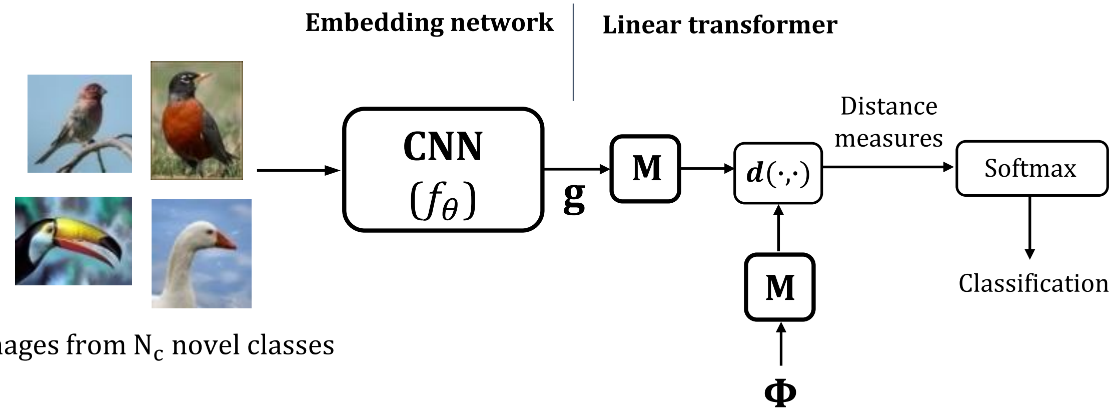
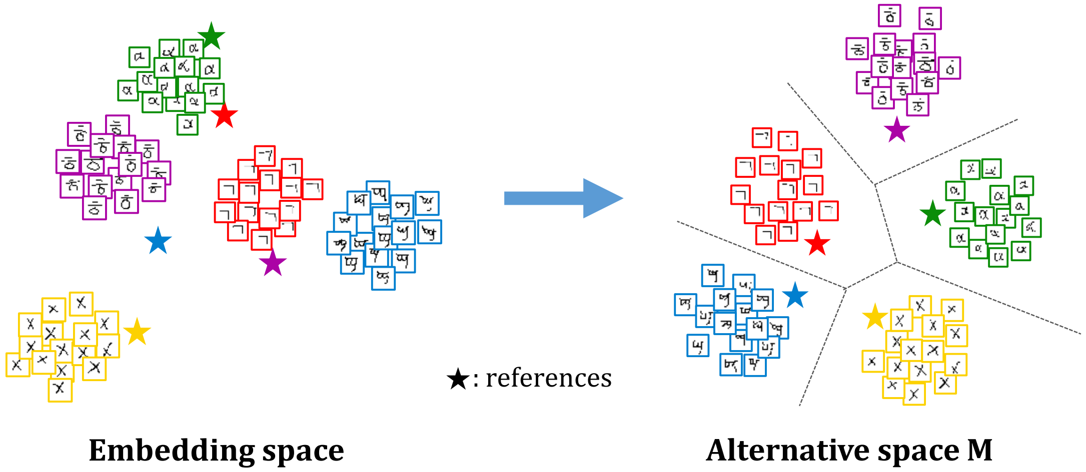
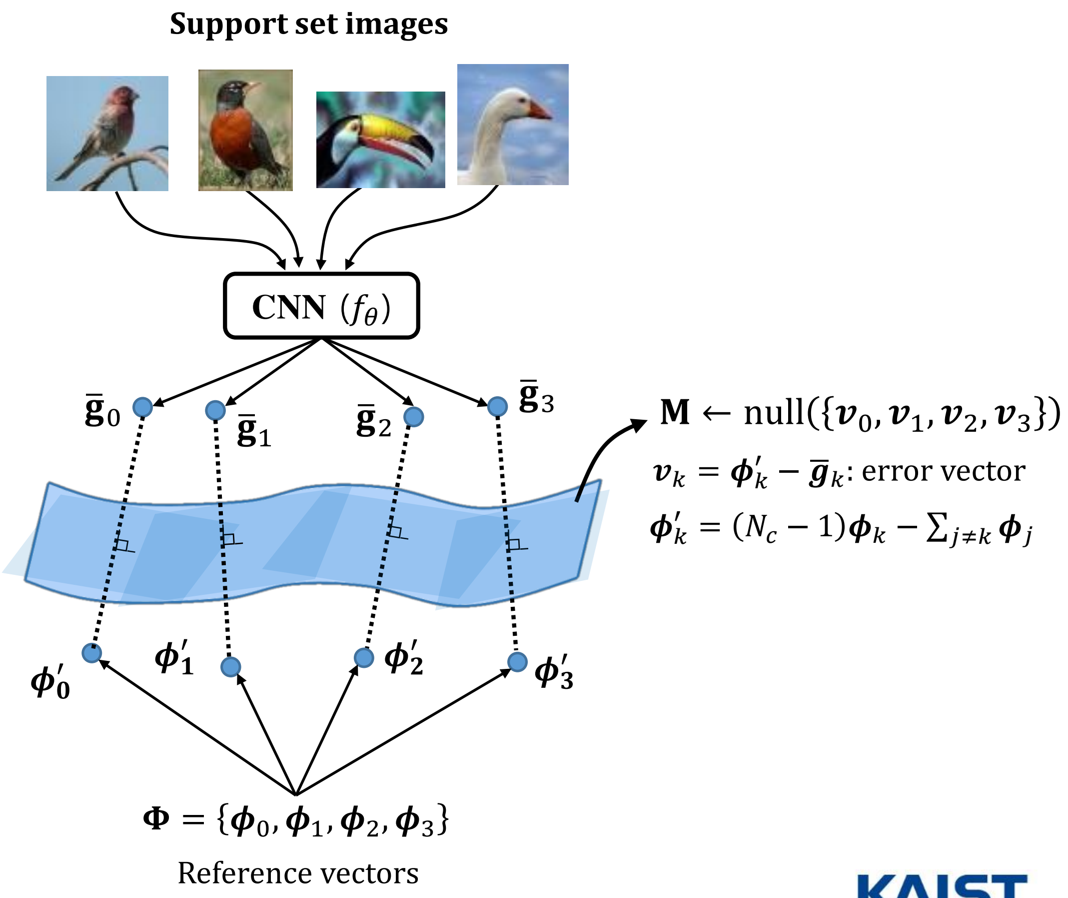
Meta-Learner with Linear Nulling

Sung Whan Yoon
Postdoctoral Researcher
Jun Seo
Ph. D. Student
Jaekyun Moon
Professor

- An <u>embedding network</u> is combined with a <u>linear transformer</u>.
- The linear transformer carries out <u>null-space projection on an alternative classification space.</u>
- The projection space M is constructed to <u>match the network output with a special set of reference vectors.</u>







OBOE: Collaborative Filtering for AutoML Initialization

Chengrun Yang, Yuji Akimoto, Dae Won Kim, Madeleine Udell
Cornell University

Goal: Select models for a new dataset within time budget.

Given: Model performance and runtime on previous datasets.

