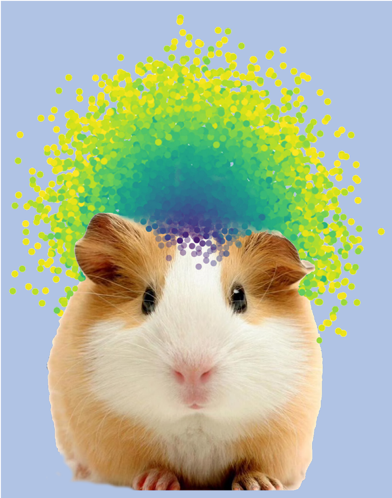


CAVIA
Fast Context Adaptation via Meta-Learning

Luisa Zintgraf, Kyriacos Shiarlis, Vitaly Kurin, Katja Hofmann, Shimon Whiteson

Problem Setting: Few-Shot Learning

Given distributions over tasks: $p_{\text{train}}(\mathcal{T}), p_{\text{test}}(\mathcal{T})$
- Meta-learn how to adapt fast to any tasks from p_{train}
- Evaluate generalisation ability on tasks from p_{test}

Background: MAML (Finn et al., 2017)

Idea: Learn network initialisation s.t. at test time, only few gradient steps are necessary to perform well.

Inner loop:
- Sample batch of tasks, $\mathcal{T}_i \sim p_{\text{train}}$
- For each task:
  - Get train/test data: $\mathcal{D}_i^{\text{train}}$, $\mathcal{D}_i^{\text{test}}$
  - Adapt model: $\theta_i = \theta - \alpha \nabla_\theta \sum_{(x,y)\in\mathcal{D}_i^{\text{train}}} \mathcal{L}_{\mathcal{T}_i}(f_\theta(x), y)$

Outer loop:
- Update initial parameters for good test performance by backpropagating through inner loop update

$$\theta \leftarrow \theta - \beta \nabla_\theta \sum_i \sum_{(x,y)\in\mathcal{D}_i^{\text{test}}} \mathcal{L}_{\mathcal{T}_i}(f_{\theta_i}(x), y)$$

CAVIA

Many tasks & benchmarks don't require generalisation beyond *task identification*. In this case, we shouldn't update all parameters!

Hence we separate the network into task-specific parameters ϕ, and shared parameters θ, trained similarly to MAML as follows.

Inner loop update:

For each task i, update context parameters:

$$\phi_i = \phi_0 - \alpha \nabla_{\phi_0} \sum_{(x,y)\in\mathcal{D}_i^{\text{train}}} \mathcal{L}_{\mathcal{T}_i}(f_{\{\phi_0,\theta\}}(x), y)$$

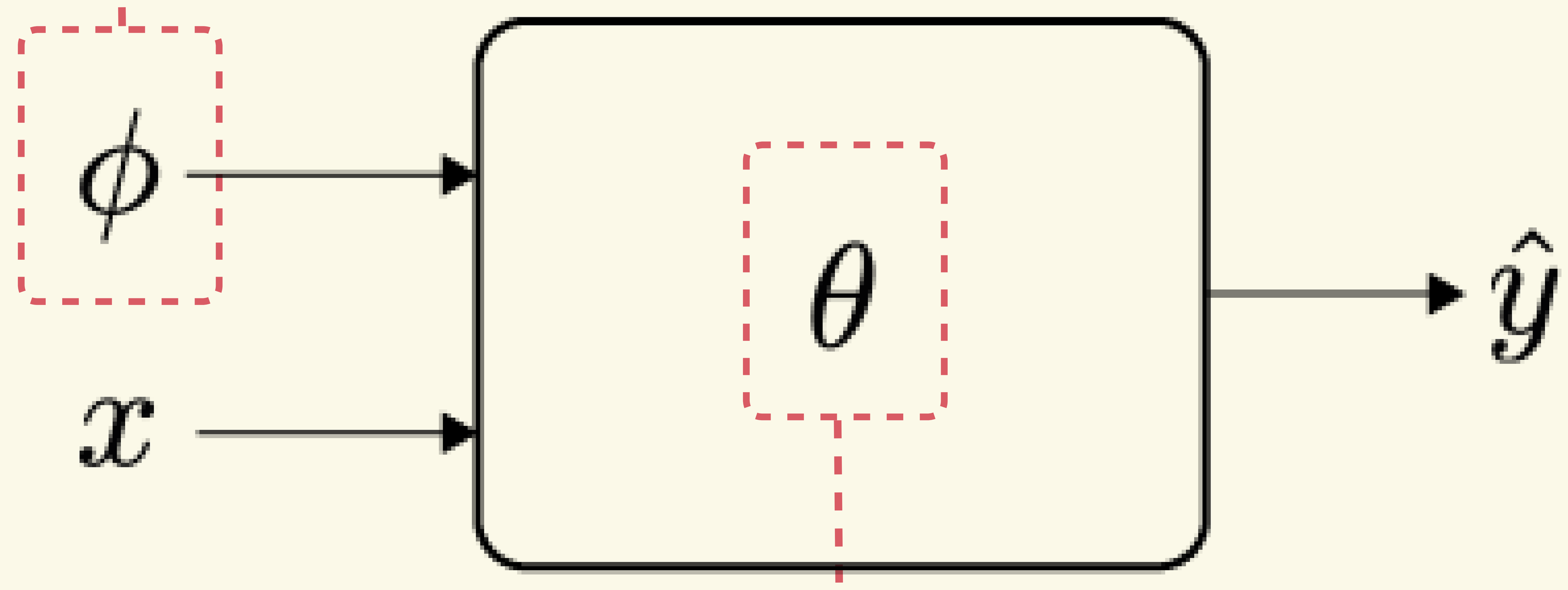
Outer loop update:

