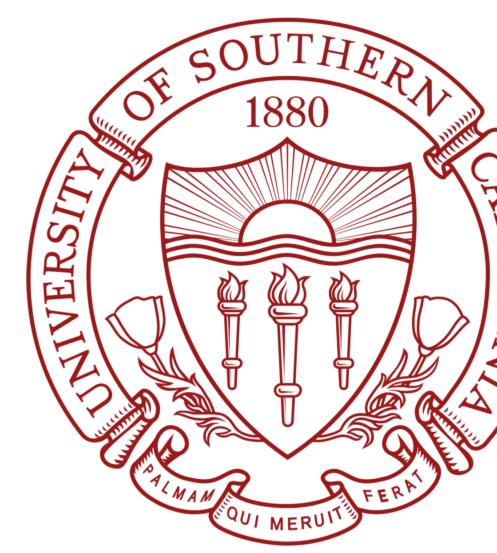


Toward Multimodal Model-Agnostic Meta-Learning



Risto Vuorio^{1*}, Shao-Hua Sun², Hexiang Hu² & Joseph J. Lim²

University of Michigan¹ University of Southern California²



Introduction & Model Overview

Multimodal Task Distributions

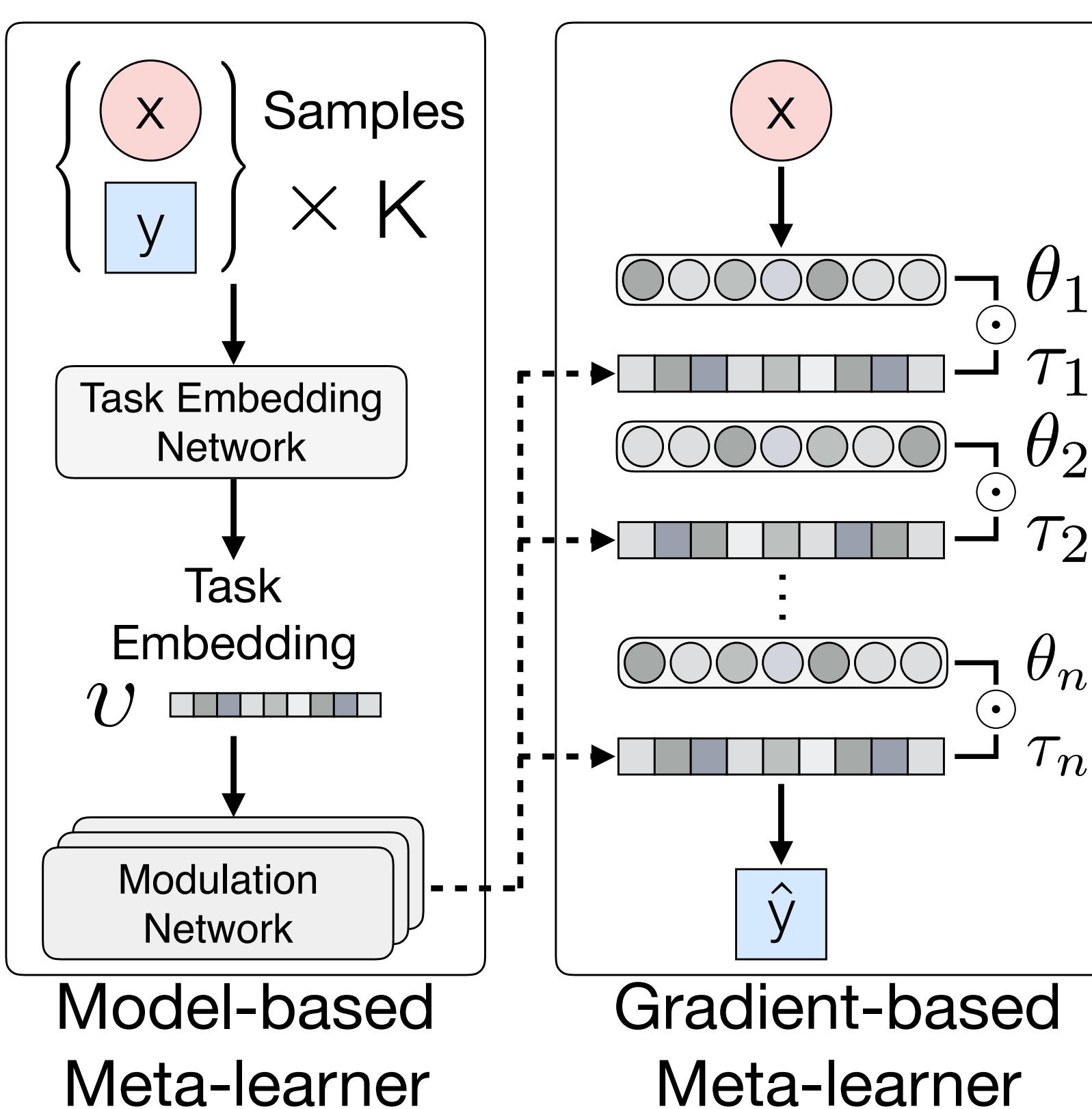
- Knowledge transferable between modes

The limitation of the MAML family

- Learns a single initialization, which is suboptimal for far apart modes

Multi-Modal MAML (MuMoMAML)

- Model-based meta-learner
 - Identify task modes
- Gradient-based meta-learner
 - Fast adaptation through gradient steps



Algorithm 1 META-TRAINING PROCEDURE.

```

1: Input: Task distribution  $P(\mathcal{T})$ , Hyper-parameters  $\alpha$  and  $\beta$ 
2: Randomly initialize  $\theta$  and  $\omega$ .
3: while not DONE do
4:   Sample batches of tasks  $\mathcal{T}_j \sim P(\mathcal{T})$ 
