significant.

[1] Zintgraf, Luisa M., et al. "Fast context adaptation via meta-learning." (2018).

[2] Harrison, James, et al. "Control Adaptation via Meta-Learning Dynamics." (2018).

[3] Nagabandi, Anusha, et al. "Learning to adapt in dynamic, real-world environments through meta-reinforcement learning." (2018).

[4] Faust, Aleksandra et al. "Aerial Suspended Cargo Delivery through Reinforcement Learning." (2013).

Minor comments:
- PETS uses an ensemble of dynamics models. It is unclear if your method also uses an ensemble of learned models $p\_\theta()$ or just a single model.
- Confidence intervals in tables are not defined - are they 95% confidence? +/- standard deviation?
- Plots of tracking error should (1) compare against baseline approaches over time, and (2) present results averaged over the 5 trials, to make clear that the presented results were not cherry picked.
- In the full demonstration, was Dtest cleared upon transitions between the different payloads? If not, how is the system able to properly adapt with a dataset containing data corresponding to different z values?

3. Overall score:
Fair (a paper that is on its way to making a good contribution but not there yet)

4. Confidence score:
You are absolutely certain about your assessment and very familiar with the related work

Reviewer #4

1. Summary of Contributions: Please briefly list three things this paper contributes (e.g., system demonstration or development, theoretical, methodological, algorithmic, empirical; bridging fields; or providing an important critical analysis). For each contribution, briefly state the level of significance (i.e., how much impact will this work have on researchers and practitioners in the future?). If you cannot think of three things, please explain why. Not all good papers will have three contributions.

This paper addresses the challenge of optimally controlling robot behavior while interacting with objects for scenarios in which aspects of the dynamics model (e.g. physical properties of a payload object) cannot be known a priori. In particular, this work focuses on quadcopter flight trajectories with tethered payloads of varying physical properties (e.g. payload mass and tether length). Three explicit contributions are listed:

1. A model-base meta-learning approach and training procedure for the specific task

2. Some comparative experimentation for quantitative analysis (e.g. Table 1)

3. A system demonstration of the approach for a complex action sequence. The quadcopter moved to a payload at rest, connected a tether, transported the payload to a goal location, detached the tether, and moved away from the goal location.

2. Detailed comments: Please provide a thorough review of the submission, including its originality, quality, clarity, and significance.

Writing / Paper Organization:
This paper is well written and coherently organized. The introduction clearly establishes the technical gap that prevents desired robot behavior (rapid in-flight adaptation to a priori unknown payloads) and articulates the intuition of the proposed solution. Figure 1 does a great job providing a visualization of the problem scenario and making clear the scenario that will be investigated later on. Overall, this introduction is great. The authors might consider including a brief higher-level motivating statement about why it is desirable to have quadcopters carrying payloads (i.e. what is a compelling real-world use case?).

The rest of the paper's organization is great. Nothing seems out-of-place or rushed. All of the figures were effective and well utilized within the body. In general, the paper was clear about its goals and was a pleasure to read.

As a minor comment, it is worth noting that there are some spacing issues throughout the document. For example, see the white space around the header for Section III or the white space around equation 5 and equation 6.

Another minor comment, some of the references to previous equations are just included as a parenthesis; it would improve clarity to say "Equation (3)" rather than just "(3)" The second paragraph of Section IV.F. has some examples.

A final minor comment: The sentence after Equation 6 seems to just suddenly begin ("…called

the evidence lower bound"), almost as though a piece of a sentence is missing.

Related Works: I am not aware of any omitted work that should be included. The related work is well structured and serves to do more than just list related approaches. By including relevant limitations of the cited related work, the authors further reinforce the technical gap stated in the introduction and make it clear how their proposed solution is different.

Technical Approach: I did not notice any technical errors. I found Figures 2 & 3 to be very effective. The included algorithm blocks were also useful, although Algorithm 1 might not be necessary. Referencing the relevant equations in the algorithm block is a nice feature.

Experiments: My main criticism of the paper is the scope of the experimentation. Since the training data included flights with varying tether lengths between 18cm and 30cm, I was hoping to see tests with more variation. Also, the training data included payloads of differing masses, but it does not appear that the experiments varied the mass. The paper would benefit from further experimentation over different ways to vary the unknown physical properties of the payload.

Another comment about the experiments: I think it would be interesting (and probably useful) to characterize the significance of the specific unknown dynamics variable (tether length). It may not be obvious to a novice quadcopter pilot how significant of a difference an 18cm tether would be as compared to a 30cm tether. One might imagine a model in which the tether length is assumed to be known a priori; how does the performance of that model compare when the value is correctly provided as 30cm versus when the value is incorrectly provided as 18cm for an actual 30cm tether?