Update shared parameters θ using test loss from individual tasks

$$\theta \leftarrow \theta - \beta \nabla_\theta \sum_i \sum_{(x,y)\in\mathcal{D}_i^{\text{test}}} \mathcal{L}_{\mathcal{T}_i}(f_{\{\phi_i,\theta\}}(x), y)$$

References
Finn, Chelsea, Pieter Abbeel, and Sergey Levine. "Model-agnostic meta-learning for fast adaptation of deep networks." ICML 2017.
Garnelo, M., Rosenbaum, D., Maddison, C. J., Ramalho, T., Saxton, D., Shanahan, M., ... & Eslami, S. M. (2018). Conditional neural processes. ICML 2018.

Context parameters:
Task-specific, the only thing that's updated test time.
Represent task embedding.



Network parameters:
Shared across tasks, fixed at test time.
Trained in outer loop only.

Less prone to overfitting compared to MAML

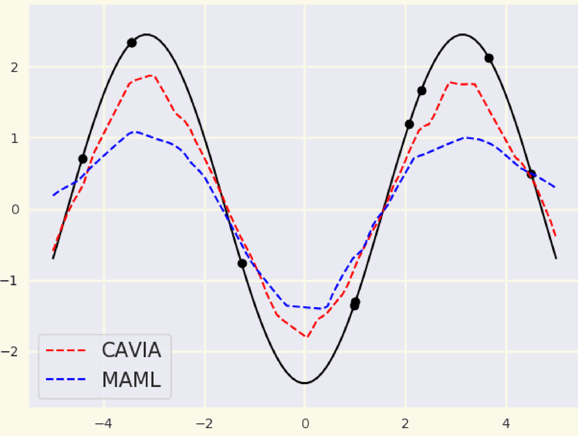
Context parameters are interpretable / reusable

Easy to parallelise

Sine Curve Regression

Task = sine curve with randomly chosen amplitude and phase.

Before Update:

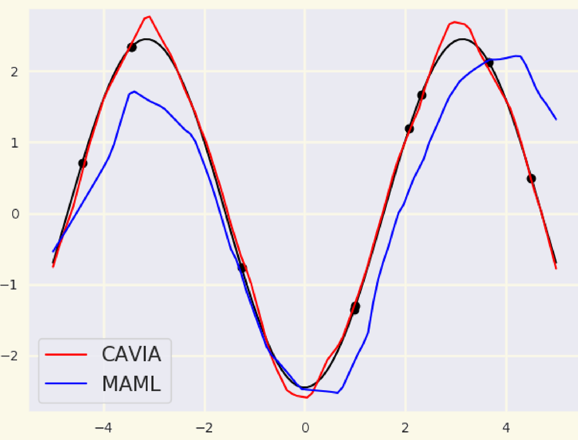


Gradient Update:

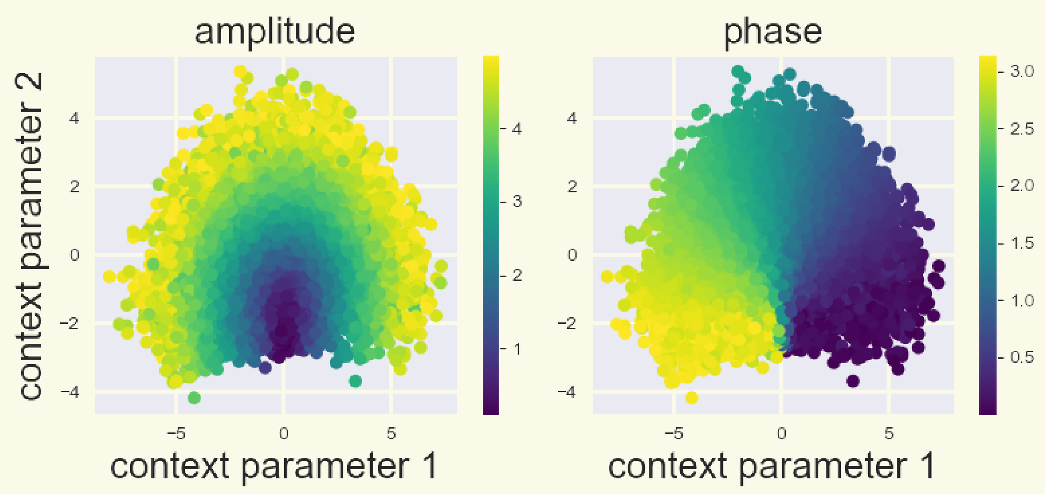
MAML: ~ 1500 params

CAVIA: 2 params

After Update:



Visualisation of the learned context parameters:



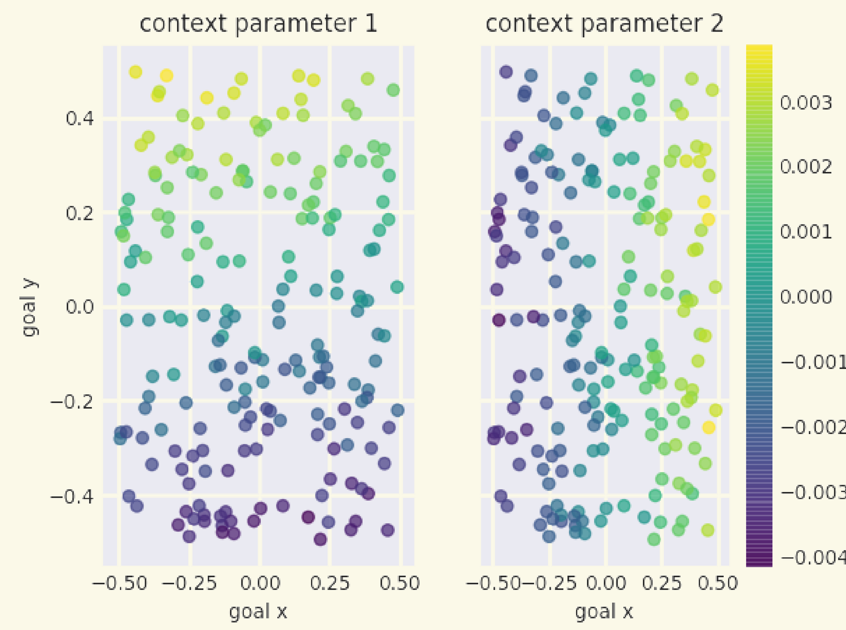
Few-Shot Classification

- Mini-ImageNet benchmark
- Number of context parameters: 100
- Increased network size leads to:
  - overfitting for MAML
  - better performance for CAVIA

Method	5-way accuracy 1-shot	5-shot
Matching Nets (Vinyals et al., 2016)	46.6%	60.0%
Meta LSTM (Ravi & Larochelle, 2017)	$43.44 \pm 0.77\%$	$60.60 \pm 0.71\%$
Prototypical Networks (Snell et al., 2017)	$46.61 \pm 0.78\%$	$65.77 \pm 0.70\%$
Meta-SGD (Li et al., 2017)	$50.47 \pm 1.87\%$	$64.03 \pm 0.94\%$
REPTILE (Nichol & Schulman, 2018)	$49.97 \pm 0.32\%$	$65.99 \pm 0.58\%$
MT-NET (Lee & Choi, 2018)	$\mathbf{51.70} \pm 1.84\%$	-
VERSA (Gordon et al., 2018)	$\mathbf{53.40} \pm 1.82\%$	$\mathbf{67.37} \pm 0.86$
MAML (32) (Finn et al., 2017a)	$48.07 \pm 1.75\%$	$63.15 \pm 0.91\%$
MAML (64)	$44.70 \pm 1.69\%$	$61.87 \pm 0.93\%$
CAVIA (32)	$47.24 \pm 0.65\%$	$59.05 \pm 0.54\%$
CAVIA (128)	$49.84 \pm 0.68\%$	$64.63 \pm 0.54\%$
CAVIA (512)	$\mathbf{51.82} \pm 0.65\%$	$65.85 \pm 0.55\%$
CAVIA (512, first order)	$49.92 \pm 0.68\%$	$63.59 \pm 0.57\%$