5:   for all  $j$  do
6:     Infer  $\tau = g(\{x, y\}; \omega)$  with  $K$  samples from  $\mathcal{D}_{\mathcal{T}_j}^{train}$ 
7:     Evaluate  $\nabla_{\theta} \mathcal{L}_{\mathcal{T}_j}(f(x; \theta, \tau); \mathcal{D}_{\mathcal{T}_j}^{train})$  w.r.t the  $K$  samples
8:     Compute adapted parameter with gradient descent:

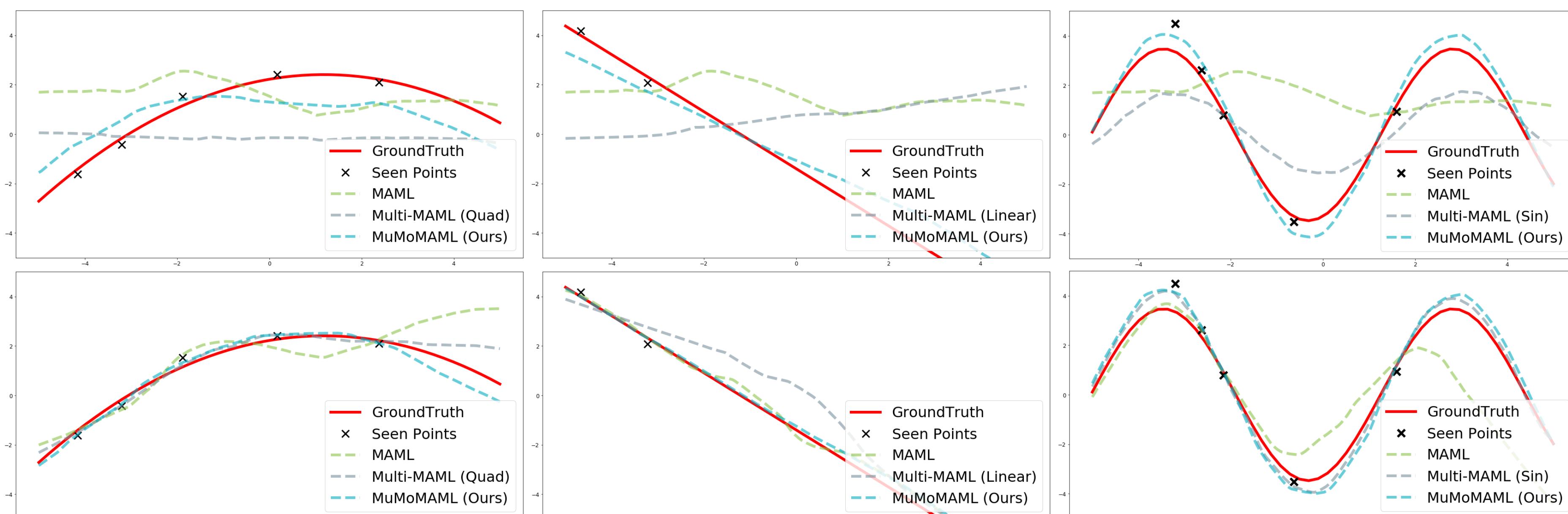
$$\theta'_{\mathcal{T}_j} = \theta - \alpha \nabla_{\theta} \mathcal{L}_{\mathcal{T}_j}(f(x; \theta, \tau); \mathcal{D}_{\mathcal{T}_j}^{train})$$

9:   end for
10:  Update  $\theta \leftarrow \theta - \beta \nabla_{\theta} \sum_{T_j \sim P(\mathcal{T})} \mathcal{L}_{\mathcal{T}_j}(f(x; \theta', \tau); \mathcal{D}_{\mathcal{T}_j}^{val})$ 
11:  Update  $\omega \leftarrow \omega - \beta \nabla_{\omega} \sum_{T_j \sim P(\mathcal{T})} \mathcal{L}_{\mathcal{T}_j}(f(x; \theta', \tau); \mathcal{D}_{\mathcal{T}_j}^{val})$ 
12: end while