Approach:

- **low rank** dataset-by-model collaborative filtering matrix
- **predict model runtime** using polynomials
- **classical experiment design** for cold-start
- missing entry imputation for model performance prediction

Performance:

- cold-start: high accuracy
- model selection: fast and perform well

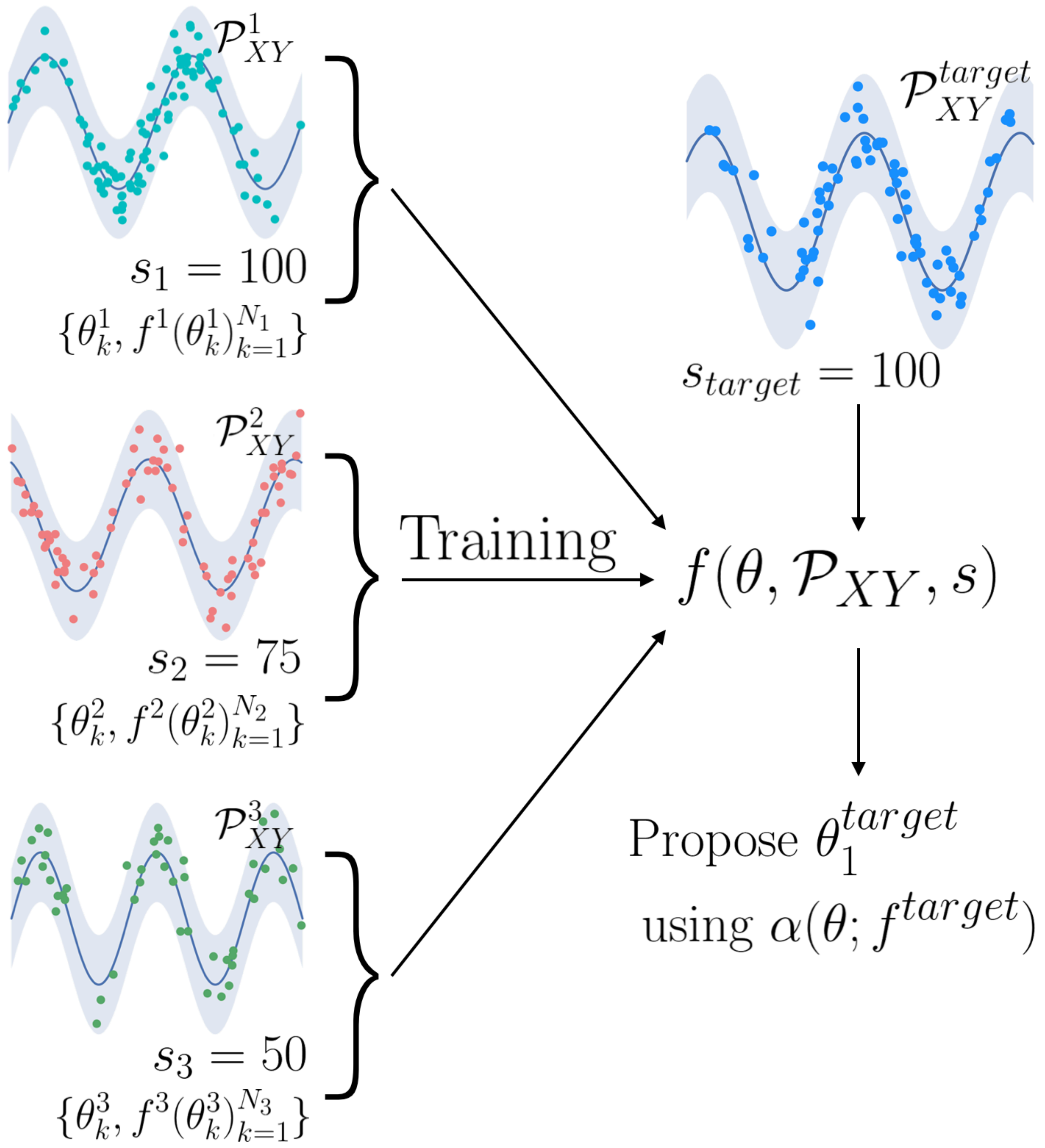
Backpropamine: meta-learning with neuromodulated Hebbian plasticity

- **Differentiable plasticity**: meta-learning with Hebbian **plastic** connections
    - Meta-train both the baseline **weight** and **plasticity** of each connection to support efficient learning in any episode
- In nature, plasticity is under real-time control through **neuromodulators**
    - The brain can decide **when** and **where** to be plastic
- **Backpropamine** = **Differentiable plasticity + neuromodulation**
    - Make the **rate** of plasticity a real-time **output** of the network
    - During each episode, the network effectively learns by self-modification
- Results:
    - Solves tasks that non-modulated networks cannot
    - Improves LSTM performance on PTB language modeling task

UBER AI Labs

Hyperparameter Learning via Distributional Transfer

Ho Chung Leon Law[1], Peilin Zhao[2], Junzhou Huang[2] and Dino Sejdinovic[1]
[1]University of Oxford and [2]Tencent AI Lab



Goal (hyperparameter selection):
Optimise f^{target} (target objective) w.r.t θ:

$$\theta^*_{target} = \text{argmax}_{\theta\in\Theta} f^{target}(\theta)$$

Scenario:

- We have n potentially related tasks f^i, $i = 1, \dots n$
- For these tasks, we have $\{\theta^i_k, f^i(\theta^i_k)\}_{k=1}^{N_i}$ from past runs

Method:

- Assume training data D_i comes from distribution $\mathcal{P}^i_{XY}$
- Transfer information using embeddings of $\mathcal{P}^i_{XY}$
- Jointly model θ, $\mathcal{P}_{XY}$ and sample size s