Regarding the results, was it surprising that the unknown variables approach outperformed the known variables approach? It seems particularly interesting that providing true/correct structure to the learning problem results in worse performance. I would like to see some more discussion about this outcome; do the authors have an intuition about why this is the case?

Comments / General Question:
I had a few questions while reading the paper that might be worth addressing.

The data collection initially involves manually piloting the quadcopter along "random trajectories". Are there particular properties that this random movement by the manual operator is trying to capture? Is there something about this initial step that would require the operator to have expert knowledge? Could this be automated?

Regarding data collection, the authors state "We chose to not include any explicit information about the drone in the state representation to avoid any overly burdensome requirements, such as a motion capture system." I was surprised by this comment and hope the authors will expand on it more. For example, why not track the position of the drone in the same way that the position of the payload is tracked?

Part of the motivation for the proposed approach is "the need for rapid adaptation in order to cope with sudden dynamics changes when pickup up payloads". I'm interested in how this

relates to the sample frequency during data collection, which is stated as being one sample per 0.25 seconds. How was this time interval chosen? Would it need to be higher for other plausible scenarios?

3. Overall score:

Very Good (a solid paper that makes an important contribution)

4. Confidence score:

You are fairly confident in your assessment

Reviewer #5

1. Summary of Contributions: Please briefly list three things this paper contributes (e.g., system demonstration or development, theoretical, methodological, algorithmic, empirical; bridging fields; or providing an important critical analysis). For each contribution, briefly state the level of significance (i.e., how much impact will this work have on researchers and practitioners in the future?). If you cannot think of three things, please explain why. Not all good papers will have three contributions.

- Experimental demonstration: Presenting a meta-learning model-based reinforcement learning (MBRL) algorithm in real-world experiments of a quadrotor transporting a suspended load. Having real-world robotic experiments of RL and learning approaches is highly relevant in itself and important to advance research in RL for robotics. However, I have some doubts about the specific setup and conclusions drawn from the experiments (see comments below).

- Algorithmic: There might be some advances in the presented meta-learning MBRL algorithm; however, these do not become clear due to an insufficient discussion of related work on this aspect.

2. Detailed comments: Please provide a thorough review of the submission, including its originality, quality, clarity, and significance.

Overall evaluation: The paper presents interesting real-world experiments of meta-learning MBRL, which is a nice achievement. However, the overall contribution of this paper does not come out well. If the main contribution is in solving the application problem (quadrotor transport of suspended loads) as the current presentation suggests, it is questionable whether the presented method (based on deep learning models) is well chosen and superior to more standard approaches (e.g., online estimation of payload and tether + optimal control). If the main contribution is supposed to be in the method itself, it does not become clear what the novelty of the method is compared to existing work. The contributions could be made clearer by improving the overall presentation, being more specific in the discussion of related work, and presenting experimental comparisons. Please see my comments below for further details.

Main comments:

1) My main concern is that I am not convinced that using a deep learning approach (i.e., a very expressive model) is most suitable for the considered problem. The uncertainty considered in the quadrotor task stems from an unknown payload and unknown tether length. Hence, this should be captured well by a parametric model that models exactly these parameters as uncertain. So, I do not see why a deep learning model is beneficial here, and I did not see any reasoning in the paper for why it would be. For example, an experimental comparison of the deep learning approach with a parametric approach (e.g., adaptive control, online system identification + some form of optimal control) would be very insightful. This might prove or disprove the usefulness of the proposed method for the adaptation task. As the paper puts its focus on the quadcopter application (rather than the general method), this seems important.

2) The discussion of related work also does not well support the choice of method. In Sec. II,

existing work on quadrotor transport of suspended payloads is discussed. Bottom line of this discussion is that these methods require prior identification of key physical parameters (payload, tether length), which is in contrast to the work herein. While this is true, these methods apparently show that parametric models are sufficient for solving the considered control task. Hence, considering adaptation of parametric models (specifically of payload and tether length) seem like a natural choice. Are there such approaches in literature? The discussion of related work then continues by pointing out that parametric approaches require some domain knowledge. While I would agree that this is a valid concern for some applications, it seems that such domain knowledge is easily available in the problem at hand (uncertain tether and payload).