```

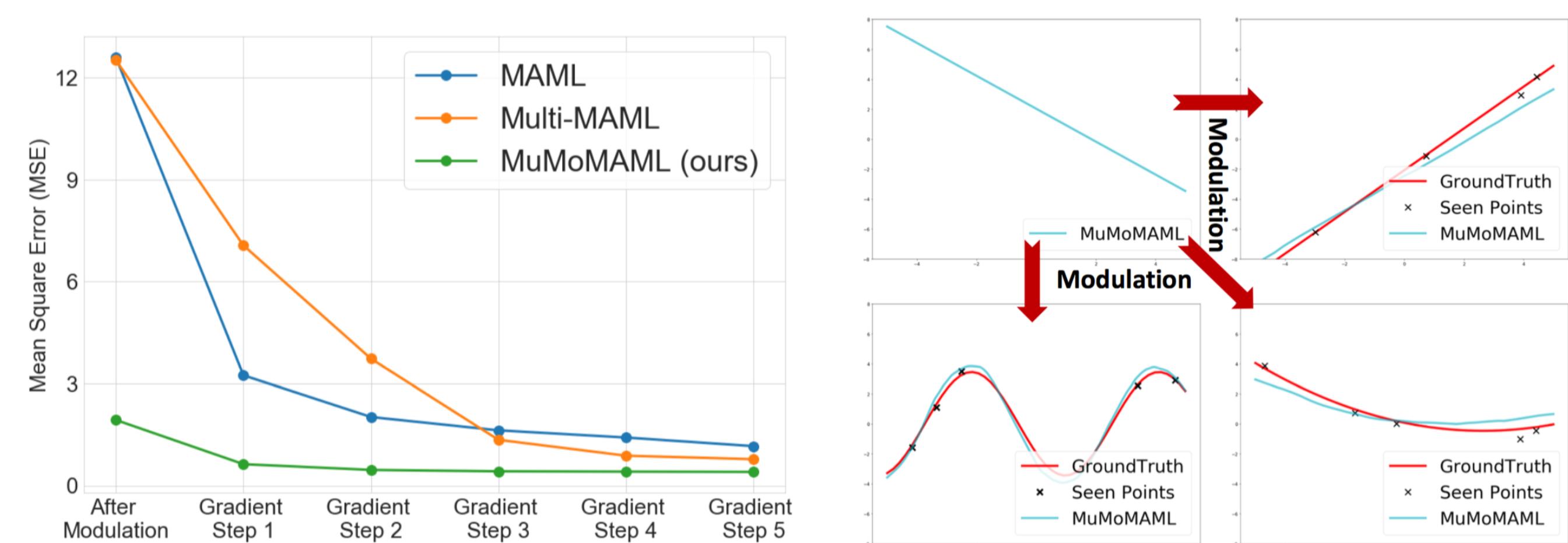
Few-shot Regression

Configuration		Two Modes (MSE)		Three Modes (MSE)	
Method	Modulation	Post Modulation	Post Adaptation	Post Modulation	Post Adaptation
MAML [5]	-	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