Toward Multimodal Model-Agnostic Meta-Learning

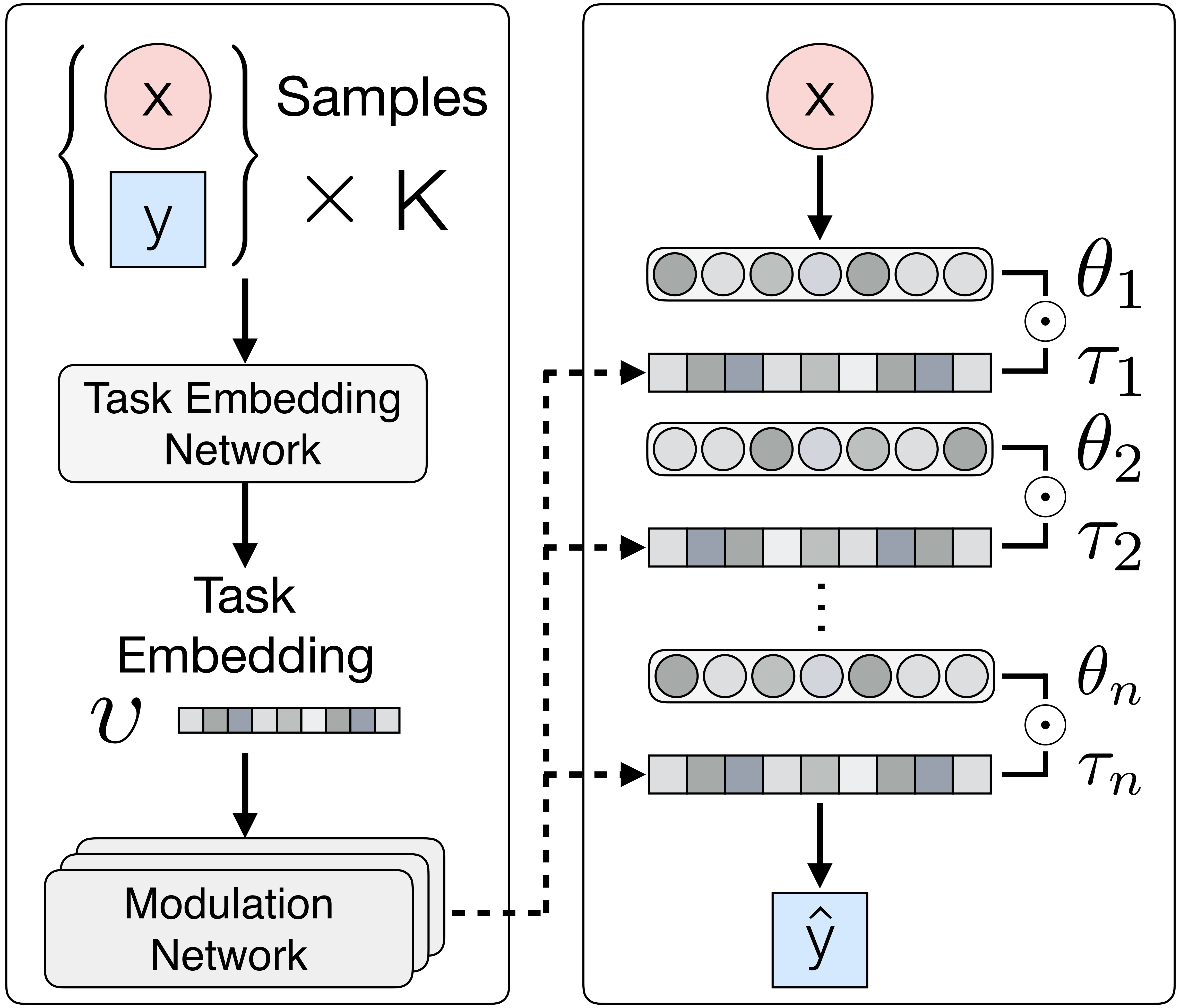
Risto Vuorio[1], Shao-Hua Sun[2], Hexiang Hu[2] & Joseph J. Lim[2]
University of Michigan[1] University of Southern California[2]

The limitation of the MAML family

- One initialization can be suboptimal for multimodal task distributions.

Multi-Modal MAML

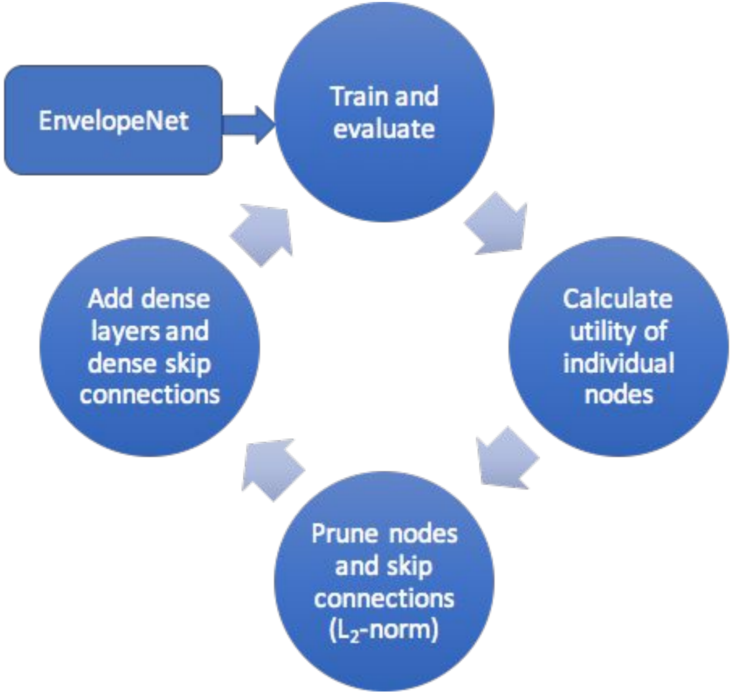
1. Model-based meta-learner computes task embeddings
2. Task embeddings are used to modulate gradient-based meta-learner
3. Gradient-based meta-learner adapts via gradient steps



Model-based Meta-learner

Gradient-based Meta-learner

Fast Neural Architecture Construction using EnvelopeNets

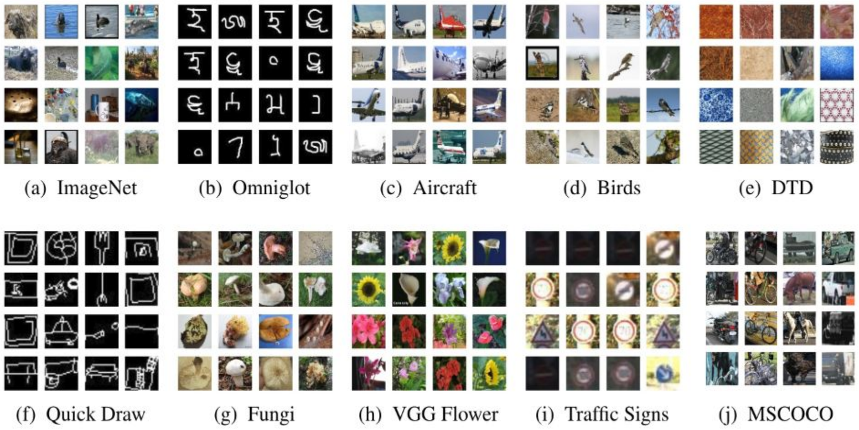


1. Finds architecture for CNNs in ~0.25 days
2. Based on the idea of utility of individual nodes.
3. Closely aligns with a theory of human brain ontogenesis.

