MuMoMAML: Model-Agnostic Meta-Learning for Multimodal Task Distributions

Anonymous authors
Paper under double-blind review

Abstract

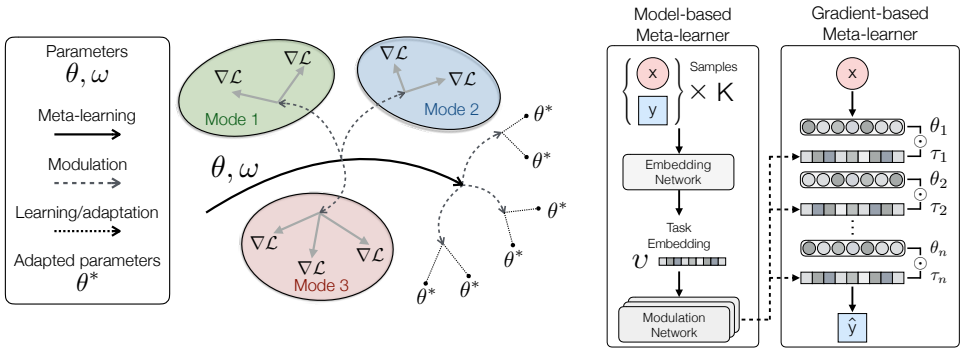
Gradient-based meta-learners such as MAML (Finn et al., 2017) are able to learn a meta-prior from similar tasks to adapt to novel tasks from the same distribution with few gradient updates. However, such frameworks seek a common initialization shared across the entire task distribution, substantially limiting the diversity of the task distributions that they are able to learn from. In this paper, we aim to augment existing gradient-based meta-learners with the capability to identify the modes of a task distribution and adapt quickly through gradient updates given tasks sampled from a multimodal task distribution. Specifically, we propose a multimodal MAML algorithm (MuMoMAML), which is able to modulate its meta-learned prior according to the identified task modes, allowing further fast adaptation. We evaluate the proposed algorithm on a diverse set of problems including regression, few-shot image classification, and reinforcement learning. The results demonstrate the effectiveness of our model in efficiently acquiring a meta-learned prior under a multimodal task distribution.

1 Introduction

Humans are capable of effectively utilizing prior knowledge to acquire new skills as well as adapt to new environments. For example, imagine that we are learning how to carve on a snowboard on a mountain that we just arrived at. While we only know the basics of snowboarding and barely know the terrain, we can still accomplish this feat quickly. Even when the skill that we are interested in learning is only vaguely related to the set of skills that we have acquired in the past, we are usually able to exploit the relationship among skills and generalize. For the snowboarding example, our knowledge about skiing and skateboarding can play the role that prepares us for learning snowboarding skills more efficiently.

Similarly, can machines rapidly master a novel skill based on a variety of related skills they have already acquired? Recent advances in meta-learning offer machines a way to learn from a distribution of tasks and adapt to a new task from the same distribution using few samples. Model-based meta-learning approaches (Duan et al., 2016; Wang et al., 2016; Munkhdalai & Yu, 2017; Mishra et al., 2017) propose to recognize the task identity from a few sample data and make the appropriate predictions by adjusting a model's state (*e.g.* RNN's internal states). Those methods demonstrate improved performance at the expense of engineering hand-designed architectures, yet the optimal strategy of designing a meta-learner for an arbitrary range of tasks may not be obvious to the humans. On the other hand, model-agnostic gradient-based meta-learning frameworks (Finn et al., 2017; 2018; Kim et al., 2018; Lee & Choi, 2018; Grant et al., 2018) seek an initialization of model parameters such that a small number of gradient updates will lead to fast learning on a new task, offering the flexibility in the choice of models.

While most of the existing gradient-based meta-learners rely on a single initialization, different modes of a task distribution can require substantially different parameters, making finding an initialization that is a short distance away from all of them infeasible. When the modes of a task distribution are disjoint, we might be able to obtain a set of separate meta-learners to master each mode. However, this not only requires additional identity information about the modes, which is not always available or is ambiguous when the modes are not clearly disjoint, but also eliminates the possibility of associating transferable knowledge across different modes of a task distribution. To



(a) (b)

Figure 1: (a) Diagram of multimodal model-agnostic meta-learning (MUMOMAML) algorithm, which aims to acquire a meta-learned prior parameterized by θ and ω, enabling fast adaptation to a task sampled from a multimodal distribution. Areas of different color indicate modes of tasks. (b) Overview of our model. Our model consists of two components: a model-based and gradient-based meta-learner. The former strives to identify the mode of a task distribution from a few samples and modulate the prior accordingly; the latter performs gradient updates to effectively yield better performance.

overcome this issue, we aim to develop a meta-learner that is able to acquire a meta-learned prior and adapt quickly given tasks sampled from a multimodal task distribution.

To this end, we leverage the strengths of the two main lines of existing meta-learning techniques: model-based and gradient-based meta-learning. Specifically, we propose to augment gradient based meta-learners with the capability of generalizing across a multimodal task distribution. Instead of learning a single initialization point in the parameter space, we propose to first estimate the mode of a sampled task by examining task related samples. Given the estimated task mode, our model then performs a step of *model-based adaptation* to modulate the meta-learned prior in the parameter space to fit the sampled task. Next, from this model-based adapted meta-prior, a few steps of *gradient-based adaptation* are performed towards the target task to progressively improve the performance on the task. This main idea is illustrated in Figure 1.

The main contributions of this paper include a meta-learner that is able to acquire a meta-learned prior and adapt quickly given tasks sampled from a multimodal task distribution by taking advantage of both model-based and gradient-based meta-learning. We evaluate the effectiveness of our proposed model on a diverse set of different learning problems, including classification, regression, and reinforcement learning. We demonstrate that our model offers a better generalization ability compared to state-of-the-art meta-learners especially when task distributions are multimodal.

2 Related Work

We review the several families of meta-learning, which our work builds upon:

Optimization-based meta-learning. The approaches that learn to optimize a learner model are known as optimization-based meta-learning. Pioneered by Schmidhuber (1987); Bengio et al. (1992), optimization algorithms with learned parameters have been proposed, enabling the automatic exploitation of the structure of learning problems. From a reinforcement learning perspective, Li & Malik (2016) represent an optimization algorithm as a learning policy. Andrychowicz et al. (2016) train LSTM optimizers to learn update rules from the history of gradients, and Ravi & Larochelle (2016) train a meta-learner LSTM to update a learner's parameters for few-shot image classification. Unlike these methods, our algorithm does not require additional parameters to learn update rules; instead, we augment our meta-learner with the ability to infer the modality of tasks.