3) Unclear contribution of the presented algorithm/method: It remains unclear to me whether, and to what extent, the presented meta-learning algorithm is a contribution in itself. Which aspects are novel or potentially advance the state-of-the-art in meta-RL? Neither in the introduction, nor in related work, this is discussed. From the presentation of the method in Sec. IV, I get the impression that the used algorithm is essentially a combination of known methods. (This is also suggested by the fact that no code is provided, but a reference to an implementation of a prior algorithm is given instead.) If there is a significant contribution in the algorithm, this should be specifically pointed out (e.g., in the related work). If this is the case, the authors might want to consider making the algorithm and method the focus, rather than the application.

4) Related to my comments above, I find the statement in Sec. IV "The primary challenge is that this interaction is difficult to model" questionable. As the source of uncertainty is very clear here, I would expect that one can well model the suspended payload as a nonlinear dynamical system with a few uncertain/adapting parameters.

5) Further below, the statement "our method is general and is applicable to any robotic system that interact with the environment under changing conditions" is very bold. While the method is developed for general MDPs, whether it is actually applicable to real-world robotic system would have to be shown. Making such a statement based on a single experimental system is not justified and should be avoided in my opinion.

6) As variational approximation, a Gaussian with diagonal covariance is chosen. How well of an approximation is this? This is not discussed. Likewise for other variational approximations done in the paper.

7) It is hard to evaluate the achieved performance in experiments (e.g., Fig. 4) because of a missing ground truth. For example, it would be helpful to see an optimal control approach for a quadrotor model with known payload and tether length as ground truth. PID is obviously not a good ground truth as it is based on heuristic tuning. As is, I have difficulty judging the performance; also between the contestant methods. Again, as the paper focuses on the application problem, it seems important to be able to judge the achieved performance.

8) In Fig. 4, why is the performance of the proposed method with unknown parameters better than with known parameters? This seems counter-intuitive.

9) I find the conclusion "Our approach is able to successfully complete the full task (Q5) due to

our online adaptation mechanism" questionable. Obviously, the task is completed (with is nice!). However, whether this is *due to* the online adaptation is questionable. We do not see how the quadrotor would have performed without the adaptation. The argument that the tracking error reduces over time (as mentioned in the caption of Fig. 7), I do not find fully convincing either, as this could also be the effect of a vanishing transient. Similarly, "the quadcopter must re-adapt online to be able to successfully follow the specified trajectories" is a statement that is, in my opinion, not supported by the shown experiments.

Minor comments:

10) In the second paragraph of the introduction, "learning" is contrasted with "system identification" in the sense that the latter requires a domain expert. I'm not quite following this characterization. Do the authors mean that most of system identification requires some form of structured models and these needs to be found by a domain expert? If yes, this should be explained. Further, applying sophisticated ML algorithms such as deep learning typically also requires some model selections and experience, and thus some domain expertise.

11) Sec. IV-A: It is interesting that no explicit information about the quadrotor was included in the states for learning. I would have expected that the state of the quadrotor is relevant for making accurate predictions about how the payload behaves. Have the authors experimented with this and also tried to include quadrotor state information? It might be worth expanding this discussion.

12) Sec. IV-B: Instead of including z_k as a "special" auxiliary variable, one could also consider augmentation of the state, i.e. include z_k into s_k. This seems more standard, but might lead to larger models (also increased output dimension). I suggest to comment on this.

3. Overall score:
Fair (a paper that is on its way to making a good contribution but not there yet)

4. Confidence score:
You are fairly confident in your assessment

RSS Rebuttal

We are glad that R4 noted that "the related work is well structured and serves to do more than just list related approaches. By including relevant limitations of the cited related work, the authors further reinforce the technical gap stated in the introduction and make it clear how their proposed solution is different," but we believe there was a misunderstanding on the part of R3 and R5 regarding our contribution. Our goal is not to propose a general improvement to the state-of-the-art in meta-reinforcement learning, but to study applications of meta-reinforcement learning to quadcopter control with suspended payloads. We believe that such a scope is reasonable: robotics is an applied field, and applications of methods that have previously been explored primarily in more limited simulated settings to complex real-world robotics tasks---such as the suspended payload control task we consider, which has noise, lag, real-time constraints, and is partially observable---are of interest to the research community.

R5 stated that the experiments do not show that online adaptation is even necessary for the suspended payload transportation task. We believe that our experiments do show this: Table 1 indicates that MBRL without adaptation completely fails at following all desired trajectories.

R5 also questions the value of applying deep learning methods to quadcopter control. While this is a point worthy of discussion, we believe it is inappropriate to reject a paper simply because the reviewer believes that a particular class of techniques is not promising. This belief clearly does not reflect any consensus in the community, and exploring risky techniques is important to make meaningful research progress.