Meta-Dataset: A Dataset of Datasets for Learning to Learn from Few Examples

Eleni Triantafillou, Tyler Zhu, Vincent Dumoulin, Pascal Lamblin, Kelvin Xu, Ross Goroshin, Carles Gelada, Kevin Swersky, Pierre-Antoine Manzagol, Hugo Larochelle

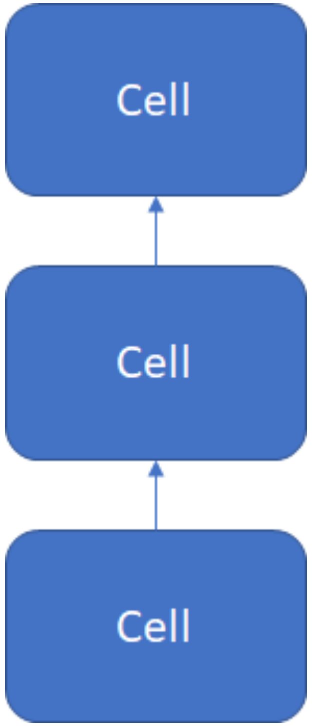
- New benchmark for **few-shot classification**
- Two-fold approach**:**
  1. **Change the data**
     - Large-scale
     - Diverse
  2. **Change the task creation**
     - Introduce imbalance
     - Utilize class hierarchy for ImageNet



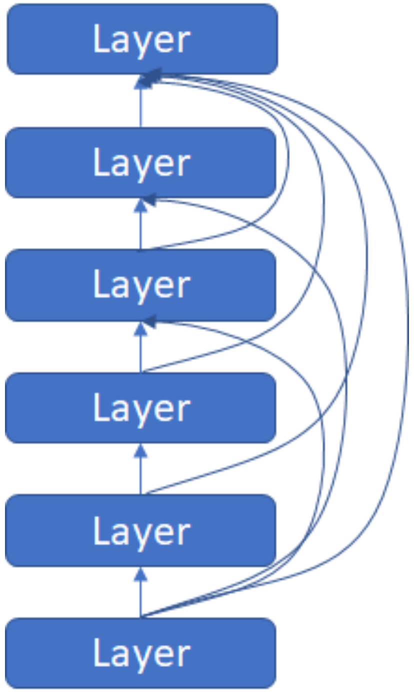
- Preliminary results on: baselines, Prototypical Networks, Matching Networks, and MAML.
- Leveraging data of multiple sources remains an open and interesting research direction!

Macro Neural Architecture Search Revisited

Hanzhang Hu[1], John Langford[2], Rich Caruana[2], Eric Horvitz[2], Debadeepta Dey[2]
[1]Carnegie Mellon University, [2]Microsoft Research



Cell Search: applies the found template on predefined skeleton.



Macro Search: learns all connections and layer types.

Cell Search: the predefined skeleton ensures the simplest cell search can achieve 4.6% error with 0.4M params on CIFAR 10.

Key take-away: macro search can be competitive against cell search, even with simple random growing strategies, if the initial model is the same as cell search.

Microsoft Research

THE ROBOTICS INSTITUTE

AutoDL 2019

Help Automating Deep Learning

Join the AutoDL challenge!

https://autodl.chalearn.org

AutoDL challenge design and beta tests

Zhengying Liu*, Olivier Bousquet, André Elisseeff, Sergio Escalera, Isabelle Guyon, Julio Jacques Jr., Albert Clapés, Adrien Pavao, Michèle Sebag, Danny Silver, Lisheng Sun-Hosoya, Sébastien Tréguer, Wei-Wei Tu, Yiqi Hu, Jingsong Wang, Quanming Yao

Modular meta-learning in abstract graph networks for combinatorial generalization

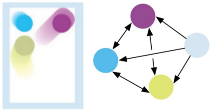
Ferran Alet, Maria Bauza, A. Rodriguez, T. Lozano-Perez, L. Kaelbling

code&pdf:alet-etal.com

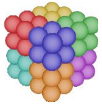
Combinatorial generalization: generalizing by reusing neural modules

Graph Neural Networks

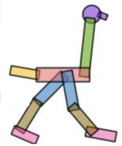
Nodes tied to entities



Objects

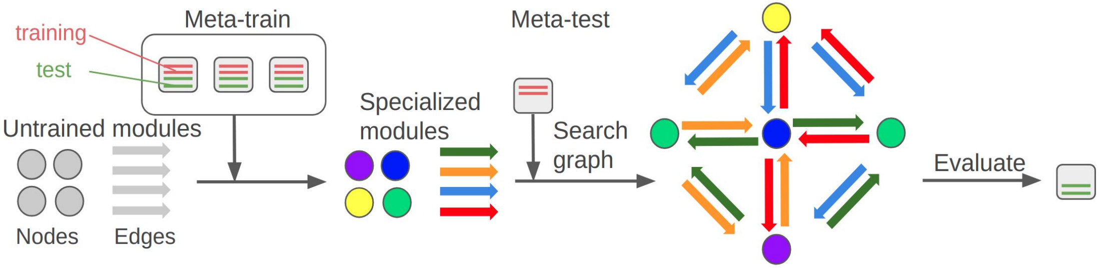


Particles



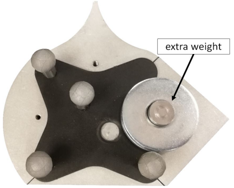
Joints

Modular meta-learning



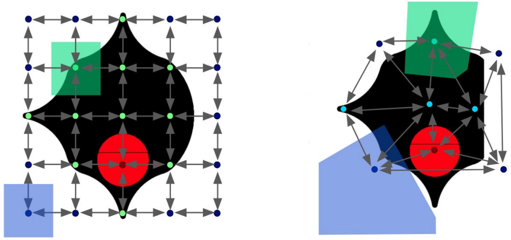
We introduce: Abstract Graph Networks

nodes are not tied to concrete entities



OmniPush dataset

Graph Element Networks



Cross-Modulation Networks For Few-Shot Learning

Hugo Prol[†], Vincent Dumoulin[‡],
and Luis Herranz[†]

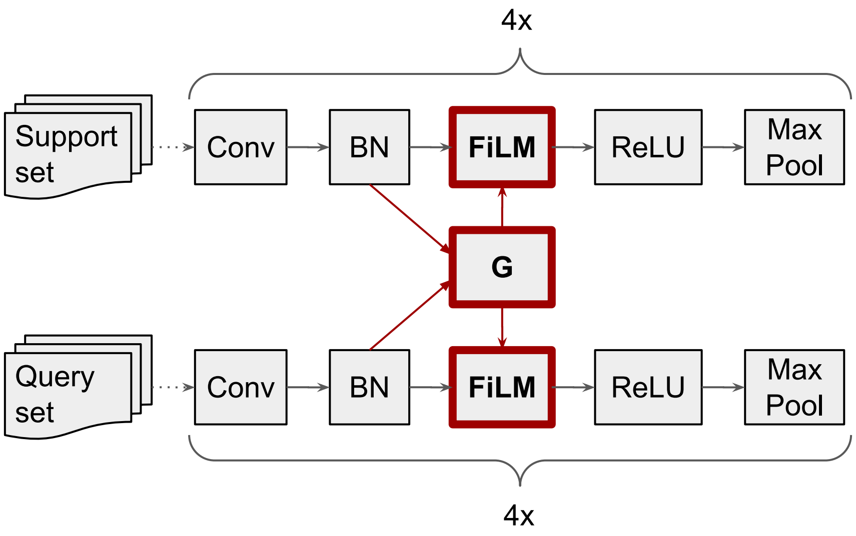
[†] Computer Vision Center, Univ. Autònoma de Barcelona
[‡] Google Brain

Key idea: allow support and query examples to interact at each level of abstraction.