Model-based meta-learning. Model-based meta-learning frameworks learn to recognize the identities of tasks and adjust the model state (*e.g.* the internal state of an RNN) to fit the task. Santoro et al. (2016) train a network with an external memory that is able to assimilate new samples and leverage this data to make accurate predictions. Duan et al. (2016) represent a fast RL algorithm as an RNN and learn it from data. Munkhdalai & Yu (2017) propose to learn task-agnostic knowl-

edge and use it to shift its fast parameters for rapid generalization. Unlike these methods, during the meta-training phase, our model does not maintain an internal state but seeks an initialization, allowing fast adaptation with a few gradient steps during meta-testing phase.

Gradient-based meta-learning. Finn et al. (2017) and its extensions Finn et al. (2018); Kim et al. (2018); Lee & Choi (2018); Grant et al. (2018), known as gradient-based meta-learning, aim to estimate a parameter initialization among the task-specific models, that provides a favorable inductive bias for fast adaptation. With the model agnostic nature, appealing results have been shown on a variety of learning problems. However, assuming tasks are sampled from a concentrated distribution and pursuing a common initialization to all tasks can substantially limit the performance of such methods on multimodal task distributions where the center in the task space becomes ambiguous. In this paper, we propose to first identify the mode of a sampled task, in a procedure which is similar to model-based meta-learning approaches Santoro et al. (2016); Mishra et al. (2017). Then, we modulate the meta-learned prior in the parameter space to make the model better fit to the mode and take gradient steps to rapidly improve the performance on the task afterwards.

3 PRELIMINARIES

In this paper, we propose a meta-learner that is able to quickly master a new skill based on a variety of relevant but vaguely related skills that it has already acquired. Specifically, we pursue the goal of rapidly adapting to a novel task sampled from a multimodal task distribution $\mathcal{T} \sim p(\mathcal{T})$ by discovering and utilizing the common underlying structure of the tasks. We consider a target dataset $\mathcal{D}$ consisting of samples extracted from a multimodal task distribution. The dataset is split into meta-training and meta-testing sets, which are further divided into task specific training $\mathcal{D}_{\mathcal{T}}^{\text{train}}$ and validation $\mathcal{D}_{\mathcal{T}}^{\text{val}}$ sets. A meta-learner learns about the underlying structure of the task distribution through training on the meta-training set and is evaluated on meta-testing set.

Next we introduce the model-agnostic meta learning (MAML) algorithm (Finn et al., 2017) upon which our work builds. Given a target task $\mathcal{T}$, two sets of data points are sampled, one for task-training $\mathcal{D}_{\mathcal{T}}^{\text{train}}$ and the other for task-validation $\mathcal{D}_{\mathcal{T}}^{\text{val}}$. The MAML algorithm performs one or few steps of gradient updates on the $\mathcal{D}_{\mathcal{T}}^{\text{train}}$, from its initial parameter θ, to optimize for the generalization performance on $\mathcal{D}_{\mathcal{T}}^{\text{val}}$ by optimizing:

$$\min_{\theta} \sum_{\mathcal{T} \sim p(\mathcal{T})} \mathcal{L}(\theta - \alpha \nabla_{\theta} \mathcal{L}(\theta, \mathcal{D}_{\mathcal{T}}^{\text{train}}), \mathcal{D}_{\mathcal{T}}^{\text{val}}) \tag{1}$$

4 METHOD

Our goal is to develop a meta-learner that is able to identify the modes of a multimodal task distribution and adapt quickly given a novel task. To this end, we propose to learn a model-based meta-learner (parameterized by ω) that generates a set of task specific parameters τ for our meta-learned prior parameters. Next, this modulated prior learns to adapt to a target task rapidly through gradient-based optimization (see Figure 1 (b)). As discussed above, our gradient-based meta-learner consists of two sets of parameters, with one of them τ being task-specific (inferred by the model-based meta-learner), and the other set θ being task-agnostic: $\hat{y} = f\big(x; \theta, \tau\big)$, where f can represent a generic function with two sets of parameters θ and τ. In the context of this paper, we utilize neural networks as the instantiation of the meta-learner.

In this section, we first introduce our model-based meta-learner as well as a variety of modulation operators in section 4.1. Then we describe the learning details for MUMOMAML in section 4.2 and present the implementation details in section 4.3.

4.1 MODEL-BASED META-LEARNER

Supposing we have K data points $\{x_k, y_k\}_{k=1,\ldots,K}$ sampled from D^{train}, the task embedding model encodes the training data points into a vector that encodes the characteristics of a task. Specifically, we learn a generator model g to take a sequence of paired data x and its labels y as input, and generate the task-specific parameters as: $\tau_i = \mathrm{g}_i(\{x, y\}_{k=1,\ldots,K}; \omega)$. Note that our modulation is

applied block-wise to the gradient-based meta-learner. We use an index i to represent the task-specific parameters τ for a block i. A block in the neural network might contain one or more layers, which is parameterized by θ_i. In the rest part of this paper, we omit the index i as the default configuration for brevity.

During this process, the task encoder model takes few shots of data and embeds them into an intermediate representation υ that summarizes the task characteristics. Next, multiple generation functions take this shared task representation υ as input and predict the task-specific parameters τ for each block in the gradient-based meta-learner model. Based on these block-wise parameters τ, a variety of modulation operations can be used for updating the task-specific prior information in parameter θ, in a block-wise manner. We formalize this procedure as:

$$\phi = \theta \odot \tau$$

Where ϕ represents the modulated prior parameters for gradient-based meta-learner, and $\odot$ represents a general modulation function. In this paper, we investigated some of the most representative modulation operations and discuss them as follows.

Attention based modulation (Mnih et al., 2014; Vaswani et al., 2017) has been widely used in modern deep learning models and has proved its effectiveness across various tasks (Yang et al., 2016; Mnih et al., 2014; Zhang et al., 2018; Xu et al., 2015). Inspired by the previous works, we employed attention to modulate the prior model. In concrete terms, attention over the outputs of all neurons (Softmax) or a binary gating value (Sigmoid) on each neuron's output is computed by the model-based meta-learner. These parameters τ are then used to scale the pre-activation of each neural network layer $\mathbf{F}_\theta$, such that $\mathbf{F}_\phi = \mathbf{F}_\theta \otimes \tau$. Note that here $\otimes$ represents a channel-wise multiplication.