In regard to comparisons with system identification methods (R3 and R5): While we agree that such comparisons would be valuable, we note that there is an important difference in assumptions. Our method does not assume any known model of the system or even known 3D localization, which would be required for system identification methods. The only state input is the 2d image-space position of the payload -- the model is not even aware of the pose of the quadcopter. Under these assumptions, we are not aware of any system identification approach that would be applicable. Furthermore, our method is general and does not assume that the tether length or mass is the parameter that varies, and in principle our method could be applied to any other variation without any change to the method.

Regarding generalization (R3 and R4): we note that our method already outperforms the methods that we compare to, by a large margin, even on in-distribution tasks. So while we agree that these additional experiments would be valuable and we will add them, we believe that the current experiments do already illustrate the value of our method over the prior

approaches that we compare to.

CoRL Reviews

Reviewer #1

3. Please summarize the paper, state what you think the contribution is, comment on its strengths and weaknesses, and give advice for improving the paper.

This paper focuses on the UAV navigation problem with unknown suspended payload. The proposed approach aims to solve this problem by using meta-learning to learn the altered dynamics of the system with unknown payload, and to use a model-based predictive control to adapt the system's plan and control to follow a given trajectory or complete given pickup/drop-off task. The proposed approach has been applied on a real vehicle and tested on various trajectories and tasks. The paper also provides performance comparison with other adaptive and non-adaptive techniques and demonstrates tracking improvement. The paper is well-organized and well-written. The approach has been explained well with the diagrams and algorithms. There are some points that can be considered to improve the quality of the paper and to provide clarifications:

1) During the section 4.1 Data collection, it is mentioned that during different tasks 3D printed payloads weighing between 10g and 15g are attached to the vehicle. However, during testing, the weights of the payloads are not mentioned. Are they same as during training? It would be good if the information of the payload weights and how they can be affecting the dynamics can be included.

2) The string lengths during the test (21cm and 27cm) are bounded by training values (18cm and 30cm). Does the testing conditions, i.e., the string length during testing need to be bounded with the training values? If yes, it would be good if that can be mentioned. If it is not necessary, then some insight about the performance of the approach for larger or smaller strings could be useful.

3) While collecting data D_train, what are the issues need to be considered for this approach to work? Does it need to contain data from different maneuvers to be accurately extract the model parameters \theta^* ? Since the performance during the testing would still depend on how accurate these dynamic model parameters are, some insight about how the training should be performed would be useful to understand.

4) It is mentioned in the paper that even with unknown variables, the approach would work. In fact the results show that it almost always works better with unknown variables, which is not typically expected. It would be a clarification if it can be explained further.

Minor comments:

1) In Figure 4, to be able to see the improvement over time, it would be useful to show the start point and the direction of movement.

2) Figures should be refered as Figure instead of Fig.

3) Line 213, it should be includes, not include.

4. Was the video submission useful? What was good about it? What could be done to improve it? (Write N/A if there was no video submission)

The video submission was useful to see how this approach was implemented on a real vehicle. It provides further explanation and tests. It was particularly useful how the experiment was set up with the external camera, and how the best trajectory was picked at runtime.

5. If code was submitted, we ask that you perform a sanity check. We do not ask that you run the code. But if you do, please comment on this too. Authors have been instructed to provide a readme file that points to the part of the code that needs to be peer-reviewed. Please, read this part of the code and answer the following questions: a) Does the code appear to implement the algorithm described in the paper? b) Did the readme file allow you to identify relevant parts of the code? c) Were you provided with data to replicate the results? Please, remember that code is for peer-review only and that you are not allowed to use this code outside the review process. (Write N/A if there was no code submission)

N/A

9. Comment on the potential impact of the paper to real robotics problems (taking COVID-19 impact into account). Please check the supplement material for the COVID-19 response.

The paper includes extensive real robot experiments. The approach presented in this paper could be used for UAV delivery applications.

12. Overall recommendation:

Weak Accept

13. Rate your level of confidence in this recommendation:

High

Reviewer #2

3. Please summarize the paper, state what you think the contribution is, comment on its strengths and weaknesses, and give advice for improving the paper.

This paper applies meta-RL to the problem of controlling a quadrotor with partially, yet structured, unknown dynamics/disturbances. The unknown dynamics/disturbances come from attaching a suspended load with unknown length/weight to the quadrotor. The paper proposes an online system identification approach where data is collected during training and used to characterize a distribution over the unknown dynamics/disturbances. This distribution is then used in real-time to identify the dynamics that best match the data collected in real-time. The paper is well written in general, but I have the following concerns.