Extending the feature extraction pipeline of Matching Networks:

☆ Channel-wise affine transformations: $\mathrm{FiLM}(\mathbf{x}) = (1+\gamma)\odot\mathbf{x}+\beta$

☆ Subnetwork G predicts the affine parameters γ and β



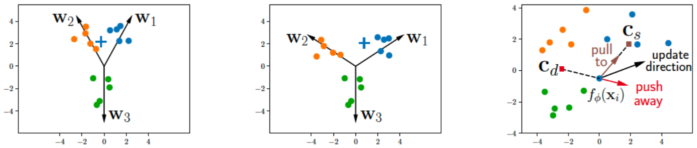
Large Margin Meta-Learning for Few-Shot Classification

The University of Hong Kong[1], The Hong Kong Polytechnic University[2]

Yong Wang[1], Xiao-Ming Wu[2], Qimai Li[2], Jiatao Gu[1], Wangmeng Xiang[2], Lei Zhang[2], Victor O.K. Li[1]

Large Margin Principle

$$\mathcal{L} = \mathcal{L}_{\text{softmax}} + \lambda * \mathcal{L}_{\text{large-margin}}$$



(a) (b) (c)

Fig. 1: Large margin meta-learning. (a) Classifier trained without the large margin constraint. (b) Classifier trained with the large margin constraint. (c) Gradient of the triplet loss.

One Implementation: Triplet Loss

$$\mathcal{L}_{\text{large-margin}} = \frac{1}{N_t}\sum_{i=1}^{N_t}\left[\| f_\phi(\mathbf{x}_i^a) - f_\phi(\mathbf{x}_i^p) \|_2^2 - \| f_\phi(\mathbf{x}_i^a) - f_\phi(\mathbf{x}_i^n) \|_2^2 + m\right]_+ .$$

Case study

- Graph Neural Network (GNN)
- Prototypical Network (PN)

Analysis

After rearrangement:

$$\mathcal{L}_{\text{large-margin}} = \frac{1}{N_t}\left(\sum_{\mathbf{x}_s\in S_s} \| f_\phi(\mathbf{x}_i) - f_\phi(\mathbf{x}_s) \|_2^2 - \sum_{\mathbf{x}_d\in S_d} \| f_\phi(\mathbf{x}_i) - f_\phi(\mathbf{x}_d) \|_2^2\right) + \text{const.}$$

The gradient:

$$\begin{aligned}\frac{\partial \mathcal{L}_{\text{large-margin}}}{\partial f_\phi(\mathbf{x}_i)} &= \frac{2}{N_t}\left(\sum_{\mathbf{x}_s\in S_s}(f_\phi(\mathbf{x}_i) - f_\phi(\mathbf{x}_s)) - \sum_{\mathbf{x}_d\in S_d}(f_\phi(\mathbf{x}_i) - f_\phi(\mathbf{x}_d))\right)\\ &= -\frac{2|S_s|}{N_t}\left(\frac{1}{|S_s|}\sum_{\mathbf{x}_s\in S_s} f_\phi(\mathbf{x}_s) - f_\phi(\mathbf{x}_i)\right) - \frac{2|S_d|}{N_t}\left(f_\phi(\mathbf{x}_i) - \frac{1}{|S_d|}\sum_{\mathbf{x}_d\in S_d} f_\phi(\mathbf{x}_d)\right)\\ &= \underbrace{-\frac{2|S_s|}{N_t}(\mathbf{c}_s - f_\phi(\mathbf{x}_i))}_{\text{pull towards its own class}} - \underbrace{\frac{2|S_d|}{N_t}(f_\phi(\mathbf{x}_i) - \mathbf{c}_d)}_{\text{push away from other classes}} .\end{aligned}$$

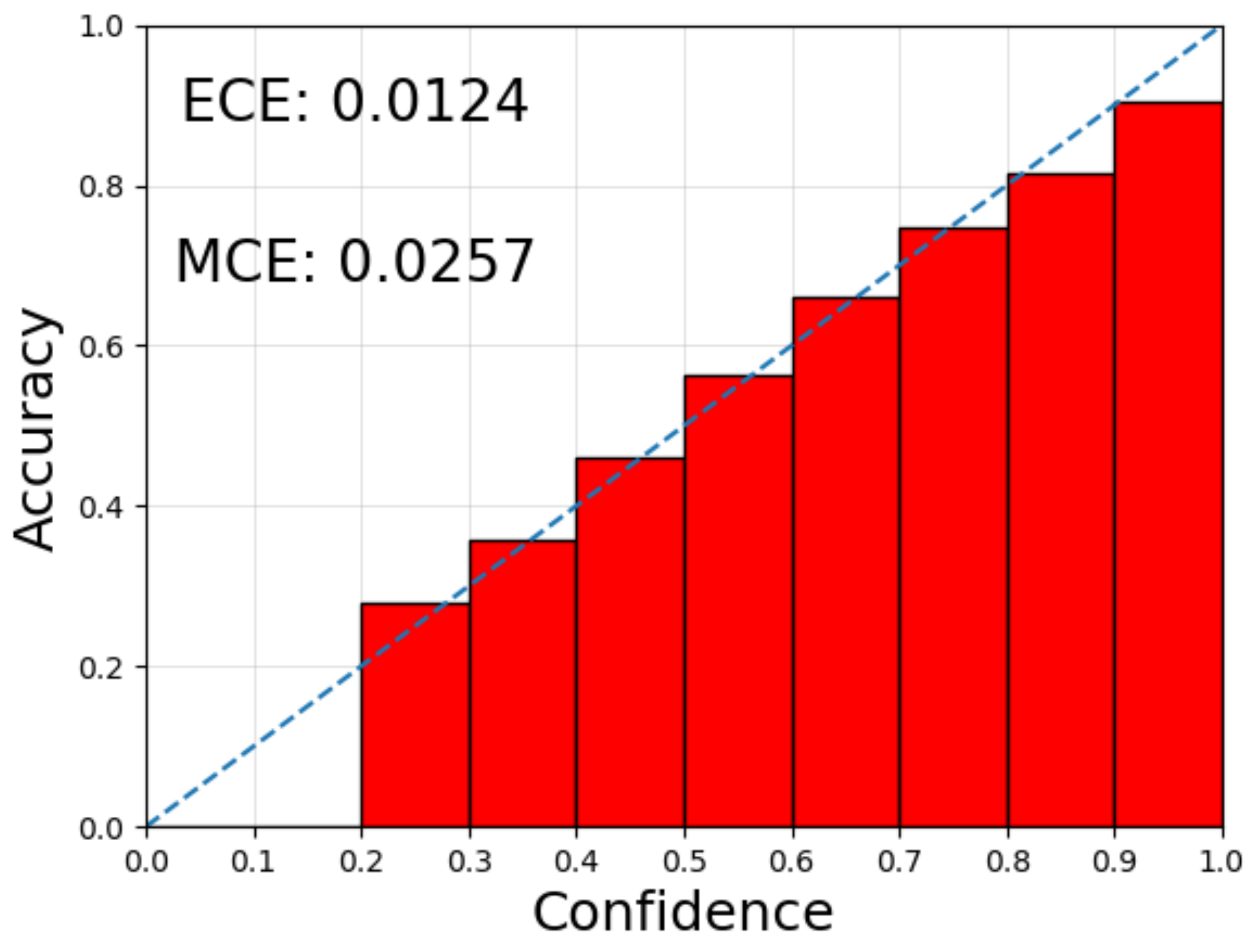
Features

- We implement and compare several of other large margin methods for few-shot learning.
- Our framework is simple, efficient, and can be applied to improve existing and new meta-learning methods with very little overhead.