Feature-wise linear modulation (FiLM) (Perez et al., 2017) proposed to modulate neural networks for achieving the conditioning effects of data from different modalities. We adopt FiLM as an option for modulating our gradient-based meta-learner. Specifically, the parameters τ are divided in to two components τ_γ and τ_β such that for a certain layer of the neural network with its pre-activation $\mathbf{F}_\theta$, we would have $\mathbf{F}_\phi = \mathbf{F}_\theta \otimes \tau_\gamma + \tau_\beta$. It can be viewed as a more generic form of attention mechanism. Please refer to Perez et al. (2017) for the complete details.

4.2 Gradient-based Fast Adaptation

With the modulated prior model parameterized by θ and τ, we perform the gradient-based meta-learning step to fine-tune our model to better fit the target task $\mathcal{T}_i$. We expect that our model can fit tasks sampled from different modes efficiently and generalize across modes of the task distribution. The concrete training procedure is described by the following pseudo-code:

Algorithm 1 Meta-Training Procedure for MuMoMAML.

Input: Task distribution $P(\mathcal{T})$, Hyper-parameters α and β
Randomly initialize θ and ω.
while not DONE **do**
  Sample batches of tasks $\mathcal{T}_j \sim P(\mathcal{T})$
  for all j **do**
    Infer the parameter $\tau = g(\{x, y\}; \omega)$ with K samples from $\mathcal{D}^{\text{train}}_{\mathcal{T}_j}$
    Evaluate $\nabla_\theta \mathcal{L}_{\mathcal{T}_j}\left(f(x; \theta, \tau); \mathcal{D}^{\text{train}}_{\mathcal{T}_j}\right)$ with respect to the these K samples
    Compute adapted parameter with gradient descent: $\theta'_{\mathcal{T}_j} = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_j}\left(f(x; \theta, \tau); \mathcal{D}^{\text{train}}_{\mathcal{T}_j}\right)$
  end for
  Update $\theta \leftarrow \theta - \beta \nabla_\theta \sum_{T_j \sim P(\mathcal{T})} \mathcal{L}_{\mathcal{T}_j}\left(f(x; \theta', \tau); \mathcal{D}^{\text{val}}_{\mathcal{T}_j}\right)$
  Update $\omega \leftarrow \omega - \beta \nabla_\omega \sum_{T_j \sim P(\mathcal{T})} \mathcal{L}_{\mathcal{T}_j}\left(f(x; \theta', \tau); \mathcal{D}^{\text{val}}_{\mathcal{T}_j}\right)$
end while

Note that the τ are not updated in the inner loop. The model-based learner is responsible for finding a good task-specific initialization through modulation. The gradient-based learning phase shown in the inner optimization loop is responsible for fitting the target task $\mathcal{T}$ with one or few updates.

Table 1: Model's performance on the **multimodal 5-shot regression** with two or three modes. A Gaussian noise with $\mu = 0$ and $\sigma = 0.3$ is applied. The three modes regression is in general more difficult (thus higher error). In Multi-MAML, the GT modulation represents using ground-truth task identification to select different MAML models for each task mode. MuMoMAML (wt. FiLM) outperforms other methods by a margin.

Configuration		**Two Modes (MSE)**		**Three Modes (MSE)**	
Method	Modulation	Post Modulation	Post Adaptation	Post Modulation	Post Adaptation
MAML (Finn et al., 2017)	-	15.9255	1.0852	12.5994	1.1633
Multi-MAML	GT	16.2894	0.4330	12.3742	0.7791
MuMoMAML (ours)	Softmax	3.9140	0.4795	0.6889	0.4884
MuMoMAML (ours)	Sigmoid	1.4992	0.3414	2.4047	0.4414
MuMoMAML (ours)	FiLM	1.7094	**0.3125**	1.9234	**0.4048**

During the meta-testing phase, with a sampled task $\mathcal{T}_i$ and a few data samples $\mathcal{D}^{\text{train}}_{\mathcal{T}_i}$, we perform an adaptation step that corresponds to the inner loop of the meta-training. Our model-based meta-learner first identifies the mode a target task belongs to based on the input data samples and then infers the parameter τ to modulate the prior model θ. Next, we perform gradient updates on the prior model $f(x; \theta, \tau)$ and compute the adapted parameters θ'. Finally, we evaluate our adapted model $f(x; \theta', \tau)$ on the meta-test set to measure the performance of the model.

4.3 Implementation Details

For the model-based meta-learner, we used Seq2Seq (Sutskever et al., 2014) encoder structure to encode the sequence of $\{x, y\}_{k=1,...,K}$ with a bidirectional GRU (Chung et al., 2014) and use the last hidden state of the recurrent model as the representation for the task. We then apply a one-hidden-layer multi-layer perception (MLP) for each layer in the gradient-based learner's model to generate the set of task-specific parameters τ_i, as described in the previous section. We implemented our models for three representative learning scenarios – regression, few-shot learning and reinforcement learning. The concrete architecture for each task might be different due to each task's data format and nature. We discuss them in the section 5.

5 Experiments

In this section, we evaluate the proposed MuMoMAML in a variety of tasks, including regression, few-shot image classification and reinforcement learning. To investigate the hypotheses about the behaviors of gradient-based meta-learners and our proposed model, we design experiments with multimodal task distributions for regression and reinforcement learning tasks. To shed some light on the task distributions whose modes are not discrete or obvious, we also compare our model against other baselines on few-shot image classification.

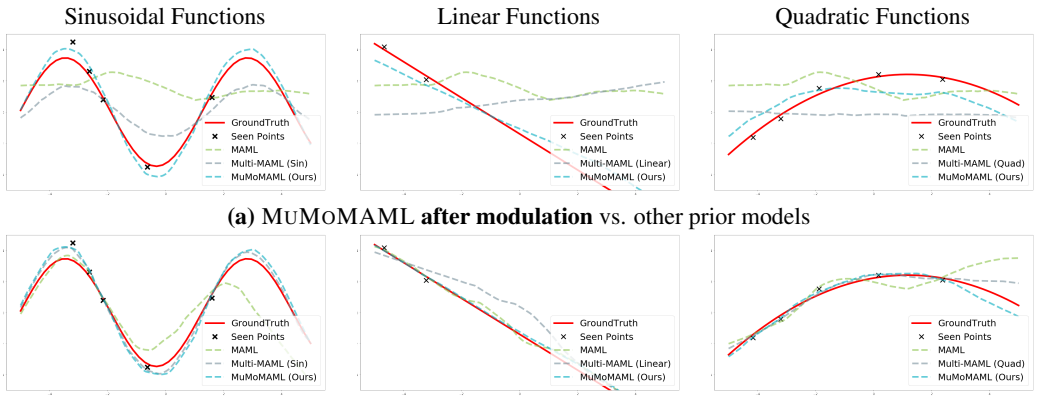
5.1 Regression

We investigate our model's capability of exploiting multimodal task distributions and rapidly adapting on a few-shot regression tasks, where a few input/output pairs $\{x_k, y_k\}_{k=1,...,K}$ sampled from a function are given and a model is asked to predict L output values $y^q_1, ..., y^q_L$ from input queries $x^q_1, ..., x^q_L$. We set up two different regression settings with two task modes (sinusoidal and linear functions) or three modes (all three functions). Please see Supplementary Material for details.