1- The novelty of the proposed framework is limited. Using neural networks to represent the dynamics of physical systems has been used extensively in the literature (which is well captured in the references of this paper). The paper provides no insights on how to use the specifics of the problem under consideration.

2- The paper main contribution is mainly experimental. Nevertheless, the experiments fail to answer the critical question of why to use a data-driven approach if adaptive model-based control techniques can adapt to these unknown dynamics/disturbances. In other words, why learning is needed in the first place if model-based techniques could achieve it (with typically strong theoretical guarantees)?

4. Was the video submission useful? What was good about it? What could be done to improve it? (Write N/A if there was no video submission)

The video is very informative and shows the algorithms in action.

5. If code was submitted, we ask that you perform a sanity check. We do not ask that you run the code. But if you do, please comment on this too. Authors have been instructed to provide a readme file that points to the part of the code that needs to be peer-reviewed. Please, read this part of the code and answer the following questions: a) Does the code appear to implement the algorithm described in the paper? b) Did the readme file allow you to identify relevant parts of the code? c) Were you provided with data to replicate the results? Please,

remember that code is for peer-review only and that you are not allowed to use this code outside the review process. (Write N/A if there was no code submission)

N/A

9. Comment on the potential impact of the paper to real robotics problems (taking COVID-19 impact into account). Please check the supplement material for the COVID-19 response.

The paper applies the proposed framework to real robotic systems.

12. Overall recommendation:

Borderline

13. Rate your level of confidence in this recommendation:

Moderate

CoRL Rebuttal

We thank the reviewers for their detailed feedback!

In regard to the experimental evaluation, we would first like to highlight the "extensive real robot experiments" (R1), which show that our method (a) outperforms state-of-the-art model-based (meta-)reinforcement learning algorithms and commonly-used feedback controllers (Table 1), (b) is able to better generalize to payloads not seen at training time (Table 2), and (c) is able to perform end-to-end payload transportation (Fig 6), navigate around obstacles (Fig 8), greedily follow a target (Fig 9), and follow trajectories indicated by a "wand"-like interface (Fig 10). We believe these results successfully evaluate the central claim of our paper: that our meta-learning system enables a quadcopter to adapt to and control a priori unknown suspended payloads better than prior methods.

Second, we would like to highlight the novelty and applicability of our work. While our method combines ideas from previous papers (as is true of many successful robotics papers), to our knowledge our work is the first to demonstrate model-based RL for quadcopter suspended payload control with meta-learning. As stated by R1, "the approach presented in this paper could be used for UAV delivery applications," and we believe this application is of interest to the robotics community.

Responses to specific comments are detailed below.

R2 asked why a data-driven approach is needed if adaptive model-based control techniques can perform the task. We note that adaptive model-based controllers typically require: a priori knowledge of the quadrotor dynamics and suspended payload equations, 3d state estimation of both the quadrotor and suspended payload, and camera parameters in order to perform this state estimation. These assumptions can be a significant impediment to deployment in the real-world. In contrast, our method only requires the pixel location of the payload and the quadrotor's commanded actions to learn to adapt to and control the suspended payload. More generally, we believe that learning-based approaches that remove assumptions on known models are of interest to the CoRL community.

R2 asked what the novelty was in regard to other methods that use neural network dynamics models. We do not claim the use of NN dynamics models is novel, and we will clarify this in the manuscript. Our novel contribution is in leveraging neural network dynamics models in conjunction with meta-learning for the task of controlling an aerial robot with a suspended payload.

R2: Regarding the paper not leveraging "the specifics of the problem under consideration," we believe the generality of our approach is actually a significant strength! Our approach only requires the (1) pixel location of the payload and (2) commanded actions in order to learn to adapt to and control the suspended payload. This makes our approach off-the-shelf and easily applicable to many suspended payload tasks.

R1 asked whether the test-time string lengths need to be bounded by the training values. Technically, they do not have to be. However, even though our experiments are only for string lengths within the training values, our approach performed significantly better compared to other methods. We will clarify this in the paper.

R1 asked what should be considered when collecting training data. This is a great question, which we will clarify in the revised version: all that is required is that a diverse number of state and action sequences be visited. These sequences do not need to be the same as what will be asked at test time -- for example, none of our training data contained executions of the circle, square, or figure 8 patterns.

R1: Regarding why the unknown variable approach sometimes performs better, it could be possible that the known latent values were not optimal with respect to optimizing the KL divergence loss.

R1: We will clarify in the paper that the payload weights were the same during training and testing.

R1: We will fix all writing issues for the final version, thank you!