Reinforcement Learning

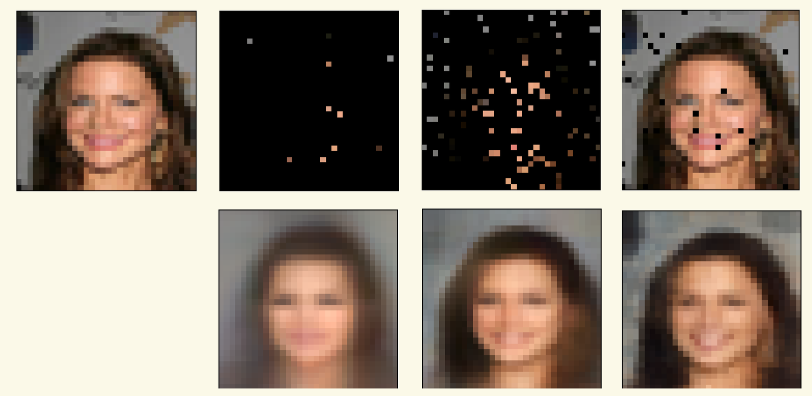
2D Navigation: Task = goal position.
- Visualisation of learned context parameters:



MuJoCo Cheetah (Vel & Dir):
- Performs similar to MAML, with only 50 context parameters

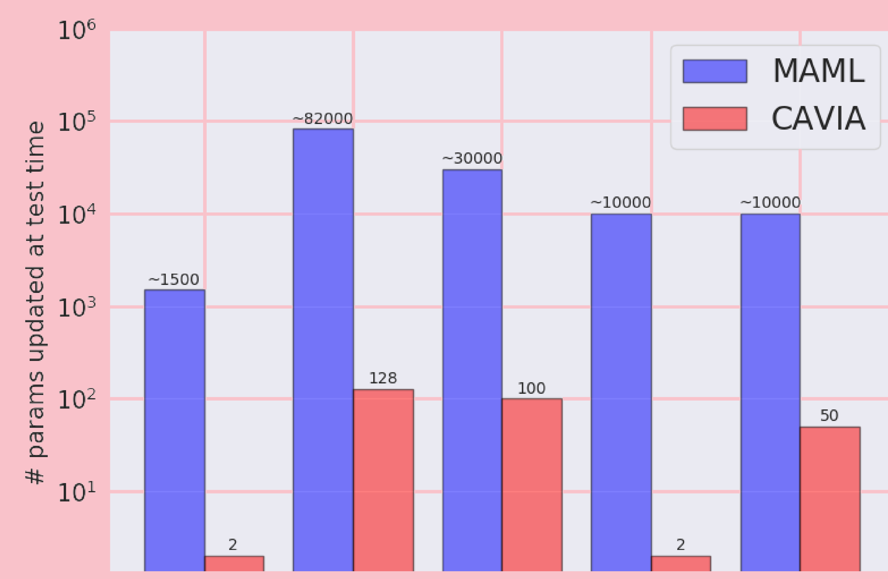
Image Completion: CelebA

- Challenging regression task (Garnelo et al. 2018)
- Number of context parameters: 128
- CAVIA learns low-level image embeddings solely via backward pass



	Random Pixels			Ordered Pixels		
	10	100	1000	10	100	1000
CNP*	0.039	0.016	0.009	0.057	**0.047**	0.021
MAML	0.040	0.017	**0.006**	0.055	**0.047**	0.007
CAVIA	**0.037**	**0.014**	**0.006**	**0.053**	**0.047**	**0.006**

Number of adapted paramers



Conclusion

- When only task identification is required, don't update entire net!
- Possible weakness of current benchmarks: adaptation required is sometimes small