Besides the baseline MAML model, we proposed Multi-MAML, which consists of M (the number of modes *i.e.* two or three) copies of MAML models and are selected for query based on ground-truth task-mode labels. This baseline is proposed to serve as a upper-bound performance of MAML algorithm for scenarios that we have true labels of the task modes.

The quantitative results are shown in Table 1 and Figure 3. Multi-MAML outperforms MAML, showing that MAML's performance degrades on multimodal task distributions. The marginal gap between the performance of our model in two and three mode settings indicates that MuMoMAML is able to identify the mode of a sampled task as well as quickly adapt to it. Also, MuMoMAML consistently achieves better results than Multi-MAML, which demonstrates that our model is able to discover and exploit transferable knowledge across different modes to improve its performance.

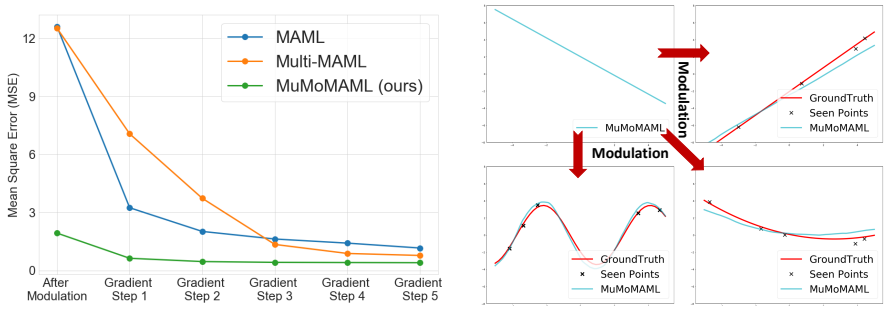
The qualitative results are shown in Figure 2, visualizing the predicted functions. We observe that MuMoMAML is able to effectively modulate its meta-learned prior to fit a sampled task (see Figure 3 (a)), which greatly eases the optimization procedure for our gradient-based learner to adapt (see



(a) MUMOMAML **after modulation** vs. other prior models

(b) MUMOMAML **after adaptation** vs. other posterior models

Figure 2: Qualitative Visualization of Regression on Three-modes Simple Functions Dataset. **(a)**: We compare the predicted function shapes of modulated MUMOMAML against the prior models of MAML and Multi-MAML, before gradient updates. Our model can fit the target function with limited observations and no gradient updates. **(b)**: The predicted function shapes after five steps of gradient updates, MUMOMAML is qualitatively better. More visualizations in Supplementary Material.



(a) Performance upon updates

(b) Effect of modulation

Figure 3: **(a)** Comparing the models' performance with respect to the number of gradient updates applied. For MUMOMAML, we report the performance after modulation for gradient step 0. **(b)** A demonstration of the modulation on prior model by our model-based meta-learner. With the FiLM modulation, MUMOMAML can adapt to different priors before gradient-based adaptation.

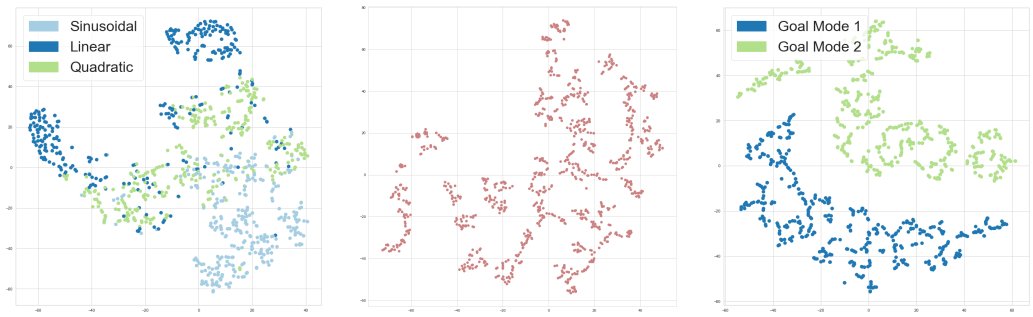
Figure 3 (b)). To gain some insights of the task embeddings produced by our model, we perform tSNE Maaten & Hinton (2008) visualization on the predicted embedding vectors υ as Figure 4 (a). It shows that our model is able to capture the mode structure in the embeddings, allowing the performance gain from modulation.

As an ablation study, we found that FiLM as a modulation method achieve better results comparing to Sigmoid and Softmax. We therefore use FiLM for further experiments. Please refer to Supplementary Material for additional results.

5.2 FEW-SHOT IMAGE CLASSIFICATION

The task of few-shot image classification consider a problem of classifying images into N classes with a small number (K) of labeled samples available. To evaluate our model's capability in this task, we conduct experiments on OMNIGLOT, a widely used handwritten character dataset of binary images. The results are shown in Table 2, demonstrating that our method achieves comparable or better results against state-of-the-art algorithms. Please refer to Supplementary Material for details.

To gain insights to the task embeddings υ produced by our model, we again randomly sampled 2000 tasks and employ tSNE to visualize the υ in Figure 4 (b). While we are not able to clearly distinguish the modes of task distributions, we observe that the distribution of the produced embeddings is not uniformly distributed or unimodal, potentially indicating the multimodal nature of this task.



(a) Regression (b) Few-shot image classification (c) Reinforcement learning

Figure 4: tSNE plots of the task embeddings produced by our model from randomly sampled tasks; marker color indicates different modes of a task distribution. The plots (a) and (c) reveal a clear clustering according to different task modes, which demonstrates that MUMOMAML is able to infer the mode from a few samples and produce a meaningful embedding υ. (a) Regression: the distance among distributions aligns with the intuition of the similarity of functions (*e.g.* a quadratic function can sometimes be similar to a sinusoidal or a linear function while a sinusoidal function is usually different from a linear function) (b) Few-shot image classification: we observe a embedding manifold with some sub-structures appearing. However, it is not intuitive to understand them directly. (c) Reinforcement learning: the embeddings for 2D navigation goals sampled from two Gaussian distributions environment are cleanly separated.