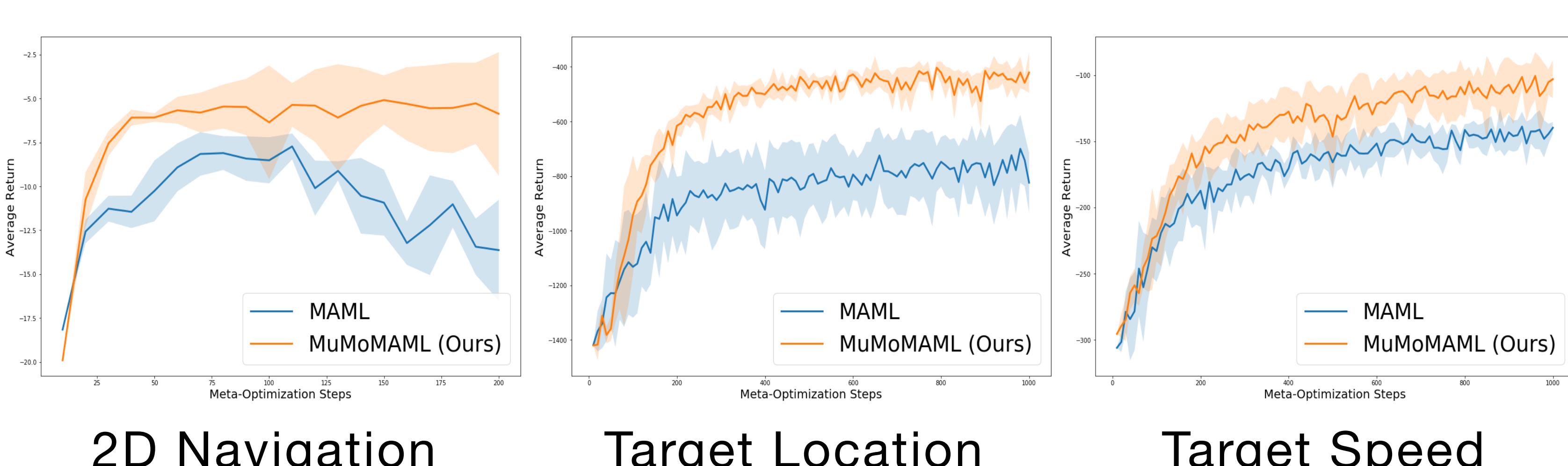
Few-shot Image Classification

Method	OMNIGLOT	
	5 Way Accuracy (in %) 1-shot	20 Way Accuracy (in %) 5-shot
Siamese nets [10]	97.3	98.4
Matching nets [28]	98.1	98.9
Meta-SGD [12]	99.5	99.9
Prototypical nets [23]	97.4	99.3
SNAIL [14]	99.1	99.8
T-net [11]	99.4	-
MT-net [11]	99.5	-
MAML [5]	98.7	99.9
MuMoMAML (ours)	99.7	99.9