Amortized Bayesian Meta-Learning

Sachin Ravi & Alex Beatson
Department of Computer Science, Princeton University

- Lot of progress in few-shot learning but under controlled settings
- In real world, relationship between training and testing tasks can be tenuous
  - Task-specific predictive uncertainty is crucial
- We present gradient-based meta-learning method for computing task-specific approximate posterior
- Show that method displays good predictive uncertainty on contextual-bandit and few-shot learning tasks



The effects of negative adaptation in Model-Agnostic Meta-Learning

Université de Montréal

Mila
Université de Montréal
McGill

Tristan Deleu, Yoshua Bengio

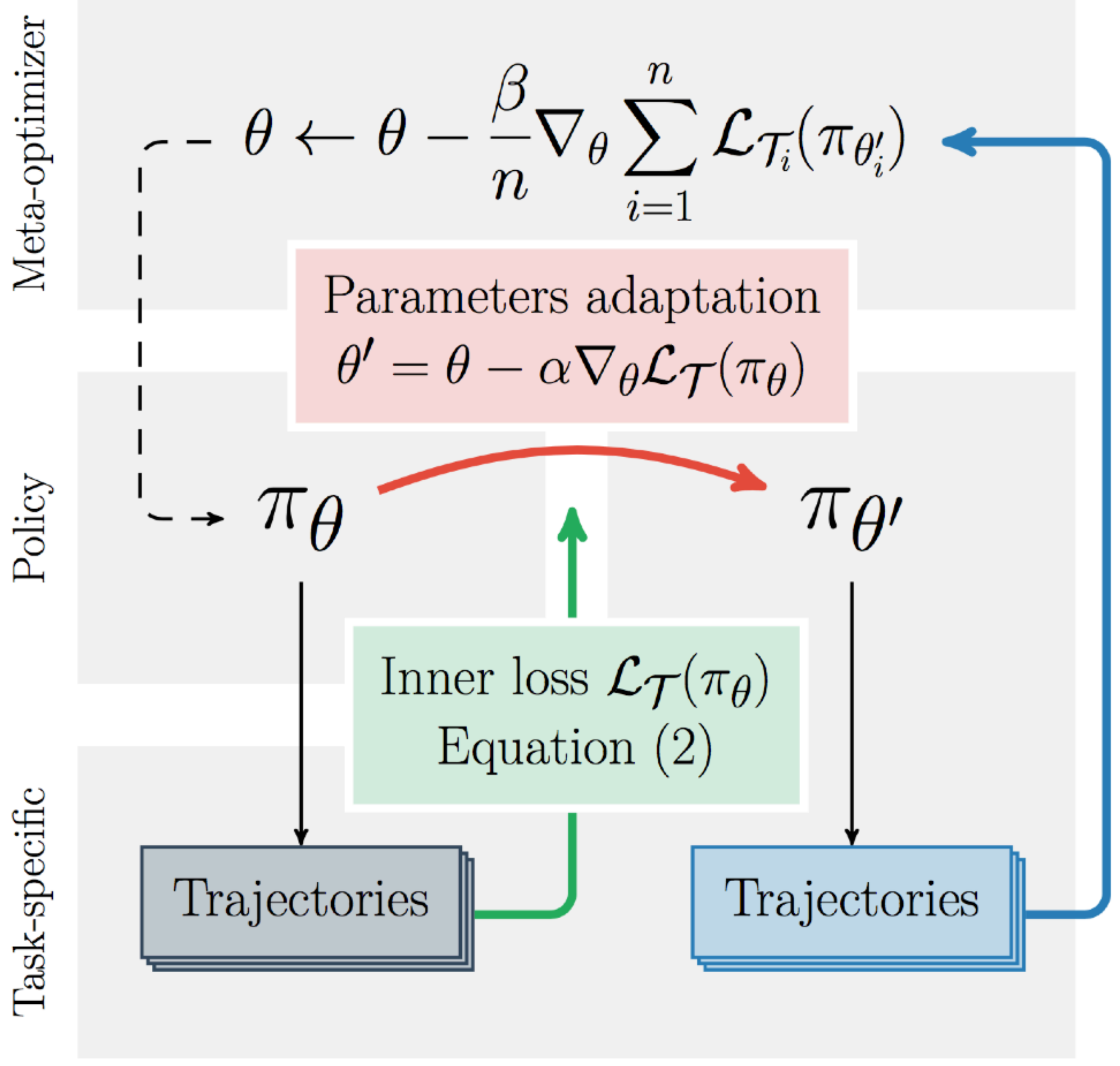
- The advantage of meta-learning is well-founded under the assumption that **the adaptation phase does improve the performance** of the model on the task of interest

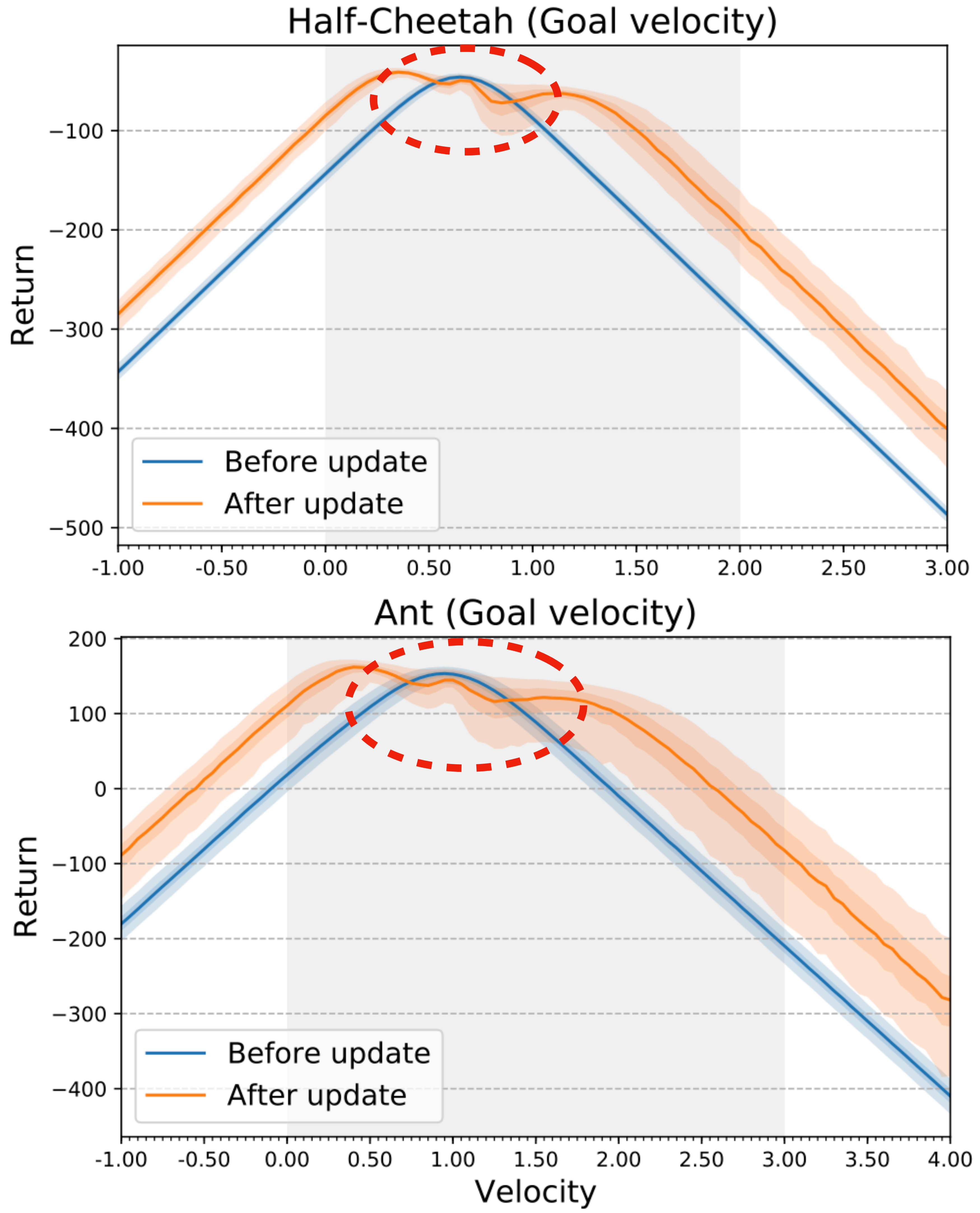
- Optimization: maximize the performance after adaptation, **performance improvement is not explicitly enforced**