Table 2: 5-way and 20-way, 1-shot and 5-shot classification accuracies on OMNIGLOT Dataset. For each task, the best-performing method is highlighted. MUMOMAML achieves comparable or better results against state-of-the-art few-shot learning algorithms for image classification.

	5 Way Accuracy (in %)		**20 Way Accuracy (in %)**	
Method	1-shot	5-shot	1-shot	5-shot
Koch et al. (2015)	97.3	98.4	88.2	97.0
Vinyals et al. (2016)	98.1	98.9	93.8	98.5
Snell et al. (2017)	97.4	99.3	96.0	98.9
SNAIL Mishra et al. (2017)	99.1	99.8	**97.6**	**99.4**
T-net Lee & Choi (2018)	99.4	-	96.1	-
MT-net Lee & Choi (2018)	99.5	-	96.2	-
MAML (Finn et al., 2017)	98.7	**99.9**	95.8	98.9
MUMOMAML (ours)	**99.6**	**99.9**	97.2	**99.4**

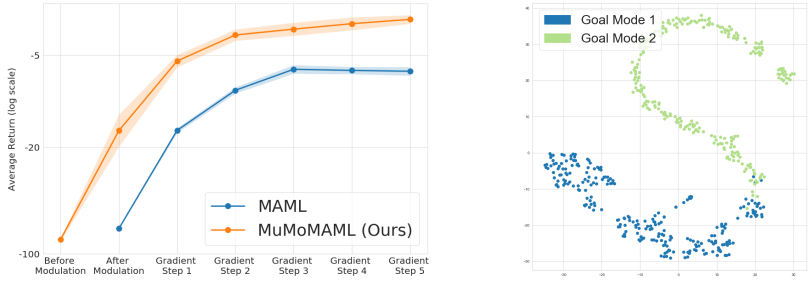
5.3 REINFORCEMENT LEARNING

In the context of reinforcement learning (RL), we are interested in verifying if MUMOMAML is able to learn from a multimodal distribution of tasks and rapidly adapt to a sampled task given only a minimum amount of interaction with an environment. Specifically, we consider a Markov decision process (MDP) with horizon H formed by a distribution of tasks T_i and each task contains an initial state distribution $p_i(s)$, a transition distribution $p_i(s_{t+1}|s_t, a_t)$, and a reward function $R_i(s_t, a_t)$. Each rollout contains $(s_1, a_1, ..., s_H)$ and its corresponding rewards $R(s_t, a_t)$ and K rollouts are available for K-shot RL.

To analyze our model's performance on an MDP formed by a multimodal distribution of tasks, we designed and conducted experiments on a 2D navigation environment similar to Finn et al. (2017). In the environment, an agent is trained to navigate to a goal location sampled from a distribution. The agent takes its current location as input and has an action space limited to a 2D vector representing the moving orientation. At each time step, it receives a reward defined as the negative squared distance to the goal. Instead of uniformly sampling goals from a square like Finn et al. (2017), we sample goals from M Gaussain distributions that are far away from each others, which makes our environment suitable for our analysis but more challenging.

To adapt MUMOMAML to the RL setting, we used a single trajectory collected based on a meta-prior policy interaction with environment to produce τ for modulating its own policy. Next, the gradient-based meta-learner operates on the modulated policy and seeks to rapidly adapt to the environment – approach the goal. We refer the reader to A.3 for more details.

We compare MUMOMAML against MAML. In all experiments, we set $H = 100$, and $M = 2$. The average returns are shown in Figure 5 (a), demonstrating that our model is able to achieve a



(a) Average return (b) Interpolation between two goal modes

Figure 5: (a) A comparison between MuMoMAML (orange) and MAML (blue) on a 2D navigation task. For MuMoMAML the **Before Modulation** return corresponds to the return of the first trajectory sampled for computing the task embedding. (b) A tSNE plot of task embeddings of interpolated goals between the two centers of the task modes in the 2D navigation task.

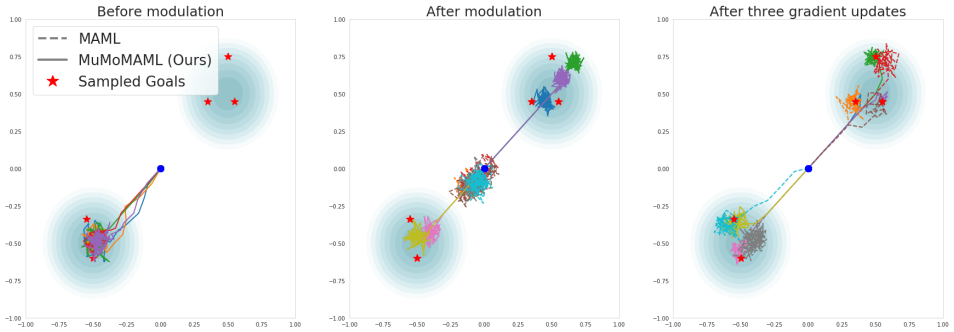


Figure 6: Sample trajectories in the 2D navigation environment. In the before modulation plot on the left, the MuMoMAML policy (line) moves randomly around one of the modes. In the middle plot, before gradient updates are applied, MuMoMAML navigates rapidly to the correct modes, while MAML policy (dashed line) has not yet observed any rewards. On the right, after three gradient updates, both have found good policies for the goals, while MuMoMAML converges on the goals in fewer steps.

significant performance gain from modulation and can consistently outperform MAML. Qualitative comparisons are shown in Figure 6. While MuMoMAML performs reasonably solely based on the modulation, MAML's meta-prior does not exhibit meaningful behaviors. After taking a few gradient updates, both the models are able to reach to goals while MuMoMAML converges in fewer steps. Please refer to Supplementary Material for additional results.

To reveal the correlation between a sampled goal and produced embedding υ, we sample 1000 goals and operate the meta-learned prior policy in the environment, to collect one trajectory each goal, feed the trajectory to our model-based learner, and produce an embedding. As shown in Figure 4 (c), a clear separation of goals sampled from two goal modes indicates that our meta-learned policy effectively provides useful trajectories and our model-based learner is able to infer the mode as well as produce meaningful embeddings. To further examine this hypothesis, we uniformly sample goals along a straight line from the center of a goal mode to another and produce embeddings. As shown in Figure 5, the clean curved structure demonstrate the competence of our model-based meta-learner.