Reinforcement Learning

Training Curves

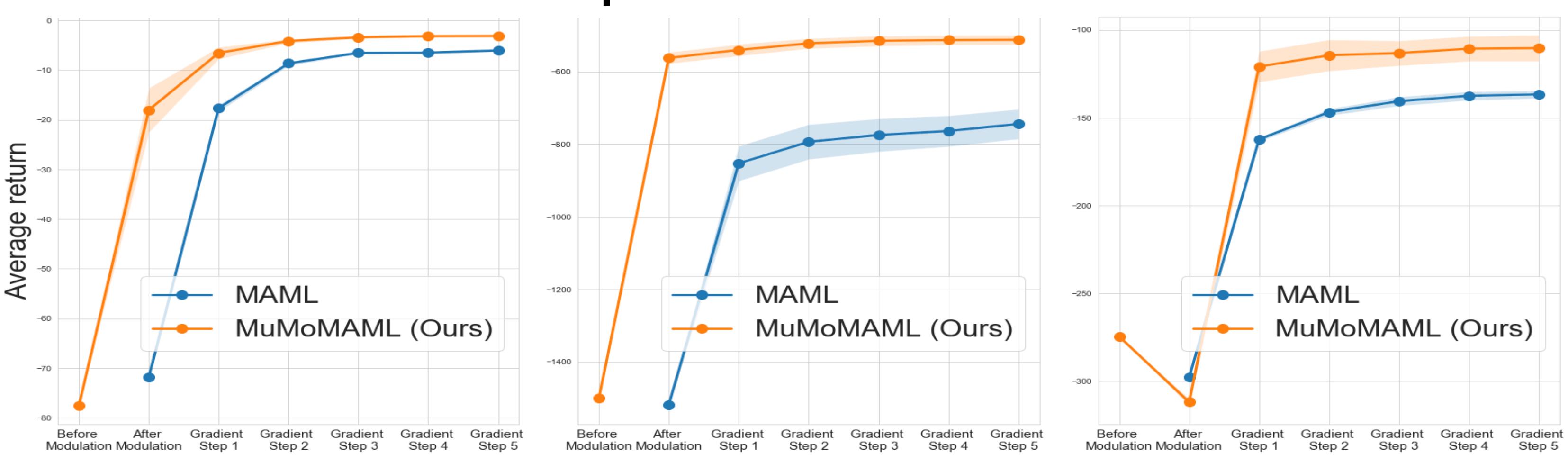


2D Navigation

Target Location

Target Speed

Adaptation Curves

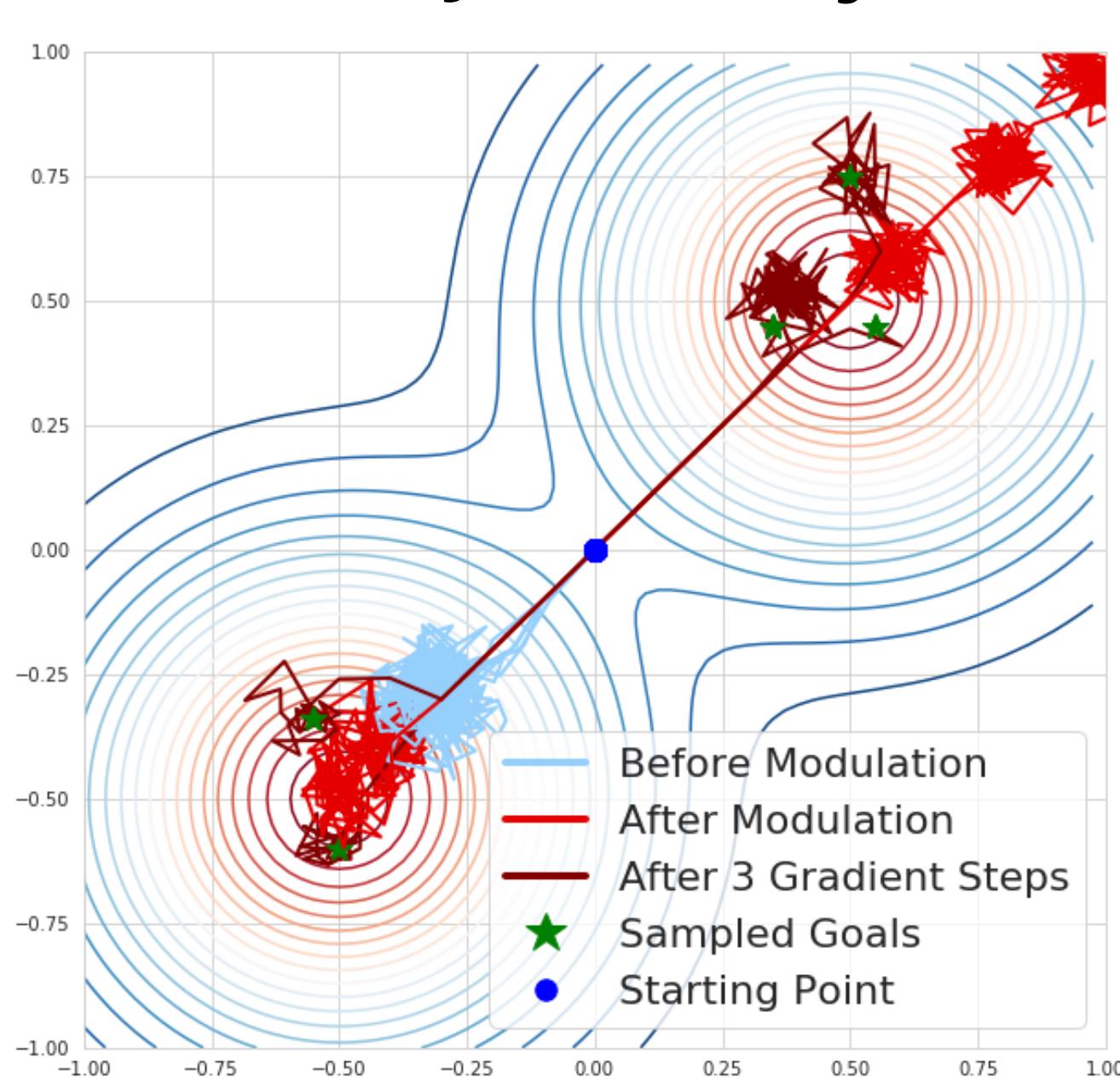


2D Navigation

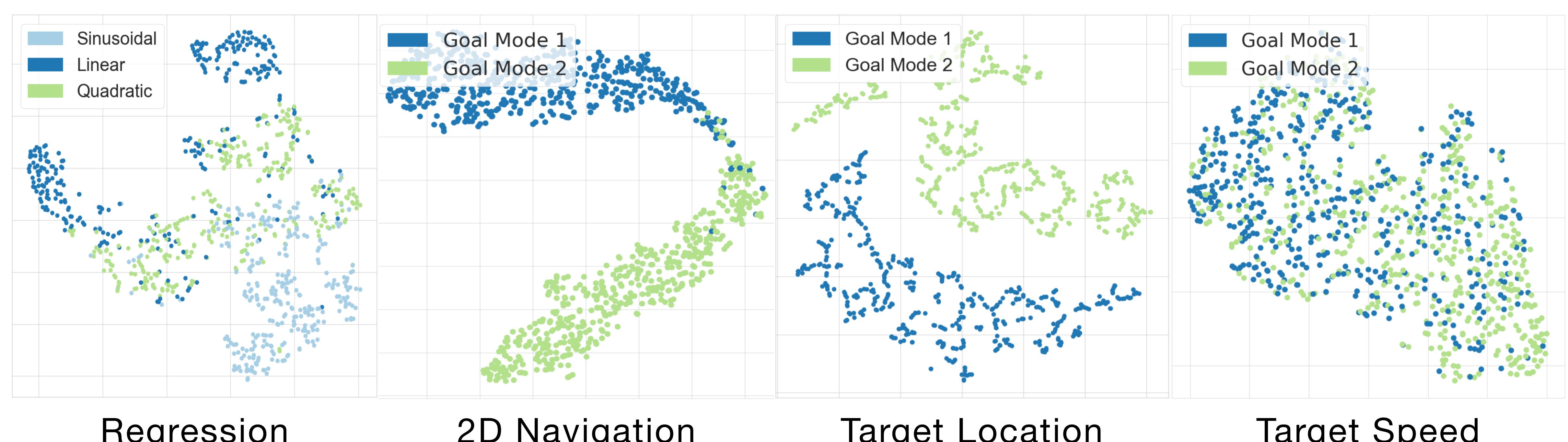
Target Location

Target Speed

2D Navigation Trajectory



Task Embedding tSNE Plots



*The work was done while Risto Vuorio worked at SK T-Brain. The authors are grateful for the computation resources provided by the company.