$$\min_{\theta} \mathbb{E}_{\mathcal{T}\sim p(\mathcal{T})}[\mathcal{L}(\theta'_{\mathcal{T}}; \mathcal{D}'_{\mathcal{T}})]$$

- We show empirically that performance **can decrease** after adaptation in MAML. We call this **negative adaptation**

- How to fix this issue? Ideas from **Safe Reinforcement Learning**





Mitigating Architectural Mismatch During the Evolutionary Synthesis of Deep Neural Networks

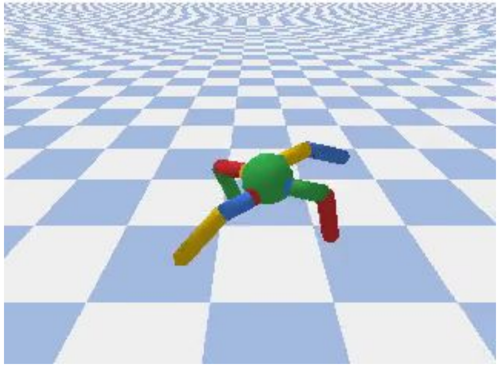
Audrey G. Chung, Paul Fieguth, Alexander Wong

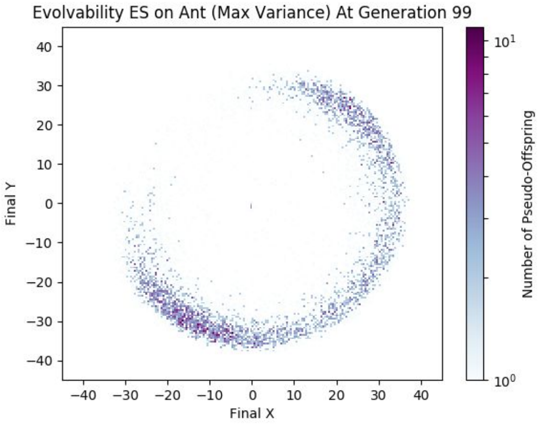
- *Evolutionary deep intelligence* for increasingly efficient networks
- Preliminary study into the effects of architectural alignment
- Like-with-like mating policy via gene tagging system
- Resulting networks are comparable:
  - Restricts search space exploration?
  - Compensated with training epochs?
  - ???

Evolvability ES: Scalable Evolutionary Meta-Learning

By Alexander Gajewski, Jeff Clune, Kenneth O. Stanley, and Joel Lehman

- Evolvability ES is a meta-learning algorithm inspired by Evolution Strategies [1]
- Surprisingly, Evolvability ES finds parameters such that at test time, **random** perturbations result in diverse behaviors
- In a simulated Ant locomotion domain, adding Gaussian noise to the parameters results in policies which move in many different directions



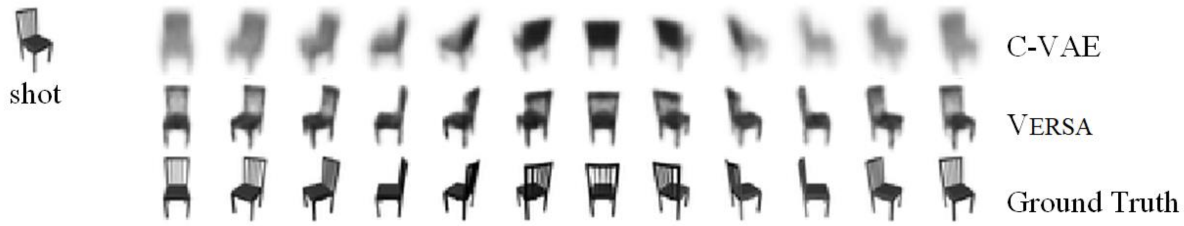


[1] Salimans et al., Evolution Strategies as a Scalable Alternative to Reinforcement Learning, 2017.

Consolidating the Meta-Learning Zoo

A Unifying Perspective as Posterior Predictive Inference

- **Novel**: Probabilistic, amortized, multi-task, meta-learning framework.
- **Meta-learning**: Learns how to learn a classifier or regressor for each new task.
- **Unifies**: MAML, Meta-LSTM, Prototypical networks, and Conditional Neural Processes are special cases.
- **State of the art**: Leading classification accuracy on 5 of 6 Omniglot & *mini*ImageNet tasks.
- **Efficient**: Test-time requires only forward passes, no gradient steps are needed.
- **Versatile**: Robust classification accuracy as shot and way are varied at *test*-time.
- **High quality 1-shot view reconstruction:**