6 Conclusion

We present a novel approach that is able to leverage the strengths of both model-based and gradient-based meta-learners to discover and exploit the structure of multimodal task distributions. Given a few samples from a target task, our model-based learner first identifies the mode of the task distribution and then modulates the meta-learned prior in a parameter space. Next, the gradient based meta-learner efficiently adapts to the target task through gradient updates. We empirically observe that our model-based learner is capable of effectively recognizing the task modes and producing embeddings that captures the structure of a multimodal task distribution. We evaluated our proposed model in few-shot regression, image classification and reinforcement learning, and achieved superior generalization performance on tasks sampled from multimodal task distributions.

REFERENCES

Marcin Andrychowicz, Misha Denil, Sergio Gomez, Matthew W. Hoffman, David Pfau, Tom Schaul, Brendan Shillingford, and Nando de Freitas. Learning to learn by gradient descent by gradient descent. *arXiv:1606.04474 [cs]*, June 2016. URL http://arxiv.org/abs/1606.04474. arXiv: 1606.04474.

Samy Bengio, Yoshua Bengio, Jocelyn Cloutier, and Jan Gecsei. On the optimization of a synaptic learning rule. In *Preprints Conf. Optimality in Artificial and Biological Neural Networks*, pp. 6–8. Univ. of Texas, 1992.

Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.

Yan Duan, John Schulman, Xi Chen, Peter L Bartlett, Ilya Sutskever, and Pieter Abbeel. Rl˜2: Fast reinforcement learning via slow reinforcement learning. *arXiv preprint arXiv:1611.02779*, 2016.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. *arXiv:1703.03400 [cs]*, March 2017. URL http://arxiv.org/abs/1703.03400. arXiv: 1703.03400.

Chelsea Finn, Kelvin Xu, and Sergey Levine. Probabilistic Model-Agnostic Meta-Learning. *arXiv:1806.02817 [cs, stat]*, June 2018. URL http://arxiv.org/abs/1806.02817. arXiv: 1806.02817.

Erin Grant, Chelsea Finn, Sergey Levine, Trevor Darrell, and Thomas Griffiths. Recasting Gradient-Based Meta-Learning as Hierarchical Bayes. *arXiv:1801.08930 [cs]*, January 2018. URL http://arxiv.org/abs/1801.08930. arXiv: 1801.08930.

Taesup Kim, Jaesik Yoon, Ousmane Dia, Sungwoong Kim, Yoshua Bengio, and Sungjin Ahn. Bayesian Model-Agnostic Meta-Learning. *arXiv:1806.03836 [cs, stat]*, June 2018. URL http://arxiv.org/abs/1806.03836. arXiv: 1806.03836.

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

Gregory Koch, Richard Zemel, and Ruslan Salakhutdinov. Siamese neural networks for one-shot image recognition. In *ICML Deep Learning Workshop*, volume 2, 2015.

Yoonho Lee and Seungjin Choi. Gradient-Based Meta-Learning with Learned Layerwise Metric and Subspace. In *International Conference on Machine Learning*, pp. 2933–2942, July 2018. URL http://proceedings.mlr.press/v80/lee18a.html.

Ke Li and Jitendra Malik. Learning to Optimize. *arXiv:1606.01885 [cs, math, stat]*, June 2016. URL http://arxiv.org/abs/1606.01885. arXiv: 1606.01885.

Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

Nikhil Mishra, Mostafa Rohaninejad, Xi Chen, and Pieter Abbeel. A Simple Neural Attentive Meta-Learner. *arXiv:1707.03141 [cs, stat]*, July 2017. URL http://arxiv.org/abs/1707.03141. arXiv: 1707.03141.

Volodymyr Mnih, Nicolas Heess, Alex Graves, et al. Recurrent models of visual attention. In *Advances in neural information processing systems*, pp. 2204–2212, 2014.

Tsendsuren Munkhdalai and Hong Yu. Meta Networks. *arXiv:1703.00837 [cs, stat]*, March 2017. URL http://arxiv.org/abs/1703.00837. arXiv: 1703.00837.

Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. FiLM: Visual Reasoning with a General Conditioning Layer. *arXiv:1709.07871 [cs, stat]*, September 2017. URL http://arxiv.org/abs/1709.07871. arXiv: 1709.07871.

Sachin Ravi and Hugo Larochelle. Optimization as a Model for Few-Shot Learning. November 2016. URL https://openreview.net/forum?id=rJY0-Kcll.

Adam Santoro, Sergey Bartunov, Matthew Botvinick, Daan Wierstra, and Timothy Lillicrap. Meta-Learning with Memory-Augmented Neural Networks. In *International Conference on Machine Learning*, pp. 1842–1850, June 2016. URL http://proceedings.mlr.press/v48/santoro16.html.

Jurgen Schmidhuber. Evolutionary principles in self-referential learning. on learning now to learn: The meta-meta-meta...-hook. Diploma thesis, Technische Universitat Munchen, Germany, 14 May 1987. URL http://www.idsia.ch/~juergen/diploma.html.

John Schulman, Sergey Levine, Philipp Moritz, Michael I. Jordan, and Pieter Abbeel. Trust Region Policy Optimization. *arXiv:1502.05477 [cs]*, February 2015. URL http://arxiv.org/abs/1502.05477. arXiv: 1502.05477.

Jake Snell, Kevin Swersky, and Richard S. Zemel. Prototypical Networks for Few-shot Learning. *arXiv:1703.05175 [cs, stat]*, March 2017. URL http://arxiv.org/abs/1703.05175. arXiv: 1703.05175.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pp. 3104–3112, 2014.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention Is All You Need. *arXiv:1706.03762 [cs]*, June 2017. URL http://arxiv.org/abs/1706.03762. arXiv: 1706.03762.

Oriol Vinyals, Charles Blundell, Tim Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. In *Advances in Neural Information Processing Systems*, pp. 3630–3638, 2016.

Jane X Wang, Zeb Kurth-Nelson, Dhruva Tirumala, Hubert Soyer, Joel Z Leibo, Remi Munos, Charles Blundell, Dharshan Kumaran, and Matt Botvinick. Learning to reinforcement learn. *arXiv preprint arXiv:1611.05763*, 2016.

Ronald J Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3-4):229–256, 1992.

Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visual attention. In *International conference on machine learning*, pp. 2048–2057, 2015.

Zichao Yang, Xiaodong He, Jianfeng Gao, Li Deng, and Alex Smola. Stacked attention networks for image question answering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 21–29, 2016.

Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena. Self-Attention Generative Adversarial Networks. *arXiv:1805.08318 [cs, stat]*, May 2018. URL http://arxiv.org/abs/1805.08318. arXiv: 1805.08318.

A ADDITIONAL EXPERIMENTAL DETAILS

A.1 REGRESSION

Setups. To form multimodal task distributions, we consider a family of functions including sinusoidal functions (in forms of $A \cdot \sin w \cdot x + b + \epsilon$, with $A \in [0.1, 5.0]$, $w \in [0.5, 2.0]$ and $b \in [0, 2\pi]$), linear functions (in forms of $A \cdot x + b$, with $A \in [-3, 3]$ and $b \in [-3, 3]$) and quadratic functions (in forms of $A \cdot (x - c)^2 + b$, with $A \in [-0.15, -0.02] \cup [0.02, 0.15]$, $c \in [-3.0, 3.0]$ and $b \in [-3.0, 3.0]$). A Gaussian observation noise with $\mu = 0$ and $\epsilon = 0.3$ is added to each data point sampled from the target task. In all the experiments, K is set to 5 and L is set to 10. We report the mean squared error (MSE) as the evaluation criterion. Due to the multimodality and uncertainty, this setting is more challenging comparing to (Finn et al., 2017).

Models and Optimization. In the regression task, we trained a 4-layer fully connected neural networks with the hidden dimensions of 100 and ReLU non-linearity for each layer, as the base model for both MAML and MUMOMAML. In MUMOMAML, an additional model with a Bidirectional GRU of hidden size 40 associated with multiple linear layers is trained to generate τ and to modulate each layer of the base model. We used the same hyper-parameter settings as the regression experiments presented in Finn et al. (2017) and used Adam Kingma & Ba (2014) as the meta-optimizer. For all our models, we train on 5 meta-train examples and evaluate on 10 meta-val examples to compute the loss.

A.2 FEW-SHOT IMAGE CLASSIFICATION

Setups. In the few-shot learning experiments, we used OMNIGLOT, a dataset consists of 50 languages, with a total of 1632 different classes with 20 instances per class. Following Santoro et al. (2016), we downsampled the images to 28×28 and perform data augmentation by rotating each member of an existing class by a multiple of 90 degrees to form new data points of a given class.

Models and Optimization. Following prior works (Vinyals et al., 2016; Finn et al., 2017), we used the same 4-layer convolutional neural network and applied the same training and testing splits from Finn et al. (2017) and compare our model against baselines for 5-way and 20-way, 1-shot and 5-shot classification.

A.3 REINFORCEMENT LEARNING

Environment. In the bimodal 2D navigation environment the goals are sampled from one of two bivariate Gaussians with means of $[0.5, 0.5]$ and $[-0.5, -0.5]$ and standard deviation of 0.5. Each mode is selected with equal probability. In the beginning of each episode, the agent starts from the origin. The agent's observation is its 2D-location and the reward is the negative squared distance to the goal. The agent does not observe the goal directly, instead it must learn to navigate there based on the reward function alone. The episodes terminate after 100 steps or when the agent comes to the distance of 0.01 from the goal.

Models and Optimization. The first trajectory from the environment is sampled using the unmodulated, unadapted model and used for computing the task embedding. We experimented with sampling more trajectories for this purpose, but found no improvement over using only one. The batch of trajectories used for computing the first gradient-based adaptation step is sampled using the modulated model and the batches after that use the modulated and adapted parameters from the previous update step. For the gradient adaptation steps, we use the vanilla policy gradient algorithm Williams (1992). As the meta-optimizer we use the trust region policy optimization algorithm Schulman et al. (2015). With respect to the gradient-based adaptation we follow the meta-learning procedure described in Finn et al. (2017).

The models are trained for one gradient adaptation step with a batch size of 20 trajectories. We use 20 tasks for the meta-batch. We chose the hyperparameters of MUMOMAML and MAML through random search. For MAML we use the inner loop update step size 0.03 and the discounting parameter γ of 0.95, for MUMOMAML we use step size 0.01 and γ 0.99. MAML uses a two-layer MLP model with hidden size 32 and ReLU activations. For MUMOMAML, we use the same size

MLP in addition to the model-based meta-learner, which consists of an RNN for the embedding network and an MLP for the modulation network. The RNN model used is the GRU with hidden size 8. The number of hidden units in the embedding network is 8.

B ADDITIONAL EXPERIMENTAL RESULTS

For the better understanding of our paper, we provide additional results in this section.

B.1 ADDITIONAL QUALITATIVE RESULTS FOR REGRESSION

Additional qualitative results for MUMOMAML after modulation are shown in Figure 7 and additional qualitative results for MUMOMAML after adaptation are shown in Figure 8.

B.2 ADDITIONAL QUALITATIVE RESULTS FOR REINFORCEMENT LEARNING

Additional trajectories sampled from the 2D navigation environment are presented in Figure 9.

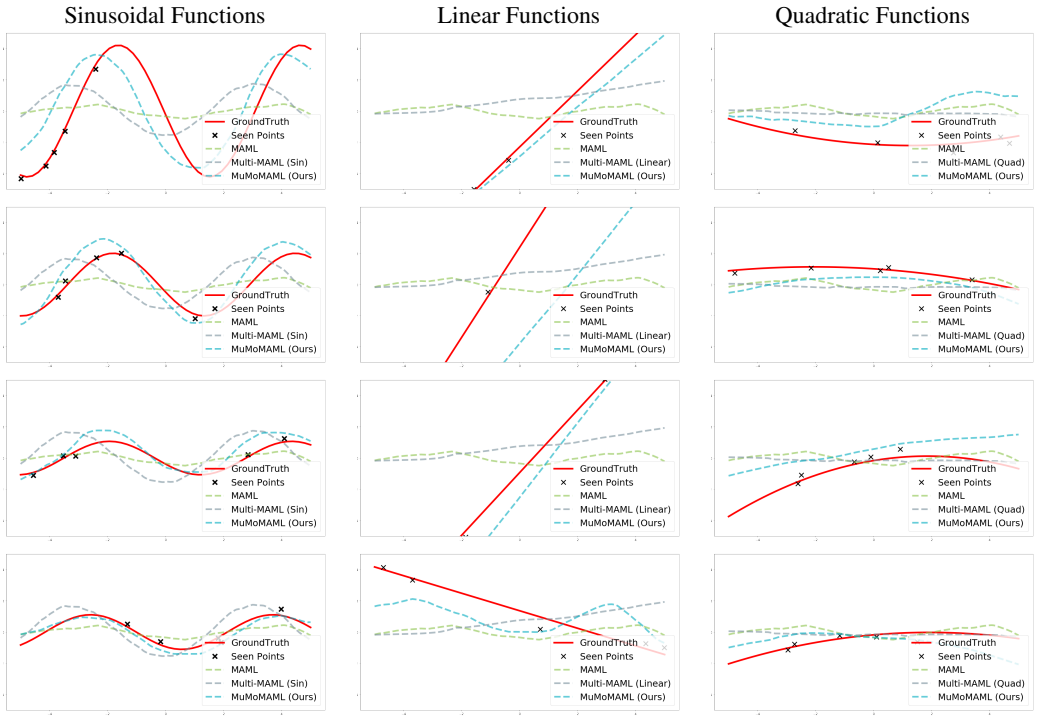


Figure 7: Additional qualitative results of the regression tasks **(a)**: MUMOMAML **after modulation** vs. other prior models.

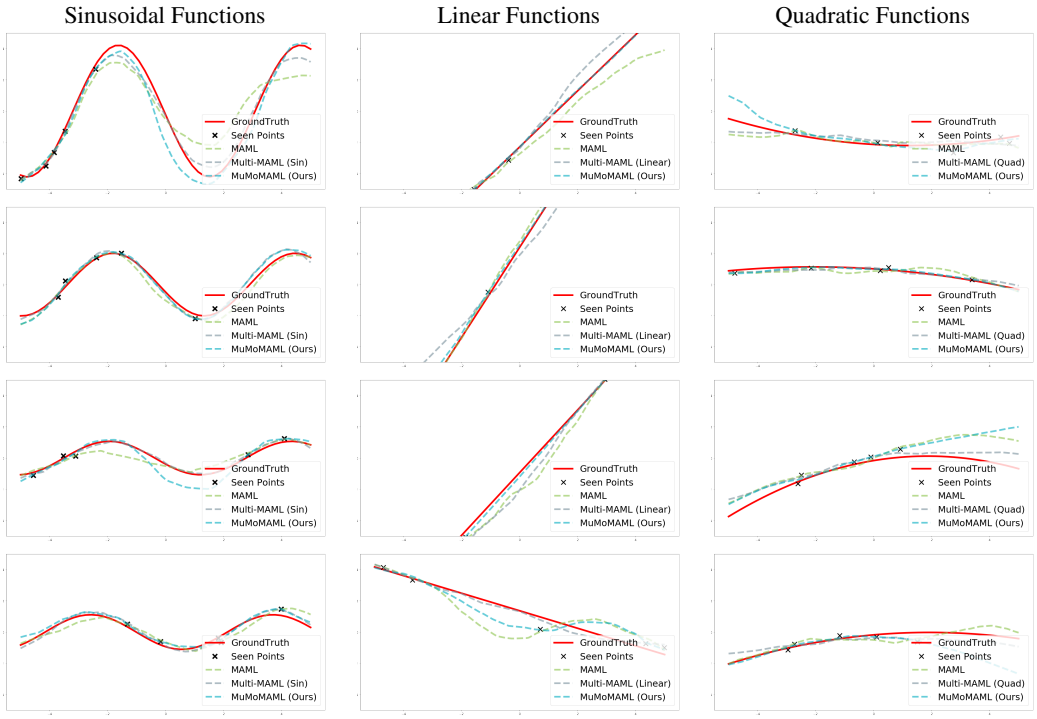


Figure 8: Additional qualitative results of the regression tasks **(b)**: MuMoMAML **after adaptation** vs. other posterior models.

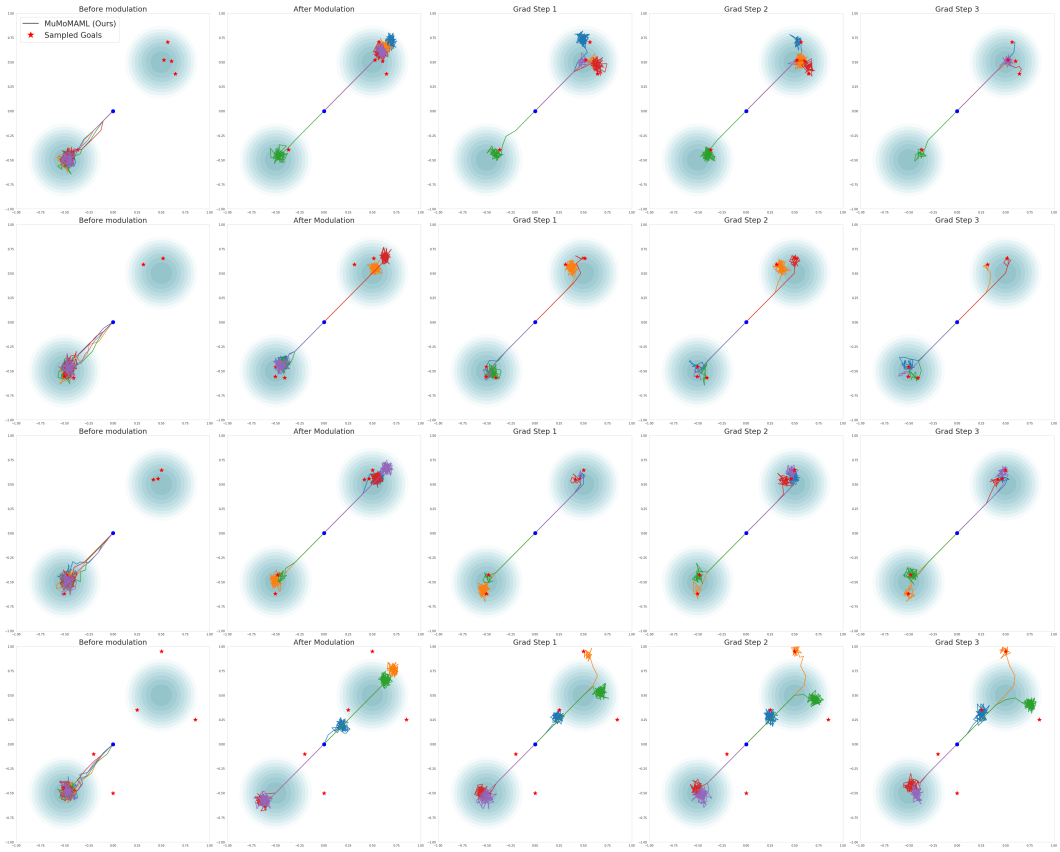


Figure 9: Additional trajectories sampled from the 2D navigation environment with MuMoMAML. The first four rows are with goals sampled from the environment distribution, where MuMoMAML demonstrates rapid adaptation and is often able to locate the goal exactly. On the fifth row, trajectories are sampled with extrapolated goals. The agent is left farther away from the extrapolated goals after the modulation step, but the gradient based adaptation steps then steadily recover the performance.