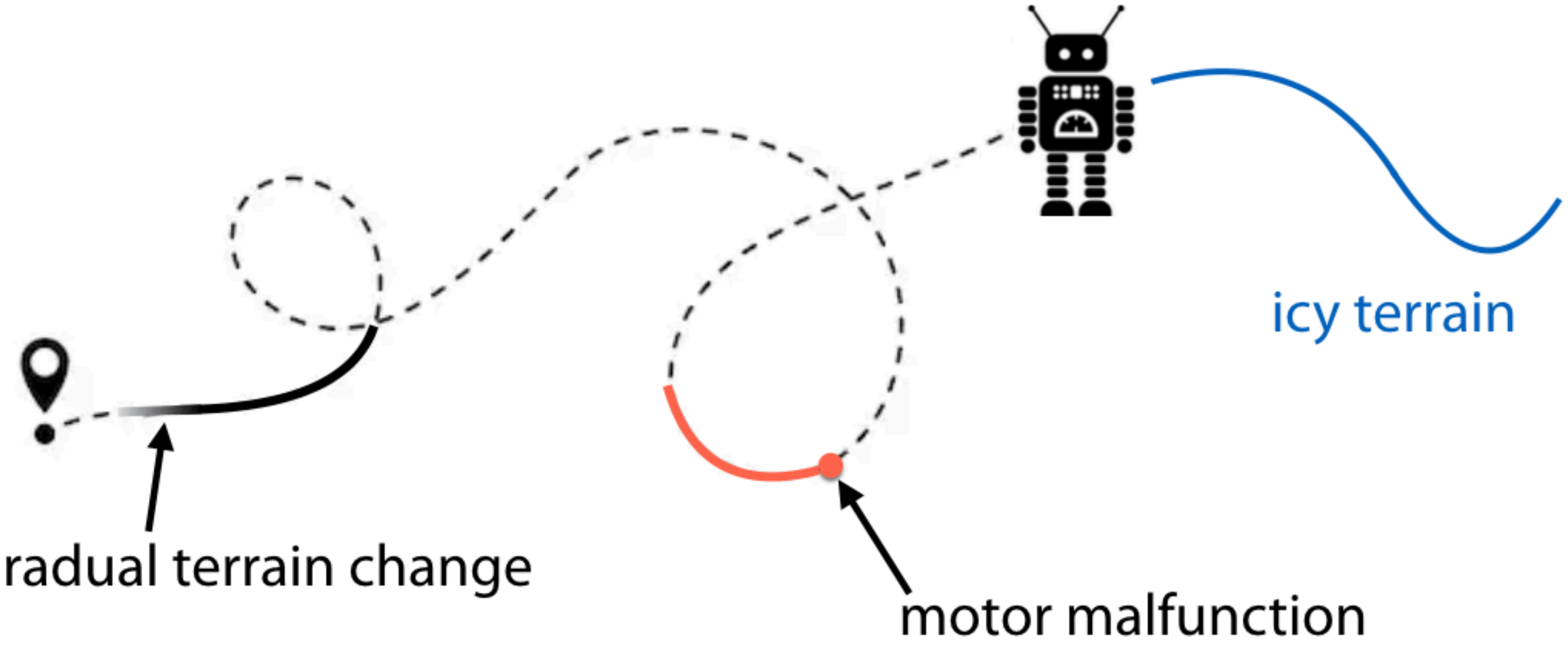


UNIVERSITY OF CAMBRIDGE

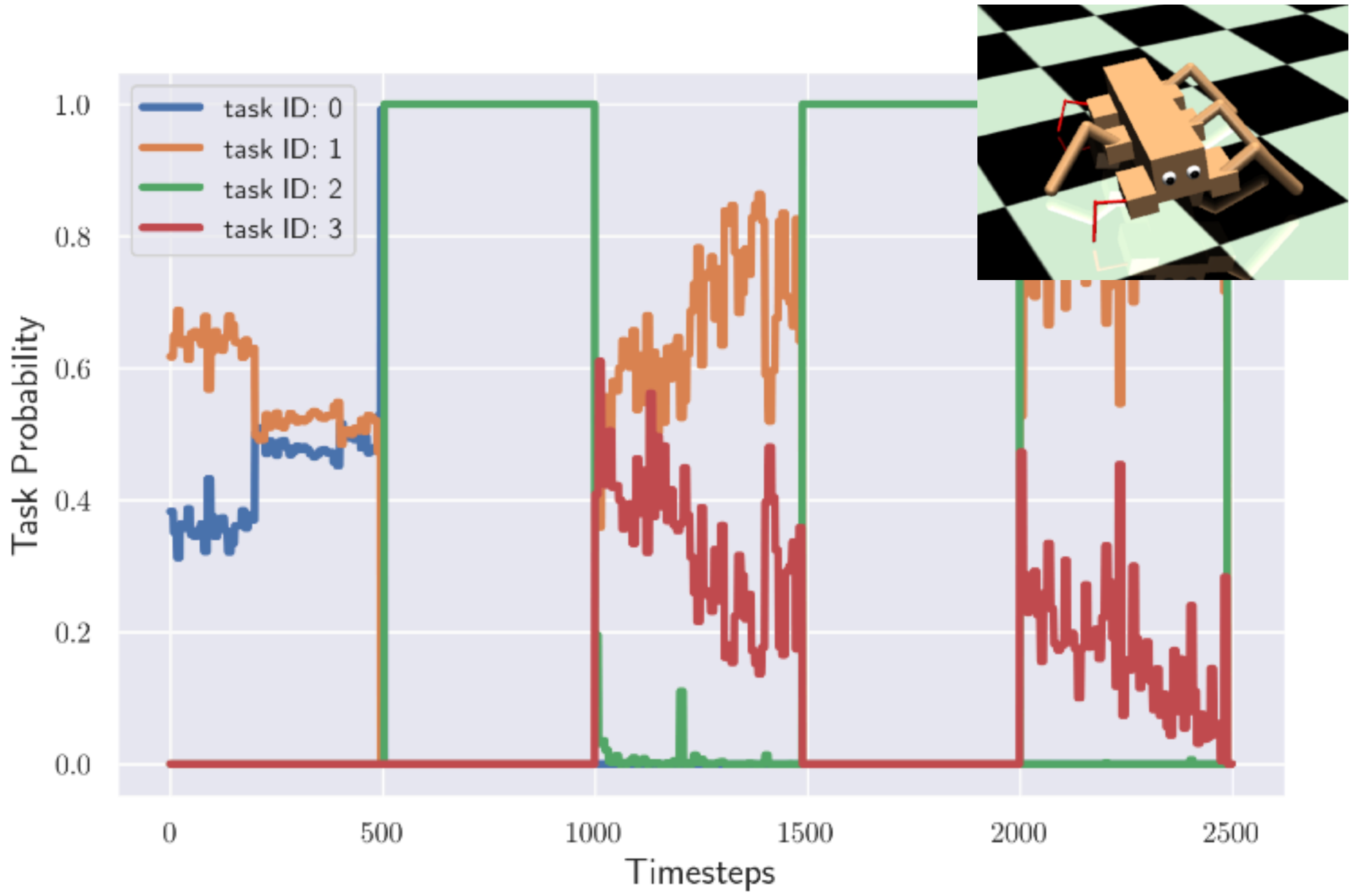
Jonathan Gordon[1], John Bronskill[1], Matthias Bauer[1,2], Sebastian Nowozin[3], Richard E. Turner[1,3]
[1]University of Cambridge, [2]MPI for Intelligent Systems, Tübingen, [3]Microsoft Research

Deep Online Learning via Meta-Learning: Continual Adaptation for Model-Based RL

Anusha Nagabandi, Chelsea Finn, Sergey Levine



Can we use meta-learning for effective online learning?



Our method can:

- Reason about non-stationary latent distributions over tasks.
- Recall past tasks

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH